2021-08-01

Examination of Enrollment Patterns and Behaviors of High School Graduating Seniors at a Hispanic Serving Institution

Gustavo Monzon

University of Texas at El Paso

EXAMINATION OF ENROLLMENT PATTERNS AND BEHAVIORS OF HIGH SCHOOL GRADUATING SENIORS AT A HISPANIC SERVING INSTITUTION

GUSTAVO MONZÓN

Doctoral Program in Educational Leadership and Administration

APPROVED:

Arturo Olivárez, Jr. PhD

Jesus Cisneros, PhD

Teresa Cortez, EdD

John Hadjimarcou, PhD

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

Dedication

To my mother, Maria Evangelina Monzon and to my father, Pablo C. Monzon.

My siblings Martin, Elvia, Pablo, Alberto, and Angelica.

To Marisa. It is my privilege to be your husband.

And to my sons, Marq and Nico.

EXAMINATION OF ENROLLMENT PATTERNS AND BEHAVIORS OF HIGH SCHOOL GRADUATING SENIORS AT A HISPANIC SERVING INSTITUTION

by

GUSTAVO MONZÓN, M.B.A

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF EDUCATION

Educational Leadership and Foundation

THE UNIVERSITY OF TEXAS AT EL PASO

August 2021

Acknowledgements

I am able to present this work because of the love, inspiration, and support I have received from generous and kind people who have given their time, wisdom, and patience. It is my privilege to work with Dr. Arturo Olivarez and to receive his guidance in bringing cohesion to ideas and concepts.

I am very grateful to Dr. Teresa Cortez for being the most influential mentor and guide throughout this endeavor. You have inspired me from the start and I have persisted because of your un-wavering support.

I also acknowledge other esteemed members of my dissertation committee. Dr. Cisneros and Dr. John Hadjimarcou, for advancing the challenging work of access and excellence at The University of Texas at El Paso. If not for such values, I would not have seen opportunities in higher education.

I am grateful to Dr. Jerry D. Hubbard. You welcomed me at The University of Texas at Arlington and upon learning about my status in the doctoral program, you made a commitment to see me through to completion. Thank you.

Lastly, Dr. Marisa Pierce, your love of life, joy of new experiences, and desire to bring out the best in people shine brightly in this world. I am grateful for the many things we have experienced together, and for that one resounding message of encouragement you say to remind me that if anyone can do it, I can.

Abstract

Institutions of Higher Education (IHE) are seeking new methods of improving efficiency and effectiveness in recruiting new undergraduate students. The investment of resources can be optimized through the use of predictive modeling. For example, the IHE may consider using predictive modeling as a means of determining where it may achieve optimal marginal returns on investment of marketing and operational resources. Through an understanding of factors influencing college choice, and the impact these factors have on the college choice process and subsequent participation in post-secondary education, it is hypothesized that the application of predictive modeling may support IHE's in advancing opportunities for students of color and of low socioeconomic backgrounds, through the examination of key variables, utilizing predictive modeling. College Choice Theory provides the lens for theory-based model development. The college choice model reflected in this study is the Three-Phase Model, where each phase is influenced by a dynamic set of individual and organizational characteristics and attributes. The study will add to the body of knowledge, the development and use of a predictive model aimed at identifying important student factors intended to support Institutions of Higher Education with a mission of access.

Table of Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

Central to the advancement of economies is the development and use of human capital (OECD, 1996). It is stipulated that industry is driven by the creation and use of knowledge. The unfortunate irony is that institutions of higher education (IHE), the very same institutions tasked with developing the human capital needed to drive a knowledge-based economy, may be considered to be laggards in the creation and use of knowledge for the benefit of improving institutional efficiency and effectiveness. The use of business intelligence and business analytics to support decision making and resource management in higher education remains a challenge as institutions work towards developing an information infrastructure and the institutional capacity for use of advanced applications and associated management information (Goldstein & Katz, 2005). It is in this context upon which this study is based; the use of advanced analytics in strategic enrollment management.

1.1 ADVANCED ANALYTICS IN HIGHER EDUCATION

Advanced analytics may be defined as the application of analytic techniques used to obtain a better understanding of the business problem and optimizing solutions (Bose, 2009). One example for the application of advanced analytics in higher education is predictive modeling in strategic enrollment management; more specifically, predictive modeling in the application of undergraduate admissions marketing. The use of predictive modeling in higher education is increasingly promoted as a tool for informed decision making and optimizing the use of limited resources. In practice, however, the development and use of predictive modeling in higher education is relatively uncommon given the skills required to formulate accurate and reliable models from which to base decisions supporting strategic enrollment strategies and tactics

(DesJardins, 2002; Goldstein & Katz, 2005). Design and development of predictive models requires a capacity in statistical analysis and associated analytical software, development of database solutions, access to data, and a functional understanding in the use of the model's output.

1.2 Predictive Modeling Defined

Predictive modeling is a statistical method used to predict outcomes based on a specified set of input data. This research is concerned with the use of predictive modeling as a means of supporting decision making among enrollment managers within institutions of higher education. The modeling strategy applied in this study will allow us to examine and identify which of the proposed independent variables help us explain the enrollment behavior of the student admitted at a target institution. It is hypothesized that the application of predictive modeling, utilizing logistic regression, can support institutions of higher education in advancing opportunities for students of color and of low socioeconomic backgrounds.

1.3 The Undergraduate Enrollment Funnel

The role of any Office of Undergraduate Admissions, in the context of higher education and four-year institutions, is primarily one of developing the entering class, while considering the underlying goals of increasing enrollment, and ensuring diversity. The processes managed by an admissions office are often characterized as having a linear progression and described in *stages*, where the institution will invest resources intended to support, assist, or compel the aspiring student to realize one stage in order to progress to the next. The enrollment funnel is the typical representation of this process (see Figure 1).

The enrollment funnel is viewed from the institutional perspective. Segmentation of a target population is based on the individual's present state within a stage of the enrollment funnel. Enrollment managers will refer to the enrollment funnel as a model for guiding marketing activities in an effort to increase yield. Yield is calculated as the number of students matriculated for a given term, over the number admitted for that same term. The model represented in Figure 1.1 demonstrates five stages, however, it is noted that similar models have been found to contain additional stages. Regardless of the stages included in a model, its core function is to support enrollment managers in planning communication and marketing, and to assess operational performance at each stage.

The stages in the enrollment funnel may be defined within a context in which the student has realized an aspiration to attend college. The stages included in this funnel are: Suspect, Inquiry, Applicant, Admitted, and Enrolled. The *Suspect* stage is the set of aspiring students who have a demonstrated interest in going to college, but may not be aware of the target institution and its program offerings. Institutions acquire *Suspect* leads by means of student search service providers.

The *Inquiry* stage is the set of students who have demonstrated interest towards the target institution by means of personal contact with a campus representative, participating in a campus visit, or simply requesting information about educational opportunities at the target institution. The *Applicant* stage consists of the set of students who have submitted the application for admission. Students having a complete admissions file permitting an admissions decision, are considered to be in the Admitted stage. The culmination of resource investment throughout the preceding stages is enrollment. The *Enrolled* student is considered to be admitted and registered for classes.

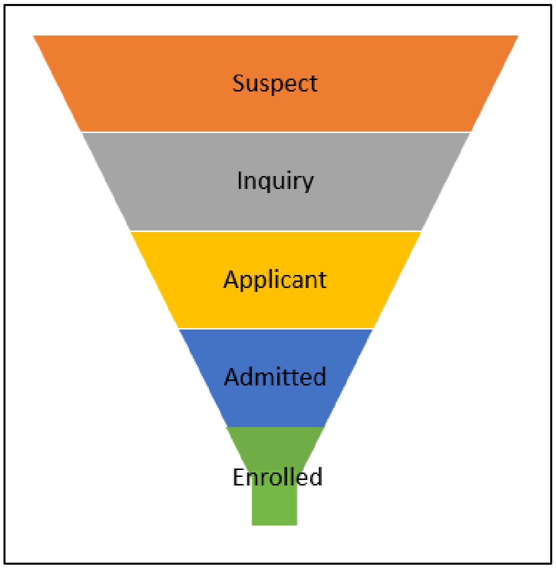

Figure 1. 1 Undergraduate Enrollment Funnel

The IHE's ability to segment a population of students into those outlined by the enrollment funnel, requires a moderate skill set in data analysis and assumes access to data. The institutional capacity to perform the analysis is typically provided by offices of institutional research, information technology teams, and when available, within the enrollment management team.

The segmentation of the prospective student population becomes a critical function for the enrollment management team, as communication strategies are focused on these target populations, with the intention of compelling the prospective student to advance through the stages of the enrollment funnel, and the underlying goal being matriculation. Guiding the student across stages requires the use of marketing strategies that make use of multiple channels of communication; one method is multi-channel marketing (See Figure 1.2). Multi-channel marketing refers to the planning and use of a variety of communication channels to deliver the institutional brand and call to action. This method of marketing involves communication tactics that assume a cost to implement, deploy, and maintain.

For example, a university decides to acquire a listing of students who fit an institutional profile of qualified applicants. The acquisition of the list is through a nationally recognized service provider, at a cost of forty-three cents per name. The university has identified and purchased a list containing 10,000 leads for a total cost of $4,300. Assuming the university plans to mail institutional marketing collateral to each lead, the balance of the work for this target market may include preparation of the dataset, preparation of communication collateral, postage, and service fees if utilizing a mailing services provider. The objective of such a campaign is the conversion and yield of a *suspect* pool. IHE's may also have access to customer relationship management (CRM) systems, permitting the use and management of digital channels for communication. While this may appear to be a viable alternative to direct mail and generation of print collateral, this calculation is not as simple.

CRM involves other investments calling upon a cross-section of institutional resources and functions, including the development of institutional capacity for implementation, maintenance, and use of CRM. In addition, the enrollment manager must be aware of the reach potential inherent in the choice of channel. While the distribution of emails may have a lower cost as compared to direct mail, emails can be filtered, discarded, or remain unopen by the recipient. Industry reports show average email open rates at 30.9% for School and Education, according to industry analysts (IBM Marketing Cloud, 2016).

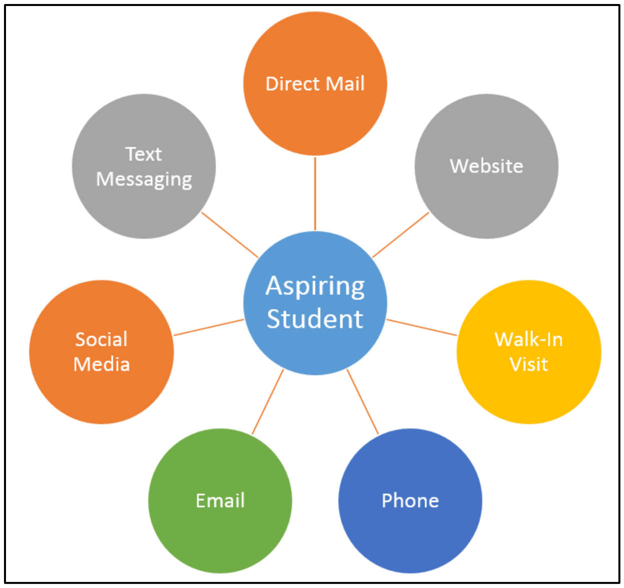

Figure 1. 2 Multi-channel Marketing and Types of Communication Medium

At the same time, direct mail is subject to the family's mobility or mail is simply undeliverable as addressed. In order to achieve enrollment goals, astute enrollment managers understand that effective marketing strategies require the use of many channels to reach their target population "where they are" and on demand. Cost is a significant factor affecting strategic enrollment management strategies, as many IHE's are working within environments having limited resources. Where available, resources are focused on those tactics which are expected to generate a greater yield of enrolled students.

A second example is the investment of resources intended to compel the aspiring student who has inquired at the target institution, to complete the application process. The IHE typically makes an additional investment of resources to drive students in the *Inquiry s*tage to complete the admissions application, and working towards producing a complete admissions file. The progression then shifts towards conversion of applied to admitted, where resources continue to be invested in the conversion process. Once admitted, resources are then focused on enrollment activities, where a concerted effort and further investment of resources is made to increase yield.

In each of these phases, the use of advanced analytics, such as predictive modeling, may have a substantial impact on the IHE's ability to maximize yield, while improving operational efficiency and increasing effectiveness in the use of limited resources.

1.4 Background of the Problem

Institutions of higher education have a need to improve efficiency and effectiveness in recruiting new undergraduate students. This need arises from funding cuts facing public IHEs, increasing accountability and competing demand for limited institutional resources (Mitchell & Leachman, 2015). The investment of resources can be optimized through the use of predictive modeling. For example, the IHE may consider using predictive modeling as a means of determining where it may achieve optimal marginal returns on investment of marketing and operational resources within the *Admitted* stage of the enrollment funnel. However, many IHE's do not have the resources or institutional capacity needed for development and use of predictive modeling, associated analysis, and the ability to understand and use predictive modeling.

The use of predictive modeling to inform higher education enrollment strategies is hindered by a limited capacity among the enrollment management team. While some institutions may choose to contract external resources with expertise in the development and use of predictive models, the cost of such services is often a barrier (DesJardins, 2002). The development and maintenance of models intended to support enrollment management strategies is also hindered by a deficit in the literature explicitly related to model development, analysis and use of output by enrollment managers (DesJardins, 2002; Thomas, Reznik, & Dawes, 1999).

There exists another concern with the literature informing solutions for and use of predictive modeling in strategic enrollment management, inasmuch as researchers tend to propose solutions focused on recruiting the best students. Thomas, Reznik, and Dawes make this point in their research, stating “that focusing recruitment efforts on high-probability “hot prospects” without controlling for students’ academic credentials will not produce the outcome most desired in college admissions (Thomas et al., 1999).” At the same time, DesJardins closely follows this point as he includes student quality as a consideration of enrollment goals (DesJardins, 2002). Indeed, the research that is available to higher education administrators will often focus on acquiring high achieving students (Bruggink & Gambhir, 1996; DesJardins, 2002; Sampath, Flagel, & Figueroa, 2009; Thomas et al., 1999).

Nevertheless, this demonstrates the adaptability of predictive modeling and the importance of the institutional mission and enrollment goals, towards the development and use of a model. It is within this context that this study will focus on the use of predictive modeling to support an increase in yield from the admitted stage to the enrolled stage; this is referred to as a Yield Model. As previously discussed, Yield is calculated as the number of students matriculated for a given term, over the number admitted for that same term. Within the context of the enrollment funnel, a Yield Model attempts to quantify the enrollment rate of those students who are admitted by the target institution.

Adding further context to this study, the research will attempt to predict enrollment out-of-sample, giving enrollment managers a tool to further segment the pool of admitted students in order to improve upon recruitment strategies. The institution selected for this study is a Hispanic serving, Carnegie high-research-activity, urban university, committed to serving a 21st century student demographic. This is a departure from the literature informing the development and use of advanced analytics by IHEs, as population samples used in these studies are largely represented by a White demographic (DesJardins, Dundar, & Hendel, 1999; DesJardins, 2002; Goenner & Pauls, 2006).

1.5 Statement of the Problem

There exists a paradox in the aspirations held by students of color and of low socioeconomic backgrounds, and their attainment of a post-secondary education (Hill & Torres, 2010; Roderick, 2006). It may be stated, *aspiring* to attain a post-secondary education implies a *choice process* in which student characteristics and attributes, as well as institutional characteristics, factor into the student's college choice process. There is a need to understand factors influencing college choice, and the impact these factors have on the college choice process and subsequent participation in post-secondary education, in particular, for graduating high school seniors aspiring to attend college, who are also students of color and of low socioeconomic backgrounds. It is hypothesized that the application of predictive modeling may support IHE's in advancing opportunities for students of color and of low socioeconomic backgrounds, through the examination of key factors and the use of logistic regression analysis.

The set of factors considered in this study are parental educational attainment, average of class size, high school curriculum rigor, student's academic performance (quartiles), average enrollment yield by high school, participation in the compulsory admissions application intake process, college credit earned, participation in Advanced Placement programs, selection of STEM related academic area of interest, and college readiness as determined by Texas Success Initiative assessment results and qualifying exemptions. Analysis of the set of factors will provide a binary result indicating the outcome of the probability the student will enroll at the target institution.

1.6 Purpose of the Study

The interest of this research is to expand the college choice model to include factors related to participation in compulsory college application strategies. *Compulsory college* *application strategies* refer to policy requirements for high school graduation and a collaboration among higher education and high school partners. Some independent school districts require that high school seniors complete a minimum of one admissions application to a postsecondary institution as a condition in satisfying high school graduation requirements. These policies are supported by a collaboration between high school administrators and regional partners in higher education and are intended to facilitate the college application process taking place during the first four months of the high school senior year. The study will add to the body of knowledge, the development and use of a predictive model intended to support the strategic enrollment management function in higher education, by identifying important student factors influencing student choice. This study will focus on the application and use of a predictive model using logistic regression in an effort to inform the types of intervention and marketing strategies, as a means of increasing enrollment of undergraduate freshmen, specifically, students of color and of low socioeconomic backgrounds.

1.7 Research Questions

The Hossler and Gallagher (1987) college choice model identified for this study puts forth a model composed of three phases in which the student must engage if the intent is enrollment in higher education; these phases are: predisposition, search, and choice. The model is considered to be comprehensive in its approach to explaining the college choice process of students aspiring to attend college. However, compulsory college application strategies employed by some school districts may generate a specious representation of participation in the *college choice process*.

The college choice process is important as it affords a student the opportunities to recognize an aspiration to attend college, embark upon a search for and development of a college choice set, and assimilate the acquired knowledge and experience into the selection of a college to attend. The researcher hypothesizes that compulsory college application strategies disable the efficacy of the college choice process, observed as participation in post-secondary education.

In analyzing this issue, the researcher will consider the following research questions:

1. Do factors selected for the regression model, such as *Parental Educational Attainment*, *Average Class Size, Average High School Yield, Diploma Type, Quartiles, Participation in Compulsory Application Intake, Earned College Credit*, *Participation in Advanced Placement programs, STEM related Academic Areas of Interest,* and *TSI College Readiness* have an influence on the decision to enroll at the target institution?
2. Of the proposed set of student factors, which are the most predictive (important) factors of the likelihood for a student to enroll at the target institution?
3. Does the final predictive model for the likelihood for a student to enroll perform similarly using a hold-out data-set, thus suggesting a generalizable model?

1.8 Research Hypothesis

College choice is typically described as a process in which the student must develop an awareness and understanding of factors salient to the student's decision to enroll at a target institution. The following hypothesis is presented:

Student aptitude factors will show to be most important in predicting the likelihood of the student enrolling at the target institution. The set of student aptitude factors considered in this study include *Quartiles, Diploma Code, College Credit Earned,* and participation in *Advanced Placement Programs.*

1.9 Factors Related to Student College Choice

The body of knowledge regarding student college choice is widely represented by the Three-Phase Model of student college choice developed by Hossler and Gallagher. Through this work, Hossler and Gallagher provide the foundation to the often cited three-phase model of student college choice, expounding a purpose of enhancing goals of access and choice, while affecting efficient use of resources. The model draws from earlier research on the subject of student college choice, each providing important insight into the attributes, disposition, activities and decision-making behavior of students as they make decisions on applying to and attending college. The model consists of three-phases: Predisposition, Search, and Choice. Each of these are influenced by individual and organizational factors (Hossler & Gallagher, 1987). This study is primarily concerned with the third phase of the process: Choice.

The literature presents a set of categories intended to frame those factors influencing college choice. These factors relate to student attributes and characteristics, pre-college experiences, and family background. In addition, organizational characteristics become relevant as the pre-college experience is dependent upon the high school a student has attended (Perna, 2006). The selection of variables considered for this study is based on the research presenting the theory of college choice. Data sources contributing to the sample data will include the state's common application for admission and the student's academic achievement record. The state's common application for admission is a state mandated common application form for undergraduate students seeking admission to an institution of higher education. The application, in its current form, is supported by a consortium of public and private IHEs across the state, providing access and opportunity across two and four-year institutions. The application for freshman-level students consists of nine sections. These sections serve to collect important information used by admissions officers in providing an admissions decision. The major sections are biographical information, educational background, self-reported test scores, residency information, extracurricular and volunteer activities, employment information, and institution specific questions. An example of the set of independent variables available through the common application include ethnicity, gender, residency, Family Income, Number in Household, and Family Obligation.

