Predicting Nontraditional Freshman Retention Using Pre-enrollment Data

A Thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts at George Mason University

by

Bolorchimeg Boldbaatar
Bachelor of Science
Minnesota State University, 2005

Director: Dimiter Dimitrov, Professor
College of Education and Human Development

Spring Semester 2014
George Mason University
Fairfax, VA

DEDICATION

This thesis is dedicated to my family, my husband Bat-Erdene and my four kids Anu, Temuulen, Nomin and Ermuun, who supported me unconditionally and believed in me no matter what. I would also like to dedicate this thesis to my father and my father-in-law. One of them encouraged me to go back to school again and one of them supported me financially and made this all possible. Thank you both; I'm here today because of your help.

ACKNOWLEDGEMENTS

First of all, I would like to acknowledge my thesis committee. Thank you to my committee chair, Dr. Dimiter Dimitrov, for infinite support and guidance throughout this process. Having you as my chair gave me the courage and comfort that I needed. Thank you so much to my committee members, Dr. Kelly Schrum and Dr. Lori Bland. Without your insights and feedback, I would have never accomplished this much in such a short time. I'm very grateful that you agreed to work with me and support me. Second, I would like to thank my former statistics teacher at PITT, Dr. Najeeb Shafiq, who helped open my eyes to educational research and inspired me to study in this field. I would also like to thank my oldest daughter Anu, who stepped up to take responsibility of household chores and take care of her brothers and sister. Without your help, I would not have accomplished this successfully. I am proud to share my achievements with you - my straight 4.0 GPA and finishing my degree on-time. I also hope that this experience has made you want to accomplish more than I achieved and that you've learned from my mistakes along the way.

TABLE OF CONTENTS

Page

List of Tables vii

List of Figures viii

List of Equations ix

Abstract x

Chapter One 1

Definition of Nontraditional Students 2

Background of Study 5

Problem Statement 8

Purpose of Study 12

Hypothesis and Research Questions 14

Significance of Study 15

Limitations 15

Key Terms 16

Chapter Two 17

Student Retention Theories 17

Study of First-Year Retention 23

Predictive Variables 24

Characteristics of Nontraditional Students 27

Chapter Three 32

Setting of the Study 32

Sample 32

Variables 35

Research Design 37

Logistic Regression 39

Chapter Four 43

Description of the Dependent Variable 43

Description of Independent Variables 44
Missing Data Handling 48
Logistic Regression Analysis 49
Chapter Five 63
Discussion of the Findings 63
Limitations 68
Directions for Future Research 70
Conclusions 71
Appendix A 75
References 76

LIST OF TABLES

Table Page
Table 1. Attrition Rate of Different Institutions 9

Table 2. List of Predicted Variables 13

Table 3. Characteristics of Nontraditional Students Statistics 33

Table 4. Predictive Variables from NCES: 88/2000 35

Table 5. Levels of Nontraditional Students 44

Table 6. Percentage of Nontraditional and Traditional Students in High School Variables 45

Table 7. Percentage of Nontraditional and Traditional Students in College Variables 46

Table 8. Logistic Regression Analysis of Student Retention as a Function of Ten High School Variables 50

Table 9. Logistic Regression Analysis of Student Retention as a Function of Ten Variables in the Last Model 52

Table 10. Classification Table for 10 Variable Model 53

Table 11. Comparison of models of nontraditional and traditional students 56

Table 12. Classification Table 58

Table 13. Comparison of Two Models 60

LIST OF FIGURES

Figure Page

Figure 1. Attrition across the Student Lifecycle based on 21 selective universities..... 8

Figure 2. Percentage of institutions that track their retention and completion of nontraditional students. 11

Figure 3. Comparison of retention rates between traditional and nontraditional students 34

Figure 4. Blockwise logistic regression 38

Figure 5. Hierarchical logistic regression 39

Figure 6. Classification table interpretation. 42

Figure 7. Missing value statistics 48

Figure 8. ROC curve for the last model 54

Figure 9. Relationship of retention and predictors 55

Figure 10. Correlation Matrix 57

Figure 11. The percentage of explanation of each variable in the last model.. 61

Figure 12. Relationship of type of degree variable with levels of nontraditional students and traditional students 62

Figure 13. Logistic regression, random forest, and ensemble in HS_GRADE and HS_UNITS variables 72

LIST OF EQUATIONS

Equation Page

Equation 1 Logistic regression model..... 41

Equation 2. Odds for success..... 41

Equation 3. Odds Ratio Equation..... 42

ABSTRACT

PREDICTING NONTRADITIONAL FRESHMAN RETENTION USING PRE-ENROLLMENT DATA

Bolorchimeg Boldbaatar, M.A.

George Mason University, 2014

Thesis Director: Dr. Dimiter Dimitrov

This thesis studies the pre-enrollment variables that predict nontraditional freshman retention and includes data from the National Education Longitudinal Study (NELS: 88-2000). Pre-enrollment variables were selected based on the availability of variables in the beginning of the fall term. This thesis examines the pre-enrollment of 10 high school and 15 college variables, and levels of nontraditional students. Blockwise logistic regression was used to determine most predictive variables. The findings identified three high school variables (high school grade, attendance, and Carnegie units) and seven college variables (enrollment status, number of math and science classes, type of degree, hours of employment, campus job, grants, and marital status). After controlling these variables, it was found that the level of nontraditional students has a unique effect on student retention. The result of this finding supports previous research of nontraditional student retention: The higher the number of nontraditional characteristics that students possess, the less they are likely to be retained. These findings are discussed

in terms of measurement of the variables, handling of missing data, and logistic regression analysis combined with data visualization. Implications of future research emphasize the importance of studying nontraditional students without any age restriction.

CHAPTER ONE

> *"The word "college" tends to call to mind images of fresh-faced young students studying, living and, yes, partying on or near leafy suburban campuses. But that picture only describes a small fraction of the nation's 18 million undergraduates…*" (Casselman, 2013)

This quote from a recent *Wall Street Journal* article by Ben Casselman demonstrates that only 29 percent of today's undergraduate students are now comprised of traditional students. This type of student enrolls in college full-time directly after high school, depends on his or her parents for financial support, and has limited work and family obligations. However, most of today's students do not fit this characterization. In fact, the 2000 National Center for Educational Study (NCES) found that 73 percent of undergraduates identified with at least one characteristic of nontraditional students (Choy, 2002). Among these students, almost half (47 percent) are independent students who are 24 years of age or older, married, and have dependents (Center for Law and Social Policy [CLASP], 2011). Forty-six percent of them are enrolled part-time and 40 percent are low-income students, so they are most likely to work full-time (32 percent) or part-time (43 percent) to support their undergraduate studies (CLASP, 2011).

Even though the population of nontraditional students is growing larger and their retention and completion issues are rising, the studies focused on these types of students are few and limited. There are two main reasons. First, the definition of nontraditional students is constantly shifting, so it is difficult to identify such students. Thus, this study

attempts to combine the various definitions of nontraditional students from literature on higher education (see section entitled “Definition of Nontraditional Students”). Second, calculating nontraditional students is far more complex than calculating traditional students. Institutions of higher education regularly report their first-time, full-time students. They do not, however, typically report students who do not fit this category, such as students with transfer credits, students who enroll part-time, or those who begin in the spring semester. Usually these large, diverse groups are left out of national longitudinal data sets, so it is more challenging for researchers to accomplish meaningful analysis (Advisory Committee on Student Financial Assistance [ACSFA], 2012). The complex characteristics of nontraditional students are further complicated because one student can fall into several different subgroups. For example, part-time students usually fall under full-time employment and financially independent subgroups at the same time. For these reasons, defining nontraditional students is very important to establish a common understanding for this study.

Definition of Nontraditional Students

Literature on higher education fails to define nontraditional students in a consistent way. The researchers who study nontraditional students usually define only one or several subgroups of these students. After comprehensive review, Kim (2002) combined three distinctive definitions of nontraditional students: age, background, and at-risk characteristics.

The Traditional Definition: The Age Criterion

The most common criterion for nontraditional students is age. Adult students who are 25 years of age or older are most commonly studied among nontraditional students. These students demand a distinctive learning curriculum, flexible scheduling, online learning options, and other services such as childcare, longer library hours, and availability of night or Saturday classes. However, age alone fails to identify important characteristics of these students, such as patterns of attendance, reasons for pursuing a college education, challenges, and resources (ACSFA, 2012). For example, younger students may also take on regular, full-time work during the academic year and can have legal dependents.

The Second Definition: Student Background Characteristics

The general understanding of nontraditional students refers to students who are historically underrepresented in, or traditionally excluded from, higher education. Examples include minority students, veterans, students with disabilities, and recent immigrants (Schuetze & Slowey, 2002). Other background characteristics include socioeconomic status, ethnicity, first-generation students, and employment status (Rendón, 1994). These types of students are unfamiliar with a university environment, struggle to adjust to university regulations, terms, and policies, and have a hard time being socially engaged in the school atmosphere. In this study, three background characteristics of nontraditional students are included: economically disadvantaged students, first-generation students, and racial/ethnic minority students (including Black, Hispanic, and Native American students).

The Third Definition: Characteristics of At-Risk Students

One of the most widely used definitions for nontraditional students comes from the NCES study of nontraditional students' persistence in higher education. In that study, nontraditional students are defined through seven possible at-risk characteristics: Delaying college enrollment, attending part-time, working at least 35 hours per week, financial independence, single parenthood, having dependents, and failure to receive a high school diploma (Choy, 2002). If the students have one of these nontraditional characteristics, they are considered "minimally nontraditional"; if students have two or three of these characteristics, they are considered "moderately nontraditional"; and if students have more than four of these characteristics, they are considered "highly nontraditional."

Along with the three definitions above, the ACSFA's 2012 report also defines a nontraditional student as any student who fails to fit into the traditional student template. The committee provides an additional list of subgroups and labels that help define nontraditional students (see Appendix A). The report further summarizes nontraditional students as "categorized across the dimensions of age, marital status, family size and composition, level and type of employment, and educational preparation and goals," and states that "this population – often referred to as 21st century or contemporary students – consists of many subgroups, each with unique circumstances, educational needs, and goals," (ACSFA, 2012, p.5).

While this broad term helps to identify nontraditional students, this definition is very general, making it difficult to study nontraditional students. Thus, this study will

only consider seven of the above characteristics of nontraditional students: first generation, race/ethnicity, low socioeconomic status, married, part-time enrollment, full-time employment, and having a non-standard high school diploma.

Background of Study

Undergraduate retention rate is one of the key measurement tools for higher education institutions. Along with completion rate, retention rate is a success outcome used to rank the quality of universities and colleges. Despite the increased effort of institutional practices and development of retention theories, the first-to-second year retention rate has remained mostly consistent for all post-secondary institutions over the past 20 years. According to a College Board study (2009), the average first-to-second year retention rate at both public and private institutions was 73.7 percent in 2007, which is lower than the 1988 rate of 74.5 percent (p. 4).

In recent years, both internal and external pressures have forced institutions to attempt to increase their student retention rates (Hossler, Ziskin, Moore, & Wahungu, 2008). Three main external pressures of the higher education industry have influenced or soon will be influencing the retention of nontraditional students.

Big Data and Predictive Analytics

Following the examples of many businesses today, higher education institutions have started using predictive analytics and data mining to accelerate the process of analyzing student data and to help institutions make data-driven decisions. One of the

most common retention practice today is to identify possible at-risk students who are likely to drop out after their midterm grades are posted and to intervene to promote retention. The problem with this solution is that it is usually too late to retain at-risk students. Using predictive analytics, the chances of targeting the right at-risk students at the right moment will significantly increase (Nandeshwar, Menzies, & Nelson, 2011) and administrators, faculty, and students can see real-time data of students' risk of dropping out.

State Funding Decline

Since the beginning of the 2008 economic downturn, states have cut their funding for state universities and colleges. As a result, many public and some private universities and colleges have been forced to increase tuition significantly. This phenomenon has placed pressure on institutions by questioning their accountability, raising issues of cutting back on resources without exhausting quality, and retaining more students than enrolling. The burden of this change mostly falls on the shoulders of students and their parents. Institutions that could offset their higher costs by increasing traditional student enrollments now have to look elsewhere. Lane (2013) suggested that "many higher education institutions are bound to lose enrollments unless more significant attention is paid to nontraditional students or recruiting students from outside of the region" (p.1).

Changing Demographic and Mobility

A central challenge for higher education institutions is changing student demographics. Adding to the complicated characteristics of nontraditional students is the

fact that these students have become very mobile. Students who enroll on a part-time basis can easily switch to full-time, or vice versa. It is much simpler to transfer from one institution to another. Students can take a break as many times as they want and they can still come back and continue their studies. This kind of freedom allows students to "swirl" between institutions, within institutions, and from semester to semester (Borden, 2004).

Many theories and models are dedicated to the importance of the relationship between students and institutions. For example, the most important and widely cited theory is the Student Interaction Theory (Tinto, 1975). This theory suggests that early and continued institutional commitment will impact both academic and social integration within the university (Berger, Ramirez & Lyons, 2012) and also establishes the basic foundation of how students interact with the institution. As colleges continue to diversify, the retention of minority groups, first-generation students, and those from lower socioeconomic backgrounds needs to be studied. Researchers focused on nontraditional students have revealed that the factors that are influential to heterogeneous students (different individuals distinguished on the basis of their values on the predictive factors) are different than factors for homogeneous students (Hu & St.John, 2001; Johnson et al, 2007). There are more factors raised as retention becomes more complex. "Many of the reasons that students leave college are outside of Tinto's model: finances, poor academic performance, lack of family or social/emotional encouragement, difficult personal adjustment" (De Witz, Woosley, & Walsh, 2009 p.20).

In an effort to retain students, universities need to identify at-risk students who are likely to drop out before graduation. To identify at-risk students as early as possible, the practice of predictive analytics is one approach that needs to be implemented. In order to predict retention, institutions need to determine the most predictive factors that contribute to their student retention. Many of these variables will be addressed in this study to explore nontraditional freshman student retention.

Problem Statement

Over the last four decades, great efforts have been made to improve student retention based on numerous studies and theoretical models. However, despite these efforts, retention rates have remained mostly consistent. Many of these studies show that the first year of college is the most critical to retaining college students (Tinto, 1993). The following graph shows the comparison of attrition rates of students in their first year of college compared to other years in college (Figure 1). Unlike traditional students,

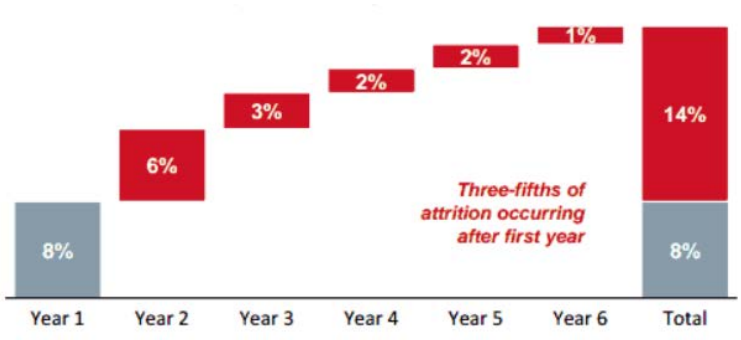

Figure 1. Attrition across the Student Lifecycle based on 21 selective universities.
Source: Education Advisory Board [EAB], 2012

nontraditional students have a number of obligations and responsibilities beyond school. Thus, they are more vulnerable during their first year of college and are more likely to drop out (NCES, 1999). More than one-third of attrition occurs in the first year of college. However, an attrition rate of eight percent is only telling one side of the story because it is based on only a few selected universities. The following table is adopted from Tinto's book *Leaving College* and shows average attrition rates for different types of institutions (Table 1). For open enrollment institutions, the attrition rate reached 45.5 percent. These institutions usually enroll large numbers of nontraditional students, so that can be one of the reasons of higher attrition rate.