From the major sections available through the application for admissions, we are able to examine elements of the college choice model across the stages in predisposition, search, and choice. This is developed further in Chapter 2.

1.10 Assumptions

This study assumes that data for those variables representing factors influencing college choice are accurately captured through data collection mechanisms such as the common application and the academic achievement record. Another assumption made by the researcher was that the data provided was the most accurate or valid set available at the time of the collection of the dataset. In addition, and although yet unproven, it is assumed that the data set accurately reflects the patterns of enrollment for this particular population of secondary students.

1.11 Delimitations

The study focuses on a single university, with a special emphasis on one category of student ethnic background. Competing theories not considered for this study include Human Capital Theory and Prospect Theory, each positing an economic determination made by the individual, however, not accounting for economic and social stratification (Beattie, 2002; Levy, 1992). Statistical models not considered for this study are multiple regression, probit regression and linear probability models.

1.12 Limitations

Limitations to this study may result from the selection of certain variables originating from the common application as it is lacking in the psychometric measurement of student ability. Examples of such data include results from the College Board's SAT and the ACT exam. While SAT/ACT composite scores may contribute to this study and may be captured in the admissions application process, the collection of this assessment data is considered outside the scope of the data found in ApplyTexas and Academic Achievement Record. Furthermore, the College Board and ACT prohibit third-party sharing of proprietary data.

However, this limitation may be mitigated through the use of other state assessments such as the Texas Success Initiative assessment data available through the academic achievement record. Other limitations are in the institution's decision to exclude demographic variables from the data set. Excluded fields include gender, ethnicity, and high school identifier.

In addition, the present study was able to secure permission from the participating institution for one academic year, 2015-2016.

1.13 Organization of the Study

The study is organized in five chapters. Chapter Two describes the theoretical framework used to explain the college choice process. The Three-Phase College Choice Model is also discussed, highlighting its strengths and its weaknesses. In addition, the chapter will provide a discourse on academic efficiency and effectiveness, providing an understanding of organizational constraints impacting the development of institutional capacity in the use of advanced analytics. Chapter Two concludes with a review of strategic enrollment management (SEM) and the use of advanced analytics supporting SEM strategies. Chapter Three presents the methodology applied in this study, beginning with a description of the population and the setting.

A component of the setting is a presentation of institutional data that demonstrates a need for increased efficiency in enrollment strategies. The chapter then provides a presentation of design and analysis, data collection, and selection of variables. This is followed by a discourse on the logistic regression model design, including assumptions of binomial logistic regression and validation methods. Chapter Four presents results of the binomial logistic regression analysis. The chapter begins with a review and test of assumptions, this is followed by an iterative analysis of utilizing binomial logistic regression and the development of the model's final form. The chapter proceeds with a method for validating the results of the model and its application to out-of-sample data. The chapter concludes with a report of main findings. The study closes with a discussion of results and conclusions in Chapter 5.

1.14 Summary of Chapter

The development and use of advanced analytics by institutions of higher education to support strategic enrollment management strategies is hindered by a lack of information infrastructure and the institutional capacity for use of advanced applications and the information these applications provide. A review of the literature yields few resources informing higher education researchers and administrators in the development and use of predictive modeling to support enrollment management in higher education. Moreover, of the studies that are available, it is found these are typically supported by data largely representing a White demographic. The theoretical framework explicating the study is the Theory of College Choice, typically viewed as a comprehensive model attempting to explain the complexity of the college choice process and the factors impacting the choice decision. This study is intended to add to the body of knowledge regarding the use and development of a theory-based predictive model using binomial logistic regression, with a focus on students of color and of low socioeconomic status.

Chapter 2: Review of the Literature

INTRODUCTION

The literature review supporting this study is intended to provide an understanding of the problem that is the paradox of aspiration to participation, and the use of predictive analytics to support the advancement of participation in post-secondary education by students of color and of low socioeconomic status. The review examines the literature on college choice and Hossler and Gallagher's (1987) foundational work, the Three-Phase College Search model. The research will provide a discourse on contemporary applications of the three-phase model, its deficiencies, and theoretical factors influencing college choice. These factors support the selection of variables hypothesized as affecting the student's decision to matriculate.

The significance of the study is based upon the premise that institutions of higher education will benefit from the use of advanced analytics, however, research shows that IHE's are lagging in the use of advanced analytics (Goldstein & Katz, 2005). The ability of higher education to develop, sustain, and use the resources needed for supporting and furthering the aspiring student's intentions to attend a two-year or four-year institution exists at a basic level. A review of *academic efficiency and effectiveness in higher education* will provide an understanding of Higher Education's capacity to effectively manage sophisticated enrollment strategies, where the utilization of advanced methods in predictive modeling supporting strategic enrollment management plans is atypical.

The review then explores *strategic enrollment management* (SEM), a nascent concept in higher education developed through use of principles of business administration and strongly dependent on business intelligence and analytics. Closely related to SEM is the use of *predictive modeling in higher education* and conceptual frameworks, such as the *Three-Phase College*

Choice Model. The review will present the relationship among each of these and their contribution to the development and use of predictive analytics in higher education.

2.1 College Choice Theory

The body of research on College Choice is well established, building upon sociological perspectives related to status attainment. While status attainment models have been used to explore the development and achievement of educational aspirations, they have not been used to explore the complex process of college choice (Hossler, Schmit, & Vesper, 1999). College choice theory has provided for this gap in the research, offering a well-developed breadth of knowledge to the college choice process.

2.1.1 Chapman's Model on Student College Choice

Are colleges taking adequate steps to positively and proactively influence student college choice? Do IHEs recognize moments of opportunity in which to appropriately engage the aspiring student in order to provide information that goes beyond updating promotional copy and glossy marketing collateral? According to David Chapman, this effort is not sufficient and moreover, contends most university administrators are not aware of the factors influencing college choice (Chapman, 1981). The result is an investment in poor performing communication strategies resulting in a low yield of student enrollment. Chapman stresses those institutions failing to support college recruitment strategies with a systematic model which considers factors understood to influence student choice, may "overlook ways to increase the effectiveness of their recruiting or, conversely, overestimate the influence of recruiting activities in which they do engage" (Chapman, 1981).

Chapman observes, although there exists considerable research towards understanding factors affecting a student's aspiration to pursue post-secondary education, there is a lack of research in student college choice. Chapman's contribution to the body of research is a model of student college choice, qualifying the patterns of influence presented by the model, as limited to traditional age students.

The model presented by Chapman puts forth a set of factors contained within student characteristics and external influences. The set of student characteristics include *social economic status* and *academic aptitude*, while the set of external influences include *significant persons*, *fixed college characteristics*, and *college efforts to communicate with students*. Expanding on both sets of factors, the author presents an argument for information that is timely, relevant, and actionable. Indeed, there are certain factors that are not easily modified or controlled by administrators seeking to influence student choice, such as location, availability of academic programs, and cost of attendance; however, understanding influencing factors and the extent to which they can be utilized to influence student choice is an improvement upon false assumptions at the foundation of prevailing recruitment strategies and practice (Chapman, 1981). Chapman asserts the existence of a disconnect between IHE administrators' understanding of factors influencing the college choice process and the timely, relevant, and actionable information needed by the student to provide for a fully informed decision.

2.1.2 Tierney's Model on Student Choice Clusters

Tierney adds to this argument a notable dimension, the empirical characterization of student college choice sets. The ability to group prospective students according to a corresponding characterization of institution type provides administrators and policy makers an understanding of factors influencing the student's college choice (Tierney, 1983). Tierney's analysis of student's who submitted ACT test scores to one or more colleges provides insight into a preselection process in which students engage at the time they take the college entrance exam. This preselection process considers neither an admissions decision nor an account of financial aid availability for the selected institutions of higher education. Through his study of suburban Philadelphia County students, Tierney posits that disconnects may exist between a process in which the student preselects among those post-secondary educational opportunities that best fit a set of factors to which they subscribe and policies intended to promote equality in educational opportunities, such as admissions and financial aid, as these do not come into effect at the selection stage of the college choice process. Tierney's research provides evidence for the empirical characterization of student college choice sets. Indeed, by forming clusters of institutions by type and using factors such as selectivity, size, cost, and location, Tierney presents several important findings: (1) Self-Selection – a process of self-selection is occurring among test-takers, (2) Timing – students are making decisions based on incomplete information when submitting test score reports, (3) Location – the distance between home and the college matters to students.

2.1.3 Hurtado College Access and Choice Model

Of interest to this study are barriers to access to and participation in higher education, in particular, when considering factors of race and ethnicity. Hurtado, Inkelas, Briggs and Rhee (1997) examine the progress to access and the barriers impeding progress to access to higher education. They contend there is disagreement among researchers on the impact of policies intended to affect participation rates in higher education among ethnic/racial groups, and whether gains were achieved among historically under-represented minorities or whether ground has been lost among these groups. The authors cite a climate of change as a result of a perceived disparity adversely affecting White and Asian students as a result of these policies.

The theory guiding this study is based on the theoretical model advanced by Hossler and Gallagher, positing student college choice is comprised of three stages: predisposition, search, and choice. Each stage contributing to the next, culminating with the student selecting the preferred path and timing for their post-secondary learning. The stages are influenced by “the students’ backgrounds, attributes, activities, and institutional characteristics…” (Hurtado, Inkelas, Briggs, & Rhee, 1997).

The study makes use of the *National Education Longitudinal Study (NELS:89/92)* and the *Beginning Postsecondary Student Longitudinal Study (BPS:90/92)*. The researchers conducted an analysis across racial/ethnic groups in order to understand differences in college choice. Hurtado et al found secondary students are adequately informed about the need for a post-secondary education, however, they contend this is not demonstrated by the student in the course of the college choice process. Key findings disclose the presence of barriers when racial/ethnic and family income are considered. The results show,

> *“In reality, only a small number of African American and Latino students meet the criteria of "equality" along these dimensions necessary for college, and being strategic about educational opportunity is perhaps the only way these few students can succeed. Our results primarily showed that large proportions of African American (45%) and Latino students (47%) do not even apply to college during the 12th grade, nor do approximately one-fifth to over one-quarter among these groups (respectively) who were identified as high achievers on 8th-grade cognitive tests”(Hurtado et al., 1997)*.

Hurtado et al provide that a limitation to the study is the lack of the NELS third follow up study, providing the choice of institution type selected by the student. This may have been the most important piece of this study since it would provide a deeper understanding of the relationship among factors in each stage of the student college choice process. While we understand college choice sets are influenced by predisposition, family income, race/ethnicity, the question remains, "how and to what extent do these factors affect the student's college choice sets and what are the implications for access to college?"

Hurtado et al., (1997) conclude that additional research in the development of models focused on the predisposition phase is needed if we are to gain a better understanding of the student's predisposition, with respect to race/ethnicity.

As previously stated, the body of research on student college choice commonly depicts the college choice process as consisting of three stages, *Pre-disposition*, *Search*, and *Choice*. We have explored factors influencing the college choice process, and reviewed the effects of race and ethnicity on the college choice process. It is necessary to understand the role of habitus and cultural capital on the college choice process in order to develop a deeper understanding of factors influencing student choice, in particular, the psychosocial dimensions influencing choice.

2.1.4 Nora Model – Role of Habitus and Cultural Capital

In the study, *The Role of Habitus and Cultural Capital in Choosing a College, Transitioning from High School to Higher Education, and Persisting in College Among Minority and Nonminority Students*, Nora expands upon the research in college choice with the addition of an analysis of psychosocial dimensions (factors), specifically, habitus and cultural capital. The extent to which these factors influence student college choice is the focus of their research. The conceptual framework for this research proposed that students use other dimensions (psychosocial factors) to choose a college and further states that such dimensions may also be used to understand a student's decision to re-enroll at that same college (Nora, 2004). Through the use of factor analysis, Nora reveals eight habitus factors (Personal Acceptance, Personal and

Social Fit, Academic Interests, Early Influences, Approval by Others, Family Encouragement, Intuition, and Family Expectation) and four cultural capital factors (Academic Self-Esteem, Leadership Experiences, Extrafamilial Encouragement, and Institutional Support) affecting student choice. The results affirm the influence psychosocial factors have on college choice and the contribution these factors have on satisfaction of choice and persistence. Nora asserts the findings reveal that psychosocial dimensions are more influential towards a student's college choice, than are previously established factors, such as high school academic performance, preparation, and experiences, as well as institutional attributes. This is a powerful finding as it reveals a student's leaning towards college choice resulting from psychosocial dimensions are not easily changed. In addition, there exists opportunity in the use of these factors to further understand the student college choice, choice satisfaction, and persistence.

2.1.5 The Three-Phase College Choice Model

The college choice model of interest for this study is the Three-Phase Model developed by Hossler and Gallagher (1987); the model consists of three-phases: Predisposition, Search, and Choice. Each of these phases are influenced by individual and organizational characteristics and attributes. The Three-phase model implies a linear approach to the decision-making process undertaken by students and it provides an understanding of the effect attributes and characteristics bring about within the process, shaping student decision and choice across each phase.

The body of knowledge regarding student college choice is widely represented by the Three-Phase Model of student college choice developed by Hossler and Gallagher. Through this work, Hossler and Gallagher provide the foundation to the often cited three-phase model of student college choice, offering the purpose of enhancing the goals of access and choice, while affecting efficient use of resources. The model draws from earlier research on the subject of student college choice, each providing important insight into the attributes, disposition, activities and decision-making behavior of students as they make decisions on applying to and attending college. The model consists of three-phases: Predisposition, Search, and Choice. Each of these are influenced by individual and organizational factors (Hossler & Gallagher, 1987).

According to Hossler and Gallagher, *the Predisposition phase* is subject to individual factors, such as student characteristics, significant others, and educational activities (Hossler & Gallagher, 1987). In addition, there are influential organizational factors at the pre-college and college level contributing to the student's intent or the extent to which a student aspires to attend college. The set of student characteristics includes background characteristics and the authors assert the appearance of a positive correlation with college attendance and indication of cumulative effects on student college choice (p.210). Hossler and Gallagher also find socioeconomic status (SES) to be one of the most important background characteristics, as students with high SES are more likely to actualize enrollment in college than are students in low SES (P.210). The set of student characteristics also considers student achievement or ability, as a factor that is understood to increase the likelihood of participation in higher education. Research conducted by Litten (1982) explores differences among ability groups, finding that students in higher ability groups begin a formal application process earlier than lower ability groups and apply to more institutions than peers in the lower ability groups (Litten, 1982). These factors are also considered in the Three-Phase Choice Model.

Hossler and Gallagher also consider the influence of parent, peers, and significant others. Noting that parental encouragement and increases the likelihood of going to college and selectivity of college choice in selective, non-selective, or two-year institutions (Hossler & Gallagher, 1987).

The student is expected to achieve an understanding of options for college and options that do not include attending an institution of higher education. The individual factors influencing *the Search phase* are related to the student's preliminary college values and the student's search activities. It is understood, colleges and universities may influence the student's search in this phase, by means of institutional student search activities directed at the student. The student is expected to arrive at a set of choices for college, or other options, based on the comprehensive set of influencing factors.

The culmination of the process for predisposition and search is the development of a college choice set. This becomes the third phase of the model in which the student has considered previous influencing factors, both personal and organizational, and choses a college to attend (Hossler & Gallagher, 1987).

2.2 A Critique of the Three-Phase College Choice Model

An underlying assumption to the Hossler and Gallagher (1987) college choice model is that students must have access to information in order to make informed decisions throughout the college choice process. The student must possess the capacity to acquire relevant information in a timely manner, furthermore, the student must also have the ability to process said information in order to proceed through a college bound pathway. This is the premise of the Three-Phase College Choice model, and although the model is considered to be comprehensive, a common criticism arising from the literature is that the model fails to consider other factors affecting college choice for students of color and of low socioeconomic status (Bergerson, 2009).

Bergerson finds that the underlying assumptions associated with the comprehensive college choice process often described in research literature, tend to favor students of high socioeconomic backgrounds (p. 35). As such, Bergerson presents a common point in the literature critiquing the college choice model; it fails to consider the limitations associated with social class, such as access to information, technology resources, or systemic inequities associated with race, gender, or immigration status.

It is on this point that we turn to the hypothesized use of predictive modeling to improve upon our understanding of a predominantly Hispanic population and the likelihood that the admitted student will enroll at the target institution, given the insightful contribution provided by the analysis of the model's output, and giving enrollment managers an improved toolkit for anticipating the outcome of the admitted student's choice.

While the model provides a foundation for the development and use of variables that may be considered in a logistic regression model, an accounting for the experiences of students of color and of low socioeconomic status remains unclear; from a marketing perspective, college choice literature is lacking information on factors impacting consumer behavior (Pitre, Johnson, & Pitre, 2006).

2.3 Rationale for the Use of the Three-Phase College Choice Model

The Three-Phase Model provides practitioners and researchers a sound starting point for understanding the student college choice process, however, the need for additional research in the area of pre-disposition is clear, as the issue of predisposition is compounded by race/ethnicity, socio-economic status, and factors associated with concerted cultivation and the accomplishment of natural growth.

To improve upon efficient and effective use of resources intended to increase participation in higher education, there was also a desire to maximize the use of extant data available through the admissions application (ApplyTexas) and the student's academic achievement record. The data captured through these resources are standardized and consistent in method of collection, and factors are understood to influence the college choice process across each stage, where some factors are seen to have a cumulative effect across stages and overall plans to enroll in college. Table 2.1 shows factors available in the data and understood to influence college choice across stages.

Table 2.1 Factors Considered in the Three-Phase College Choice Model

Predisposition	**Search**	**Choice**
HS or College Characteristics	**Organizational Characteristics**	**Organizational Characteristics**
High School Class Size	Admissions Requirements	Compulsory Application
High School Average Yield	**Pre-College Experience**	Intake Academic Area of Interest
Significant Others	Quartile	**Significant Others**
Parental Educational Attainment	Academic Rigor	Parental Educational Attainment
Pre-College Experience	Advanced Placement	**Pre-College Experience**
Advanced Placement	College Readiness	College Readiness
College Credit Earned	**Significant Others**	
	Parental Educational Attainment	

The factors selected for the study were limited by institutional decisions that excluded demographic and socioeconomic factors that may have contributed a greater understanding of effects associated with habitus, social capital, and socioeconomic advantages. The Three-Phase College Choice Model provides a fundamental framework for examining a choice process, and although the model can be improved upon by means of an integrated conceptual modeling approach as proposed by Perna (2006), the limitations of the study may set the stage for future integration of other theoretical constructs.

2.4 Academic Efficiency and Effectiveness

Since the early 1900's the concepts of *academic efficiency and effectiveness* in higher education have received mixed views with regard to definition, conceptualization, and application (Kent, 1912; Lindsay, 1982; McEwen & Synakowski, 1954). To understand the importance of continuity and intent associated with these concepts, we find a resounding message in the words of William Kent stated in 1912:

"Our modern educational literature, addresses of college presidents, school superintendents, proceedings of societies, etc., all show the prevailing consensus of opinion that there is something seriously wrong with our whole educational system, and that instead of getting better it is constantly tending to grow worse. There exists also a great amount of ultra-conservatism and of mental inertia relating to the subject. It is high time that something practical be done in the way of reform (Kent, 1912)."

The overarching argument, with respect to efficiency and effectiveness, is that higher education is fundamentally lacking in "a precise conceptualization for 'institutional performance'" (Lindsay, 1982). It stands to reason, failure to conceptualize models to measure institutional performance implies a failure in capacity for informed decision making.

Indeed, the concept of institutional performance, as it relates to efficiency in higher education, is both vast and complex, given an institution's mission & goals, demands and expectations from internal and external stakeholders, and the ever changing political and financial landscape of higher education. Lindsay contends there is no lack of strategies and methods intended to provide a structured and guided approach to evaluating institutional performance in areas of efficiency, effectiveness, and quality. However, a mindset of "inappropriateness" persists among some educators who view efficiency and effectiveness simply as a preventable notion (Lindsay, 1982).

The need for thoughtful and purposeful conceptualization of institutional performance evaluation methods, intended to support decision making, is without question an ever-increasing requirement placed upon administrators in higher education. As the demand on limited state resources continue to compel state governments to shift funds away from higher education, the burden of higher education costs continue to weigh upon students and their families (SHEEO, 2016); this is further compounded by the continuously increasing costs of higher education. In addition, the continued shift in momentum for the stewardship of public funds also places new exigencies for accountability from public institutions. This includes accountability for reducing costs, improving completion rates, and generally improving upon institutional efficiency and effectiveness.