Table 1. *Attrition Rate of Different Institutions*

Institution Selectivity	**SAT Scores**	**Mean %**
Highly selective	SAT >1100	8.0%
Selective	SAT 931 - 1099	17.5%
Traditional	SAT 801 – 930	26.4%
Liberal	SAT 700 - 800	32.9%
Open	SAT < 700	45.5%

Note: Adapted from *Leaving College* by Tinto, 1993, p.16

The main practice of predicting retention for at-risk students starts after fall midterm grade submission, which is about six to eight weeks into the semester. This is usually the time when administrators see the initial signals of students who might drop out. To identify such at-risk students, institutions practice predictive analytics. Many

retention studies found that college GPA is the most important variable to predict freshman student retention. However, administrators do not have such information available until the end of the semester. The current practice does not allow faculty and staff enough time to assist at-risk students before the fall semester ends. By the time a student is identified, he/she has typically already made his/her decision about leaving a college. In addition to midterm grades, administrators use student engagement surveys, such as the National Survey of Student Engagement (NSSE), or internally developed surveys to identify at-risk students. These surveys are very helpful to understand social or behavioral analysis of students. However, the problems with student surveys are that they are difficult to construct and manage and the participation rates are very low. Therefore, the survey results can be biased towards students who are more active and engaged, not the students who need support. Therefore, this study will only focus on institutional databases that are available at the beginning of the term.

When Tinto explained Table 1, he stated that "these data reported are for full-time, first-time students only, they excluded information about who are part-time, non-matriculated, or transfers from other institutions" (p.15). These publicly reported retention rates account for only first-time college students who enroll as full-time students at the beginning of the fall semester; these rates do not account for part-time students, transfer students, students who start in the spring semester, or students who return back to school after spending a period of time away. According to the American Council on Education, "about 61 percent of students at four-year schools and 67 percent of those at two-year institutions are simply excluded from the calculation" (Hartle, 2011).

The lack of recognition for these students is the main reason higher education institutions make less of an effort to retain nontraditional undergraduates. The NCES study revealed that 38 percent of nontraditional undergraduates left school in their first year as opposed to 16 percent of traditional undergraduates (Horn & Carroll, 1996). Additionally, studies from UPCEA and InsideTrack (2013) found that the percentage of institutions that track retention rates for nontraditional students has increased from 57 percent to 69 percent (Figure 2). However, of those 69 percent that claimed to track retention rates for nontraditional students, only 22 percent were able to provide their current retention rate for nontraditional students.

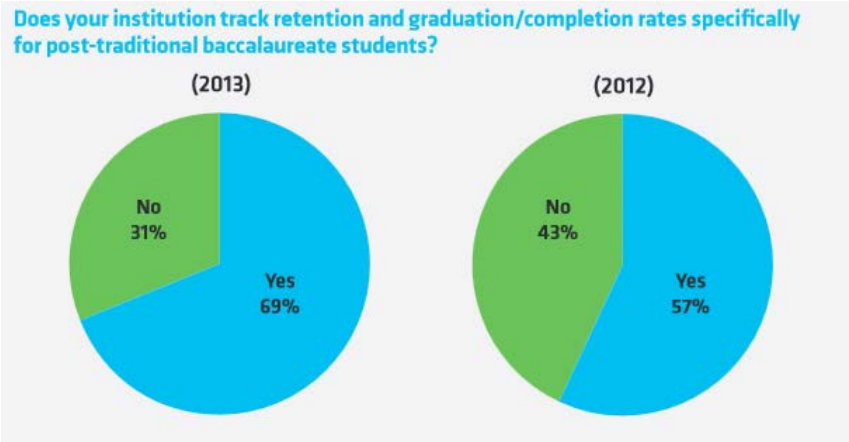

Figure 2. Percentage of institutions that track their retention and completion of nontraditional students. Source: UPCEA and InsideTrack (2013)

Unlike many institution-specific studies, Adelman's study (2006) highlights the importance of differentiating the analysis of nontraditional students from traditional student population. He suggested that researchers "not mix your 18-year-old daughter and

your 34-year-old brother-in-law in the same analytic of degree completion" (p.96). And most studies that identified predictive factors focus more on degree completion rather than retention of first-year students. Thus, there is a need to distinguish characteristics of nontraditional students when exploring predictive factors for determining freshman student retention.

Purpose of Study

The purpose of this study is to identify the predictive variables that are most important to retaining nontraditional freshman students. The characteristics of nontraditional students included in this study are:

- Three background characteristics: first generation, race/ethnicity, and low socio-economic status; and
- Four at-risk characteristics: part-time enrollment, full-time employment, marital status, and failure to receive a standard high school diploma.

These characteristics are classified into three categories: minimally nontraditional (when a student possesses one characteristic), moderately nontraditional (when a student has two or three characteristics), and highly nontraditional (when a student has more than four characteristics).

The main criterion for the selection of the variables in this study is restricted to information available for admissions offices and university administrators at the time of initial enrollment. Those variables include high school variables that institutions regularly collect during the admissions process and college variables that are available in the beginning of the fall semester, such as choice of degree, enrollment status, in-state or out-

of-state status, and financial aid information. The variables that are included in this study are presented in Table 2.

Table 2. *List of Predicted Variables*

Predictive Variables	
High School Variables	
1	Standard test score
2	Advanced placement course
3	High school grade
4	Preparation of college
5	Family income
6	Number of Carnegie units
7	Number of college exams taken
8	Participation in extracurricular activities
9	Marital status
10	Class attendance
College variables	
1	Extra-curricular activities in college
2	Remedial courses taken
3	Any college level math and science classes
4	Loans
5	Grants/Scholarships
6	Work-Study
7	No financial aid
8	Parents borrow
9	Campus job
10	Enrollment status
11	Type of tuition
12	Type of degree
13	Marital status
14	Hours of work weekly
15	Type of institution

This study uses data from the National Education Longitudinal Study of 1988 (NELS: 88/00), which followed eighth-grade students from 1988 to 2000. This national

study was chosen for this research because it is the most comprehensive, large-scaled study that is freely and publicly accessible, it captured data for first-year college student performance, and it captured a large sample of nontraditional student data.

Hypothesis and Research Questions

The following research questions and hypotheses are addressed in this study:

Research Question 1
Do the high school and college variables identified in this study predict retention of nontraditional freshman students?

Hypothesis 1:
College variables have better predictive power than high school variables alone.

Research Question 2

a) Does the level of nontraditional students predict retention of nontraditional freshman students?
b) Does the level of nontraditional students have a unique effect in the prediction of retention controlling for high school and college variables?

Hypothesis 2a:
The level of nontraditional students predicts the retention.

Hypothesis 2b:
The level of nontraditional students has a unique effect in the prediction of retention.

Significance of Study

The significance of this study is to contribute to the limited literature regarding important factors for retention of nontraditional freshman students. These studies are limited because they only use institutionally available data and compare the results with data on traditional students. Because this study uses national data of nontraditional students, it will extend the knowledge of retaining nontraditional students in general.

Limitations

Data used for this study is from NELS:88/00, a national data set conducted almost a decade ago. Student populations from that time period to the present have so many distinctive characteristics that it would not be wise to use this data alone to make generalized statements about today's student population. However, there are sufficient numbers of nontraditional students from this study that can make inference for this type of student population.

The NELS: 88/00 data is based on eighth-grade students from 1988. Due to this limitation, it is impossible to include age characteristics of nontraditional students (adult students) in this study. Older students are a large part of the nontraditional student population and this characteristic has been used as a main criterion for defining nontraditional students. Even though this is an important characteristic, far more studies have been done on age only. Therefore, exploration of other characteristics of nontraditional students is more meaningful to expand the literature on the subject.

Key Terms

Names	Definitions
Academic Integration	Academic integration represents both satisfactory compliance with explicit norms, such as earning passing grades, and the normative academic values of the institution (Kuh et al., 2006).
Attrition	Longitudinal process of interactions between the individual and the academic and social systems of the college during which a person's experiences in those systems…continually modify his goals and institutional commitments in ways which lead to persistence and/or to varying forms of dropout (Tinto, 1975).
Dropout	A student who is not enrolled, has not earned a degree, and is no longer pursuing a degree.
Nontraditional Students	Any student who fails to fit into the traditional student template (ACSFA, 2012).
Predictive Analytics	A set of business intelligence technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events (Woosley and Jones, 2011).
Retention	A rate or percentage of students who return from one enrollment period to another (Habley et al., 2012).
Social Integration	Social integration represents the extent to which a student finds the institution's social environment to be congenial with his or her preferences, which are shaped by the student's background, values, and aspirations (Kuh et al., 2006).
Underrepresented Groups	Persons who are minorities, including African American, Hispanic, Asian American, American Indian or Alaskan Native, women, persons with disabilities, and students identified as first-generation college participants.

CHAPTER TWO

The topic of student retention is well developed and a number of perspectives concerning retention have been researched. Most researchers rely on Tinto's Integration Theory, Astin's Student Involvement Theory, or Bean's Student Attrition Theory for their retention studies. This literature review takes a different approach by following the example of *Increasing Persistence* by Habley, Bloom, and Robbins (2012). Theories discussed in this book are categorized into five main perspectives: sociological, psychological, organizational, economic, and cultural. This approach helps make sense of theories and find connections with other studies. In addition to the examples of theories, the discussion of different predictive factors for student retention follows. Most of the factors examined in this study will involve the first four perspectives – sociological, psychological, organizational, and economic; this is the focus of the literature review. Moreover, the importance of studying students' first-year of college is deliberated. Finally, studies of nontraditional students and their characteristics are reviewed.

Student Retention Theories

This section explores the four different perspectives of student retention and how different theories are associated with each perspective.

Sociological Perspective

The sociological perspective is the most influential concept in student retention theories (Habley et al., 2012). The main theory that introduced the sociological perspective is Tinto's interactionist theory. Tinto's predictive model (1975) was based on his theory that integration is the key to solving student attrition problems. As students are more integrated into the social life of their institution, their commitment increases, which encourages their continued enrollment and academic achievement. There are two different dimensions of students' integration within institutions: social and academic. Academic integration occurs when a student is academically engaged with faculty and research. Social integration occurs when a student finds that the institution's social environment fits his or her preferences, which are shaped by the student's background, values, and objectives (Kuh, Kinzie, Buckley, Bridges, and Hayek, 2006). Social integration is often composed of peer-to-peer interactions and faculty-student interactions. Integration in turn is influenced by pre-college characteristics and goals, interactions with peers and faculty, and out-of-classroom factors.

Despite its popularity, empirical studies based on Tinto's model have ended with contradictory results. Half of the studies examined by Braxton and colleagues do not show a direct relationship between retention and academic integration. However, these studies do show that "increasing social integration leads to greater institutional commitment and thus greater likelihood of persistence to graduation" (Kuh et al., 2006, p12). Many of the studies are based on a single institution. Therefore, the results cannot be generalized across all institutions. Additionally, nontraditional students do not socially

engage with peers and faculty as much as traditional students because they usually do not live on campus, their part-time attendance reduces the time of socializing while on campus, and they have other obligations and social contacts through employment and family (Bean and Metzner, 1985).

Psychological Perspective

The psychological perspective is not as dominant as the sociological perspective, yet student characteristics are a significant determinant of student decisions. Before the sociological perspective was popular, the psychological perspective was the main explanation for a student's decision to leave college (Kuh et al., 2006). The lack of student attributes, skills, and motivation were considered the main reasons for a student's decision to drop out. Although introduction of the Student Interactionist Theory shifted the view of student retention to the role of institutions, individual psychological processes are still closely related to retention decisions (Habley et al., 2012). Bean and Eaton (2001) introduced a psychological model to explain such phenomenon based on four psychological theories: attitude-behavior theory, coping behavioral theory, self-efficacy theory, and attrition theory. The model did not exclude the influence of the institution as in the previous view, but added psychological factors to the academic and social integration. Psychological factors - self-efficacy, coping behavior, and locus of control - were introduced and are shaped by an institution and its representatives in the bureaucratic, academic, and social realms (Bean and Eaton, 2001).

Astin's theory of student involvement suggests that students who spend more time and energy on their academic experience and on-campus activities are more likely to stay

in college and persist. "A highly involved student is one who, for example, devotes considerable energy to studying, spends much time on campus, participates actively in student organizations, and interacts frequently with faculty members and other students," (Astin, 1987). In a later study, Astin (1997) used a sample of more than 200,000 students and examined more than 80 different types of student involvement, including place of residence, honors programs, undergraduate research participation, social fraternities and sororities, academic involvement, and student-faculty interaction. In that study, he found that "nearly all forms of student involvement are associated with greater than average changes in entering freshmen characteristics," (Astin, 1997, p. 524).

Kuh's (2005) Student Engagement Theory emphasizes that student engagement is the key for student success because it represents both the time and energy students invest in educationally purposeful activities and the effort institutions devote to using effective educational practices. His studies confirm that student engagement, including a "range of behaviors that institutions can influence with teaching practices and programmatic interventions such as first-year seminars, service-learning courses, and learning communities," has a positive relationship with, and compensatory effect on, a student's first- and last-year grades as well as persistence to the second year (Kuh, 2008).

Organizational Perspective

The organizational perspective is more closely related to institutional factors and a student's relationships with institutional members such as faculty, staff, and administrators. Institutional factors include support services, academic advising, accessibility of faculty members, availability of financial aid, institutional fit, campus

size, and facilities (Hably et al., 2012). Bean's Student Attrition Model was the first model that observed organizational effects on individual students and introduced variables such as institutional fit, institutional quality, staff/faculty relationship, campus organizations, and institutional commitment. Bean's model is important because it provides an opportunity to see "how students perceive organizational processes as perceptual precursors to persist rather than examining campus-wide retention rates …" (Berger, 2001).

There are some studies that combine organizational features from Bean's model with Tinto's interactionist model. One such study found that three organizational features - participation, communication, and fairness - were significant alternatives to social integration (Berger, 2001). Institutional culture, including habits of faculty interactions with students outside the classroom, has a significant role in Tinto's model of academic and social integration (Adelman, 2006).

Economic Perspective

The economic perspective examines how students weigh the costs and benefits of staying in college and participating in various activities. However, there are no particular retention theories that focus on the economic perspective. Most researchers refer to it as the Human Capital theory. If the cost of studying at a college exceeds the perceived financial benefit of obtaining post-secondary credentials, students are more likely to drop out (Hably et al., 2012). The variables examined for retention studies are tuition cost, loans, grants, and scholarships.