The demand for accountability in higher education is stressed through legislative mandates where we find a push for data collection that may appear to support the examination of institutional performance, but may not be entirely representative of data needed to support an institution' s mission and goals. A legislative attempt to impose efficiency and effectiveness in an overall endeavor to improve institutional performance is likely to achieve the desired result in very specific areas, however, as Lindsay explains, the use of "simple measures" are insufficient in capturing "the intangibility of the multiple objectives and outputs in education and the different values placed on them by people with differing views" (Lindsay, 1982; p. 179). We find an example of this practice in Texas higher education.

The state of Texas has created the infrastructure for coordination and reporting of institutional data from state institutions of higher education. The collection of data, through the Texas Higher Education Coordinating Board (Coordinating Board), provides for reporting on matters such as student enrollment, college readiness, course inventory, class detail, building and room assignment, faculty assignment, graduation, facilities room inventory, and admissions. These reports are utilized, for the most part, to support determination of funding allocations based on legislative mandates (Texas Higher Education Coordinating Board, 2017). The reports also serve in tracking performance in areas that are viewed as critical to the state's mission for higher education, as adopted in the strategic plan known as "Closing the Gaps" (Texas Higher Education Coordinating Board, 2016b) and most recently, the 60x30TX strategic plan (Texas Higher Education Coordinating Board, 2016a).

Since its adoption by the Coordinating Board in October of 2000, the *Closing the Gaps* strategic plan has focused on closing gaps in participation in higher education, student success, excellence in public education, and research (Texas Higher Education Coordinating Board, 2016b). A primary goal for the *Closing the Gaps* strategic plan was to increase participation in higher education, as it is widely considered to be a means of attaining the benefits of educational advancement and avoiding the adverse effect of a low-skilled workforce, and maintaining the state's capacity to sustain a growing economy (THECB, 2012). As a result, both state and federal governments have made significant policy mandates intended to increase access to higher education and promote progress towards degree completion. These policies are centered on affordability, providing additional funding to shore grants and scholarships in an effort to cover tuition, fees and books, while establishing incentives intended to promote a culture of academic and administrative efficiencies among institutions of higher education (THECB, 2012).

The *60x30TX* higher education strategic plan, adopted by the Texas Higher Education Coordinating Board in July 2015, follows the work of the *Closing the Gaps* strategic plan, continuing the effort of participation in higher education, with a focus on degree completion while emphasizing the value of higher education in the workforce (Texas Higher Education Coordinating Board, 2016a). Central to the plan is the Texas Pathways Model.

The Texas Pathways Model is described as a system-wide approach, integrating state and institutional policies, and K-16 education partner strategies and goals for student success. Texas Pathways seeks to increase attainment rates across Texas while supporting the state's higher education strategic plan, 60X30TX, calling for 60% attainment among Texans who are of 25 to 34 years of age, by 2030 (Texas Association of Community Colleges, 2017). While data collected by the THECB may be sufficient for calculating *Student Success Points* in order to award and appropriate funds based on institutional performance, it is insufficient for understanding the student's intentions or salient beliefs influencing the student's behavior, such as enrolling, persisting, and completing a post-secondary education. From an enrollment management perspective, the Coordinating Board data is important and relevant, however, not entirely indicative of performance in specific areas of interest to administrators responsible for student success and enrollment management.

At the same time, data associated with Coordinating Board reports do not keep pace with the needs of enrollment managers working in the changing landscape of higher education. The range of enrollment management performance measures and benchmarks needed to adequately direct institutional enrollment strategies subsume functions, resources, and processes across an institution's academic and service units. This researcher posits, the need for in-depth data collection, analysis, and development of business intelligence outpaces an institution's capacity to respond to public demand for academic efficiency and effectiveness. The use of business intelligence and business analytics to support decision making and resource management in higher education remains a challenge as institutions work towards developing an information infrastructure and the institutional capacity for use of advanced applications and associated management information (Goldstein & Katz, 2005).

A study conducted by the Educause Center for Applied Research highlights 380 institutions and their use of academic analytics; such use is categorized in *five stages of application*. The five stages of application include: (1) Extraction and reporting of transaction-level data; (2) Analysis and monitoring of operational performance; (3) "What-If" decision support (e.g., scenario building); (4) Predictive modeling and simulation; and (5) Automation triggers of business processes (e.g., alerts) (Goldstein & Katz, 2005). Table 2.2 summarizes the application of academic analytics by functional area and stage (Goldstein & Katz, 2005). The advancement of an institution's proficiency and use of academic analytics across these stages is an indication of the institution's development of information infrastructure and capacity for use of advanced applications. The results of the study reveal one remarkable point – a majority of institutions surveyed (70 percent) indicate the use of academic analytics was primarily in the first stage – operational or transactional reporting.

Table 2. 2 Primary Application of Academic Analytics, by Functional Area

Use	Advancement & Fund-raising	Business & Finance	Budget & Planning	Institutional Research	Human Resource	Research Administration	Academic Affairs
Stage 1: Extraction and reporting of transaction-level data	56.9%	68.4%	49.6%	48.8%	62.2%	45.0%	52.8%
Stage 2: Analysis and monitoring of operational performance	11.0%	17.0%	19.6%	28.4%	7.8%	10.3%	18.2%
Stage 3: "What-If" decision support (e.g., scenario building)	2.3%	1.9%	13.5%	4.1%	0.6%	0.9%	4.7%
Stage 4: Predictive modeling and simulation	3.1%	3.0%	9.6%	11.6%	1.1%	1.7%	5.2%
Stage 5: Automation triggers of business processes (e.g., alerts)	3.7%	2.5%	0.6%	7.1%	1.9%	1.1%	2.2%
Not active users	22.9%	7.1%	7.2%	0.0%	26.4%	41.0%	16.9%
Total	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %	100.0 %

Source: Academic Analytics: The Uses of Management Information and Technology in Higher Education; (Goldstein & Katz, 2005).

2.5 Strategic Enrollment Management

Strategic Enrollment Management (SEM) is concerned with the underlying goals associated with promoting student success and completion. Factors driving the success of these goals are best described as a system of dimensionalities related to integrated business planning within the context of the educational environment (Black, 2008). Black, strategic enrollment management consultant and practitioner, argues that institutions of higher education are in fact organizational systems that necessarily have integrated, but individual, components best managed as a “cohesive whole” (Black, 2008). Student success is not dependent on a single stakeholder or a unilaterally determined set of enrollment goals. Driving the successful management of the student lifecycle requires participation from many actors who are sufficiently informed through data in order to support operational areas affecting new student enrollment, retention, and completion.

SEM provides the foundation from which we gain an understanding about the complexity in managing enrollment in higher education and the need for informed decision making. SEM is intended to address matters of efficiency and effectiveness, but assumes an institutional capacity possessing knowledge and experience with SEM, and does not account for lacking or diminishing resources needed for the successful development and implementation of the institutional SEM plan. We find that the successful implementation of strategic enrollment management relies upon an investment of resources intended for innovative changes supporting the work of enrollment management teams (Black, 2008; Bontrager, 2004; Langston & Scheid, 2014).

However, Institutions of Higher Education are facing fiscal challenges further imposing upon the need for increased efficiency and effectiveness in order to realize, let alone maximize, the benefits of limited institutional resources (Langston & Scheid, 2014). Among the various resources utilized by an IHE to ensure student success, those invested in supporting the Strategic Enrollment Management function are paramount. Of particular importance are the investment of resources and work undertaken by offices of undergraduate admissions. The work of constructing strategies and tactics intended to deliver increased enrollment, while balancing issues of diversity, access, and academic excellence begins with Admissions. A dominating concern for the enrollment management professional is the maximization of tuition revenue through recruitment (new enrollment) and retention.

The range of enrollment management performance measures and benchmarks needed to adequately direct institutional enrollment strategies subsume functions, resources, and processes across an institution's academic and service units. The use of business intelligence and business analytics to support decision making and resource management in higher education remains a challenge as institutions work towards developing an information infrastructure and the institutional capacity for use of advanced applications and useful management information (Goldstein & Katz, 2005).

2.6 Predictive Modeling in Higher Education

A search of the ERIC, EBSCO Information Services, and Web of Science yields few results in the literature informing higher education researchers and administrators in the development and use of predictive modeling to support enrollment management in higher education. Thomas et al (1999) provide a similar assertion, and are among the few contributors to this aspect of the literature in the study, *Using Predictive Modeling to Target Recruitment: Theory and Practice* (Thomas et al., 1999).

2.7 Types of Models

The studies referenced in this section of the literature review may be categorized by the (Chapman, 1981)intended use for informing enrollment management decisions at various stages within the enrollment funnel. For example, enrollment managers may have a desire for improving the effectiveness of the institution's investment in suspect lists purchased from student search service provider; this model type is referred to as an Inquiry Model. Similarly following progression through the enrollment funnel, we also have the Applicant Model as well as the Admitted Model. The determinants for model selection are likely a matter of institutional goals, such as maximizing investment on suspect leads or increasing Yield for the benefit of college rankings. However, this assumes an institutional capacity favoring the use of advanced analytics.

While each model type provides a benefit for decision makers by answering the question, “will the student enroll at the target institution?” There are considerations for the limitations imposed by each model. The Inquiry Model, for example, is intended to predict enrollment from the suspect stage. In many cases, the limitations are in the data available for use in the model. This is typically due to institutional recruitment practices typically resulting in the collection of names, academic area of interest, high school, contact information, and referral source. Goenner and Pauls (2006) point out that such limitations in the data requires additional variables such as academic achievement, socioeconomic status, and IHE characteristic preferences. Such limitations have expanded data considerations for the application of geodemography (Goenner & Pauls, 2006). The Applicant Model follows the progression of the enrollment funnel, with an interest in predicting if the applicant will enroll at the target institution. This model uses data considered to be more robust as compared data available in the Inquiry Model. This is due to availability of data captured through the application process, which may include student ability data. The Yield Model is perhaps the model most commonly used by enrollment managers. The Yield Model is typically used to score the admitted student’s probability of enrolling at the target institution. The data available to support this model is considered to be the most complete, as the student’s application for admission will very likely contain demographic, socioeconomic, and pre-college experience factors taken from the admissions application as well as the application for financial aid.

2.8 Theoretical Considerations Across Model Types

The use of theoretical frameworks across model types is consistently driven by the researcher's desire to improve our understanding of the college choice process and the complexity in predicting behavior based on limited empirical data (González & DesJardins, 2002). As a result, we find researchers explaining student choice through the various lenses such as Human Capital Theory, Utility Theory, Gravity Model, Theory of Planned Behavior, Social Capital Theory, or Cultural Capital Theory (Bruggink & Gambhir, 1996; DesJardins, 2002; Goenner & Pauls, 2006; González & DesJardins, 2002; Leppel, 1993; Thomas et al., 1999). This continues to be a response to the frequent and common criticism regarding the comprehensive Student Choice Model – the theoretical model cannot be applied to populations of students of color or of low socioeconomic status. This highlights the need to consider the stratification within the higher education system and the impact of those factors at the root of stratification (Bergerson, 2009).

2.9 Statistical Methods Across Model Types

Logistic regression is the prevalent statistical method used in predictive modeling across the model types previously discussed. This is explained by the appropriate use of this technique given that each study is premised on whether the predictive model provides a discrete outcome suggesting group membership for a given set of predictor variables (Tabachnick & Fidell, 2013); determining whether a student will enroll or will not enroll at the target institution. Other statistical methods found in the literature apply more complicated techniques such as Bayesian Model Averaging or the non-traditional Artificial Neural Network technique (Goenner & Pauls, 2006; González & DesJardins, 2002). However, if an objective for undertaking the research is to support enrollment managers in development and use of predictive models in enrollment management, then these complex methods may become unattainable.

2.10 Summary of Chapter

This chapter attempts to present several interconnected elements, which are believed to impose upon the use and development of predictive models by institutions of higher education. College Choice Theory provides the lens for theory-based model development. In addition, it provides the foundation from which researchers attempt to build upon a comprehensive college choice model, explaining the complexity in the college choice process. While the principles of strategic enrollment management are becoming a more prevalent across IHEs, there challenges hindering the use of advanced analytics by IHEs.

Chapter 3: Methodology

The central goal of this study is to develop a predictive model intended to support enrollment managers in optimizing the use of resources in their effort to achieve institutional enrollment goals. Through the application of logistic regression, the use of predictors is intended to inform intervention and marketing strategies, as a means of increasing enrollment of undergraduate freshmen, specifically, students of color and of low socioeconomic background. This section will establish the setting for the study, the population considered for analysis, a description of the acquisition of data, an explanation of the study's design, concluding with considerations for the application of logistic regression in the analysis. The primary research questions focus on the identification of a set of predictors that best contribute to an understanding of student choice factors and the student decision making process when selecting and actualizing enrollment in a college.

3.1 SETTING

The setting for this study is a Hispanic Serving Institution (HSI) with a Carnegie Research 1 designation for very high research activity. The university is located in the southwest region of Texas. The region has more than 837,000 constituents and ranks 8th among the most populous counties in the state. In addition, the region averages nearly 12,000 graduating high school seniors each year since 2011. Average participation rates in higher education for regional graduating high school students is approximately 49% since 2011. The university matriculates on average, 19% of the region's high school graduating class. These are members of the graduating class who enroll the first fall semester after graduating from high school. The two-year community college, matriculates on average, 23% of the regional high school graduating class. The data suggest a propensity among regional graduating high school seniors to enroll in regional institutions. The university is one of two public IHEs in the region and the only four-year public institution of higher education serving the region. Total enrollment for the university has grown over the past decade from 18,918 in the fall of 2004 to 23,079 in the fall of 2014, representing an increase in total enrollment of 22% over this period. Examining enrollment by level, as shown in Table 3.1, we find the greatest increase is at the undergraduate level with a growth rate of 27% over this period.

Table 3. 1 Change in Enrollment for the Period Fall 2004 through Fall 2014

	Fall 2004			**Fall 2014**		
Level	**Female**	**Male**	**Level Total**	**Female**	**Male**	**Level Total**
Undergraduate	8,467	7,125	15,592	10,611	9,206	19,727
Graduate	1,962	1,364	3,326	1,846	1,416	3,262
Total	10,429	8,489	18,918	12,457	10,622	23,079

Note: Data acquired from the institution's Common Data Set.

At the undergraduate level, the growth rate is greater for males at 29%, followed closely by a 25% growth rate for females. Graduate enrollment appears to have experienced a drop in growth, dipping to a -2% growth rate over the same period. Current enrollment data for the university demonstrate high application counts and admission rates, however, yield rate (admitted stage to enrolled stage) shows a five-year average of 46%. Table 3.2 shows applicant count, conversion, and yield for freshman applicants for the period fall 2013 through fall 2016. The university's student population is 83% Hispanic, 6.3% White, and 2.5% African American. This is highly representative of the population in the region, where 83% of the population is Hispanic, 12% White, and 3% African American.

Table 3. 2 Undergraduate Applicant Yield Report

Status	2016	2015	2014	2013
Applied	15,204	14,086	13,221	12,603
Admitted	7,183	6,877	6,742	6,636
Enrolled	3,286	3,156	3,006	3,102
Applicant Conversion	47%	49%	51%	53%
Admitted Yield	46%	46%	45%	47%

Data Source: Office of Institutional Research Fact book and applicant dashboard;

3.2 Population and Sampling Plan

The population in this study is the high school senior graduating from a regional high school and is part of the Class of 2016. The student must have applied for admission to the target institution and must be admitted to be included in the sample. The composition of school districts in the region is in public, private, parochial, and non-traditional education systems. Nearly 94% of the entering student population indicate a high school of origin that is in urban and rural independent school districts serving the region. Females make up 54% of the population in the applicant and accepted pool of students. Hispanics represent the majority (89%) in ethnicity, with the remaining population in White (5%), African American (3%), Asian (1%), and Other (2%). Selection criteria will include all students from the set of high school in the region and who are graduating from high school in the period between December 2015 and August 2016.

3.2.1 TAC §74.11. High School Graduation Requirements

A component of the Texas Administrative Code (TAC) is Subchapter B, §74.11. High School Graduation Requirements. The administrative code provides a clear definition of requirements that must be completed by the student. These requirements include the completion of the Foundation High School Program, state assessments, and demonstrated proficiency as determined by the district. While TAC §74.11 does not include a requirement for completing an admissions application to a post-secondary institution, a majority of applications submitted and subsequently used in this study are considered part of a compulsory admissions application collection process that is driven through collaborative efforts between regional high schools and IHEs.

3.2.2 ApplyTexas and Compulsory Application for Admission

Regional high schools participate through a coordinated effort to provide graduating seniors support to complete and submit an application for admission to an institution of higher education. The Texas common application, known as ApplyTexas, is the mechanism for online collection and submission of the admissions application detail. Since this effort is coordinated and facilitated by the regional university and community college, both selection and direction tend to focus on these same institutions. Applications for admission used in the study will be selected from the set of students completing the application for admission during the period bound by the date ApplyTexas makes the application available for service (i.e., August 1, 2015) and continuing through the first week of classes (i.e., September 2016). In addition, only those records for students graduating from regional high schools are included.

3.2.3 Members in the Population

This study will consider only graduating high school seniors who are also first-time in college. The cohort used for this study graduated high school in the fall of 2016. The cohort is limited to regional graduating high school seniors admitted to a regional Hispanic serving institution, for the enrollment term being the first fall semester after graduating from high school (fall 2016).

3.2.4 Selection of the Population

Regional compulsory college application strategies used by the university result in large applicant pools with applied to admitted conversion rates declining as pool size increases as previously shown in Table 3.2. A total of 7,183 records are in the data set provided by the institution of study.

A consideration for the sampling plan is the selection of developmental and validation samples to support model validation, as further described in section 3.4, Model Validation. A simple random sample without replacement is used for the selection of both developmental and validation samples. As described by Hosmer et al, a simple random sample of cases are selected and covariate values are determined from the sample (Hosmer, Lemeshow, & Sturdivant, 2013; Menard, 2010). Table 3.3 shows the count in the developmental and validation data sets.

Table 3. 3 Frequencies for Developmental – Validation Data Sets

Sample	Frequency	Percent	Valid Percent	Cumulative Percent
Developmental	3,645	50.7	50.7	50.7
Validation	3,538	49.3	49.3	100
Total	7,183	100	100	

3.3 Data Collection

The data used for this study is extant institutional data collected through the institution's admissions application process. The data set consists of undergraduate applicants who were admitted for the fall 2016 term. The data set captures variables found in the ApplyTexas Application and the student's Academic Achievement Record (AAR). Upon securing Institutional Review Board approval, data sets were acquired through institution's office of institutional research. The data is maintained on university authorized data servers, secured with encryption technology maintained by the institution's Information Technology team.

3.3.1 Compulsory Application Intake

Facilitating the completion and submission of an online admissions application requires the consideration for the use of resources at the location of the activity. These resources include but are not limited to: counselor, faculty and student time and effort; the use of the high school computer labs; and use of instruction time. This activity also includes the use of resources by teams of university and college admissions officers and recruiters as they coordinate the intake schedule with high school counselors. The intake schedule typically starts in August approaching the fall term and concludes in December, with limited activity extending into January of the new calendar year.

3.3.2 Rolling Admissions Policy

The university's policy for admissions permits the submission of an undergraduate admissions application for a fall, spring or summer start term. In addition to receiving applications from forthcoming graduating high school seniors, applications are also received from other student types, such as: returning students, applicants for second bachelors, transfer students, and non-traditional first-time in college students. This study will consider only graduating high school seniors who are also first-time in college. The cohort used for this study graduated high school in the spring of 2016. The cohort is limited to the set of regional graduating high school seniors admitted to the institution for the enrollment term being the first fall semester after graduating from high school (fall 2016). A total of 7,183 records are in this data set.

3.4 Research Design

This study will apply a binomial logistic regression research design as a means of understanding those factors hypothesized as related to college choice. Since it is our desire to predict a discrete outcome, that is, whether a student is enrolled or the student is not enrolled, then binomial logistic regression is the appropriate statistical technique considered for this study. Logistic regression permits the use of a mixed-set of independent variables (IV) that may include dichotomous, discrete, or continuous variables. This is anticipated for the variables selected for this study. It is important to note several assumptions regarding logistic regression that differ from linear regression and general linear models. Logistic regression does not assume a linear relationship between the dependent variable and the independent variable. In addition, a dichotomous dependent variable is assumed. Furthermore, logistic regression assumes that the probability of an event occurring is, $P = Y(1)$.

As previously stated, the researcher will consider the following research questions:

1. Do factors selected for the regression model, such as *Parental Educational Attainment*, *Average Class Size, Average High School Yield, Diploma Type, Quartiles, Participation in Compulsory Application Intake, Earned College Credit*, *Participation in Advanced Placement programs, STEM related Academic Areas of Interest,* and *TSI College Readiness* have an influence on the decision to enroll at the target institution?
2. Of the proposed set of student factors, which are the most predictive (important) factors of the likelihood for a student to enroll at the target institution?
3. Does the final predictive model for the likelihood for a student to enroll perform similarly using a hold-out data-set, thus suggesting a generalizable model?

3.5 Selection of the Study Variables

The data utilized for this study are derived from two sources considered part of the admissions application process. Sources for the secondary data are the admissions application and the student's individual academic achievement record. While other variables may be available through the student information system, these were selected for the consistency in delivery, that is, all admitted students must submit the state's common application for admission and the state's academic achievement record. As previously stated, the purpose of this study is to provide enrollment managers a method of developing an analytic tool utilizing the data that is available from the admissions application process.

3.5.1 Categorization of Variables

Family Background defines a set of characteristics about the student's *family background* and hypothesized to impact student choice. Variables in these categories include both *father* and *mother's educational attainment*.

The set of variables in the Pre-College Experience category include factors related to the student's readiness for college as defined by their academic achievement through pre-college experiences. Variables used in the study include *high school grade point average*, *high school rank*, *high school program rigor (diploma code), earned dual credit*, *earned college credit, Texas Success Initiative assessment outcomes*, *Advanced Placement* program participation, and participation in regional admissions application intake programming.

High School characteristics are representative of the external detail related to the high school attended by the student. Organizational characteristics such as social composition, quality, curriculum and programs, and resource availability are understood to influence student college choice (Kidd, 2016; Vrontis, Thrassou, & Melanthiou, 2007). Variables included in this category are *Average High School Yield* and *School Class Size*.

Educational Aspiration attributes are also included in the set of independent variables. These are defined as factors related to the student's academic area of interest. The variable included in this category is STEM.

3.5.2 Description of the Study Variables

The variables considered for this study are presented in Table 3.4. The discrete dependent variable (DV) will indicate whether the student has enrolled at the target institution or has not enrolled at the target institution. These are coded as Enrolled = 1, Not Enrolled = 0. The selected variables, hypothesized to affect student choice, are selected from the data sources previously mentioned.

Table 3. 4 Description of the Variables in the Study

Variable Name	**Description**
Enrolled	DV - Student Enrolled (0 = Did not enroll, 1 = enrolled)
pQuartile	Student's rank position in quartiles determined from high school percentile;
AvgOfClassSize	Graduating class size average for student's campus;
DiplomaCode	Categories are in (0 = Distinguished, 1 = Recommended)
hsAverageYield	Percent yield (admitted to enrolled) from specified high school;
ClgRdy	College Ready; ((0 = No, 1 = Yes)
CompulsoryIntake	Submitted admissions application during compulsory application intake period (0 = No, 1 = Yes)
ColCrdEarned	Earned college credit, any program (0 = No, 1 = Yes)
FEdAttain	Father's educational attainment (1 = College-Beyond, 2 = High School, 3 = Middle School, 4 = Other)
MEdAttain	Mother's educational attainment (1 = College-Beyond, 2 = High School, 3 = Middle School, 4 = Other)
AP	Participation in Advanced Placement Program (0 = No, 1 = Yes)
STEM	Academic area of interest in STEM program (0 = No, 1 = Yes)
TsiOvrall ColRdy	TSI overall college ready satisfied (0 = No, 1 = Yes)
Tsimath ColRdy	TSI math college ready satisfied (0 = No, 1 = Yes)
Tsiwriting ColRdy	TSI writing college ready satisfied (0 = No, 1 = Yes)
TsireadColRdy	TSI reading college ready satisfied (0 = No, 1 = Yes)

Percentile (Quartiles) is recorded from the AAR. In accordance with the Texas Education Agency's minimum standards for the AAR, percentile (rank) is subject to change, however, data should be maintained as accurately as possible at all times. Rank and class size determine student percentile. Percentile is a determinant for admission at many IHEs and a determinant for automatic admission to publicly funded Texas IHEs for students graduating from a Texas high school and who are in the top ten percent of their graduating class. Percentile (rank) is a metric by which a student will self-assess if they are a fit for a choice-list institution (Chapman, 1981; Nora, 2004). **Quartiles** (pQuartile) is a calculated field derived from *percentile* which is recorded from the academic achievement record (AAR), Percentile is associated with academic performance as determined by the student's grade point average against other members of the graduating class. GPA has been found to have a strong influence on the student's aspiration to participate in higher education (Chapman, 1981; Hossler & Stage, 1992). GPA, therefore rank and percentile, can impose limits or provide opportunities in the selection of institutions considered by the student.

Average Class Size (AvgOfClassSize) is recorded from the AAR. The data in high school class size is normalized to provide the mean class size for each student within a given high school code. High school characteristics are in the set of factors that influence the decision to enroll (Hossler & Stage, 1992). This considers the hypothesis that resource rich organizations have a capacity to provide "superior intellectual and material resources (Paulsen, 1990)."

Diploma Code (DiplomaCode) is recorded from the AAR. All Texas high school graduating seniors complete a curriculum with rigor that is categorized as Minimum High School Program, Recommended High School Program, and Distinguished Achievement Program. Each category is differentiated by the rigor in curriculum, with the Recommended High School

Program and Distinguished Achievement Program considered to be college preparatory programs. Curriculum is understood to influence the college choice process, particularly in the decision to apply and enroll to a target institution (DesJardins, Ahlburg, & McCall, 2006; Hossler & Stage, 1992; Paulsen, 1990).

High School Average Yield (hsAverageYield) is a calculated field determined by the historic average yield for each high school considered in the study. Viewed as an organizational characteristic, high schools with historically high admitted to enrolled yield are hypothesized to have greater odds of enrollment than students from high schools with low historic yield (DesJardins, 2002).

College Ready (ClgRdy) is a calculated field determined by the target institution from a set of criteria that may be applied in qualifying a student as "college ready" in accordance with Texas College Readiness standards. The standards are intended to align K-12 and higher education curriculum, and provide for a seamless transition from high school to college, by "articulating a baseline knowledge" needed for successful participation in college (Texas Higher Education Coordinating Board, 2009).

Compulsory Application Intake (CompulsoryIntake) is calculated from the application submission date against the institution's census day. Public IHEs are required to submit institutional enrollment data captured on the 12^{th} class day of the fall and spring terms. Census day was selected as students are allowed to enroll in the period after classes begin and leading to census. There is interest in understanding if participation in compulsory application intake programming yields an effect on the decision to enroll.

College Credit Earned (ColCrdEarned) is recorded from ApplyTexas, the state's common application for admission to state IHEs. This is self-reported data that is further validated by the student's academic transcript. Participation in curriculum contributing towards college credit such as Advanced Placement, Dual Credit, or other credit by exam is seen as building upon the student's awareness and understanding of the expectations in the level of knowledge and skills required to succeed in college (Adelman, 2006; An, 2013).

Parental Educational Attainment (FEdAttain, MEdAttain) is recorded from ApplyTexas. The self-reported data records both mother and father's educational attainment with categories in "College-Beyond", "High School", "Middle School", and "Other". Parental educational attainment is considered an important factor impacting student choice. According to Hossler et al, parental education has a direct effect on developing the student's aspirations for participation in higher education and has a greater impact on the student's decision to enroll in college (Hossler et al., 1999).

Advanced Placement (AP) participation is recorded from the AAR. Considered a factor related to the student's academic preparation, college enrollments are expected to be greater for students who participate in Advanced Placement programs (Perna, 2006).

Science, Technology, Engineering, and Math (STEM) is recorded from ApplyTexas as academic area of interest and is coded from the student's selection of the "college" offering the curriculum in the academic area of interest. Considered a factor related to self-efficacy within the context of pre-college experience and academic preparation, specifically readiness in math and science, the actualization of choice has been found to be influenced by the student's personal expectations for performance in the academic area of interest (Wang, 2012).

Texas Success Initiative (TSI) variables are considered for math, writing, reading, and overall status. The variables (TsiOvrall ColRdy, Tsimath ColRdy, Tsiwriting ColRdy, TsireadColRdy) present whether the student has satisfied TSI standards determined through a combination of qualifying assessments and/or experiences, such as SAT/ACT and state assessments, completion of specific college level course work, or military service. The variables hold the condition of "Satisfied" or "Not Satisfied" for each variable. The status of "Satisfied" or "Not Satisfied" is determined from a set of conditions captured from the AAR and transfer credit transcripts. As a measure of academic achievement, TSI readiness is considered a factor influencing college choice (Hurtado et al., 1997).

3.6 Binomial Logistic Regression

The goal of this analysis is to correctly predict if the admitted student will enroll or will not enroll at the target institution. This is to be carried out in several steps, starting with determining if a relationship is found between the dependent variable and the predictor variables. For example, this study will determine if enrollment at the target institution can be predicted on "Earned College Credit" or "High School Rank". We will approach this directly, using the standard form for the logistic regression model, generally expressed as

$$\log(\frac{Y}{1-Y}) = B_0 + B_1X_1 + B_2X_2 + B_2X_3 + \cdots + B_iX_i + \epsilon \qquad (3.1)$$

Where we calculate the ratio of the probability the student will enroll (Y) to the probability that the student will not enroll (1-Y). Student attributes, family background, pre-college experiences, and high school characteristics are represented in the selected predictor variables $(X_1, X_{2,}, X_{3,} \dots, X_i)$; the estimated coefficients are represented by $(B_1, B_2, B_3, \dots, B_i)$, with standard error ϵ. This form of the probability equation is best suited for analyzing dichotomous dependent variables (Menard, 1995).

The significance of coefficients is tested using the Wald test, providing an understanding as to whether the explanatory variables contribute to the model or if they can be removed since the given explanatory variable is not statistically significant. The Wald test is determined by

$$W_i = B_i / SE_{B_i} \quad (3.2)$$

where B_i is the predictor's coefficient and $SE_{B_i}$is the coefficient's standard error.

Further informing the selection of predictor variables and whether the model is specified in a proper manner, is Goodness-of-Fit test. This is an important step as it determines the model's ability to correctly classify those students who will enroll and those who will not enroll.

This step is then followed by a calculation of the probability for enrollment for the fall 2016 cohort data, a process referred to as "scoring the data" (DesJardins, 2002). The hypothesized outcome is that the model will tend towards specificity given that we expect compulsory college application strategies will not increase enrollment of admitted students.

3.6.1 Theoretical Issues Concerning Logistic Regression Analysis

While logistic regression may have the benefit of few restrictions and the power to analyze discrete, dichotomous, and continuous variables, cautions must be taken to avoid causal inferences (Tabachnick & Fidell, 2013)

Tabachnick & Fidell assert that a common practice among researchers is to develop a model using many predictor variables and eliminating those that are shown to be not statistically significant (Tabachnick & Fidell, 2013). This concern is mitigated when predictor variables are qualified through research based theoretical models. College Choice Theory is the theoretical model guiding the selection of predictor variables for this study.

3.6.2 Practical Issues Concerning Logistic Regression Analysis

Binomial logistic regression considers a set of assumptions about the study's design and how data fits the binomial logistic regression model (Laerd Statistics, 2020; Tabachnick & Fidell, 2013). The first four assumptions relate to the study's design and the measurements chosen for the study. The remaining assumptions relate to how the data fits the model and its use in providing valid results. The steps that follow will guide us through the study's design and documents how each assumption is satisfied.

3.7 Assumptions of a Binomial Logistic Regression

Binomial logistic regression considers a set of assumptions that are critical to the process and require testing of the data to ensure it may be used in binomial logistic regression analysis. The purpose of the test of assumptions of a binomial logistic regression is to allow for an understanding of the accuracy of the model's predictions. It is also important to understand how well the data fits the binomial logistic regression model, and how much of the variation in the dependent variable is explained by the independent variable(s).

The study design is considered a retrospective cohort study as historical data is used and the actualization of enrollment is known. As such, the first set assumptions are related to the study design and the measurements to be made in the study. These assumptions are as follows: (1) there is one dependent variable that is dichotomous, (2) the set of independent variables are measured on a continuous and nominal scale, (3) there exists independence of observations and the categories are mutually exclusive in the dichotomous dependent variable and the nominal independent variables, and (4) there are at least 15 cases for each independent variable (DeMaris, 1992; Laerd Statistics, 2020; Tabachnick & Fidell, 2013).

The second set of assumptions support the determination of how well the data fits the binomial logistic regression model. These assumptions are as follows: (5) there exists a linear relationship between continuous predictors and the logit transformation of the dependent variable, (6) the set of independent variables are uncorrelated, (7) there should be no significant outliers (DeMaris, 1992; Laerd Statistics, 2020; Tabachnick & Fidell, 2013). Tests of these assumptions are presented in Chapter 4.

3.8 Model Validation

Validation is an assessment that is considered when the objective is to predict the outcome of a new set of admitted students, that is, predicting if the admitted student will enroll or not enroll for the term in which the student has applied for admission. The concern is the fitted model's performance tends to perform optimistically on the developmental data set (Hosmer et al., 2013).

For a study such as that currently investigated, one approach is to use the *Class of 2016* data set for development and the *Class of 2017* data set for validation. However, the availability of data limits us to the use of the *Class of 2106* data set. Given this condition, the *Class of 2016* data set will be randomly split into a "developmental" sample to determine the model's results and a "validation" or "hold out" sample to mitigate concerns of bias in results and determine the model's effectiveness in correctly predicting results from another sample (DesJardins, 2002; Hosmer et al., 2013).

3.9 Summary of Chapter

The development of a logistic regression model is expected to predict if the admitted student will enroll or will not enroll at the target institution. Binomial logistic regression is selected as the statistical method for this design, given the dichotomous resultant of the dependent variable and mixed set of independent variables. The study will make use of historical data to provide the development sample as well as the validation sample. The research undertaken here is an attempt to understand the effect of those variables hypothesized as being influential factors in the college choice process.

Chapter 4: Results

INTRODUCTION

Results of the study are presented in this chapter and are intended to provide an understanding of the procedures related to binomial logistic regression analysis used to determine which of the independent variables have a statistically significant effect on the dependent variable and how well the model predicts the dependent variable. The chapter begins with a set of diagnostics used to determine if the data meets the set of assumptions associated with binomial logistic regression. The chapter continues with a discussion of the use of IBM SPSS Statistics to perform the logistic regression procedure and the use of output to report the results of the regression analysis.

As previously stated, College Choice Theory provides the lens for theory-based model development. The college choice model reflected in this study is the Three-Phase Model, where each phase is influenced by a dynamic set of individual and organizational characteristics and attributes. The Three-phase model suggests a linear approach to the decision-making process undertaken by students, providing an understanding of the effect attributes and characteristics provoke within the choice process, and how these shape student decision and choice across each phase.

A common criticism arising from the literature is that the model fails to consider other factors affecting college choice for students of color and of low socioeconomic status (Bergerson, 2009). Bergerson presents a common point in the literature critiquing the college choice model; it fails to consider factors associated with social class. It is on this point that we turn to the hypothesized use of predictive modeling to improve upon our understanding of the likelihood that the admitted student will enroll at the target institution, given the contribution provided through analysis of the model's results. The expectation is an improved toolkit that may be used by enrollment managers to influence the outcome of the admitted student's choice.

4.1 Test of Assumptions

As previously stated, binomial logistic regression considers a set of assumptions that are critical to the analysis and require testing of the data to ensure the assumption is satisfied. The purpose of the test of assumptions of a binomial logistic regression is to allow for an understanding of the accuracy of the model's predictions.

4.1.1 Assumption 1 – One Dependent Variable that is Dichotomous.

The dependent variable, *EnrolledFall2016*, is dichotomous providing two mutually exclusive outcomes for enrollment, "Yes, the student enrolled" or "No, the student did not enroll."

4.1.2 Assumption 2 – Measured on a Continuous Scale

The set of independent variables are measured in continuous, nominal, or scale. The independent variables selected for the model are continuous, nominal, and scale.

4.1.3 Assumption 3 – Independence of Observations

There exists independence of observations and the categories are mutually exclusive in the dichotomous dependent variable and the nominal independent variables. In this study, we see the dependent variable (DV), *EnrolledFall2016*, can only be in "Yes, enrolled" or "No, not enrolled." Similarly, nominal independent variables (IV) have independence of observations and categories are mutually exclusive. For example, the independent variable, *ColCrdEarn*, exists in one of two states, *"Yes, College Credit Earned"* or *"No, College Credit Not Earned."* Likewise, the *STEM* variable exists in either, "*Yes, STEM Related*" or "*No, Not STEM Related*." (See Table 3.4 – *Definition of the Variables* for the full set of independent variables and definitions.)

4.1.4 Assumption 4 – A Minimum of Cases

There are at least 15 cases for each independent variable. The model considers 18 independent variables. Therefore, this requirement may be satisfied with a minimum of 270 cases. The total number of cases present in the data set is 7,183. The number of cases in the developmental data set is 3,645 and the number of cases in the validation data set is 3,538.

Hosmer, Lemeshow, and Sturdivant remind us that the goal of an analysis using a logistic regression model is to "find the best fitting and most parsimonious, clinically interpretable model to describe the relationship between an outcome (dependent or response) variable and a set of independent (predictor or explanatory) variables" (Hosmer et al., 2013; Laerd Statistics, 2020). This study enfolds this goal as it considers predictors viewed as student choice factors in a three-phase student choice process. The second set of assumptions relate to how the data fits the model and its use in providing valid results.

4.1.5 Assumption 5 – Linear Relationship in the Logit Transformation

In logistic regression, there is an assumption of a linear relationship between continuous predictors and the logit transformation of the dependent variable (Tabachnick & Fidell, 2013). Testing for a linear relationship between the continuous independent variables and the logit transformation of the dependent variable is assessed using the Box-Tidwell method (Box & Tidwell, 1962). The test of linearity is conducted with respect to the logit of the dependent variable, *EnrolledFall2016***.**

The Box-Tidwell (1962) procedure requires the use of all terms in the model, the identification of categorical and continuous variables, and the creation of the interaction term for each continuous variable in the model. IBM SPSS Statistics 25 is used to run the Box-Tidwell (1962) procedure. The results for the test of the linearity assumption are shown in Table 4.1.

Table 4. 1 Box-Tidwell Test

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
AvgOfClassSize by ln_AvgOfClassSize	0.00	0.00	1.25	1.00	0.26	1.00	1.00	1.01
hsAverageYield by ln_yield	-0.03	0.08	0.15	1.00	0.69	0.97	0.83	1.13

A statistically significant interaction term indicates the covariate for the interaction term is not linearly related to the logit of the dependent variable, and would be considered to fail the assumption of linearity. The *p*-value at which statistical significance is accepted is $p < 0.05$. The interaction terms included are *AvgOfClassSize* by *ln_AvgOfClassSize*, and *hsAverageYield* by *ln_yield*. Table 4.1 shows no interaction term is statistically significant. Based on this assessment, all continuous independent variables were found to be linearly related to the logit of the dependent variable (Menard, 2010).

4.1.6 Assumption 6 – Test for Multicollinearity

Testing for multicollinearity among the set of independent variables. The purpose of this test is to eliminate the impact of independent variables that are highly correlated among independent variables. The presence of highly correlated independent variables confounds our ability to understand which independent variables contribute to the variance explained in the dependent variable, creating problems when fitting and interpreting results from the logistic regression model. In general, multicollinearity may cause problems in fluctuation and precision of the coefficient estimates of the model (De Sá, 2007; Tabachnick & Fidell, 2013).