Combined Perspectives

There are models that combine some of the previous concepts to explain some important perspectives. Bean and Metzner (1985) introduced the Conceptual Model of Nontraditional Student Attrition that defined a nontraditional student as "older than 24, or does not live in a campus residence or is a part-time student, or some combination of these three factors; is not greatly influenced by the social environment of the institution; and is chiefly concerned with the institution's academic offerings" (Bean & Metzner, 1985). The model has four sets of variables explaining the persistence of nontraditional students: academic performance, intent to leave, background and defining variables, and environmental variables. An environmental variable is one factor that is different from previous models and includes finances, hours of employment, family responsibilities, and transfer opportunities. In addition to these four factors, compensatory effects are also highlighted, revealing that environmental factors compensate for poor academic variables. For example, a student with strong family or employment support is likely to persist despite poor academic grades or uncertainty of major. However, this effect does not work the other way around; when a student cannot arrange childcare or find adequate work schedules, he/she will not stay regardless of a good academic record. Therefore, environmental support compensates for weak academic support, but academic support will not compensate for weak environmental support. Even though the Conceptual Model of Nontraditional Student Attrition is well established in the literature, it is limited to certain types of nontraditional students (ACSFA, 2012).

Study of First-Year Retention

Research toward student retention found that it is important to study first-year retention because this is the most critical time period for freshman persistence. Tinto (1993) revealed that approximately three-fourths of most attrition happens some time during the freshman year. And most of these students leave during the first six weeks of the fall term (ACT, 1992). Thus, many early retention studies were highly focused on students' first-year persistence and the transition to college. However, data from these studies were drawn from traditional students attending large residential universities. Therefore, researchers involved in these studies could not "speak to the experience of students in other types of institutions…and of students of different gender, race, ethnicity, income, and orientation," (Tinto, 2007, p.3).

There are not many comprehensive studies of nontraditional student retention that focus on students' first-year of college. The National Center for Education Statistics' 1992-93 nontraditional undergraduates study found that nontraditional students are more than twice as likely to drop out than their traditional counterparts in their first year of college (NCES, 2012). The study also determined that the gap is closed considerably once students reach their second year. Among degree-seeking students, associate degree students are less likely to persist in their first year and far more likely to be highly nontraditional when compared to their counterparts pursuing bachelor degrees (NCES, 2012).

Predictive Variables

Generally, administrators divide those students who are at risk of dropping out into three categories: academically challenged, not socially engaged, and financially at-risk (Beaudoin & Kumar, 2012). Academic failure of students is usually the result of factors such as college unpreparedness, class absenteeism, and insufficient academic support. Older students are more likely to have a difficult time becoming socially engaged in the school atmosphere. “About two-thirds of incoming students said they had “some” or “major” concern about their ability to pay for their education,” (Zernike, 2010, p.1). Therefore, predictive variables are categorized into three factors in this literature review: academic, financial, and social.

Academic Factors

Many retention studies focus on academic factors because these factors are easier to collect and more readily available. Academic factors prior to college include most or all high school data such as class rank, high school GPA, and SAT/ACT scores. Research suggests that a student’s pre-college academic preparedness has a significant correlation with the student’s persistence (Caison, 2007). High school GPA, SAT score, taking AP and high-level math courses, and preparing for college exams are also significant variables when predicting retention (Kuh, et al., 2006). Adelman’s 1999 study on NS78 discovered that the level of math preparation in high school is the single most important variable in college preparation. Earlier studies of NELS:88 found that students who obtained help with the college application process had a stronger chance of

persistence (NELS, 2012). Some studies suggest that there is a correlation between high school GPA and a student's decision to leave college, but Seidman argues that this is not a good predictor (2012). Astin and his colleagues (2003) found that high school GPA and test scores account for only 12 percent of the variance in retention.

Academic factors after a student's entry into college include college GPA, fewer than 20 credits completed in the first-year, number of remedial courses taken, and level of math courses taken. From these variables, college GPA is the single most important variable (Tinto, 1997; Cabrera et al., 1993; Hu and St. John, 2001). DesJardins, et al. (2002) found that college GPA is a more powerful predictor than students' previous academic background. They concluded that it does not matter what academic resources students possess prior to college, but it does matter how they use these resources during their college years. Herzog (2006) found that a student's first-year performance in math is the second most important predictor after the college GPA. Students who took remedial courses in their first year showed that they are not completely prepared for college. However, Atwell and fellow researchers (2006) found that students who successfully completed those remedial courses had a higher chance of graduating than similar students who did not take those courses.

Financial Factors

Studies on financial factors typically include financial aid information. However, prior college status requires different variables. For such purposes, research includes factors such as students' socioeconomic status, family income information, and first-generation status. Walsh (1997) found that students' socioeconomic status is the most

important factor in retention and graduation rates. The higher the socioeconomic status, the more likely students are to be retained and graduate. Other studies show that low-income students (Hu and St. John, 2001; Choy, 1999) and first-generation students (Ishitani, 2003; Chen & Carroll, 2005) have a higher risk of dropping out of college after their first year.

Financial aid, grants, loans, and work-study information are part of college-related factors. Most of these factors are very effective predictors, but financial aid proved to be the most predictive factor of all. Students with financial aid have a higher chance of being retained than students who have no financial aid. Work-study programs are also effective forms of financial aid. Students in work-study programs are more likely to be engaged academically and socially and to stay in college.

Social Factors

Predicting whether or not students are socially challenged before a term begins is a difficult task. Most studies use student surveys to determine this information. These surveys often include questions about whether students feel they belong to the school and whether they are satisfied with their social life on campus. There are no attempts to predict whether or not a student is socially engaged prior to college entry. However, this study attempted to identify social factors using students' extracurricular activities in high school, in/out of state status, and marital status. Participation in co-curricular activities at school is a significant measure that a student is socially engaged in school. Students who are engaged in such activities and athletics are retained at higher rates (Beaudoin & Kumar, 2012). Students with other obligations, such as spouses, dependents, or working

full/part-time have a higher chance of dropping out of school (NCES, 2012). Students who are studying far away from their home often have a harder time adjusting to a different environment, so homesickness and loneliness can be reasons for leaving school for out-of-state students. Additionally, studies showed that low-income and first-generation students are more socially vulnerable (Torres, 2003).

Characteristics of Nontraditional Students

There are not many studies on the retention of nontraditional students. The main reason is that nontraditional students possess a very wide range of characteristics. There's a tendency to perceive nontraditional students only as adult students (ACSFA, 2012). However, nontraditional students have many other characteristics:

Enrollment criteria
- Delayed enrollment
- Part-time enrollment

Financial and family status
- Financially independent
- Full-time employment while enrolled
- Have dependents
- Single parent

High school graduation status
- Did not receive standard high school diploma (Horn, 1996).

Adelman's study on national student completion (1999) was influential among the educational community. His findings were important to studies of first-year retention. Most important of all, his study confirmed that nontraditional student characteristics are a large part of today's students and that educators should not miss calculating these

characteristics. For example, he found that 64.8 percent of the students he studied who attended a four-year college also attended more than one institution at some time; 26 percent attended more than two institutions; more than 60 percent of the students in the sample enrolled during summer terms; and 75 percent of students at four-year institutions were working.

Racial/Ethnic Differences in Students

Many retention studies reference African and Hispanic American students when discussing racial and ethnic groups. These studies show that these groups have lower retention rates than White students. They also have lower odds of completing high school and enrolling in college (Kuh et al., 2006). Carter and Wilson (1997) reported that the high school completion rate was 77 percent for African Americans, 57 percent for Hispanics, and 82 percent for Whites. On the other hand, college participation rates were 35 percent for both African and Hispanic American students and 43 percent for Whites. If this gap between education levels of Whites and other racial and ethnic populations continues to increase, the completion rate will decline over the next 15 years (Kuh et al., 2006).

Studies show that Hispanic and African American students are less likely to be retained than White students. However, these group differences diminish greatly after controlling for factors like high school achievement and socioeconomic status (Hanover Research [HR], 2011). Hispanic and African American students are less likely than White students to take advanced math courses and college placement exams (Hoffman, Llagas, & Snyder, 2003).

First-generation Students

Approximately one-third of college students are characterized as first-generation students; they are more likely to be female, to be older, to have lower incomes, to be married, and to have dependents (Kuh et al., 2006). Earlier studies found that first-generation students are less likely to persist and eventually graduate than their peers whose parents attended college. Chen & Carroll (2005) demonstrated that first-generation students tend to earn fewer credits in their first year, take more remedial classes, are more likely to repeat courses, and are less likely to choose their major in their first year. Those are their likely characteristics and the reasons why they tend to drop out. Compared to their peers, these students receive less financial and family support (HR, 2011). A demanding high school curriculum can narrow the college persistence gap for first-generation students (Nuñez & Cuccaro-Alamin, 1998), especially if they graduate high school in the top quartile. These students perform similarly to other students in terms of their college grades and remedial coursework. But on balance, even after controlling for socioeconomic status, institution type, and enrollment patterns, first-generation status still has a negative effect on degree completion.

Low Socioeconomic Status

Among the student background variables, Astin (1997) found that socioeconomic status is the best predictor for completing a bachelor's degree after controlling for academic ability. It was evident in many studies that no matter how academically qualified a student is, he/she is less likely to persist if he/she cannot afford to pay his/her

tuition. Choy (1996) revealed that only 49 percent of high school completers from low-income families attend college, while 63 percent of middle-income students and 78 percent of high-income students attend a college.

Part-time Students

According to the NCES 92-93 study, more than half of all undergraduates were enrolled part-time (NCES, 2012). These students are less likely to be retained than their full-time counterparts. According to 2010 statistics from the National Center for Higher Education Management Systems (NCHEMS), the retention rate for full-time students is 78.7 percent, while the retention rate for part-time students is only 44.6 percent. The recent Clearinghouse (2012) study shows that older students who enrolled exclusively part-time actually had a higher completion rate than part-time students of a traditional age. The report also showed that younger adult students have a higher tendency to register full-time than their older counterparts.

Full-time Working Students

Even though the majority of adult students work full-time while enrolled in college, more than one-third of younger students work full-time while enrolled part-time and more than one-fifth of younger students work full-time while enrolled full-time (Kasworm, Polson & Fishback, 2002). Research suggests that working adults may have the advantage of employer support and flexibility in the workplace, opportunities that are not present for younger working students (Kasworm, 2003). Students who work full-time may seem financially independent, but they are more likely obliged to do so in order to support their families and finance their college attendance.

Married Students

There are not many studies toward the retention of younger married students. Studies of adult students, however, found that divorce is most likely a reason for older students going back to school (Brown, 2002). However, that characteristic is almost non-existent in the NCES:88/00 data. Younger married students consist of almost seven percent of all college undergraduates (Campus Explorer, 2013). If older married students are counted toward this number, the number is likely to increase. Compared to traditional students, married students have many different experiences and challenges, such as the financial burden of two or more people, altered social experiences, and balancing school work and family life.

Students With No Standard High School Diploma

This group of students represents those students who did not initially graduate from high school, receive a GED, or take other tests that would enable them to enroll in college. These students usually consist of a small number of all undergraduates, but are considered to be at a higher-risk of dropping out of college. According to the GED Testing service (2011), about 43 percent of GED test-takers pursued further education in 2004, yet only 28.7 percent stayed in the first-to-second year of their post-secondary education.

CHAPTER THREE

This study explores logistic regression analysis in the prediction of nontraditional freshman student retention. This chapter will describe the setting of the study, samples, variables, research design, models, and data analysis.

Setting of the Study

This study is a secondary analysis on data from the NELS: 88 to 2000. The NELS:88 base-year survey consisted of some 25,000 eighth-grade students from a random national sample of 1,052 public and private schools. The first and second follow-up studies were surveys of the representative samples of tenth- and twelfth-grade students. The third follow-up survey took place during 1994 and the sample of 12,144 students was surveyed after many entered post-secondary institutions. This study was especially focused on the second and third study of NELS:88. NELS data were chosen since it is a nationally representative sample and covers post-secondary student retention.

Sample

After matching the students from the second and third surveys, the sample of 8,024 students were characterized in this study. When surveyed the third time, it was found that 68 percent (5,574) of the students in this sample were retained in their first-year post-secondary education.

Nontraditional students in this study will only have the following seven characteristics:

1. Background characteristics:
    a. low socioeconomic status;
    b. ethnicity: African American and Hispanic origins; and
    c. first-generation students.
2. At-risk characteristics:
    a. attends part-time;
    b. works at least 35 hours a week;
    c. married; and
    d. lacks a standard high school diploma.

The following table shows the number of enrolled students with these characteristics and the number of students who are retained (Table 3).

Table 3. *Characteristics of Nontraditional Students Statistics*

Nontraditional student characteristics	Number of students	Percentage of total	Number of retained	Retention rate
First-generation	1,968	24%	1,192	61%
African American and Hispanic students	1,612	20%	1,009	63%
Low socioeconomic status	1,232	15%	718	58%
Worked full-time	1,705	21%	737	43%
Attended part-time	1,243	15%	511	41%

Married	684	8%	267	39%
No standard high school diploma	270	3%	68	25%

Fifty-eight percent of the sampled students with these seven characteristics have at least one nontraditional characteristic. From these nontraditional students, 49 percent are minimally nontraditional, 49 percent moderately nontraditional, and two percent highly nontraditional. Traditional students, on the other hand, have a much higher retention rate than these nontraditional students (Figure 3). Students with at-risk characteristics are less likely to be retained.

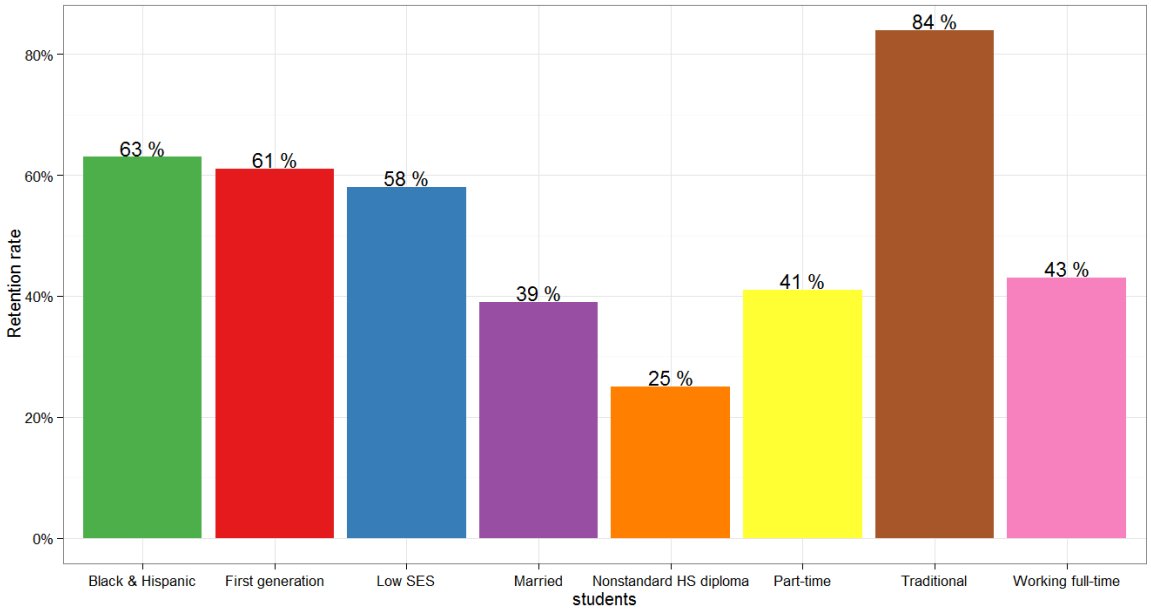

Figure 3. Comparison of retention rates between traditional and nontraditional students

Variables

More than 7,500 variables were collected overall for the NELS:88/00 study. However, this study overlooks the most important variables for nontraditional freshman students. The dependent variable was a binary variable: Retention. Retention in this study is defined as students who are retained or remained at their first institution for more than 12 months. The predictor variables were selected based on the criteria of whether they were regularly collected by institutions without any additional survey questionnaires. These variables can be collected from student applications, initial enrollment status, and later by different institutional sectors such as financial aid and employment. The variables are categorized as *high school* or *college* depending on when the variables were collected. The following variables were considered for this study (Table 4).

Table 4. *Predictive Variables from NCES: 88/2000*

Predictive Variables			
High School Performance		**Scale**	**Description**
HS_STD	Standard test score	Interval	Average standard test score of Reading, Math, Science, and History
HS_AP	Advanced Placement course	Nominal	0/1 indicating if student took an advanced placement course
Grade	High School GPA	Interval	Average grade of English, Math, Science, Social Studies, Computer and Foreign Language. Numbers indicate 1 being mostly As and 13 being F
HS_POST	College Preparation Test	Categorical	Number of college exams taken

HS_REM	Remedial classes	Categorical	Number of remedial classes
HS_UNITS	Carnegie Units	Interval	Number of units in English, Math, Science, Social Studies, Computer and Foreign Language.
SES	Socioeconomic status	Nominal	0/1 indicating low socioeconomic status
FAMINC	Family annual income	Categorical	1 being none, 2 being less than $1,000 and so on, 15 being more than $200,000 annual income
GEN	First-generation student	Nominal	0/1 indicating first generation student
HS_ECA	Participation in extracurricular activities	Categorical	Number of extracurricular activities involved during high school
HS_ATTND	High school attendance	Categorical	Number of times missing school, 0 = never, 1 = 1-2 times, …4 = 10-15 times
HS_MARST	Marital status	Categorical	1=single 2=married 3=divorced 4=widowed 5=marriage like
HSSTAT	High school status	Categorical	1= received high school … 6= did not graduate high school

College Performance		**Measure**	**Description**
C_REM	Remedial courses taken	Categorical	Number of remedial classes /English and Math/
C_REG	Regular courses	Categorical	Number of regular courses taken /Math, Physics, Chemistry, Biology/
TYPEDEGREE	Degrees chosen	Categorical	1=none 2=Certificate 3=AS 4=BS 5=Other
TYPEINST	Type of Institution	Categorical	1=private for profit … 6=public 4 year
LOANS	Loans	Nominal	0/1 indicating if received loans

GRANTS	Grants/Scholarships	Nominal	0/1 indicating if received grants or scholarships
NO_FIN	No financial aid	Nominal	0/1 indicating if received no financial aid
WORKSTUDY	Work-Study	Nominal	0/1 indicating if received workstudy
HRSWORK	Work hours	Numeric	Number of hours work weekly
PARBORROW	Parents Borrow	Nominal	0/1 indicating if parents borrowed to pay students' tuition
C_ECA	Extra-curricular activities	Categorical	Number of extracurricular activities
ENRLSTAT	Enrollment status	Categorical	1=full-time 2=part-time 3=less than part-time
MARSTAT	Marital status	Categorical	1=single 2=married 3=divorced 4=widowed 5=marriage like
TYPETUIT	In-state/out-of-state	Nominal	0/1 indicating in-state and out-of-state
LEVELS	Nontraditional student status	Categorical	0=traditional 1=Minimally nontraditional 2=Moderately nontraditional 3=Highly nontraditional

Research Design

This study performed logistic regression analysis and used SPSS 21 to answer the research questions. The research questions addressed in this study are:

1. Do the *high school* and *college* variables identified in this study predict retention of nontraditional freshman students?

2. Does the level of nontraditional students predict retention of nontraditional freshman students and does this prediction have a unique effect controlling for *high school* and *college* variables?

In this study, predictor variables are grouped into two blocks: *high school* and *college*.

To answer the first research question, this study employed blockwise logistic regression to identify significant variables from both blocks (Figure 4). First, the variables in the *high school* block were analyzed using logistic regression. Stepwise selection specifies the most significant variables. The variables that survived the selection from the first block were added to the *college* block and the logistic analysis was conducted again for another stepwise selection.

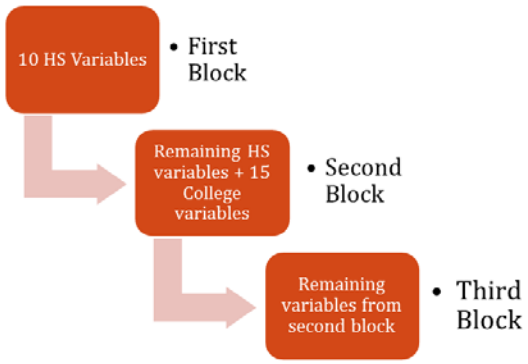

Figure 4. Blockwise logistic regression

To answer the first part of the second research question, this study used logistic regression with the levels of nontraditional students. For the second part of the question, hierarchical logistic regression was used with all survived *high school* and *college*

variables from the regression model of the first research question and the levels of nontraditional students (Figure 5).

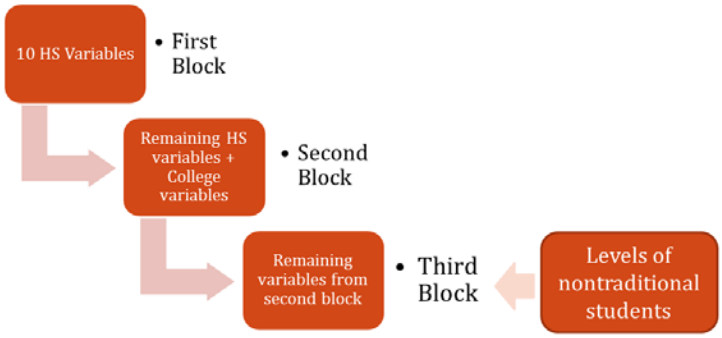

Figure 5. Hierarchical logistic regression

Logistic regression

Many retention researchers employed binary logistic regression to study predictive variables (Bogard, Helbig, Huff & James, 2011). In fact, many national and private research organizations rely on logistic regression as a main statistical technique to predict student retention (American College Testing [ACT], 2010; Adelman, 2006). Logistic regression only analyzes how the predictive variables affect the response variables, so the outcome shows the list of predictive factors of overall retention for the duration of a study. Among the studies on student retention that use logistic regression, most of them reported a correct classification rate that is between 60 to 80 percent (Nandeshwar et al., 2003).

Logistic regression is the most exercised model in the study of student retention. Binary logistic regression is commonly used when the dependent variable, retention, is

dichotomous (1=Retained, 0=Not Retained). "Logistic regression deals with relationships among variables (not mean differences) with one variable being the dependent (i.e. outcome or response) variable while the other(s) is/are the independent (predictor or explanatory) variable(s)" (Huck, 2004, pp.438-439).

Logistic regression analysis provides predicted probabilities of retention for a combination of the independent variables (Pyke & Sheridan, 1993). Selecting independent variables that are significant to the prediction of student retention is the main consideration for logistic regression models. Too many variables in a model may result in overfitting; too few variables in a model may result in underfitting. Therefore, choosing the right set of variables in a logistic regression model is important to the quality of the model.

In this study, the blockwise selection method was used for the selection of the predictor variables to enter high school and college variables separately. Under the blockwise selection approach, a stepwise selection is used with the items of the first block and the items that are most predictive are combined with the items of the second block, then stepwise selection is used again (Dimitrov, 2009). The best predictive items from that selection are combined with the items of the third block and so forth until the procedure stops with the lack of statistical significance for additional items. Under hierarchical regression, the study ran a stepwise selection by separate factors then combined the best predictive items from all selections into a logistic regression model for final selection (Dimitrov, 2009).

When retention is successful, the combination of optimum predictor variables yields the best prediction of probability of the dependent variable (P(Y=1)). The logit form of the logistic regression model is presented below:

Equation 1 Logistic regression model

$$P(Y = 1) = B_1X_1 + B_2X_2 + ... + B_mX_m + A. \quad (1)$$

Where $B_1, B_2 \ldots B_m$ are regression coefficients, and $X_1, X_2 \ldots X_m$ are the set of predictors. To find the best prediction of probability of success, a combination of predictor variables needed to be chosen for the model based on empirical theory or practice.

An important aspect of conducting logistic regression is finding the odds and odds ratio. Odds for success are based on the ratio of the probability for success to the probability of failure (Equation 2). Thus, the odds are interpreted as the probability of success that is a number of times greater than the probability of failure.

Equation 2. Odds for success

$$\text{ODDS} = \frac{P}{1-P} = e^X \quad (2)$$

"Odds ratio (OR) is a measure of association between an exposure and an outcome," (Szumilas, 2010). For this study, the OR is used to compare the relative odds of the occurrence of the outcome of interest (student retention) given exposure to the variable of interest (levels of nontraditional students) (Figure 6). The OR can also be used to

determine a particular outcome and to compare the magnitude of various risk factors for that outcome (Szumilas, 2010). For example, if the OR is equal to one, it means that the exposure does not affect the odds of the outcome. If the OR is more than one, then the higher the odds are of the exposure affecting the outcome; if lower than one, the lower the odds of the exposure affecting the outcome.

OBSERVATION		PREDICTION: Retained (1)	PREDICTION: Not Retained (0)
	Retained (1)	**Hit** a	**False Alarm** b
	Not Retained (0)	**Miss** c	**Correct rejection** d

Figure 6. Classification table interpretation.

Odds ratio equation is illustrated in Equation 3.

Equation 3. Odds Ratio Equation

$$\text{Odds Ratio} = \text{OR} = \frac{a/b}{c/d} \quad (3)$$

Logistic regression has gained a reputation in data mining techniques because it does not include any of the assumptions associated with multiple regression (Raju, 2012), which are: a linear relationship between the dependent and independent variables, normal distributions of predictor variables, or equal variances of the predictor variables within groups.

CHAPTER FOUR

This chapter provides the results from the data analysis related to the research questions addressed in this study. The first section presents the descriptive statistics of the data sample and describes the methods used for handling of missing data. In the next section, where logistic regression analysis was employed, the steps are explained and summarized in detail.

Description of the Dependent Variable

In this study, freshman students who stayed with their post-secondary institutions for more than 12 months were considered as retained and students who dropped out of their post-secondary institutions within 12 months were considered as not retained. NELS 88 data does not provide information about whether students were retained in their first year or not, so the number of months students were enrolled with their institutions of higher education were considered as a decision variable. Students who persisted beyond 12 months were coded to a 1 and students who studied up to 12 months were coded to a 0. Additionally, students were classified into levels of nontraditional students, which adopted from NCES 1996 study. Students who possessed one nontraditional student characteristic were coded a 1 and classified as minimally nontraditional; students who possessed two or three nontraditional student characteristics were coded as a 2 and classified as moderately nontraditional; and students who possessed more than 4

nontraditional student characteristics were coded a 3 and classified as highly nontraditional. Retention of each level of student is presented in Table 5.

Table 5. *Levels of Nontraditional Students*

Levels of nontraditional students	**0**	**1**	**Total**
1 Minimally nontraditional	714	1,555	2,269
2 Moderately nontraditional	897	1,107	2,004
3 Highly nontraditional	224	142	366
Total	1,835	2,804	4,639

Description of Independent Variables

Research studies on college student retention usually reflect theoretical models to determine predictive variables included in those studies and often use surveys combined with demographic data to evaluate validity of those theories (Pittman, 2008). In this study, the selection of the predictive variables was not only based on the theoretical models but also domain knowledge and generally depended on the availability of the NELS 88 data. The independent variables were chosen based on whether the variable was available and regularly collected by institutions when a student first started in the fall semester. The variables are categorized into two main groups: *high school* and *college*. *High school* variables have twelve predictive variables that include high school grades,

standard test scores, number of Carnegie units, attendance, extracurricular activities, marital status, number of college preparation exams taken, advanced placement classes taken, family income, and number of remedial courses taken (Table 6). It is worth noting that those students classified as highly nontraditional are distinctively different from the rest of the group.

Table 6. *Percentage of Nontraditional and Traditional Students in High School Variables*

Student types Variables	Minimal Nontraditional	Moderate Nontraditional	High Nontraditional	Traditional	Total
HS_ATTEND – number of times missing school					
0 = never	10.6	9.0	5.0	11.2	10.3
1 = 1-2 times	32.7	30.6	30.4	35.1	33.1
2 = 3-6 times	34.4	32.8	36.3	35.2	34.4
3 = 7-9 times	11.4	14.0	11.2	11.0	11.8
4 = 10-15	6.7	7.1	12.5	4.7	6.2
5 = over 15	4.1	6.4	4.6	2.8	4.1
HS_REM – number of remedial classes					
0 – no class	82.1	71.7	66.6	85.7	80.5
1 – English or Math	8.9	12.8	18.0	6.7	9.2
2 – both English&Math	9.0	15.6	15.4	7.6	10.3
HS_MARST – high school marital status					
0 – not married	97.5	94.0	88.6	98.5	96.8
1 - married	2.5	6.0	11.4	1.5	3.2
HS_AP – advanced placement classes					
0 - no	55.6	66.8	69.0	42.3	53.1
1 - yes	44.4	33.2	31.0	57.7	46.9
HS_POST – number of college exams taken					
0	14	27.2	43.5	6.4	15.4
1	50	44.9	37.7	53.5	49.7

2	30	21.8	15.2	33	28.6
3	5.7	5.6	3.3	6.8	6.0
4	0.3	0.4	0.3	0.2	0.3

College variables include variables such as enrollment status, type of degree, marital status, number of extracurricular activities, type of tuition, remedial courses taken, grants received, and loans received (Table 7). As the number of nontraditional student characteristics increases, their chances of taking remedial courses increases, and the number of students taking regular classes decreases. These students are more likely to be married or have marriage-like relationships. Their enrollment status has a higher chance of being part-time or less than part-time.