The diagnostic methods used for this analysis will be to test for high correlation coefficients, and high variance inflation factors (VIFs). Except for perfect correlations which are considered rare, researchers frequently cite there are no firm rules for thresholds indicating a risk for or the existence of a serious problem resulting from high correlations. However, the recommended rule of thumb indicating a serious problem from multicollinearity is a Pearson correlation coefficient between two predictor variables with a value that is greater than 0.70 or a variance inflation factor with a value that is greater than 10 (Menard, 1995; Midi, Sarkar, & Rana, 2013; Myers, 1986; Senaviratna & Cooray, 2019).

A test for multicollinearity is performed by testing for high correlation between predictors. The threshold for the correlation coefficient magnitude selected for the test of risk of multicollinearity is 0.80 or higher and will be considered together with the results in the variance inflation factor.

Conducting a correlation analysis, we find there is a high correlation coefficient between the predictors in *College Ready* and *TSI Overall College Ready* (0.99), *College Ready* and *TSI Math College Ready* (0.82), and *TSI Overall College Ready* and *TSI Math College Ready* (0.82). These results are shown in Table 4.2.

Table 4. 2 Testing for Multicollinearity – Correlations

Correlations			
	ClgRdy	TsiOvrallColRdy	TsimathColRdy
ClgRdy	1.00	0.99	0.82
TsiOvrallColRdy	0.99	1.00	0.82
TsimathColRdy	0.82	0.82	1.00

A second test for multicollinearity is performed by determining the Tolerance or Variance Inflation Factor for the set of independent variables. The cutoff values considered a risk or a serious problem for multicollinearity are in tolerance values less than or equal to 0.10 or variance Inflation factor values greater than 10 (De Sá, 2007). These values are shown in Table 4.3. Reviewing the values for variance inflation factor in the set of collinearity statistics, we find a similar set of independent variables demonstrating high risk or a serious problem for multicollinearity. Selecting a variance inflation factor threshold ≥ 10, we find *College Ready* has a VIF of 143.75, TSI Overall College Ready has a VIF of 147.28. To reduce the risk of multicollinearity the independent variables in *College Ready* and *TSI Overall College Ready* are selected to be dropped from the model.

Table 4. 3 Testing for Multicollinearity – Tolerance and VIF

Model	Collinearity Statistics	
	Tolerance	VIF
pQuartiles	0.68	1.46
AvgOfClassSize	0.69	1.44
ClgRdy	0.01	143.75
FEdAttain	0.77	1.31
MEdAttain	0.77	1.29
DiplomaCode	0.72	1.39
hsAverageYield	0.62	1.62
CompulsoryIntake	0.93	1.07
ColCrdEarned	0.84	1.19
AP	0.84	1.20
STEM	0.89	1.13
TsiOvrallColRdy	0.01	147.28
TsimathColRdy	0.34	2.96
TsiwritingColRdy	0.67	1.50
TsireadColRdy	0.44	2.27

4.1.7 Assumption 7 – Test for Significant Outliers

Testing for significant outliers, high leverage points/highly influential points. The purpose of identifying outliers is to mitigate the problematic influence resulting from cases showing a strong deviation from the fitted regression curve (De Sá, 2007). Using case diagnostics in IBM SPSS Statistics 25, the binomial logistic regression results for cases not fitting the model are posted to the *casewise list* and considered outliers. The casewise list provided by IBM SPSS Statistics 25 is shown in Table 4.4. The casewise list shows those cases where the standardized residual is greater than 3.0. It is proper practice to review each case identified in the casewise list to determine why a case is unusual and decide on its removal. There are two cases with a standardize residual greater than 3.0; these are kept in the analysis.

Table 4. 4 Outliers - Casewise List

Case	Selected Status[a]	Observed	Predicted	Predicted Group	Temporary Variable		
		EnrolledFall2016			Resid	ZResid	SResid
505	S	Y**	0.01	N	0.99	9.78	3.03
2658	S	N**	0.99	Y	-0.99	-18.83	-3.43

4.2 Binomial Logistic Regression – Iterative Analysis

An iterative approach is taken to demonstrate the *Enter Method,* a process considered a common option when using binary logistic regression models in statistical programs and the default option for binary logistic regression in IBM SPSS Statistics 25. This process is intended to provide an understanding of the effects of independent variables and whether they contribute significantly to the outcome in “enrolled” or “not enrolled.” The *Enter Method* is a statistical technique that uses simultaneous input (single step) of all IVs into the model. It is understood that other statistical techniques can be used, such as the *Forward Selection (likelihood ratio) stepwise regression* technique. *Forward Selection (likelihood ratio)* is a stepwise technique that adds and removes predictor variables into the model if statistical criteria are met (Tabachnick & Fidell, 2013). The technique is viewed as a controversial procedure given that interpretation or meaning of the predictor variables is set aside and instead subordinated to the statistics calculated from the sample used for the analysis.

4.2.1 Binomial Logistic Results – First Iteration

Upon completion of the assumption diagnostics the first iteration of the binomial logistic regression analysis is performed using IBM SPSS Statistics 25. The data are further examined for missing cases and the impact on the general quality of the data given the remaining variables. The case processing summary reveals a total of 3,400 cases are processed, where 1,156 are included in the analysis and 2,244 are in missing cases as shown in Table 4.5. Results for *categorical variable coding* are shown in Appendix A and results for *variables in the equation, first iteration* are shown in Appendix B.

Table 4. 5 Case Processing Summary – First Iteration

Unweighted Cases		N	Percent
Selected Cases	Included in Analysis	1,156	34.0
	Missing Cases	2,244	66.0
	Total	3,400	100.0
Unselected Cases		0	0.0
Total		3,400	100.0

An analysis of these results shows the categorical variables in *Father's Educational Attainment* and *Mother's Educational Attainment* are largely contributing to the set of missing cases. The categorical variables in *FEdAttain* and *MEdAttain* are in categories *FEdAttain*, *FEdAttain(1)*, *FEdAttain(2)*, *FEdAttain(3)*, *MEdAttain, MEdAttain(1)*, *MEdAttain(2)*, *and MEdAttain(3)*, as shown in Appendix A.

The baseline comparison (reference) category for both *FEdAttain* and *MEdAttain* is *College-Beyond.* Reviewing results in Appendix B for *FEdAttain*, we find there is not a significant overall effect contributed by *FEdAttain* (*Wald=6.518, df=3, p>0.05*). Similar results are shown for *MEdAttain*, where we find there is not a significant overall effect contributed by *MEdAttain* (*Wald=2.115, df=3, p>0.05*). The effect of these variables confounds overall results of the model. The independent variables in *Father's Educational Attainment* and *Mother's Educational Attainment* are also removed from the set of variables in the model.

4.2.2 Binomial Logistic Results – Second Iteration

The first iteration of the binomial logistic regression model considered several variables now excluded from the model due to issues of multicollinearity or yielding confounding effects. Analysis of the reduced model utilizing the *Enter Method*, second iteration, demonstrates the effect of IVs upon the predictive probability of enrollment. Results for *variables in the equation, second iteration* are shown in Appendix C. We find *AvgOfClassSize* (*p = 0.49), DiplomaCode (p = 0.21),* and *TsireadColRdy (p = 0.33)* are not statistically significant. These variables are removed and the model is further reduced.

4.2.3 Binomial Logistic Results – Third Iteration

The second iteration of the binomial logistic regression model contained IVs found not statistically significant. Analysis of the reduced model utilizing the *Enter Method*, third iteration, demonstrates the effect of IVs upon the predictive probability of enrollment. Results for *variables in the equation, third iteration* are shown in Appendix D. We find *TsiwritingColRdy (p=0.08)* is not statistically significant. The variable is removed and the model is further reduced.

4.2.4 Binomial Logistic Regression Model – Final Set of Independent Variables

The third iteration of the binomial logistic regression model contained an additional independent variable found not statistically significant. The final form includes the set of several independent variables shown in Table 4.7. Results for the final form are presented in the section that follows.

Table 4. 6 Final Set of Independent Variables in the Study

Variable Name	**Description**
pQuartile	Student's rank position in quartiles determined from high school percentile;
hsAverageYield	Percent yield (admitted to enrolled) from specified high school;
CompulsoryIntake	Submitted admissions application during compulsory application intake period (0 = No, 1 = Yes)
ColCrdEarned	Earned college credit, any program (0 = No, 1 = Yes)
AP	Participation in Advanced Placement Program (0 = No, 1 = Yes)
STEM	Academic area of interest in STEM program (0 = No, 1 = Yes)
Tsimath ColRdy	TSI math college ready satisfied (0 = No, 1 = Yes)

4.3 Binomial Logistic Regression Results – Final Form

Prior to arriving at this stage of the analysis, the procedures used to prepare for the binomial logistic regression required a test of assumptions that provides an understanding of how well the regression model fits the data and qualifies the accuracy of the model's predictions. This was followed by an iterative process demonstrating the interaction of IVs and the effect on the model after elimination of variables that are not statistically significant. The sections that follow offer a summary of the resulting final model using IBM SPSS Statistics 25 binomial logistic regression output.

4.3.1 Data Coding

Analysis of results begins with data coding, a process intended to provide an understanding of the data, with respect to cases in the sample, dependent variable encoding, and categorical variable coding. The case processing summary shown in Table 4.8 shows a total of 3,400 cases processed, where 3,236 (95.2%) are included in the analysis and 164 (4.8%) are in missing cases.

Table 4. 7 Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	3,236	95.2
	Missing Cases	164	4.8
	Total	3,400	100.0
Unselected Cases		0	0.0
Total		3,400	100.0
a. If weight is in effect, see classification table for the total number of cases.			

Dependent variable encoding is shown in Table 4.9. The output allows us to verify the coding applied to the dependent variable. In this case, the dependent variable is "*EnrolledFall2016"* and coded as "Yes, Enrolled = 1" and "No, Not Enrolled = 0." Variable encoding data is verified to be correctly coded.

Table 4. 8 Dependent Variable Encoding

Original Value	Internal Value
No	0
Yes	1

Categorical variable encoding is shown in Table 4.10. This presents frequency counts only for independent variables that are categorical. Tabachnick and Fidell (2013) remind us that "As with all statistical techniques, power increases with sample size" (p.444). It is undesirable to have low frequency counts among categorical variable codings as issues of over-fitting due to small sample size may arise. The analysis shows no risk of low frequency counts in categorical variable codings.

Table 4. 9 Categorical Variable Codings

		Frequency	Parameter coding		
			(1)	(2)	(3)
pQuartiles	1st Quartile	1,043	0	0	0
	2nd Quartile	903	1	0	0
	3rd Quartile	749	0	1	0
	4th Quartile	541	0	0	1
DiplomaCode	DistingAchieve	435	0		
	Recommended	2,801	1		
CompulsoryIntake	No	261	0		
	Yes	2,975	1		
ColCrdEarned	No	1,974	0		
	Yes	1,262	1		
TsireadColRdy	No	1,313	0		
	Yes	1,923	1		
STEM	No	1,807	0		
	Yes	1,429	1		
TsiwritingColRdy	No	644	0		
	Yes	2,592	1		
TsimathColRdy	No	1,586	0		
	Yes	1,650	1		
AP	No	2,520	0		
	Yes	716	1		

4.3.2 Baseline Analysis

The analysis reviewed in this section gives a view of a *null model*, which is a model without the introduction of predictor variables and only the intercept is considered. Table 4.11 shows that overall, cases are coded as “Not Enrolled” or 62.7%, meaning the Enrollment Manager will guess correctly nearly 63% percent of the time if the guess is the student will not enroll.

Table 4. 10 Null Model Classification Table

Observed			Predicted		
			EnrolledFall2016		Percentage Correct
			No	Yes	
Step 0	EnrolledFall2016	No	2,035	0	100.0
		Yes	1,211	0	0.0
	Overall Percentage				62.7
a. Constant is included in the model.					
b. The cut value is .500					

Other detail provided in baseline analysis are *Variables in the Equation-null model,* showing only the constant (B_0) was included in the model, and *Variables not in the Equation-null model,* showing the set of predictor variables excluded from the null model. While viewed as not of interest to researchers, this detail is presented simply to guide the reader through the IBM SPSS Statistics 25 output as shown in Appendices 5 and 6.

4.3.3 Model Fit

We now evaluate the output from IBM SPSS Statistics 25, Block 1. These sections are *Omnibus Tests of Model Coefficients*, *Variance Explained – Nagelkerke R Square (Model Summary),* and the *Hosmer and Lemeshow Test.*

The output, *Omnibus Tests of Model Coefficients*, aids our understanding of the overall statistical significance of the model and whether the current model outperforms the null model in how well it will predict categories in *Enrolled* and *Not Enrolled.* Table 4.12 shows the significance value less than 0.001, indicating the current model outperforms the null model (IBM Corp, 2015; Strand, Cadwallader, & Firth, 2011).

Furthermore, the Omnibus Tests of Model Coefficients uses chi-square tests to determine if there is a significant difference between Log-likelihoods in the new model against the null model. A decreased *-2 Log likelihood* (-2LL) in the new model against the null model is an indication of improvement in explaining the variance given the introduction of the predictor variables (Strand et al., 2011).

Table 4. 11 Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	3302.65	9	0.000
	Block	3302.65	9	0.000
	Model	3302.65	9	0.000

Giving our attention to results in Table 14.12, referencing *Step 1, Model*, the resulting chi-square is significant (*chi square=3302.65, df=9, p<.000*), indicating an improvement in the new model over the null model.

To understand how much of the variation in the outcome is explained by the model, we use R^2 values found in Table 4.13; these values are referred to as pseudo R^2 values. Cox & Snell R Square (0.638) and Nagelkerke R Square (0.871) offer two values that may be referenced to understand the variation in the outcome, that is the variance in the dependent variable associated with the predictor variable.

The closer the R Square value approaches or equals to one, the more variation is explained by the model. Nagelkerke R Square, adjusted from Cox & Snell, is often given preference over Cox & Snell since the theoretical limitation of Cox & Snell is less than one (IBM Corp, 2015; Laerd Statistics, 2020; Strand et al., 2011). Referencing Nagelkerke R Square, we find the model explains 87.1% of the variation in the outcome (Strand et al., 2011).

Table 4. 12 Variance Explained – Nagelkerke R Square (Model Summary)

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	985.778[a]	0.638	0.871
a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.			

The Hosmer and Lemeshow goodness of fit test results indicate the model is a good fit to the data as *p=0.840 (>.05)*, as shown in Table 4.14. A statistically significant result would indicate a poor fitting model.

Table 4. 13 Hosmer and Lemeshow Goodness of Fit Test

Step	Chi-square	df	Sig.
1	4.184	8	0.840

4.3.4 Category Prediction

Binomial logistic regression analysis is used to determine which of the independent variables have a statistically significant effect on the dependent variable and how well the model predicts the dependent variable. SPSS will classify the actualization of enrollment (the event) as occurring if the probability is greater than or equal to 0.5. If the probability of the event occurring is less than 0.5, then SPSS will classify the event as not occurring, that is, the student did not enroll (Laerd Statistics, 2020; Strand et al., 2011). A classification table is used to assess coding designations for predicted against actual; this is shown in Table 14.15.

Table 4. 14 Classification Table

Observed			Predicted		
			EnrolledFall2016		Percentage Correct
			No	Yes	
Step 1	EnrolledFall2016	No	1942	93	95.4
		Yes	41	1170	96.6
	Overall Percentage				95.9
a. The cut value is .500					

The classification table shows the cut value is .500, and as previously stated, those cases with a probability greater than or equal to .500 are classified as "Yes, Enrolled" and those cases with a probability less than .500 are classified as "No, Did Not Enroll." Previously, the null model showed that 62.7% of cases would be correctly classified if the Enrollment Manager guessed the student did not enroll (refer to Table 4.11). The addition of predictor variables demonstrates the model correctly classifies 95.9% of cases overall as shown in the row labeled "Overall Percentage" in Table 14.15. This measure is referred to as the *percentage accuracy in classification*.

Additional measures associated with the classification of data are in *sensitivity*, *specificity*, *positive predictive value*, *negative predictive value*, and the *receiver operating characteristic* or *ROC curve*.

Sensitivity is a measure that considers the observed characteristic, "Yes, Enrolled," correctly predicted by the model. This value (96.6%) is found in Table 14.15 in the column *Percentage Correct* for the corresponding row in *EnrolledFall2016* and *Yes*.

Specificity is a measure that considers the observed characteristic, "No, Did Not Enroll," correctly predicted by the model. This value (95.4%) is found in Table 14.15 in the column *Percentage Correct* for the corresponding row in *EnrolledFall2016* and *No*.

The *positive predictive value* is the number of correctly predicted cases against the total number of cases in the observed characteristic, "Yes, Enrolled." This is calculated from data in the column, *Yes*, under *Predicted, EnrolledFall2016. Positive predictive value* is (1170 ÷ (1170+93) x 100) or 92.6%.

The *negative predictive value* is the number of correctly predicted cases against the total number of cases in the observed characteristic, "No, Did Not Enroll." This is calculated from data in the column, *No*, under *Predicted, EnrolledFall2016. Negative predictive value* is (1942 ÷ (1942+41) x 100) or 97.9%.

Summarizing these results, we show *sensitivity* was 96.6%, *specificity* was 95.4%, *positive predictive value* was 92.6% and *negative predictive value* was 97.9%. We noted these are measures that are based on a cut value of .500, where those cases with a probability greater than or equal to .500 are classified as "Yes, Enrolled" and those cases with a probability less than .500 are classified as "No, Did Not Enroll." The idea of using a cut value is noted to be arbitrary, as multiple cut values may be selected to classify the admitted student as enrolled, thus impacting *sensitivity* and *specificity* (Mandrekar, 2010). To improve our understanding of the model's power to correctly discriminate and classify the admitted student as *enrolled* or *not enrolled*, we turn to the ROC curve.

The ROC curve is described as having two components, the plotted curve resulting from *Sensitivity* against *1 minus specificity* for all cut values considered for the curve, and the diagonal line (chance curve) indicating no discrimination or classification by chance (Laerd Statistics, 2020; Mandrekar, 2010). Figure 4.1 shows the configuration of the plotted curve against the chance curve.

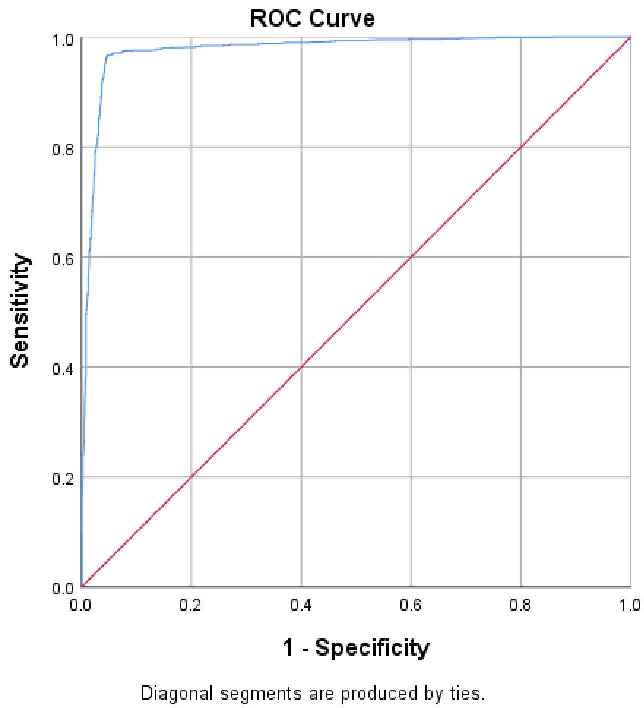

Figure 4. 1 Receiver Operating Characteristic (ROC) Curve

Summarizing the diagnostic accuracy of the test, we review the output for area under the ROC curve shown in Table 14.16. With a possible range of 0.5 to 1.0, we find that area under the curve is .977. While there is "no 'magic' number, only general guidelines" (Hosmer et al., 2013, p. 177), an area under the curve that is $\geq$0.9 (ROC $\geq$ 0.9) is considered outstanding discrimination (Hosmer et al., 2013).

The model explained 87.1% (Nagelkerke R Square) of the variance in actualizing enrollment for the term admitted and correctly classified 95.9% of cases. The area under the ROC curve was .977 (95% CI, .972 to .982), which is an excellent level of discrimination according to Hosmer et al. (2013).

Table 4. 15 Area Under the Curve

Test Result Variable(s):				
			Asymptotic 95% Confidence Interval	
Area	Std. Error[a]	Asymptotic Sig.[b]	Lower Bound	Upper Bound
0.977	0.003	0.000	0.972	0.982
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a. Under the nonparametric assumption				
b. Null hypothesis: true area = 0.5				

4.4 Results of the Model, Final Form

We now consider the final set of predictor variables and the contribution each has upon the model. It is our desire to understand the interplay among the independent variables in *pQuartile*, *hsAverageYield*, *CompulsoryIntake*, *ColCrdEarned*, *AP*, *STEM*, and *TsimathColRdy*, upon the predicted actualization of *enrollment* (DV). This detail is shown in Table 14.17, *Variables in the Equation, Final Form,* showing values for the regression coefficient (B), the Wald statistic, degrees of freedom, the statistical significance of the test (Sig.), the odds ratio (Exp(B)) and associated confidence intervals (C.I. for EXP(B)).

4.4.1 Variables in the Equation

The Wald statistic is used to establish the statistical significance for each predictor variable, in addition, the result for the test of statistical significance is given in the value within the column labeled *Sig.* The results in Table 14.17 show that all predictor variables in the *final form* of the model are statistically significant. The remainder of this section will expand upon these results.

4.4.2 pQuartile

The categorical IV, *pQuartile,* has the baseline comparison dummy variable as *first quartile.* Subsequently, *pQuartile(1), pQuartile(2),* and *pQuartile(3)* are in *second quartile*, *third quartile*, *and fourth quartile*, respectively. The results for pQuartile show a significant overall effect (*Wald=41.298*, *df=3*, *p<.05*). The *B* coefficients for the other terms in *pQuartile* are significant and positive, indicating that placement within a quartile is associated with increased odds of actualizing enrollment, since predicted probability is of membership in "*Yes, Enrolled*."

Table 4. 16 Variables in the Equation, Final Form

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	pQuartiles			41.30	3.00	0.00			
	pQuartiles(1)	1.28	0.26	24.83	1.00	0.00	3.58	2.17	5.91
	pQuartiles(2)	1.28	0.28	21.45	1.00	0.00	3.60	2.09	6.19
	pQuartiles(3)	1.87	0.34	30.67	1.00	0.00	6.47	3.34	12.52
	hsAverageYield	-0.03	0.01	10.47	1.00	0.00	0.97	0.96	0.99
	CompulsoryIntake(1)	-0.87	0.33	6.83	1.00	0.01	0.42	0.22	0.80
	ColCrdEarned(1)	6.54	0.22	851.48	1.00	0.00	689.33	444.40	1069.25
	AP(1)	0.86	0.22	15.71	1.00	0.00	2.37	1.55	3.63
	STEM(1)	0.68	0.19	12.74	1.00	0.00	1.98	1.36	2.88
	TsimathColRdy(1)	0.62	0.22	8.31	1.00	0.00	1.87	1.22	2.86
	Constant	-3.87	0.52	55.30	1.00	0.00	0.02		

a. Variable(s) entered on step 1: pQuartiles, hsAverageYield, CompulsoryIntake, ColCrdEarned, AP, STEM, TsimathColRdy. *(Output is from IBM SPSS Statistics 25)*

To understand the effect upon the actualization of enrollment for the student who places in a given quartile, we examine each of the categories in *pQuartiles*. For the case in which a student places in the second quartile (*pQuartile(1)*), we find a significant effect (*Wald=24.830*, *df=1*, *p<.05*). In addition, the log odds for *pQuartile(1)* is *1.275*, with corresponding odds ratio (OR) of 3.58 and a 95% confidence interval of *[2.168, 5.913]*.

Hence, when all other IVs are kept constant, the log odds change is 1.275. Expanding on this result, the odds ratio indicates the odds from placement in *pQuartile(1)* will increase the odds for *not* enrolling by 3.58. Students in the second quartile have increasing odds [OR=*3.58*] for *not* enrolling than students in the first quartile (baseline comparison IV).

A similar effect is seen for placement in *pQuartile(2)*, where we find a significant effect (*Wald=21.452*, *df=1*, *p<.05).* The log odds for *pQuartile(2)* is *1.281*, with corresponding *OR=3.600* and a 95% confidence interval of *[2.094, 6.190]*. Hence, when all other IVs are kept constant, the log odds change is *1.281*. Expanding on this result, the odds ratio indicates the odds from placement in *pQuartile(2)* will increase the odds for *not* enrolling by *3.60*. Students in the third quartile have greater odds [OR=*3.600*] for *not* enrolling than students in the first quartile (baseline comparison IV).

For placement in *pQuartile(3),* we find a significant effect (*Wald=30.675*, *df=1*, *p<.05).* The log odds for *pQuartile(3)* is *1.867*, with corresponding *OR=6.467* and a 95% confidence interval of *[3.340, 12.519]*. Hence, when all other IVs are kept constant, the log odds change is *1.867*. Expanding on this result, the odds ratio indicates the odds from placement in *pQuartile(3)* will increase the odds for *not* enrolling by *6.467*. Students in the fourth quartile have greater odds [OR=*6.467*] for *not* enrolling than students in the first quartile (baseline comparison IV).

4.4.3 hsAverageYield

The continuous IV, *hsAverageYield* demonstrates a significant effect (*Wald=10.469*, *df=1*, *p<.05).* In addition, the log odds for *hsAverageYield* is -*.028*, with corresponding odds ratio of .972 and a 95% confidence interval of *[.956, .989]*. To understand the effect upon the actualization of enrollment for the student who belongs to a low or high enrollment yield campus, we examine the effect as yield increases. We find the odds of actualizing enrollment are .972 the odds of lower yield campus. Therefore, students from high enrollment yield campuses are more likely to enroll than students from low enrollment yield campuses.

4.4.4 CompulsoryIntake(1)

The categorical IV, *CompulsoryIntake(1),* has the baseline comparison dummy variable as *"No, Did not participate in compulsory application intake."* The continuous IV, *CompulsoryIntake(1)* demonstrates a significant effect (*Wald=6.827*, *df=1*, *p<.05)*, with corresponding odds ratio of .419 and a 95% confidence interval of *[.218, .804]*. To understand the effect upon the actualization of enrollment for the student who participated in compulsory application intake, we examine this condition. We find the odds of actualizing enrollment are .419 < 1. This indicates that the odds of actualization of enrollment occurring in the "Yes, Did Participate" category are lower than the odds of actualization of enrollment occurring in the "No, Did Not Participate" category. Therefore, students who participated in the application intake process are less likely to enroll than students who did not participate in the application intake process.

4.4.5 ColCrdEarned(1)

The categorical IV, *ColCrdEarned(1),* has the baseline comparison dummy variable as *"No, Did not earn college credit."* The continuous IV, *ColCrdEarned (1)* demonstrates a significant effect (*Wald=851.481*, *df=1*, *p<.05)*, with corresponding odds ratio of 689.331 and a 95% confidence interval of *[444.403, 1069.248]*. To understand the effect upon the actualization of enrollment for the student who earned college credit while in high school, we examine this condition. We find the odds of actualizing enrollment are 689.331 the odds of the student who did not earn college credit. Therefore, students who earned college credit are more likely to enroll than students who did earn college credit.

4.4.6 AP(1)

The categorical IV, *AP(1),* has the baseline comparison dummy variable as *"No, Did not participate in Advance Placement programming."* The continuous IV, *AP(1)* demonstrates a significant effect (*Wald=15.713*, *df=1*, *p<.05)*, with corresponding odds ratio of 2.370 and a 95% confidence interval of *[1.547, 3.631]*. We examine the condition for the student who participated in Advanced Placement programming to understand the effect upon the actualization of enrollment. We find the odds of actualizing enrollment are 2.370 the odds of the student who did not participate in Advanced Placement programming. Therefore, students who participated in Advanced Placement programming are more likely to enroll than students who did not participate in Advanced Placement programming.

4.4.7 STEM(1)

The categorical IV, *STEM(1),* has the baseline comparison dummy variable as *"No, Did not select a STEM related academic area of interest."* The continuous IV, *STEM(1)* demonstrates a significant effect (*Wald=12.738*, *df=1*, *p<.05)*, with corresponding odds ratio of 1.981 and a 95% confidence interval of *[1.361, 2.884]*. We examine the condition for the student who selected a STEM related academic area of interest to understand the effect upon the actualization of enrollment. We find the odds of actualizing enrollment are 1.981 the odds of the student who did not select a STEM related academic area of interest. Therefore, students whose academic area of interest is in a STEM related field are more likely to enroll than students whose academic area of interest is not in a STEM related field.

4.4.8 TsimathColRdy(1)

The categorical IV, *TsimathColRdy(1),* has the baseline comparison dummy variable as *"No, Did not satisfy TSI Math College Readiness."* The continuous IV, *TsimathColRdy(1)* demonstrates a significant effect (*Wald=8.311*, *df=1*, *p<.05)*, with corresponding odds ratio of 1.868 and a 95% confidence interval of *[1.221, 2.856]*. We examine the condition for the student who satisfied *TSI Math College Readiness* to understand the effect upon the actualization of enrollment. We find the odds of actualizing enrollment are 1.868 the odds of the student who did not satisfy *TSI Math College Readiness*. Therefore, students who satisfied *TSI Math College Readiness* are more likely to enroll than students who did not satisfy *TSI Math College Readiness*.

4.5 Comparison of Model Iterations

A comparison of model iteration results previously described are shown in Table 14.18. We find each logistic regression model iteration is statistically significant with final form showing, $\chi^2(9) = 3302.65, p<.05.$ In addition, each iteration demonstrates the adequacy of each iterative model as the Hosmer Lemeshow test is not statistically significant in each, with final form showing, $p = .840.$ The variance explained (Nagelkerke R^2) across each iteration shows 77.6% in the first iteration and 87.1% in final form with classification at 95.9% also in final form.

Table 4. 17 Comparison of the Iterative Models

		First Iteration	Second Iteration	Third Iteration	Final Form
Omnibus Tests of Model Coefficients	Chi-square	892.633	3299.421	3299.554	3302.654
	df	19	13	10	9.0
	Sig.	0.000	0.000	0.000	0.000
Hosmer and Lemeshow Test	Chi-square	3.025	14.812	7.548	4.184
	df	8	8	8	8
	Sig.	0.933	0.063	0.479	0.840
Model Summary	-2 Log likelihood	473.182	978.625	982.331	985.778
	Cox & Snell R Square	0.538	0.639	0.639	0.638
	Nagelkerke R Square	0.776	0.872	0.871	0.871
Classification Accuracy	Model Fitting Data	94.0	95.9	95.9	95.9
Significant Variables		pQuartiles	pQuartiles	pQuartiles	pQuartiles
		Compulsory Intake	hsAverage Yield	hsAverage Yield	hsAverage Yield
		College Credit Earned	Compulsory Intake	Compulsory Intake	Compulsory Intake
		STEM	College Credit Earned	College Credit Earned	College Credit Earned
			AP	AP	AP
			STEM	STEM	STEM
			Tsimath ColRdy	Tsimath ColRdy	Tsimath ColRdy
			Tsiwriting ColRdy		

4.6 Validation

As previously stated, validation is an assessment that is considered when the objective is to predicting the actualization of enrolling for a new set of admitted students. Validation is important for our understanding of the fitted model's performance as there is a tendency to perform optimistically on the developmental data set (Hosmer et al., 2013). Frequencies for the hold out sample data set are shown in Table 3.3, Frequencies for Developmental – Validation

Data Sets. Results of the fitted model are applied to the hold out data to predict the actualization of enrollment of future cohorts.

The logistic regression model applied to validation data shows it was statistically significant, $\chi^2(9) = 3088.27, p<.05.$ The model explained 84.8% (Nagelkerke R^2) of the variance in predicting actualization of enrollment, with the model correctly classifying 95.9% of cases. Table 4.19 provides results for variables in the equation.

Table 4.18 Variables in the Equation – Validation Data-set

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	pQuartiles			24.90	3	0.00			
	pQuartiles(1)	0.71	0.23	9.90	1	0.00	2.04	1.31	3.17
	pQuartiles(2)	1.00	0.26	14.76	1	0.00	2.73	1.64	4.56
	pQuartiles(3)	1.48	0.32	20.89	1	0.00	4.38	2.33	8.26
	hsAverage Yield	-0.03	0.01	15.72	1	0.00	0.97	0.96	0.98
	Compulsory Intake(1)	-0.73	0.29	6.33	1	0.01	0.48	0.27	0.85
	ColCrdEarn(1)	6.13	0.21	877.04	1	0.00	461.14	307.28	692.04
	AP(1)	1.21	0.22	30.42	1	0.00	3.34	2.17	5.12
	STEM(1)	0.35	0.18	3.88	1	0.05	1.42	1.00	2.03
	TsiMath ColRdy(1)	0.38	0.21	3.26	1	0.07	1.46	0.97	2.20
	Constant	-3.19	0.45	49.30	1	0.00	0.04		

a. Variable(s) entered on step 1: pQuartiles, hsAverageYield, CompulsoryIntake, ColCrdEarn, AP, STEM, TsiMathColRdy.

The predictive model's final form applied ten variables. All variables in the equation are found to be statistically significant, with the exception of TsiMathColRdy which is not statistically significant *(p*=.071). We find the model yields results that suggest it may be generalized to provide accurate predictions using new data.

ROC curve analysis for the hold out sample shows an excellent level of discrimination, with area under the curve as 0.977, (95% CI, .972 to .982) as shown in Table 14.20.

Table 4.19 Area Under the Curve – Hold Out Sample

Test Result Variable(s):				
			Asymptotic 95% Confidence Interval	
Area	Std. Error[a]	Asymptotic Sig.[b]	Lower Bound	Upper Bound
0.977	0.003	0.000	0.972	0.982
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a. Under the nonparametric assumption				
b. Null hypothesis: true area = 0.5				

4.7 Main Findings

A binomial logistic regression was performed to determine the effects of predictor variables considered to affect student choice and actualization of enrollment at the target institution. Predictor variables included in the model are *pQuartiles*, *hsAverageYield*, *CompulsoryIntake*, *ColCrdEarned*, *AP*, *STEM*, and *TsimathColRdy*. The logistic regression model, final form, was statistically significant, $X^2(9) = 3302.27, p<.05.$ The model explained 87.1% (Nagelkerke R^2) of the variance in predicting actualization of enrollment, with the model correctly classifying 95.9% of cases. ROC curve analysis for the diagnostic sample shows an excellent level of discrimination, with area under the curve as 0.977, (95% CI, .972 to .982) as shown in Table 14.16. Similar results were found from analysis of the hold out sample as previously described.

An examination of college choice factors shows the effect upon the actualization of choice and correctly determining the probability of enrolling at the target institution. High school performance is considered in *pre-college experiences* and evaluated through the INDEPENDENT VARIABLE, pQuartiles, which is derived from percentile. As previously stated, percentile (rank) is a metric by which a student will self-assess if they are a fit for a choice-list institution (Chapman, 1981; Nora, 2004). Also, *percentile (rank)* is considered a determinant for admission at many IHEs and a determinant for automatic admission to publicly funded Texas IHEs for students graduating from a Texas high school and who are in the top ten percent of their graduating class. Table 5.1 shows the crosstabulation of *pQuartiles* against actual enrollment outcome in the DV, *EnrolledFall2016*. The Chi Squared test for the crosstabulation shows there is a significant association between *pQuartiles* and *EnrolledFall2016,* and the relation between these variables was significant, χ^2 *(3, N = 3306) = 54.474, p = .000.*

There is strong evidence that as student academic performance steps from high performance as determined by placement in the first quartile (*percent enrolled = 42.3%)*, through low performance as determined by placement in the fourth quartile (*percent enrolled = 24.4%)*, the probability of ***not enrolling*** at the target institution ***increases***. This is further supported by results of the binomial logistic regression model, for example, pQuartiles(1) shows: *B = 1.275, SE = .256, Wald = 24.830, p < .05.* The estimated OR is [*Exp (B) = 3.580, 95% CI (2.168, 5.913)*]. From the information provided through the crosstabulation and the binomial logistic regression results in *pQuartile*, we determine the estimated OR indicates an increase in the probability of not enrolling as a student's rank steps from the first quartile through the fourth quartile.

Performance in high school is viewed as an important factor determining college choice. As colleges make admissions decisions based on high school GPA and class rank, contributing to a student's motivation to self-assess their fit at a perceived selective institution. Also, teachers and counselors tend to provide greater support to high achieving students, giving advantage that is not provided to the average student (Chapman, 1981; Hossler et al., 1999; Paulsen, 1990).

Table 4.20 Crosstabulation – pQuartiles*EnrolledFall2016

			EnrolledFall2016		Total
			No	Yes	
pQuartiles	1stQuartile	Count	650	476	1126
		% within pQuartiles	57.7%	42.3%	100.0%
	2ndQuartile	Count	569	334	903
		% within pQuartiles	63.0%	37.0%	100.0%
	3rdQuartile	Count	496	241	737
		% within pQuartiles	67.3%	32.7%	100.0%
	4thQuartile	Count	408	132	540
		% within pQuartiles	75.6%	24.4%	100.0%
Total		Count	2123	1183	3306
		% within pQuartiles	64.2%	35.8%	100.0%

This presents a dilemma for the enrollment manager responsible for increasing enrollment at a public institution with a mission of access and excellence. Students in the first quartile will have greater choice set options and perhaps, greater support, than students in the second, third, and fourth quartile. As students are strongly influenced by their peers, teachers, and counselors, it is possible for the astute enrollment manager to develop strategies and tactics that will shape the influence of these significant voices as aspiring students begin their journey to college choice. An example is outreach programming supporting development of a student's social capital, and development of institutional agents with a capacity to foster a college access mindset.

A similar crosstabulation of the independent variable, *CompulsoryIntake(1)* against the DV, *EnrolledFall2016*, shows nearly 91% (n = 3010) of the population in the development sample completed an application for admissions during the compulsory application intake period, with only 33% of that number enrolling as shown in Table 5.2. The Chi Squared test for the crosstabulation shows there is a significant association between *CompulsoryIntake(1)* and *EnrolledFall2016,* and the relation between these variables was significant, χ^2 *(1, N = 3323) = 112.672, p = .000.*

Results of the binomial logistic regression model shows: *B = -.871, SE = .333, Wald = 6.827, p < .05.* The estimated OR is [*Exp (B) = .419, 95% CI (.218, .804)*]. This suggests that participation in the compulsory intake application process does not increase the probability the student will enroll at the target institution.

Table 4.21 Crosstabulation – CompulsoryIntake(1)*EnrolledFall2016

			EnrolledFall2016		Total
			No	Yes	
CompulsoryIntake(1)	No	Count	115	198	313
		% within CompulsoryIntake	36.7%	63.3%	100.0%
	Yes	Count	2016	994	3010
		% within CompulsoryIntake	67.0%	33.0%	100.0%
Total		Count	2131	1192	3323
		% within CompulsoryIntake	64.1%	35.9%	100.0%

This appears to not support the hypothesis that a compulsory application intake strategy has an effect on increasing yield. The outcome of a sweeping activity mandating the submission of an application for all high school graduating seniors, including those who have not realized a predisposition to pursue a college education, does not result in more students enrolling at the target institution.

Investment of resources needed to support a regional compulsory application intake process should be assessed for its efficacy in supporting a student through the stages of the college choice process: predisposition, search, and choice. Furthermore, an understanding of the return on investment should be pursued. Such an approach will provide the enrollment manager with the needed context for alternative strategies that may prove more effective in increasing yield.

Students who have earned college credit while in high school are more likely to enroll than students who did not earn college credit while in high school. In the set of pre-college experiences, *college credit earned* is especially important to understand as a growing number of high school students earn college credit through dual-credit and early college high school programs. A crosstabulation of the independent variable, *ColCrdEarned(1)*, against the DV, *EnrolledFall2016,* shows that 37.9% (n = 1258) of the population in the development sample earned college credit as shown in Table 5.3, and 91% of those students actualized enrollment. In addition, we find that while 62.1% (n = 2017) did not earn college credit, and nearly 98% of those students did not actualize enrollment. This outcome closely mirrors the results found in the regression classification table. In addition, the Chi Squared test for the crosstabulation shows there is a significant association between *ColCrdEarned(1)* and *EnrolledFall2016,* and the relation between these variables was significant, χ^2 $(1, N = 3323) = 2668.527, p = .000.$

Results of the binomial logistic regression model further show the magnitude of the effect contributed by *ColCrdEarned(1)*: $B = 6.536, SE = .224, Wald = 851.481, p < .05.$ The estimated OR is [*Exp (B) = 689.331, 95% CI (444.403, 1069.248)*].