Table 7. *Percentage of Nontraditional and Traditional Students in College Variables*

Student types Variables	Minimal Nontraditional	Moderate Nontraditional	High Nontraditional	Traditional	Total
REGCOURS – number of regular math and science classes					
0 = none	18.7	26.9	46.0	12.0	18.0
1	32.1	37.1	34.5	26.9	30.7
2	28.8	22.4	10.0	32.3	28.6
3	16.2	11.0	7.0	23.1	18.1
4	4.2	2.6	2.5	5.8	4.6
REMCOURS – number of remedial classes					
0 – no class	73.9	69.6	67.0	80.9	76.1
1 – English or Math	15.1	16.5	20.5	10.9	13.6
2 – both English&Math	11.0	13.9	12.5	8.2	10.3
HS_MARST – college marital status					

0 – not married	94.4	81.2	57.4	100	91.8
1 – married	2.6	9.8	22.1	0	4.2
2 – divorced	0	1.0	1.9	0	0.4
3 – marriage like	2.9	7.9	18.6	0	3.7
TYPETUIT – type of tuition					
1 – in-state	93.0	96.1	96.9	89.6	92.7
2 – out-of-state	7.0	3.9	3.1	10.4	7.3
ENRLSTAT – enrollment status					
1 – full-time	89.6	71.7	39.9	99.0	86.8
2 – part-time	7.1	17.6	36.1	0.5	8.3
3 – less than part-time	3.3	10.7	24.0	0.5	4.9
GRANTS					
1 - yes	45.8	48.1	42.5	45.4	46.0
2 - no	54.2	51.9	57.5	54.6	54.0
TYPDEGCT – type of degree					
1 – none	10.8	10.7	13.4	5.8	8.7
2 – certificate	4.8	13.4	27.9	1.4	6.6
3 – associate	27.2	36.3	36.0	15.7	25.0
4 – bachelor	50.1	30.5	13.4	72.4	53.0
5 – other	7.1	9.0	9.2	4.8	6.7
LOANS					
1 - yes	28.8	27.4	15.9	28.1	27.6
2 – no	71.2	72.6	84.1	71.9	72.4
WORKSTDY					
1 - yes	10.6	8.0	1.1	12.6	10.4
2 – no	89.4	92.0	98.9	87.4	89.6
NO_FINA – no financial aid					
1 - yes	44.1	41.3	48.5	45.9	44.4
2 – no	55.9	58.7	51.5	54.1	55.6
CAMPJOB					
1 - yes	20.7	14.8	10.4	28.4	22.0
2 – no	79.3	85.3	89.6	71.6	78.0
PARBORROW					
1 - yes	14.4	9.6	4.6	16.8	13.7

2 – no	85.6	90.4	95.4	83.2	86.3
TYPINST - type of institution					
1 – private4pro	3.6	8.2	11.7	1.2	4.0
2 -	1.3	1.3	1.3	1.0	1.2
3	0.3	1.1	2.6	0.0	0.5
4	35.8	50.3	63.1	18.4	32.9
5	18.5	9.0	3.9	31.4	21.2
6 – public4yr	40.4	30.0	17.5	48.0	40.3

Missing Data Handling

Normally, statistics packages remove any case or data record that has missing data (casewise deletion). Applying this method would eliminate 76% of the NELS:88/00 student data and would drastically limit inferences to the population as a whole (Figure 7). However, exact missing values only consist of 8% of the sample.

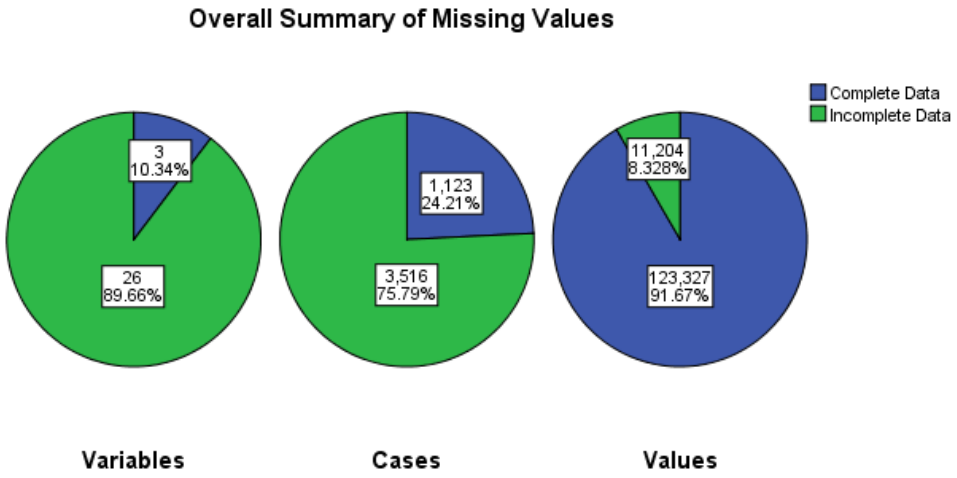

Figure 7. Missing value statistics

Excluding all cases with any missing data would clearly limit the study because students with the most financial need have the most complete financial aid information and older students have more missing high school data. One solution might be to simply remove any attribute with missing values. Like casewise deletion, however, this removes large amounts of useful model building information. Preserving information and avoiding biased model results requires imputing values for missing data. This study used SPSS's multiple imputation option of the Monte Carlo Markov Chains (MCMC) method to impute the values of missing data for five numbers of iterations. The MCMC method is a powerful and popular method of multiple imputations and can be used when data is random (IBM, 2011). Since the missing values are random, this study adopted the MCMC method to impute the missing values.

Logistic Regression Analysis

Research Question 1. **Do high school and college variables identified in this study predict retention of nontraditional freshman students?**

After missing data was imputed, logistic regression analysis was conducted using SPSS 22 software. To answer the first research question, this study employed blockwise logistic regression. The models were developed using a blockwise selection process. In the blockwise procedure, variables were added to the model one block at a time. This study executed two blockwise procedures. First, all 10 *high school* variables were included in the model and were evaluated to see if they were significant. The criteria for a variable to enter, and remain, in the model was a p-value of less than .05. All *high school* variables included in the model predicted a 62.5% retention rate of nontraditional

freshman students overall. Among those *high school* variables, only five variables were significant (Table 9).

Table 8. *Logistic Regression Analysis of Student Retention as a Function of Ten High School Variables*

				95% C.I. for Odds Ratio	
	B	Wald	Odds Ratio	Lower	Upper
HS_ATTEND	-0.14	30.36**	0.87	0.83	0.91
HS_REM	-0.05	1.79	0.95	0.87	1.04
HS_AP	0.22	9.78*	1.25	1.08	1.44
HS_ECA	-0.02	0.19	0.99	0.95	1.02
HS_MARST	-0.32	7.61*	0.73	0.54	0.98
HS_STD	-0.001	0.18	1.00	0.99	1.00
FAMINC	0.004	0.001	1.00	0.98	1.03
HS_UNITS	0.21	70.26**	1.36	1.27	1.47
HS_GRADE	-0.09	73.86**	0.91	0.89	0.93
HS_POSTTEST	0.03	0.68	1.03	0.95	1.12

** $p < 0.001$.

After this step, the five *high school* variables - high school attendance, advanced placement courses taken, marital status, Carnegie units, and high school grades – were

kept for the second step. In the second step, five survived *high school* variables were added to the fifteen *college* variables to run the next logistic analysis. The 20-variable model explains the 66.8% of the nontraditional student retention. From this logistic regression analysis, 10 predictive variables were significant and were included for the final model. The final model is statistically significant, $\chi^2(10) = 621.86$, $p < .001$, thus indicating that the last remaining 10 predictive variables predict nontraditional student retention. The Hosmer-Lemeshow statistic also showed that the data fit to the model, $\chi^2(8) = 12.65$, $p = 0.125$. The final model showed that all survived 10 variables were significant after running logistic regression with only those variables included in the model (Table 10).

In Table 10, the value of odds ratio for high school Carnegie units, EXP(B) = 1.24, indicates that the odds for nontraditional student retention increase by a factor of 1.24 for a every one-unit increase in the number of Carnegie units taken in high school, when controlling for all other predictors. At the same time, the value of odds ratio for the number of math and science courses taken in a college is 1.14, meaning that the odds for nontraditional student retention is 1.14 times more likely for each unit increase in the number of regular courses taken in college. In addition, degree level has the highest value of odds ratio. As a student's degree level increases, nontraditional students are 1.29 times more likely to be retained. Moreover, there are some variables that have a negative relationship with a student's retention. For example, each additional hour of working will decrease a student's retention by a factor of 0.98. This ratio is fairly stable because the confidence interval is small (from 0.98 to 0.99, at a 95% level of confidence). The rest of

the predictive variables have a less than 1.00 value of odds ratio, so their effects are minor.

Table 9. *Logistic Regression Analysis of Student Retention as a Function of Ten Variables in the Last Model*

	B	Wald χ^2	Odds Ratio	95% C.I. for Odds Ratio	
				Lower	Upper
HS_ATTEND	-0.11	18.72**	0.89	0.85	0.94
HS_UNITS	0.22	36.37**	1.24	1.16	1.33
HS_GRADE	-0.64	33.49**	0.94	0.92	0.96
REGCOURS	0.13	22.77**	1.14	1.07	1.22
HRSWORK	-0.02	119.27**	0.98	0.98	0.99
MARSTAT	-0.15	23.63**	0.86	0.81	0.91
ENRLSTAT	-0.12	4.07*	0.89	0.80	0.99
TYPDEGCT	0.25	72.23**	1.29	1.22	1.37
GRANTS	-0.26	14.95**	0.77	0.67	0.88
CAMPJOB	-0.33	12.02*	0.72	0.60	0.87

** $p < 0.001$.

The model correctly predicted an 81.3% retention rate for the nontraditional students in the study (Table 11). This is also known as the *sensitivity* of prediction. Additionally, 45.6% of the time, the model did not observe the predicted event. This is known as *specificity* of prediction. Overall, the prediction of nontraditional student retention was correct 3,115 out of 4,639 times, for an overall success rate of 67.1%. That is almost 5% higher than the model using 10 *high school* variables. The model also predicted wrong occurrences; for example, students might drop out after the model predicted they would not. Such an event is called a false positive and occurred in 999 of the 3,278 cases, for a false positive rate of 30.5%. Another such event is when the model predicted a student may retain, but actually did not. This is called a false negative and occurred 525 times, for a false negative rate of 38.6%.

Table 10. *Classification Table for 10 Variable Model*

Observed		Predicted		
		retention		Percentage
Step 1		0	1	correct
retention	0	836	999	45.6
	1	525	2279	81.3
Overall Percentage				67.1

Another way of measuring the model's goodness-of-fit is the receiver operating characteristic (ROC) curve. Generating the curve during logistic regression analysis is often used to evaluate if the model is a good fit. The ROC curve uses all pairs of sensitivity (true positive) and specificity (true negative) for each possible cutoff point.

The area under the ROC curve is 70.1% for the last model, which indicates that this model is a better fit (Figure 8).

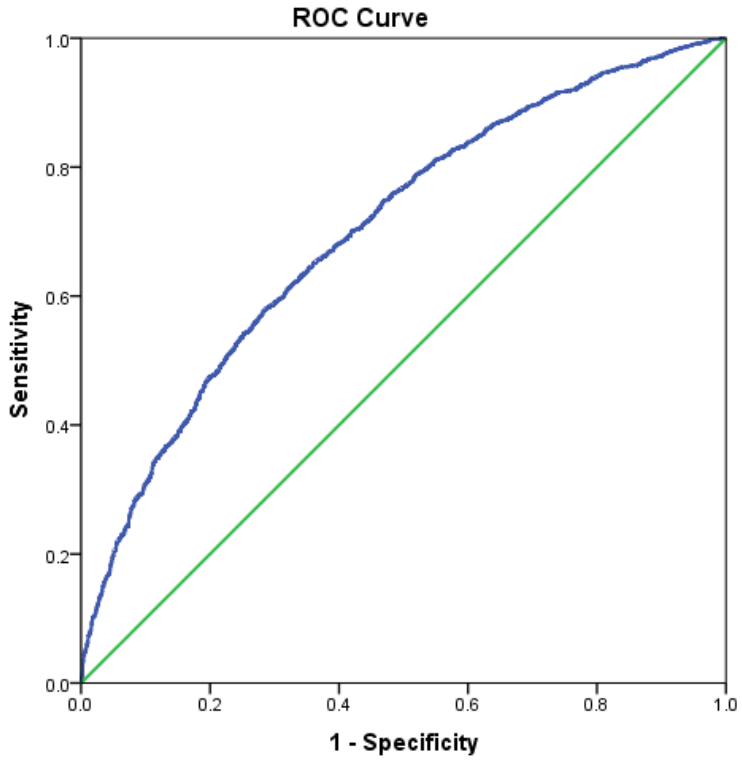

Figure 8. ROC curve for the last model

The graph below shows each predictor's effect and relationship with the retention and between each other (Figure 9). The right side column (with retention on the bottom) shows each predictor's positive or negative relationship with retention. The bottom row shows the probability of retention with each predictor and the effects of the relationship. For example, type of degree and number of regular courses has a positive relationship, while hours of work, grants, and high school attendance have a negative relationship. Also, high school grade has a negative relationship with retention. It does not make sense at first. However, it makes sense when you think about the coding of the variable. For high school grade, students with grades of mostly As were coded 1 and students with

grades of mostly Fs were coded 13. Thus, the graph shows that the higher the value of the grade variable, the lower the probability of retention; students with low high school grades are predicted to have low retention rates.