This suggests that earned college credit has a great effect on the actualization of enrollment at the target institution, as the odds of actualizing enrollment for students who earned college credit are 689 the odds of those who did not earn college credit. Indeed, the results of the *ColCrdEarned(1)*EnrolledFall2016* cross-tabulation shows this alone may be used as an indicator of the student's probability of enrolling. This begs the question, then why pursue a complex solution requiring the use of binomial logistic regression?

Table 4.22 Crosstabulation – ColCrdEarned(1)*EnrolledFall2016

			EnrolledFall2016		Total
			No	Yes	
ColCrdEarn(1)	No	Count	2017	48	2065
		% within ColCrdEarn	97.7%	2.3%	100.0%
	Yes	Count	114	1144	1258
		% within ColCrdEarn	9.1%	90.9%	100.0%
Total		Count	2131	1192	3323
		% within ColCrdEarn	64.1%	35.9%	100.0%

A compelling response is the following: As we work towards understanding the effect of factors influencing college choice and developing strategies and tactics that promote access and participation in higher education, it is imperative that we become aware of those significant factors that are also known to influence student choice. As higher education continues to experience greater demands on limited and diminishing resources, the use of predictive modeling will provide enrollment managers the capacity to make qualified decisions in resource management to support efforts to meet institutional enrollment goals by examining multiple choice factors understood to influence the student's decision to enroll. This also implies a need to adjust strategies and tactics as strategic gaps related to enrollment goals are discovered.

Students participating in Advanced Placement programs are more likely to enroll than students who did not participate in Advanced Placement programs. This is supported by results found in the crosstabulation of *AP(1)*EnrolledFall2016,* as shown in Table 5.4. The Chi Squared test for the crosstabulation shows there is a significant association between *AP(1)* and *EnrolledFall2016,* and the relation between these variables was significant, χ^2 *(1, N = 3323) = 303.179, p = .000.*

Although 80% (n = 2669) of the population in the development sample did not participate in AP programming, we also find that 71.3% did not enroll. In addition, of the 20% (n = 654) who did participate in AP programming, 65% did enroll. Results of the binomial logistic regression model shows *AP(1)* participation with: *B = .863, SE = .218, Wald = 15.713, p < .05.* The estimated OR is [*Exp (B) = 2.370, 95% CI (1.547, 3.631)*]. This suggests that the odds of *AP* participants actualizing enrollment are 2.37 the odds of non-*AP* participants.

Table 4.23 Crosstabulation – AP(1)*EnrolledFall2016

			EnrolledFall2016		
			No	Yes	Total
AP(1)	No	Count	1903	766	2669
		% within AP	71.3%	28.7%	100.0%
	Yes	Count	228	426	654
		% within AP	34.9%	65.1%	100.0%
Total		Count	2131	1192	3323
		% within AP	64.1%	35.9%	100.0%

Being mindful of this dynamic, the enrollment manager may develop outreach and communication strategies and tactics that influence the significant voices in the lives of high achieving students, while fostering a college access mindset for all.

Students with an academic area of interest in STEM programs are more likely to enroll than students who did not declare an academic area of interest in STEM programs. This is supported by results of the binomial logistic regression model showing: *B = .354, SE = .180, Wald = 3.877, p < .05.* The estimated OR is [*Exp (B) = 1.424, 95% CI (1.002, 2.025)*]. In addition, a crosstabulation of the independent variable, *STEM(1)* against the DV, *EnrolledFall2016* shows that 43% (n = 1414) of the population in the development sample selected a STEM related major, with 47% of those students enrolling. In addition, 57% (n = 1909) of the population in the development sample did not select a STEM related major, with 73% of those students not enrolling, as shown in Table 5.5. The Chi Squared test for the crosstabulation shows there is a significant association between *STEM(1)* and *EnrolledFall2016,* and the relation between these variables was significant, χ^2 *(1, N = 3323) = 152.449, p = .000.*

Table 4.24 Crosstabulation – STEM(1)*EnrolledFall2016

			EnrolledFall2016		
			No	Yes	Total
STEM(1)	No	Count	1393	516	1909
		% within STEM	73.0%	27.0%	100.0%
	Yes	Count	738	676	1414
		% within STEM	52.2%	47.8%	100.0%
Total		Count	2131	1192	3323
		% within STEM	64.1%	35.9%	100.0%

Attention is given to this factor as there exists a national interest in pursuing increased participation in STEM fields to meet workforce needs. Such an endeavor begins with increasing interest in STEM majors and associated careers (The Business Higher Education Forum, 2010). Given this context, college administrators can identify students with a demonstrated interest in STEM majors and invest additional resources to ensure the student's enrollment and success at the target institution. Moreover, the process of cultivating interest is critical in the early years of the student's education. The immediate implication is the utilization of a predictive model that yields a probability of enrollment for students declaring STEM related majors to support the institution in providing targeted and impactful interventions that support students in realizing their educational aspirations.

4.8 SUMMARY OF CHAPTER

Predictive modeling of factors understood to effect student choice was applied using a binomial logistic regression method. The method involves a test of assumptions that are intended to provide an understanding of how the regression model fits the data and to support an understanding of variation in the dependent variable explained by independent variables. An iterative approach in analysis was chosen as a method of demonstrating considerations for inclusion or exclusion of independent variables in accordance with results throughout the iterative process. Validation of the model is also presented as a method of determining if the model can be generalized and used with new data. Model results are further explained with a cross-tabulation analysis detailing validity of findings.

Chapter 5: Discussion and Conclusions

5.1 Summary of the Study

The purpose of this study was to examine college choice factors hypothesized to impact the actualization of enrollment for the student in the admitted stage of the enrollment cycle. The study was intended to provide enrollment managers with an analytical tool that may be used to support decisions for the allocation of resources and maximizing outcomes for the institutional mission emphasizing increased participation in higher education.

Considerations for the selection of predictive variables used in this study were the availability and standardization of data. This objective was achieved through the use of the state's common application for admission (ApplyTexas) and the student's academic achievement record. In addition, there is interest in understanding the effect of regional practices intended to support access to higher education, such as compulsory application intake.

As previously stated, the college choice process is important as it affords a student the opportunity to recognize an aspiration to attend college, embark upon a search for and development of a college choice set, and assimilate the acquired knowledge and experience into the selection of a college to attend (Hossler et al., 1999; Paulsen, 1990). The researcher hypothesizes that compulsory college application strategies disable the efficacy of the college choice process, observed as participation in post-secondary education.

In analyzing this issue, the researcher considered the following research questions with some observed outcomes:

Research Question 1: Do factors selected for the regression model, such as *Parental Educational Attainment*, *Average Class Size, Average High School Yield, Diploma Type, Quartiles, Participation in Compulsory Application Intake, Earned College Credit*, *Participation* *in Advanced Placement programs, STEM related Academic Areas of Interest,* and *TSI College Readiness* have an influence on the decision to enroll at the target institution?

The observed results from the examination of the full predictive model with all thirteen relevant variables indicated that only seven of the thirteen were deemed significant or with predictive value, in relation to actual student matriculation at the target institution in the term for which the student was admitted.

Research Question 2: Of the proposed set of student factors, which are the most predictive (important) factors of the likelihood for a student to enroll at the target institution?

Of the most important or predictive variables extracted from the initial full model are presented in rank order of importance: College Credit Earned, Quartiles, Advanced Placement, STEM related academic area of interest, TSI College Ready Math, Average High School Yield, and participation in Compulsory Application Intake.

Research Question 3: Does the final predictive model of the likelihood for a student to enroll perform similarly using a hold-out data set, thus suggesting a generalizable model?

The final predictive model of the likelihood to enroll yielded similar statistical results when applied to the hold-out data set across the six best final variables for both student samples examined in the study. The six best final variables are College Credit Earned, Quartiles, Advanced Placement, STEM related academic area of interest, Average High School Yield, and participation in Compulsory Application Intake.

5.2 Discussion

The use of predictive modeling by institutions of higher education may increase institutional effectiveness and efficiency in supporting institutional enrollment goals and resource management. Predictive modeling has been shown to support institutional decision-makers in developing strategies driving the achievement of institutional enrollment goals and fulfilling institutional missions of access and equity (DesJardins, 2002). However, the use of advanced analytics has been constrained due to limited information technology infrastructure and institutional capacity. This study was intended to add to the body of knowledge supporting the development and use of a binomial logistic regression model to predict whether a student will enroll or will not enroll at the target institution.

The six best final variables in the equation are in College Credit Earned, Quartiles, Advanced Placement, STEM related academic area of interest, TSI College Ready in Math, Average High School Yield, and participation in Compulsory Application Intake. Enrollment managers should consider applying this information to make informed decisions in resource allocation.

For example, the study finds that students who earn college credit while in high school have a strong likelihood to enroll in college as compared to students who did not earn college credit in high school. This factor is a significant predictor of enrolling in college, consistent with the research presented by Hossler and Gallagher (1987), suggesting a cumulative effect throughout predisposition, search, and choice. As the availability of Dual-Credit programs continues to grow and regional independent school districts support the addition of new Early College High Schools, institutions of higher education can pursue efforts to develop support structures and resources to increase participation in such impactful pre-college experiences.

The study also shows how student ability is a significant predictor of enrollment. We find that factors related to student ability (Quartiles, Advanced Placement, STEM Academic Area of Interest) were statistically significant and in line with the findings advanced by Hossler and Gallagher (1987). An examination of Quartiles shows an increase in the odds ratio as we move from highest achievers in the first quartile to low achieving students in the fourth quartile, represented as pQuartiles(3). The expectation is a greater likelihood to enroll for high achieving students. Similarly, we find students participating in Advanced Placement Programs (AP) or a demonstrated interest in STEM programs have a greater likelihood of enrolling than students who did not participate in AP programs or select a STEM program of interest.

There was an expectation for a statistically significant effect for variables in *Parental Educational Attainment*, unfortunately, available data was insufficient as the number of missing records produced confounding results. It is important to note, although eliminated from the model, parental educational attainment remains an important factor impacting student choice. According to Hossler et al, parental education has a direct effect on developing the student's aspirations for participation in higher education and a greater impact on the student's decision to enroll in college (Hossler et al., 1999).

There was also interest in understanding whether regional efforts intended to promote participation in higher education (*Compulsory Application Intake)* provide a significant effect on the likelihood to enroll in college. While statistically significant, the strength in association between the actualization of enrollment and compulsory application intake has a small effect. This is important information for enrollment managers at the target institution to consider, since an investment of resources is made each year to roll out a strategy that appears not to yield the desired effect.

The Texas Success Initiative (TSI) exists through the Texas Education Code, requiring all students planning to enroll in public institutions of higher education, to demonstrate compliance with TSI college readiness standards. Compliance may be satisfied through exemptions, satisfactory completion of college level course work, or satisfactory results in the TSI assessment. There was an expectation for a strong effect across all variables associated with TSI factors, however, factors related to Texas Success Initiative compliance did not have a statistically significant effect. While compliance to state mandated college readiness standards is required for all college bound students, the variables in TSI do not adequately support an understanding of the actualization of enrollment.

5.3 Recommendations for Practice

It is common for public IHEs to have rolling admissions policies permitting students to have a start term in fall, spring, or summer. Such a practice results in overlapping timelines, activities, and resources needed to support matriculation of students in each of these start terms. It also requires sophisticated communication strategies and planning to support students as they progress through the enrollment lifecycle. In addition, institutions with a mission of access and excellence experience greater stress on resources as acceptance rates at such institutions can be 90% or greater. With respect to the target institution considered in this study, such policies produce large cohorts of admitted students with an expected yield rate of 46%. This suggests that more than half of resources expended to convert students from admitted to enrolled become an *economic cost* through choices not made by enrollment managers, as 54% of admitted students will not enroll.

Given the set of factors shown to affect student choice, enrollment managers can examine the current use of resources and the tactics believed to increase enrollment. Time and effort used to rollout a regional application intake process is costly. Cost drivers include personnel, travel, facilities, and instruction time. However, opportunity costs must also be considered. Such costs may include alternative uses of time and effort expended by admissions and recruitment personnel. Utilizing our understanding of the effects of student ability, precollege experience, and participation in dual-credit or Early College High School, alternative efforts in outreach and communication may be explored.

Although application intake strategies are focused in a period of time in which students are creating a college choice set, there is no effort to differentiate among students who are very likely to enroll, still deciding, or considering other options and have no plans to attend college. The ability to understand factors affecting a student's decision to enroll will advance efforts to increase participation in higher education. For example, given the predictive power of *College Credit Earned,* the enrollment manager can focus resources among students who do not indicate "Yes, College Credit Earned," to support yield efforts. In addition, an understanding of the impact on the student's class standing as determined by factors in *Quartiles,* qualified and informed decisions to allocate resources to compel the undecided student to enroll.

Enrollment managers may continue efforts using a modified application intake process that provides support to complete an application for admission, however, there are alternatives that demand fewer resources and a broad-brush approach to generating application head-count. Such efforts may include outreach programs for middle school students, parents, teachers, and counselors with the intent of developing an awareness of support from significant others, consideration for pre-college experiences, and other efforts for developing a student aspiration for college.

Engaging experts in federal TRIO programs with an understanding of factors affecting student choice can also shape programming, outreach and communication. This implies additional research to provide an understanding for the use of predictive modeling supporting program initiatives. Such modeling may be considered in the scope of the student lifecycle (enrollment funnel) and may consider similar factors used to understand the actualization of enrollment.

Higher education is facing an urgent call to be responsible stewards of resources. Informed decision making is critical and requires sophisticated analytical methods to support mission critical activities impacting student success and completion. Enrollment management experts at Ruffalo Noel Levitz (RNL) recommend that predictive modeling be used in developing specific stages of the enrollment funnel (see Figure 1.1). An institution considering the use of student search services will typically buy lists without qualifying parameter decisions for the set of names to buy. It is an approach that becomes inefficient as many of the students on the list are not likely to enroll. Predictive modeling can be used to provide an understanding of factors influencing student choice for the target institution, by giving enrollment managers important information to establish search parameters consistent with statistical results for predictor variables understood to impact student choice (Ramos & Jansen, 2013).

Communication plan stratification is also recommended for the application of predictive modeling. As institutions consider various modes of communication and outreach such as direct mail, email, social media, calling campaigns, and on-site or off-site activities and events, the relevant questions to ask are, "in which activities do we make an investment, how much of an investment is required, and to whom should the activity be directed?" The ability to determine low probability or high probability students (for enrollment) will support decisions in population selection for investments in direct mail, event participation, and other activities. This is not intended to disregard low or high probability students, as other tactics may be employed to affect student choice within these groups (DesJardins, 2002; Ramos & Jansen, 2013).

An advanced approach is proposed by DesJardins (2002) involving a technique referred to as "scoring" the data set, in this case, students in the admitted pool. Once scored, cases are segmented into ten groups (deciles) containing a comparable count of cases across each group. This provides a view of students who are in the range of low probability to high probability, with sufficient information to determine those who are at the margin and who may be influenced through appropriate interventions. Understanding the probability that a student will enroll provides valuable information that can then be used to identify students for targeted communication plans, activities, and development of financial aid packages to maximize resources invested by the institution (DesJardins, 2002; Leppel, 1993; Thomas et al., 1999).

5.4 Future Research

It is understood that post-high school academic plans are shaped by high school experiences, academic achievement, family background, and organizational characteristics (Chapman, 1981; Hossler et al., 1999). In addition, it is understood that high achieving students direct more attention from parents, teachers, counselors and other significant voices in the student's life (Cabrera & La Nasa, 2000). This implies the student will possess, to some extent, a command of social and cultural capital needed to navigate the social networks and relationships to succeed in attaining the desired post-high school academic plans (Perna, 2006).

Future research may apply an integrated approach that considers new developments to College Choice modeling. The Three-Phase Model implies a sequential approach to the decision-making process undertaken by students and it provides an understanding of the effect attributes and characteristics have throughout the college choice process, shaping student decision across each phase. While the model provides for attributes, characteristics, and experiences contributing to a decision to enroll, an understanding of beliefs and attitudes contributing to the differences across individual post-secondary education plans remain unclear (Hossler et al., 1999).

It is on this point that we may turn to the Theory of Planned Behavior for an understanding of the student's intentions, factors influencing intentions, and the insightful contribution provided by the behavior-intention relationship for understanding *why* a student makes certain behavioral decisions throughout the student choice process. The Theory of Planned Behavior (TPB) has developed over decades of research related to the behavior-intention relationship and its application towards predicting behavior based on intentions has been widely used to understand decisions in self-care, medical compliance, leisure choice, and choices in consumption. The research on TPB, while having a wide-breadth of studies supporting a broad set of applications, has few studies focusing on participation in higher education. Fewer studies exploring the interplay between the college choice process and influencing factors associated with behavioral and normative beliefs, exist in the body of knowledge. In effect, current college choice models fail to capture the reasoning behind choice (Pitre et al., 2006).

Such a study would add to the body of knowledge, by demonstrating how institutions of higher education may integrate the theoretical and conceptual frameworks advanced through *College Choice Theory* and the *Theory of Planned Behavior* to support development of predictive models intended to guide strategic initiatives promoting access and participation in higher education, and improving upon the effectiveness of strategic enrollment management planning. Equally important, the research may provide an understanding of salient beliefs affecting student behavior, while informing intervention strategies intended to positively influence student behavior towards participation.

5.5 Conclusion

This study has demonstrated the application and use of predictive modeling to provide enrollment managers with analytical tools needed to support students as they navigate a complex enrollment process and to improve upon the effective and efficient use of limited resources. The results of the predictive model analysis returned a set of six important factors that explain 84.8% (Nagelkerke R^2) of the variance in predicting actualization of enrollment, with the model correctly classifying 95.9% of cases. The final six variables in the equation are College Credit Earned, Quartiles, Advanced Placement, STEM related academic area of interest, Average High School Yield, and participation in Compulsory Application Intake.

The development of the predictive model is based on College Choice Theory, specifically, the Three-Phase College Choice Model. While current literature provides strong evidence demonstrating the significance of choice factors, there is a need for instructional information providing a step-by-step process in the use of analytical software and the application of statistical methods. The development of predictive models requires a capacity in statistical analysis and associated analytical software, an ability and understanding database solutions, access to data, and a functional understanding in the use of the model's output. A noteworthy conclusion is that compulsory application intake strategies do not have the intended effect on increasing yield.

References

Adelman, C. (2006). *The toolbox revisited : Paths to degree completion from high school through college* Washington, D. C. : Office of Vocational and Adult Education, U.S. Dept. of Education, 2006. Retrieved from https://search.library.wisc.edu/catalog/9910035741402121

An, B. P. (2013). The impact of dual enrollment on college degree attainment. *Educational Evaluation and Policy Analysis, 35*(1), 57-75. doi:10.3102/0162373712461933

Beattie, I. R. (2002). Are all "adolescent econometricians" created equal? racial, class, and gender differences in college enrollment. *Sociology of Education, 75*(1), 19-43. Retrieved from http://www.jstor.org/stable/3090252

Bergerson, A. A. (2009). In Ward K., &Wolf-Wendel L.(Eds.), *College choice and access to college: Moving policy, research, and practice to the 21st century*. San Francisco, CA: Wiley Periodicals, Inc.

Black, J. (2008). Enrollment management: A systems approach. *Retrieved from Http://Www.Semworks.Net/Papers/Wp_enrollment-Management-Systems-Approach.Php#content,*

Bontrager, B. (2004). Enrollment management: An introduction to concepts and structures. *College and University Journal, 79*(3), 11-16.

Bose, R. (2009). Advanced analytics: Opportunities and challenges. *Industr Mngmnt & Data Systems, 109*(2), 155-172. doi:10.1108/02635570910930073

Box, G. E. P., & Tidwell, P. W. (1962). Transformation of the independent variables. *Null, 4*(4), 531-550. doi:10.1080/00401706.1962.10490038

Bruggink, T. H., & 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. Retrieved from http://www.jstor.org/stable/40196173

Cabrera, A. F., & La Nasa, S. M. (2000). *Understanding the college choice of disadvantaged students*. San Francisco, Calif.: Jossey-Bass Publishers.