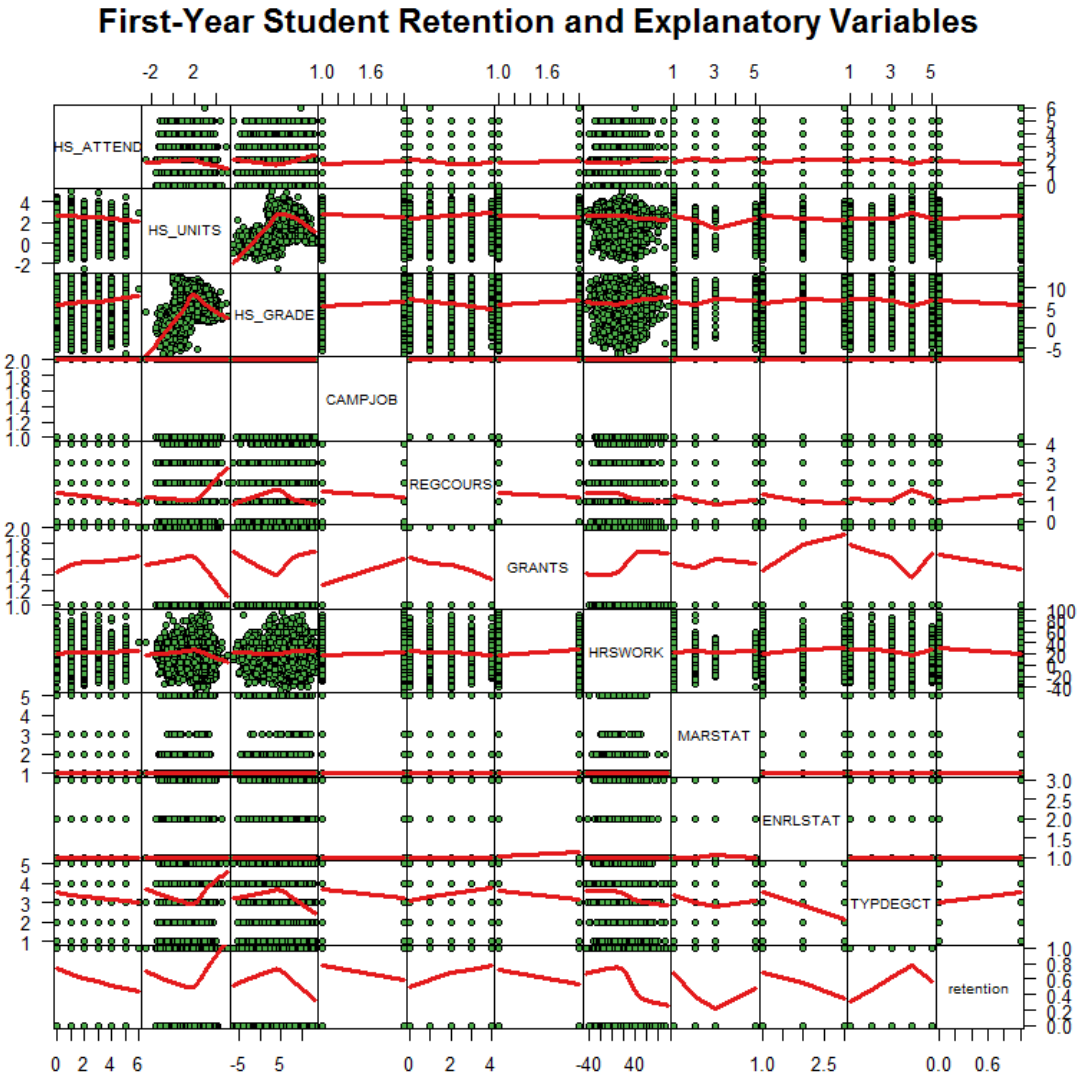

Figure 9. Relationship of retention and predictors

Moreover, using the same method on traditional students, prediction accuracy was 84.5%, showing that traditional students have a higher chance of being retained and

explained by this model. There were only minimal differences. The marital status and working hours variables were not significant for traditional students. The next table 11 shows two models for traditional and nontraditional students. The model for traditional students has eight variables compare to the model for nontraditional student has ten variables.

Table 11. *Comparison of models of nontraditional and traditional students*

	Traditional students	***Nontraditional students***
High school variables	HS_GRADE	HS_GRADE
	HS_ATTEND	HS_ATTEND
	HS_UNITS	HS_UNITS
		MARSTAT
College variables		HRS_WORK
	TYPDEGCT	TYPDEGCT
	ENRLSTAT	ENRLSTAT
	CAMPJOB	CAMPJOB
	GRANTS	GRANTS
	REGCOURS	REGCOURS

Multicollinearity in the logistic regression solution is detected by examining the standard errors for the b coefficients. A standard error larger than 2.0 indicates numerical problems such as multicollinearity among the independent variables. Standard error for each predictor in this study is lower than 1.0, which indicates there is no multicollinearity. Another test to detect multicollinearity is to see the correlation matrix for the predictors. In Figure 10, all 10 variables in the last model's correlation values are shown. If the variables are highly correlated to each other, the shades of blue and red on

the graph should appear darker. However, that is not the case. The highest correlation in the matrix is 0.37, between high school units and high school grades.

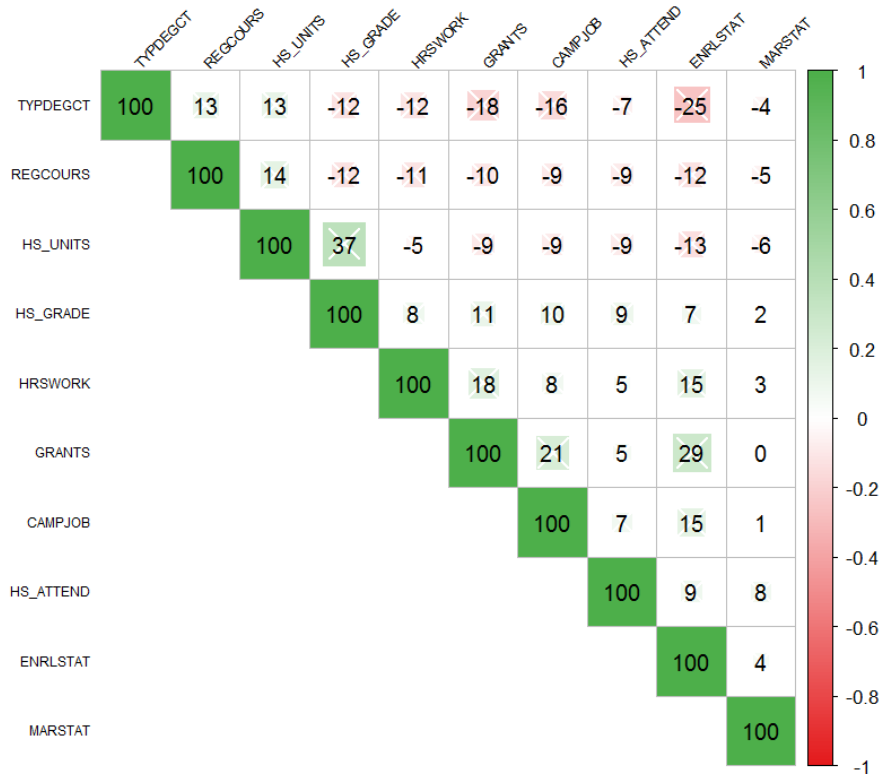

Figure 10. Correlation Matrix

Research Question 2a. Does the level of nontraditional students predict the retention of nontraditional freshman students?

The model including only the levels of nontraditional students ($\chi^2(1) = 155.5$, $p < .001$) is statistically significant, so it does predict nontraditional student retention. The Hosmer-Lemeshow statistic also shows that the data fit to the model (χ^2 (1) = 0.446, p = 0.50). Even though the model predicts the retention, it is important to evaluate accuracy.

The model correctly predicted a retention rate of 94.9% for the nontraditional students in the study (Table 12). In addition, the model did not observe the predicted event 12.2% of the time. Overall, the prediction of nontraditional student retention was correct 2,886 out of 4,639 times, for an overall success rate of 62.2%.

Table 12. *Classification Table*

Observed		Predicted		
		retention		Percentage
Step 1		0	1	correct
retention	0	224	1611	12.2
	1	142	2662	94.9
Overall Percentage				62.2

Considering only the levels of characteristics of nontraditional students, the logistic regression model is as follows:

ln(*ODDS*) = 1.383 – 0.596*Levels

This model can predict the odds that different levels of nontraditional students will stay within institutions of higher education. The odds prediction equation is $ODDS = e^{a+bX}$. If a student is minimally nontraditional, then the ODDS = $e^{1.383-0.596*1} = e^{0.787} = 2.2$, which means they are 2.2 times more likely to retain than to drop out. The probability is as follows:

$$\hat{Y} = \frac{ODDS}{1+ODDS} = \frac{2.2}{3.2} = 0.6875.$$

This model predicts that 68.75% of minimally nontraditional students will retain. The actual retention rate for this type of student is 68.5%, so this prediction is very close.

For moderately nontraditional students (level = 2), the odds are 1.2, so they are 1.2 times more likely to stay within the institutions; the model predicts that 54.5% of moderately nontraditional students will retain (actual retention rate is 55.2%). The odds ratio is 2.2/1.2 = 1.8, which means minimally nontraditional students are 1.8 times more likely to retain than moderately nontraditional students. For highly nontraditional students, the odds are 0.67 and the probability of retention is 40% (actual retention rate is 38.8%). And the odds ratio between minimally and highly nontraditional students is 3.3, so minimally nontraditional students are 3.3 times more likely to persist than highly nontraditional students. Additionally, the prediction rate moderately drops as the levels of nontraditional student characteristics increase. In other words, prediction of students' retention is more likely to drop when a student possesses more nontraditional characteristics.

Research Question 2b: Does the level of nontraditional students have a unique effect in the prediction of retention controlling for high school and college variables?

To see if the levels of nontraditional students have a unique effect on freshman student retention, hierarchical logistic regression was exercised. In hierarchical logistic regression, the order of the variables is forced into the model to see their unique contribution to the prediction of the outcome (Dimitrov, 2009). In this case, the new model was statistically significant ($\chi^2(11) = 639.03$, $p < .001$), thus indicating that there is no difference between the model with only 10 *high school* and *college* variables and the model with only the levels of nontraditional students. Therefore, the levels of nontraditional students have a unique effect on the prediction of nontraditional student

retention after controlling for *high school* and *college* variables. The Hosmer-Lemeshow statistic also shows that the data fit to the model (χ^2 (8) = 11.59, p = 0.17). In addition, after comparing each model's prediction accuracy, the percentage difference is minimal – only 0.2% (Table 13).

Table 13. *Comparison of Two Models*

Models	**Prediction accuracy**
10 survived high school and college variables	67.1%
The levels of nontraditional students controlling 10 survived high school and college variables	66.9%

When examining the individual relationship with the dependent variables, all the relationships were significant except enrollment status. The probability of the Wald statistic for the enrollment status (χ^2 (1) = 0.55, p = 0.458) was greater than the level of significance of .05. Therefore, in this model, enrollment status does not have an impact on the odds of nontraditional student retention. Exploring other variables' impact reveals relationships with the retention of nontraditional students (Figure 11). The type of degree variable has the highest positive effect, followed by grants and campus jobs. After those variables, the number of Carnegie units, regular college courses, and high school grades

follow. Levels of nontraditional students have the highest negative effect, followed by marital status, high school attendance, and hours worked weekly.

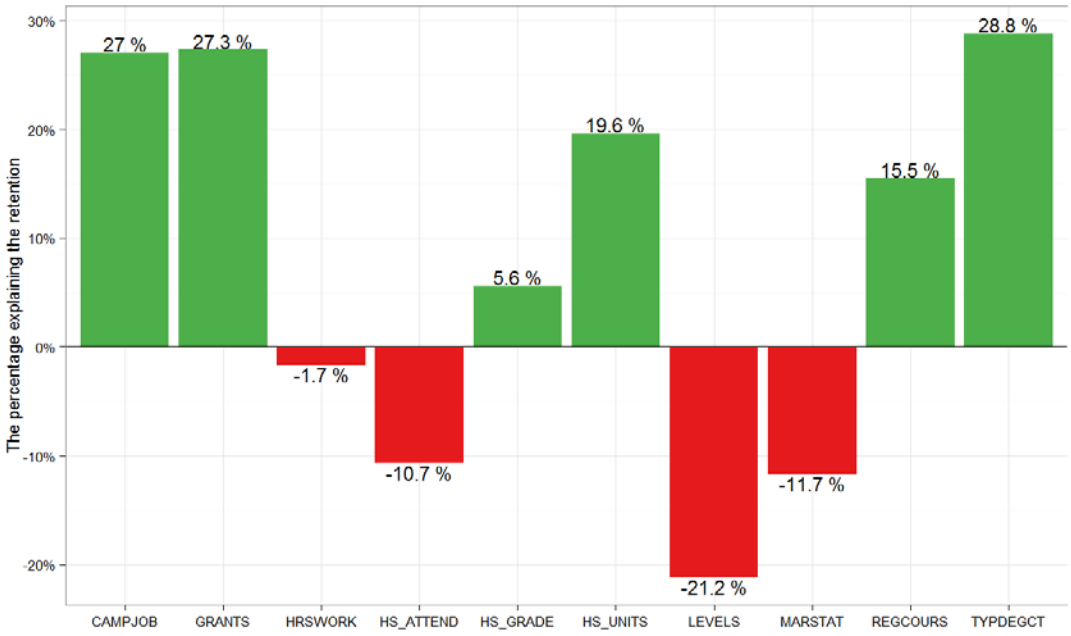

Figure 11. The percentage of explanation of each variable in the last model. The negative percentages show the negative relationship with student retention.

This graph shows the importance of visualizing the findings of a study. Without this bar graph, it is hard to notice the effects of high school grades and hours of work when compared to high school attendance and type of degree. Compare to high school attendance, high school grades have almost twice less effects. Type of degree variable consistently shows higher results. It has higher odds ratio and higher percentage explanation of the retention. Figure 12 shows how type of degree is different in different levels of nontraditional students and traditional students. Traditional students are more likely to choose to study bachelor's degree; while, highly nontraditional students are more likely to choose certificate and associate's degree. This distinctive difference

between traditional and nontraditional students was also consistent with other variables in the study.

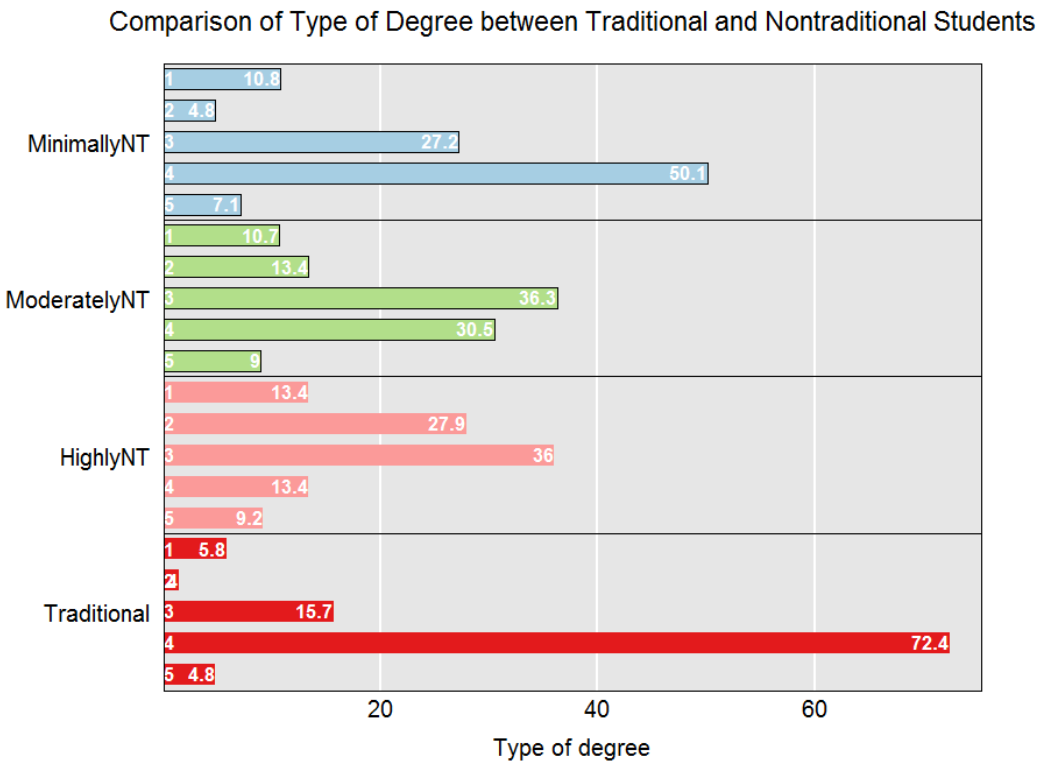

Figure 12. Relationship of type of degree variable with levels of nontraditional students and traditional students

According to literature review, the variables important to student retention can be categorized into academic, financial and social factors. High school grade, Carnegie units, regular college courses and type of degree belong to academic factor; grants, campus job and hours of work are in financial factor; high school attendance and marital status can be social factor.