Chapman, D. W. (1981). A model of student college choice. *The Journal of Higher Education, 52*(5), 490-505. doi:10.2307/1981837

De Sá, J. (2007). *Applied statistics using SPSS, STATISTICA, MATLAB and R.* (). Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-71972-4 Retrieved from http://cds.cern.ch/record/1338143

DeMaris, A. (1992). *Logit modeling : Practical applications*. Newbury Park, California: Sage Publications.

DesJardins, S. L. (2002). An analytic strategy to assist institutional recruitment and marketing efforts. *Research in Higher Education, 43*(5), 531-553. Retrieved from http://www.jstor.org/stable/40197271

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2006). An integrated model of application, admission, enrollment, and financial aid. *The Journal of Higher Education, 77*(3), 381-429. Retrieved from http://www.jstor.org/stable/3838695

DesJardins, S. L., Dundar, H., & Hendel, D. D. (1999). Modeling the college application decision process in a land-grant university. *Economics of Education Review, 18*(1), 117-132. doi:http://dx.doi.org/10.1016/S0272-7757(98)00023-5

Goenner, C. F., & Pauls, K. (2006). A predictive model of inquiry to enrollment. *Research in Higher Education, 47*(8), 935-956. Retrieved from http://www.jstor.org/stable/40197517

Goldstein, P., & Katz, R. (2005). Academic analytics: The uses of management information and technology in higher education. *Retrieved from Https://Library.Educause.Edu/Resources/2005/12/Academic-Analytics-the-Uses-of-Management-Information-and-Technology-in-Higher-Education, 8*(ECAR Key Findings)

González, J. M. B., & DesJardins, S. L. (2002). Artificial neural networks: A new approach to predicting application behavior. *Research in Higher Education, 43*(2, AIR Forum Issue), 235-258. Retrieved from http://www.jstor.org/stable/40196887

Hill, N. E., & Torres, K. (2010). Negotiating the American dream: The paradox of aspirations and achievement among Latino students and engagement between their families and schools. *Journal of Social Issues, 66*(1), 95-112. doi:10.1111/j.1540-4560.2009.01635.x

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied logistic regression.*

Hossler, D., & Gallagher, K. (1987). Studying student college choice: A three-phase model and the implications for policymakers. *College and University, 62 n3*(Spring 1987), 207-221.

Hossler, D., Schmit, J., & Vesper, N. (1999). *Going to college: How social, economic, and educational factors influence the decisions students make*. Baltimore, MD: The Johns Hopkins University Press.

Hossler, D., & Stage, F. (1992). Family and high school experience influences on the postsecondary educational plans of ninth-grade students. *American Educational Research Journal, 29*(2), 425-451. Retrieved from http://www.jstor.org/stable/1163375

Hurtado, S., Inkelas, K. K., Briggs, C., & Rhee, B. (1997). Differences in college access and choice among racial/ethnic groups: Identifying continuing barriers. *Research in Higher Education, 38*(1), 43-75. Retrieved from http://www.jstor.org/stable/40196234

IBM Corp. (2015). IBM SPSS statistics for windows, version 23.0.

IBM Marketing Cloud. (2016). *Email marketing metrics benchmark study.* IBM Marketing Cloud. Retrieved from Retrieved from https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=UVL12406USEN

Kent, W. (1912). Academic efficiency. *Science, 36*(938), 841.

Kidd, A. (2016). *School characteristics: Impact on a student's post-graduation decision* Retrieved from https://uknowledge.uky.edu/mpampp_etds/251

Laerd Statistics. (2020). Binomial logistic regression. Retrieved from https://statistics.laerd.com/premium/spss/blr/binomial-logistic-regression-in-spss-5.php

Langston, R., & Scheid, J. (2014). Strategic enrollment management in the age of austerity and changing demographics: Managing recruitment, leveraging, revenue, and access in challenging economic times. *Strategic Enrollment Management Quarterly, 2*(3), 191-210. doi:10.1002/sem3.20048

Leppel, K. (1993). Logit estimation of a gravity model of the college enrollment decision. *Research in Higher Education, 34*(3), 387-398. Retrieved from http://www.jstor.org/stable/40196054

Levy, J. S. (1992). An introduction to prospect theory. *Political Psychology, 13*(2, Special Issue: Prospect Theory and Political Psychology), 171-186. Retrieved from http://www.jstor.org/stable/3791677

Lindsay, A. W. (1982). Institutional performance in higher education: The efficiency dimension. *Review of Educational Research, 52*(2), 175-199. doi:10.2307/1170310

Litten, L. H. (1982). Different strokes in the applicant pool: Some refinements in a model of student college choice. *The Journal of Higher Education, 53*(4), 383-402. doi:10.2307/1981605

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. doi:https://doi.org/10.1097/JTO.0b013e3181ec173d

McEwen, R. W., & Synakowski, K. S. (1954). Planning college enrollment for academic efficiency. *The Journal of Higher Education, 25*(6), 301-342. Retrieved from http://www.jstor.org/stable/1977337

Menard, S. (1995). In Lewis-Beck M. S. (Ed.), *Applied logistic regression analysis*. Thousand Oaks, CA: Sage Publications.

Menard, S. (2010). *Logistic regression: From introductory to advanced concepts and applications*. Thousand Oaks, California: Sage Publications. doi:10.4135/9781483348964

Midi, H., Sarkar, S., & Rana, S. (2013). Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics, 13*, 253-267. doi:10.1080/09720502.2010.10700699

Mitchell, M., & Leachman, M. (2015). Years of cuts threaten to put college out of reach for more students. *Center on Budget and Policy Priorities,*

Myers, R. H. (1986). *Classical and modern regression with applications*. Boston, Mass: Duxbury Press.

Nora, A. (2004). The role of habitus and cultural capital in choosing a college, transitioning from high school to higher education, and persisting in college among minority and nonminority students. *Journal of Hispanic Higher Education, 3*(2), 180-208. doi:10.1177/1538192704263189

OECD. (1996). *The knowledge-based economy.* (No. OCDE/GD(96)102). Paris, France: Organisation for Economic Co-Operation and Development. (Retrieved from https://www.oecd.org/sti/sci-tech/1913021.pdf)

Paulsen, M. B. (1990). *College choice: Understanding student enrollment behavior.* (). Retrieved from NDL-OPAC

Perna, L. W. (2006). Studying college access and choice: A proposed conceptual model. In J. C. Smart (Ed.), *HIGHER EDUCATION:: Handbook of theory and research* (pp. 99-157). Dordrecht: Springer Netherlands. doi:10.1007/1-4020-4512-3_3 Retrieved from http://dx.doi.org/10.1007/1-4020-4512-3_3

Pitre, P., Johnson, T., & Pitre, C. (2006). Understanding predisposition in college choice: Toward an integrated model of college choice and theory of reasoned action. *College and University Journal, 81*(2), 35.

Ramos, C., & Jansen, B. (2013, April 9,). How predictive modeling benefits enrollment managers. Retrieved from https://www.ruffalonl.com/blog/enrollment/predictive-modeling-benefits-enrollment-managers/

Roderick, M. (2006). Closing the aspirations-attainment gap: Implications for high school reform A commentary from Chicago. *Mdrc,*

Sampath, V., Flagel, A., & Figueroa, C.A logistic regression model to predict freshmen enrollments. Paper Presented at Southeast *SAS Users Group Conference, October 2009.*

Retrieved from Http://Analytics.Ncsu.Edu/Sesug/2009/SD016.Sampath.Pdf, Birmingham, AL.

Senaviratna, N., & Cooray, T. (2019). Diagnosing multicollinearity of logistic regression model. *Asian Journal of Probability and Statistics,* 1-9. doi:10.9734/ajpas/2019/v5i230132

SHEEO. (2016). State higher education finance: FY 2015. *State Higher Education Executive Officers Association,*

Strand, S., Cadwallader, S. & Firth, D. (2011). Using statistical regression methods in education research. Retrieved from https://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/12/index.html

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics* (Sixth edition, Pearson new international edition ed.). Harlow: Pearson.

Texas Association of Community Colleges. (2017). The Texas pathways project. Retrieved from http://www.tacc.org/pages/texas-success-center/initiatives-and-programs

Texas Higher Education Coordinating Board. (2009). *Texas college and career readiness standards*

. Austin, TX: Texas Higher Education Coordinating Board.

Texas Higher Education Coordinating Board. (2016a). 60x30TX progress report - September 2016. *Retrieved from Http://Www.Thecb.State.Tx.Us/Index.Cfm?objectid=5033056A-A8AF-0900-DE0514355F026A7F,* Texas Higher Education Coordinating Board. (2016b). Closing the gaps final progress report 2016. *Retrieved from Http://Www.Thecb.State.Tx.Us/Index.Cfm?objectid=858D2E7C-F5C8-97E9-0CDEB3037C1C2CA3,*

Texas Higher Education Coordinating Board. (2017). Formula funding recommendations for the 2018-19 biennium. *Retrieved from* *http://www.thecb.state.tx.us/index.cfm?objectid=4EA741D3-C76D-FBC5-04F664C233E8802B*

The Business Higher Education Forum. (2010). *Increasing the number of STEM graduates: Insights from the U.S. STEM education & modeling project.* (). Washington, DC: BHEF. Retrieved from https://www.bhef.com/publications/increasing-number-stem-graduates-insights-us-stem-education-modeling-project

THECB. (2012). Closing the gaps by 2015.

Thomas, E., Reznik, G., & Dawes, W. Using predictive modeling to target student recruitment: Theory and practice. AIR 1999 annual forum paper. Paper presented at the *Thirty-Ninth Annual Forum of the Association for Institutional Research,* Seattle, WA.

Tierney, M. L. (1983). Student college choice sets: Toward an empirical characterization. *Research in Higher Education, 18*(3), 271-284. Retrieved from http://www.jstor.org/stable/40195537

Vrontis, D., Thrassou, A., & Melanthiou, Y. (2007). A contemporary higher education student-choice model for developed countries. *Journal of Business Research, 60*(9), 979-989. doi:http://dx.doi.org/10.1016/j.jbusres.2007.01.023

Wang, X. (2012). *Modeling student choice of STEM fields of study: Testing a conceptual framework of motivation.* Unpublished manuscript.

Glossary

1. Strategic Enrollment Management (SEM): the execution of policy and planning related to enrollment goals set by institutions of higher education. SEM is driven by the institution's mission, and is considered a cross-functional process, requiring participation from all functional areas. SEM utilizes advanced analytics and business intelligence in an effort to support and improve upon institutional efficiency and effectiveness, with a focus on recruitment, retention, and completion.
2. Yield: a critical metric for enrollment managers, yield represents the number of students who matriculate, to the number of students admitted by the institution. Yield is the observable representation of the institution's return on the investment of resources intended to optimize recruitment strategies.
3. Matriculate: a designation assigned to students who are admitted, enrolled in classes, and registered. Registration occurs when the student commits to the financial obligation incurred through enrollment in classes.
4. Enrollment Funnel: recognized as a tool utilized by enrollment managers for the purpose of segmenting a broad target population of potential students. The enrollment funnel is typically represented in the stages of the enrollment process. Such stages include *Suspect, Inquiry, Applicant, Admitted,* and *Enrolled.* The segmentation of prospective students provides enrollment managers the means to develop marketing strategies specific to these stages.
5. Data Analytics: comprised of methods, techniques, and processes utilizing data as a means of generating information to create knowledge and insight about the organization's opportunities, productivity, and performance.

Appendices

APPENDIX A.

Table A Categorical Variable Coding for Parental Educational Attainment

		Frequency	Parameter coding		
			(1)	(2)	(3)
FEdAttain	College-Beyond	519	0.00	0.00	0.00
	HS	404	1.00	0.00	0.00
	MS	145	0.00	1.00	0.00
	Other	88	0.00	0.00	1.00
MEdAttain	College-Beyond	389	0.00	0.00	0.00
	HS	417	1.00	0.00	0.00
	MS	141	0.00	1.00	0.00
	Other	209	0.00	0.00	1.00

APPENDIX B.

Table B Variables in the Equation – First Iteration

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	pQuartiles			9.27	3.00	0.03			
	pQuartiles(1)	0.62	0.33	3.65	1.00	0.06	1.86	0.98	3.52
	pQuartiles(2)	0.84	0.41	4.24	1.00	0.04	2.32	1.04	5.15
	pQuartiles(3)	1.54	0.58	7.19	1.00	0.01	4.69	1.52	14.50
	AvgOfClassSize	0.00	0.00	1.03	1.00	0.31	1.00	1.00	1.00
	FEdAttain			6.52	3.00	0.09			
	FEdAttain(1)	-0.01	0.29	0.00	1.00	0.98	0.99	0.56	1.76
	FEdAttain(2)	0.30	0.47	0.39	1.00	0.53	1.34	0.53	3.41
	FEdAttain(3)	1.78	0.72	6.09	1.00	0.01	5.95	1.44	24.51
	MEdAttain			2.11	3.00	0.55			
	MEdAttain(1)	0.12	0.31	0.15	1.00	0.70	1.13	0.62	2.05
	MEdAttain(2)	0.55	0.54	1.04	1.00	0.31	1.72	0.60	4.92
	MEdAttain(3)	-0.24	0.40	0.36	1.00	0.55	0.78	0.36	1.73
	DiplomaCode(1)	-0.45	0.37	1.46	1.00	0.23	0.64	0.31	1.32
	hsAverageYield	-0.01	0.02	0.32	1.00	0.57	0.99	0.96	1.02
	CompulsoryIntake(1)	-1.17	0.57	4.28	1.00	0.04	0.31	0.10	0.94
	ColCrdEarned(1)	6.49	0.47	193.34	1.00	0.00	658.14	263.67	1642.78
	AP(1)	0.41	0.29	2.00	1.00	0.16	1.50	0.85	2.64
	STEM(1)	0.91	0.27	11.48	1.00	0.00	2.49	1.47	4.21
	TsimathColRdy(1)	0.53	0.33	2.61	1.00	0.11	1.69	0.89	3.21
	TsiwritingColRdy(1)	-0.72	0.44	2.70	1.00	0.10	0.49	0.20	1.15
	TsireadColRdy(1)	0.07	0.38	0.04	1.00	0.85	1.08	0.51	2.26
	Constant	-3.48	1.15	9.14	1.00	0.00	0.03		

a. Variable(s) entered on step 1: pQuartiles, AvgOfClassSize, FEdAttain, MEdAttain, DiplomaCode, hsAverageYield, CompulsoryIntake, ColCrdEarned, AP, STEM, TsimathColRdy, TsiwritingColRdy, TsireadColRdy.

APPENDIX C.

Table C Variables in the Equation, Reduced Model, Second Iteration

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
pQuartiles			41.64	3	0.00			
pQuartiles(1)	1.34	0.26	26.70	1	0.00	3.81	2.30	6.34
pQuartiles(2)	1.33	0.28	22.22	1	0.00	3.77	2.17	6.54
pQuartiles(3)	1.88	0.34	29.73	1	0.00	6.53	3.33	12.83
AvgOfClassSize	0.00	0.00	0.49	1	0.49	1.00	1.00	1.00
DiplomaCode(1)	-0.36	0.28	1.59	1	0.21	0.70	0.40	1.22
hsAverageYield	-0.04	0.01	11.37	1	0.00	0.96	0.95	0.99
CompulsoryIntake(1)	-0.86	0.34	6.54	1	0.01	0.42	0.22	0.82
ColCrdEarned(1)	6.52	0.23	803.65	1	0.00	677.45	431.67	1063.15
AP(1)	0.87	0.22	15.57	1	0.00	2.39	1.55	3.68
STEM(1)	0.69	0.19	12.88	1	0.00	2.00	1.37	2.91
TsimathColRdy(1)	0.65	0.23	7.71	1	0.01	1.91	1.21	3.02
TsiwritingColRdy(1)	-0.56	0.28	4.12	1	0.04	0.57	0.33	0.98
TsireadColRdy(1)	0.24	0.25	0.93	1	0.33	1.28	0.78	2.09
Constant	-2.85	0.73	15.04	1	0.00	0.06		
a. Variable(s) entered on step 1: pQuartiles, AvgOfClassSize, DiplomaCode, hsAverageYield, CompulsoryIntake, ColCrdEarned, AP, STEM, TsimathColRdy, TsiwritingColRdy, TsireadColRdy.								

Appendix D.

Table D Variables in the Equation, Reduced Model, Third Iteration

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	pQuartiles			39.02	3	0.00			
	pQuartiles(1)	1.28	0.26	24.79	1	0.00	3.58	2.17	5.92
	pQuartiles(2)	1.25	0.28	20.43	1	0.00	3.50	2.03	6.03
	pQuartiles(3)	1.79	0.34	27.56	1	0.00	5.96	3.06	11.61
	hsAverageYield	-0.03	0.01	9.26	1	0.00	0.97	0.96	0.99
	CompulsoryIntake(1)	-0.88	0.34	6.91	1	0.01	0.41	0.21	0.80
	ColCrdEarned(1)	6.52	0.22	844.82	1	0.00	681.56	438.97	1058.21
	AP(1)	0.89	0.22	16.68	1	0.00	2.44	1.59	3.75
	STEM(1)	0.71	0.19	13.43	1	0.00	2.02	1.39	2.95
	TsimathColRdy(1)	0.72	0.22	10.31	1	0.00	2.05	1.32	3.18
	TsiwritingColRdy(1)	-0.45	0.26	2.99	1	0.08	0.64	0.38	1.06
	Constant	-3.59	0.54	43.96	1	0.00	0.03		

a. Variable(s) entered on step 1: pQuartiles, hsAverageYield, CompulsoryIntake, ColCrdEarned, AP, STEM, TsimathColRdy, TsiwritingColRdy.

APPENDIX E.

Table E Null Model, Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-0.519	0.036	204.540	1	0.000	0.595

Appendix E provides the results for the test of the null hypothesis of the null model.

H_0: B_0=0; H_A: $B_0 \neq 0$

In addition, Appendix E shows the null model (constant only) is a statistically significant predictor of the outcome (*p<0.000)*. This is true only 59.5% of the time as Step 0 shows the predicted odds of n*ot enrolling* is [Exp(B)] =0.595. That is, the predicted odds of not enrolling are 0.595, since our observed odds is 2,035/3,246=0.627. The baseline model has predictive power because of the large sample size (Strand et al., 2011).

APPENDIX F.

Table F Null Model, Variables not in the Equation

	Score	df	Sig.
pQuartiles	42.285	3	0.000
pQuartiles(1)	0.003	1	0.960
pQuartiles(2)	0.228	1	0.633
pQuartiles(3)	31.349	1	0.000
hsAverageYield	46.723	1	0.000
CompulsoryIntake(1)	82.739	1	0.000
ColCrdEarned(1)	2705.984	1	0.000
AP(1)	295.715	1	0.000
STEM(1)	198.176	1	0.000
TsimathColRdy(1)	318.785	1	0.000
Overall Statistics	2720.060	9	0.000

Vita

Gus Monzon attended The University of Texas at El Paso (UTEP), where he earned a Bachelor's of Business Administration, majoring in Marketing. He continued his education at UTEP, where he earned a Master's of Business Administration. His professional experience in higher education includes roles in Institutional Research, the El Paso Collaborative for Academic Excellence, Admissions, Recruitment, Orientation, Constituent Relationship Management, Enrollment Services One-Stop Shop, and most recently in Academic Resource Planning.

He completed training in the United States Naval Nuclear Power Program and serviced on board the USS Gurnard SSN 662. During his military service, he earned Letters of Commendation from Commanding Officer USS Gurnard and Commander Submarine Squadron 3, Good Conduct Medal, National Defense Medal, Naval Expeditionary Medal, Sea Service Ribbon (x2), and the Coast Guard Special Operations Ribbon.

He has served in local and state level committees to include the Texas Higher Education Coordinating Board: Apply Texas Advisory Committee (2007-2008); Texas Association of Collegiate Registrars and Admissions Counselors: High School Relations Committee (2005). Professional presentations include: *The Rise of the Informed Consumer and Student Return on Investment (2013).* National College Access Network (NCAN) National Conference, Co-presenter, Nashville, TN. *Texas Records Exchange Regional Workshop (2010).* The University of Texas at El Paso. *Enhancing Enrollment Services, Workflows, and Processes with Hobsons Connect (2010).* Hobsons Connect University, National Constituent Relationship Management conference. *Location, Location, Location -Applying SEM Best Practices at an Urban University (2010).* ACT 25th Annual Enrollment Planners Conference; Co-presenter, Chicago, IL.

Gustavo Monzon: gus.monzon@gmail.com