CHAPTER FIVE

This chapter argues for the importance of identifying at-risk students as early as possible and using readily available data that institutions already have in order to identify at-risk students. This section also compares new and similar findings of the current study to the literature; deliberates the results of the study, concentrating on specific groups of variables; explores future directions; and discusses the advantages of using data visualization and other methods in addition to the logistic regression.

Discussion of the Findings

The purpose of this study was to determine the important variables predicting nontraditional freshman student retention using pre-enrollment data that are available at the beginning of the fall semester. Empirical models and theories available for student retention today suggest that administrators will need to collect more factors to predict student retention. The variables they currently collect range from tens to hundreds of variables, ranging from student attendance and college grades to homesickness and motivation. Collecting such a wide variety of information requires time, money, and energy. Over the years, researchers have suggested using easily accessible data to provide comparable accuracy instead of using other more costly and difficult measures (ACT, 1992). Obtaining information early can help students find the proper support systems and programs.

The result of this study found that predicting freshman student retention using only available variables in the beginning of the semester has an overall accuracy rate of 85%. For nontraditional students, the highest predictive accuracy rate is 67.1%, while the model for traditional students predicts 84.5%. In many retention studies, the logistic regression model's classification accuracy rates range from 70% to 85% (Nandeshwar et al., 2011). In that regard, the model for this study shows similar results of predicting student retention and supports the argument that using easily obtainable data at the beginning of the fall semester could substitute for variables collected later in the semester. Even though retention literature highlights the importance of including factors that can be collected later in the semester, this study found that high school variables and basic college enrollment information successfully predicts freshman student retention.

Identifying at-risk nontraditional students starts with the definition because there are so many different types of students included in this category. Many retention studies of nontraditional students concentrate only on age criteria. However, this study shows that age alone is not a sufficient representation of nontraditional students. Due to NELS data, adult students (25 and older) could not be included in this study. Thus, the nontraditional characteristics of students in this study only apply to younger students. Yet, nontraditional students still made up more than half of the students (58%) sampled in the NELS study. Therefore, adult students alone do not completely represent nontraditional students. This indicates that studies of adult students representing nontraditional students may exclude a large amount of information regarding younger students who have a significant number of nontraditional characteristics.

Although there are evident differences between traditional and nontraditional students among the variables, the comparison of important variables predicting nontraditional and traditional students is minimal. Marital status and hours of work are important variables for predicting retention of nontraditional students, but those variables are not significant when predicting the retention of traditional students. This is understandable since traditional students are not always married and do not work full-time. Next discussion is categorized into three groups of variables: high school variables, college variables, and levels of nontraditional students.

High School Variables

Although only three out of 10 high school variables were considered as survived in the last model, those three variables have highly significant predictive capabilities. There were no differences of high school predicting variables between traditional and nontraditional students. Usually, retention studies found that high school performance measures such as high school GPA and ACT/SAT scores are good predictors. This study found that other than high school grades, high school attendance and the number of Carnegie units students have taken contain a more explanative power for predicting freshman student retention.

In this study, high school attendance (measured as number of times missing school) was a significant predictor of nontraditional student retention. There is not much literature on the relationship between high school attendance and college student retention. However, a few studies that used high school attendance as a student engagement indicator yielded mixed results (NCES, 1997; Horn & Chen, 1999). Most

high school retention studies found that school attendance is the crucial factor for passing a grade (The Education Alliance, 2009). Other studies found that college class attendance is the earliest and most accurate indicator of academic challenges (EAB, 2009). Literature of retention studies shows that high school GPA is one of the more important predictors of college freshman retention. Although it has the least effect, this study shows that high school grade is also recognized as an important variable for the prediction of first-year student retention. Although Adelman (1999) combined Carnegie units and high school curriculum intensity as a single variable in his study, this research used the Carnegie units variable as a stand-alone variable. This variable is highly predictive among other high school variables. Adelman (1999) warned, however, that this only represents the number of courses and is not equivalent to high performance standards.

College Variables

College variables include variables that are easily available in the beginning of the fall semester. The seven out of 10 variables in the last model were related to college. College variables that are more predictive are enrollment status, hours of work, marital status, number of regular courses taken, grants received, campus job, and type of degree.

Among predictive variables, the type of degree a student chooses is the best variable to use when predicting retention. Grants received and campus jobs are the next best variables and are more related to finance. Grants or scholarships are a good source of educational funding for nontraditional students. According to Beaudoin & Kumar (2012), higher education institutions typically offer grants and scholarships to traditionally underrepresented students. Studies also found that offering scholarships and financial aid

is fundamental for minority low-income and first-generation students to enroll and persist at institutions (Choy, 2001; Schneider, 2014). Also, availability of jobs on campus is a good retention tool for nontraditional students.

Those variables related to the at-risk characteristics of nontraditional students have a more negative effect on retention. Many nontraditional students are responsible for a career, family, and dependents. These responsibilities usually come before college tasks. Married students and students working full-time represent a portion of at-risk students who are most likely to dropout. Once these challenges conflict with their studies, these students are less likely to persist.

Studying part-time or full-time makes a significant difference to students. Students who choose to study part-time usually have many obligations other than college or they have difficulty affording the costs of higher education. However, this specific variable becomes less significant once the levels of nontraditional students are added to the model.

Levels of Nontraditional Students

The level of nontraditional students is an essential variable. This study shows that it has a unique effect on retention controlling for all the other variables. To study nontraditional student retention, it is important to define and identify each and every nontraditional characteristic and categorize students as minimally, moderately, and highly nontraditional. In this study, minimally nontraditional students are 68.75% more likely to retain, while highly nontraditional students are 40% more likely to retain.

This study is concentrated on seven nontraditional characteristics of freshman students that have a unique effect on predicting retention controlling for high school and college variables. Students who are highly nontraditional have the highest risk of dropping out in their first year. This result is consistent among other studies of nontraditional student retention (Horn & Carroll, 1996; NCES, 1997).

Limitations

Due to the nature of this study, there are five limitations; three of them are directly related to the use of publicly available NELS: 88/00 data. However, these limitations can be mitigated when a researcher takes into account each institution of higher education.

First, "use of a secondary data set always contains limitations, as the measurement of hypothesized constructs must be fit to archival data that were not collected with such constructs in mind," (Tyler, 2010). In this case, because NELS data was based on eighth-grade students, there was not enough information for adult students. The oldest students in this study were in their 20s. Therefore, NELS data could not represent the adult student population, which is a large part of nontraditional students today. According to NCES, the number of adult students grew by 42% between 2000 and 2010; at the same time, the number of younger students grew by only 34% (U.S. News, 2012). Additionally, there was inadequate information about student financial information in the NELS data, so it was impossible to define whether a student was financially independent or not. Thus, this characteristic was excluded from this study.

Second, some of the variables collected in the NELS data were not sufficient. For example, high school grade information was available, but it did not represent actual high school GPA. Information about the number of college preparation test scores was available, but it did not represent the actual test scores. Information about grants received was available, but it did not represent the actual grant amount. This limitation can certainly explain why some variables were not very predictive in this study, while they were significant in other studies. For example, other retention studies found that ACT/SAT scores are highly predictive (ACT, 2013). That information is not available in this study. In this study, the number of college preparation exams taken was represented in the POSTTEST variable; however, it was not significant.

Third, there was duplicate data when combining the second and third follow-up surveys. Some of the duplicate cases were exactly the same, and some had such distinct differences that there was no way to know which information was correct. In the latter situation, only the case that had the longer duration of stay was kept for the study. This may affect the results of the study. All the other duplicate cases were excluded using the "remove duplicates" function in Microsoft Excel.

Fourth, this study is limited to seven nontraditional characteristics of freshman students. Of course, these seven characteristics cannot represent the whole nontraditional student population. Therefore, the result of this study cannot be used to generalize the entire population of nontraditional students.

Finally, this study did not use the weights for the variables. That means the result of the study cannot be representative of the population. NELS study oversampled the

minority groups such as African and Hispanic Americans (NCES, 2002), so unweighted variables for these students can be overrepresented in the study.

Despite the limitations of this study, the results were useful to determine the important variables that are available in the beginning of the fall term for predicting nontraditional student retention. Individual college and universities can use their own data available in the beginning of a term to predict their nontraditional student retention.

Directions for Future Research

One of the vital points of this study was to find differences in predictive variables between nontraditional and traditional students. The result of this study found that there are only minimal differences. Due to the limitations of the NELS data, this study did not examine the adult student population. In the future, it might be viable to study the seven nontraditional characteristics in this study without any age limitations. And researchers of studies that only concentrate on adult students should be made aware that age alone should not be used to represent nontraditional students.

Many studies found that high school GPA was more predictive than other high school performance variables. However, this study found that high school attendance contains more of an explanation for nontraditional student retention. Thus, further research should consider this variable for predicting retention. Further studies should explore more variables that are available to administrators in the beginning of a fall

semester because this study was limited to the variables available in the NELS:88/2000 data.

Applying predictive analytics when identifying at-risk students is becoming a common practice for higher education administrators today. Implementing big data analysis practices, such as combining a large set of historical student data with different sources of student data to mine more accurate prediction and understand more in-depth information about each student in an institution, is a current phenomenon that will continue to gain momentum in the near future. By using such practices, institutions can predict retention and compare the rate of return on only pre-enrollment data or a combination of pre-enrollment data, college performance, and student survey results.

Many researchers dedicate their studies to different data mining techniques to compare it to the base model – logistic regression. From time to time, among comparison studies, using a logistic regression is still a very good and most robust technique for predicting retention (Nandeshwar et al., 2011). However, none of the studies tried ensemble data mining methods with logistic regression. Research on using ensemble methods for different learning algorithms and models found that ensemble selection consistently finds ensembles that outperform all other models (Caruana, Niculescu-Mizil, Crew & Ksikes, 2004). Ensembles are a better and less expensive way to find not only reliable solutions, but they also contain the superior value of combined results (Carter, 2012). According to Carter (2012), using ensemble methods achieves effective solutions for predictive modeling and has the following advantages:

- Ensemble methods have broad solution spaces (essentially multiplying the component search spaces) but search them narrowly, trying only combinations of best answers from the components.
- Ensemble methods avoid overfitting by utilizing components that read irrelevant data differently (canceling out noise) but read relevant inputs similarly (enhancing underlying signals).

As Carter (2012) described in his example, Figure 14 illustrates the relationship between two independent variables and predicts retention using logistic regression, random forest, and ensemble of two techniques. Logistic regression finds broader and direct relationships between variables, but cannot find signals or interactions between multiple variables. On the other hand, random forest is good with finding narrower signals, but can be overconfident and overfit to the noisy regions. Ensemble, on the other hand, illustrated the best of both methods by defining the relationship between high school grade and high school units.

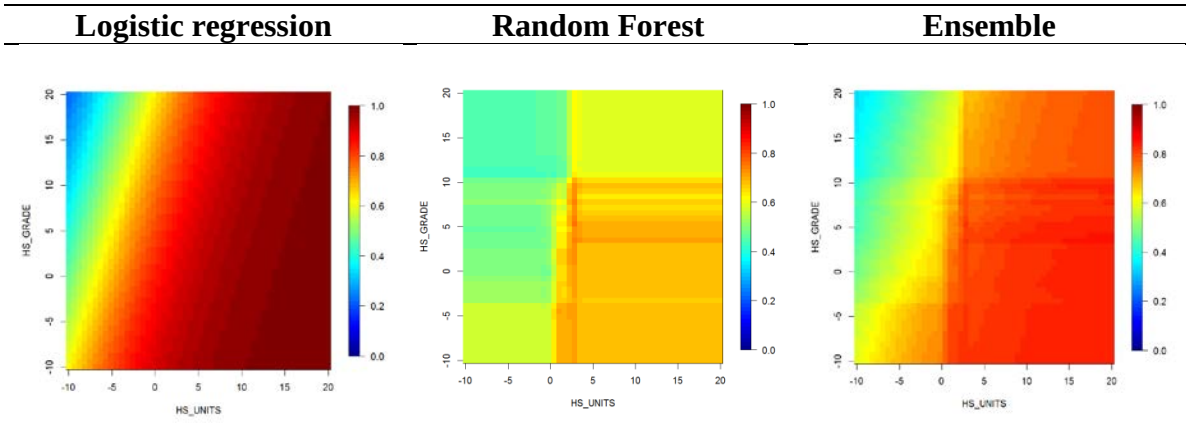

Figure 13. Logistic regression, random forest, and ensemble in HS_GRADE and HS_UNITS variables

Conclusions

Increasing numbers of nontraditional students have been matriculating at all levels of post-secondary education. Their types and characteristics are much more diverse and distinctive than traditional students. Defining and identifying this varied student population has never been easy and consistent. Studies have shown that their persistence and performance measures were not comparable with their traditional counterparts (NCES, 1997; Horn, 1996). Certainly, higher education is in need of a culture change, a change that values and validates the nontraditional student (ACSFA, 2012). Current external pressures demand higher education institutions to be efficient in retaining their students. Despite their efforts, national retention rates have remained stagnant for many years (EAB, 2009).

This study was focused on identifying those nontraditional students who are at-risk of dropping out. Most research done on nontraditional students has been directly related to adult students. Administrators, educators, and policy-makers simply refer to adult students as nontraditional students (ACSFA, 2009). However, this study found that using age-based criteria alone when studying post-secondary students eliminates large numbers of nontraditional students from research. Therefore, age should not be a critical standard for nontraditional students. College students today also possess nontraditional characteristics such as enrolling part-time, having dependents of their own, working part- or full-time, and having gaps between attending high school and college. In that sense, college administrators should identify the nontraditional characteristics on their campus

and categorize their students into three levels of nontraditional students. This approach will not eliminate those of younger students who possess nontraditional characteristics and it will be easier to target and retain those of highly nontraditional students from dropping out.

Determining important variables for predicting first-year student retention involves multiple ways of collecting student data over different times during the student college experience. Nevertheless, the reality is that administrators need to identify at-risk students as early as possible. This study attempted a more practical approach to identifying the variables that are available in the beginning of a term. The study found that even without extensive surveys or college grade information, administrators are still able to correctly predict nontraditional student retention about 70% of the time. Despite their gap in retention rates, there are only minimal differences between traditional and nontraditional students in terms of explanatory variables to retention. Among predictive variables, high school attendance is more important than high school grade and college degree type is more important than enrollment status and hours of work. In addition, those variables found important to student retention can easily fall into three factors: academic, financial and social.

APPENDIX A

Nontraditional students can include, but are not limited to, the following broad, overlapping subgroups:

- Single parents (adult single parents, teen single parents)
- Married students (divorced, widowed)
- Students with dependent children
- Students working full-time
- Part-time students
- Financially independent students
- Military personnel (active duty, reservist, and veteran)
- Adult learners (ready adult)
- Dislocated workers
- Low-income students (low-income adults)
- Working poor
- Unemployed poor
- Public assistance recipients
- Homeless students (accompanied homeless youth)
- High school non-completers (GED students)
- Historically underrepresented minorities (e.g. African American and Hispanic males)
- Distance learners (online learners)
- English as a Second Language (ESL) students
- First-generation students
- Undocumented students
- Students with disabilities (physical, mental, and learning disabilities)
- Older adults (senior citizens, retirees)
- Under-prepared students
- Students from foster care
- Orphans
- Wards of the court
- Minors

REFERENCES

Advisory Committee on Student Financial Assistance. (2012). *Pathways to success. Integrating learning with life and work to increase national college completion*. (A Report to the U.S. Congress and Secretary of Education). Retrieved December 13, 2013, from Institute of Education Sciences Web site: http://files.eric.ed.gov/fulltext/ED529485.pdf

ACT (1992). Factors Affecting Student Persistence: A longitudinal study. [ACT Research Report Series, 92-4]. Retrieved from ACT Web site: http://www.act.org/research/researchers/reports/pdf/ACT_RR92-04.pdf

ACT (2010). *What works in student retention?* (Fourth National Survey). Retrieved, January 10, 2013, from ACT Web site: http://www.act.org/ research/policymakers/pdf/droptables/CommunityColleges.pdf

ACT (2013). Differential effects on student demographic groups using ACT college readiness assessment composite score, ACT benchmarks, and high school GPA for predicting long-term college success through degree completion. [ACT Research Report Series]. Iowa City, IA.

Adelman, C. (1999). *Answers in a toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Adelman, C. (2006). The toolbox revisited: Paths to degree completion from high school through college. Washington D.C: U.S. Department of Education. Retrieved from http://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/toolbox.pdf

Allen, J., Robbins, S. B., Casillas, A. & Oh, I.-S. (2008). Third-year college retention and transfer: Effects of academic performance, motivation, and social connectedness. *Research in Higher Education, 49*, 647-664. doi: 10.1007/s11162-008-9098-3.

Astin, A. W. (1987). Assessment, value-added and educational excellence. *New Directions for Higher Education*, (59), 89-107.

Astin, A. W. (1997). How "good" is your institution's retention rate? *Research in Higher Education, 38* (6), 647-656.

Atwell, R. H., Ding, W., Ehasz, M., Johnson, S., & Wang, M. (2006). *Using data mining techniques to predict student development and retention*. Paper presented at the National Symposium on Student Retention, Albuquerque, NM.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12* (2), 155-185.

Bean, J. & Eaton, S. B. (2001). The psychology underlying successful retention practices. *Journal of College Student Retention, 3* (1), 73-89.

Bean, J. P. & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55* (4), 485-540.

Beaudoin, B. & Kumar, P. (2012). Using data to identify at-risk students and develop retention strategies. (Education Advisory Board, Custom Research Brief). Retrieved from California State University Web site: http://www20.csueastbay.edu/oaa/files/student_success/IDAtRiskStud.pdf

Beil, C., Reisen, C. A., Zea, M. C., & Caplan, R, C. (1999). A longitudinal study of the effects of academic and social integration and commitment on retention. *NASPA Journal, 37* (1), 376-385.

Berger, J. B. (2001). Understanding the organizational nature of student persistence: Empirically-based recommendations for practice. *Journal of College Student Retention, 3* (1), 3-21.

Berger, J. B., & Braxton, J. M. (1998). Revising Tinto's interactionalist theory of student departure through theory elaboration: Examining the role of organizational attributes in the persistence process. *Research in higher Education, 39* (2), 103-117.

Berger, J. B., Blanco Ramirez, G. & Lyon, S. (2012). Past to present: A historical examination of retention, p. 1-31 in A. Seidman (Ed.) *College Student Retention: Formula for Student Success*, (2nd ed.). Lanham, MD: ACE/Preager.

Berry, M., & Linoff, G. (1997). *Data mining techniques: For marketing, sales, and customer support.* New York: John Wiley & Sons.

Bogard, M., Helbig, T., Huff, G., & James, C. (2011). A comparison of empirical models for predicting student retention. Retrieved from Western Kentucky University Web site: http://wku.edu/instres/documents/comparison_of_empirical_models.pdf

Borden, V. (2004). Accommodating student swirl: When traditional students are no longer the tradition. *Change, 36* (2). 10-17.

Brown, S. M. (2002). Strategies that contribute to nontraditional/adult student development and persistence. *PAACE Journal of Lifelong Learning, 11*, 67-76.

Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College Persistence: Structural equations modeling test of an integrated model of student retention. *The Journal of Higher Education, 64* (2), 123-136.

Caison, A. (2007). Analysis of institutionally spesific retention research: A comparison between survey and Institutional datbase methods. *Research in Higher Education, 48* (4), 435-451. doi: 10.1007/s11162-006-9032-5.

Campus Explorer. (2013, March). Married college students. Retrieved from http://www.campusexplorer.com/college-advice-tips/CF0064F6/Married-College-Students/

Carter (2012). More is Always Better: The power of simple ensembles. Overkill Analytics Web site: http://www.overkillanalytics.net/more-is-always-better-the-power-of-simple-ensembles/

Caruana, R., Niculescu-Mizil, A., Crew, G., & Ksikes, A. (2004). *Ensemble selection from libraries of models*. Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.

Casselman, B. (2013, July 6). Number of the week: 'Nontraditional' students are majority on college campuses. *Wall Street Journal*. Retrieved from http://blogs.wsj.com/economics/2013/07/06/number-of-the-week-nontraditional-students-are-majority-on-college-campuses/

Center for Law and Social Policy. (2011). *Yesterday's nontraditional student is today's traditional student.* [Brochure]. Retrieved from CLASP Organization Web site: http://www.clasp.org/resources-and-publications/publication-1/Nontraditional-Students-Facts-2011.pdf

Chen, X. & Carroll, C. D. (2005). First generation students in postsecondary education. [Postsecondary Education Descriptive Analysis Report]. Retrieved from U.S. Department of Education Web site: http://nces.ed.gov/pubs2005/2005171.pdf.

Choy, S. P. (2001). Students Whose Parents did not go to College: Postsecondary Access, Persistence, and Attainment. (NCES 2001-126). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Choy, S. (2002). Nontraditional undergraduates. Findings from the condition of education 2002. National Center Foe Education Statistics. Retrieved from NCES Web site: http://nces.ed.gov/pubs2002/2002012.pdf

Clearinghouse. (2012). The overlooked student group: Adult learners. [Snapshot Report]. Retrieved from Clearinghouse Website: http://nscresearchcenter.org/wp-content/uploads/SnapshotReport4_Adult_Learners.pdf

CollegeBoard. (2009). *How colleges organize themselves to increase student persistence: Four-year institutions*. Retrieved from College Board Web site: http://professionals.collegeboard.com/profdownload/college-retention.pdf

DeBerard, M. S., Glen I, S. & Deana, C. J. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal, 38* (1), 66-80.

DesJardins, S. L., McCall, B. P., Ahlburg, D. A., & Moye, M. J. (2002). Adding a timing light to the "Tool Box". *Research in Higher Education, 43* (1), 83-110.

DeWitz, S. J., Woosley, M. L., & Walsh, W. B. (2009). College student retention: An exploration of the relationship between self-efficacy beliefs and purpose in life among college students. *Journal of College Student Development, 50* (1), 19-32.

Dimitrov, D. (2009). *Quantitative research in education: Intermediate and advanced methods.* Oceanside, NY: Whittier Publications Inc.

Education Advisory Board. (2009). Hardwiring student success. Building Disciplines for retention and timely graduation. Retrieved from The Advisory Board Company site:http://www20.csueastbay.edu/oaa/files/student_success/HardwiringStudSucc.pdf

Education Advisory Board. (2011). Supporting and retaining traditionally underrepresented students. [Custom Research Brief]. Washington, DC.

Education Advisory Board. (2012). The student Success Collaborative. Building an end-to-end solution for elevating degree completion and reducing costs. Retrieved from Indiana University Web site: http://uapp.iu.edu/InitRep/SSCfiles/SSC%20Research%20Deck%20Fall%202012a.pdf

Education Alliance. (2009). Student retention and dropout prevention. Policy recommendation and best practices. Retrieved from Education Alliance Web site: http://www.educationalliance.org/?ID=62&Resources

Fike, D. S., & Fike, R. (2008). Predictors of first-year student retention in the community college. *Community College Review, 36* (2), 68-86. doi: 10.1177/0091552108320222.

GED Testing Service. (2011). *Crossing the Bridge. GED Credentials and Postsecondary Educational Outcomes.* [Year Two Report].

Groux, C. (2012). Yesterday's Nontraditional students become today's norm. U.S. News. Retrieved from https://www.usnewsuniversitydirectory.com/articles/yesterdays-nontraditional-students-become-todays-n_12669.aspx

Habley, W. R., Bloom, J. L., & Robbins, S. (2012). *Increasing Persistence*. San Francisco, CA: Jossey-Bass.

Hanover Research (2011, May). Predicting College Student Retention. [Research Report]. Washington DC. Retrieved from http://www.hanoverresearch.com

Hartle, T. (2011). *Why graduation rate matter – and why they don't*. Retrieved from American Council on Education Web site: http://www.acenet.edu/the-presidency/columns-and-features/Pages/Why-Graduation-Rates-Matter%E2%80%94and-Why-They-Don%E2%80%99t.aspx

Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-a-vis regression. *New Directions for Institutional Research*. doi:10.1002/ir.185.

Horn, L. J., & Carroll, C. D. (1996). *Nontraditional undergraduates: Trends in enrollment from 1986 to 1992 and persistence and Attainment among 1989-90 Beginning postsecondary students*. Washington D.C: National Center For Education Statistics.

Hossler, D., Ziskin, M., Moore III, J. V., & Wahungu, P. K. (2008). *The role of institutional practices in college student persistence: Results from a policy-oriented pilot study.* Seattle, WA: Association for Institutional Research.

Hu, S., & St.John, E. P. (2001). Student persistence in a public higher education system: Understanding racial and ethnic differences. *The Journal of Higher Education, 72* (3), 265-283.

Huck, S. W. (2004). *Reading statistics and research* (4th ed.). Boston: Allyn & Bacon.

IBM. (2011). IBM SPSS Missing Values 20. [IBM Technical Report]. Retrieved from IBM Web site:

ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/20.0/en/client/Manuals/IBM_SPSS_Missing_Values.pdf

Ishitani, T. T. (2003). A Longitudinal approach to assessing attrition behavior among first generation students: Time-varying effects of pre-college characteristics. *Research in Higher Education, 44* (4), 433-449.

Johnson, D. R et al. (2007). Examining sense of belonging among first-year undergraduates from different racial/ethnic groups. *Journal of College Student Development, 48* (5), 525-540.

Kasworm, C., Polson, C., & Fishback, S. (2002). *Responding to adult learners in higher education*. Malbar, Fla.: Krieger.

Kasworm, C. (2003). Setting the stage: Adults in higher education. *New Directions For Student Services, 102*, 3-10.

Kelly, J. L., LaVergne, D. D., Boone, Jr., H. N., & Boone, D. A. (2012). Perceptions of college students on social factors that influence student matriculation. *College Student Journal,* 653-662.

Kelcey, B. 2009. Improving and Assessing Propensity Score Based Causal Inferences in Multilevel and Nonlinear Settings. Unpublished doctoral dissertation, University of Michigan.

Kim, K.A. (2002). ERIC review: Exploring the meaning of "nontraditional" at the community college. *Community College Review, 30*, 74-89.

Kuh, G.D., Kinzie, J., Buckley, J. A., Bridges, B. K., & Hayek, J. C. (2006). What matters to student success: A review of the literature. National Postsecondary Education Cooperative. Retrieved from NCES Web site: http://nces.ed.gov/npec/pdf/kuh_team_report.pdf

Kuh, G. D. et al. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79* (5), 540-559.

Lane, J. E. (2013). Are colleges in the northeast prepared for the new demographic reality? [The Nelson A. Rockefeller Institute of Government]. Retrieved from Web site: http://www.rockinst.org/observations/lanej/2013-02-Are_Colleges_in_the_Northeast_Prepared_for_the_New_Demographic_reality.aspx

Metzner, B. S. & Bean, J. P. (1987). The estimation of a conceptual model of nontraditional undergraduate student attrition. *Research in Higher Education, 27* (1), 15-35.

Nandeshwar, A., Menzies, T., & Nelson, A. (2011). Learning Patterns of student retention. *Expert Systems with Applications, 38*, 14984-14996. doi: 10.1016/j.eswa.2011.05.048.

National Center for Education Statistics (2002). Quick guide to using the NELS:88/2000 data. Taken from: Base-year to fourth follow-up. Retrieved from http://nces.ed.gov/surveys/nels88/pdf/QuickGuide.PDF

National Center for Education Statistics (2012, October). Nontraditional undergraduates. Retrieved October, 2012, from NCES Web site: http://nces.ed.gov/pubs/web/97578e.asp

National Center for Higher Education Management Systems. (2010). Retention rates – first-time college freshmen returning their second year. Retrieved March, 2014 from NCHEMS Information Center website: http://www.higheredinfo.org/dbrowser/index.php?submeasure=223&year=2010&level=nation&mode=data&state=0

Nuñez, A., & Cuccaro-Alamin, S. (1998). First-generation students: Undergraduates whose parents never enrolled in postsecondary education. [Statistical Analysis Report] National Center for Education Statistics. Washington, DC.

Pittman, K. (2008). *Comparison of data mining techniques used to predict student retention.* Doctoral dissertation, Nova Southeastern University, Ft.Lauderdale, FL.

Pyke, S. W., & Sheridan, P. M. (1993). Logistic regression Analysis of graduate student retention. *The Canadian Journal of Higher Education, 23* (2), 44-64.

Raju, D. (2012). *Predicting student graduation in higher education using data mining models: A comparison*. Doctoral Dissertation, Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama, Tuscaloosa, AL.

Rendón, L. (1994). *Beyond involvement: Creating validating academic and social communities in the community college.* University Park PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Schuetze, H. G., & Slowey, M. (2002). Participation and exclusion: A comparative analysis of nontraditional students and lifelong learners in higher education. *Higher Education 44*, 309-327.

Seidman, A. (2012). *College student retention: Formula for student success*. Lanham, MD: Rowman & Littlefield Publishers Inc. 2nd edition.

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry, 19* (3), 227-229.

Terenzini, P. T., Lorang, W. G., & Pascarella, E. T. (1981). Predicting freshman persistance and voluntary dropout decisions: A Replication. *Research in Higher Education, 15* (2), 109-127. doi: 0361-0365/81/060109-19.

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

Tinto, V. (1993). *Leaving College: Rethinking the causes and cures of student attrition.* 2nd ed. Chicago, IL: University of Chicago Press.

Tinto, V. (2007). Research and practice of student retention: What's next? College *Student Retention, 8* (1), 1-19.

Torres, V. (2003). Influences on ethnic identity development of Latino college students in the first two years of college. *Journal of College Student Development, 44* (4), 532-547.

Tyler, A. L. (2010). Examining the Stem Educational Pipeline: The influence of pre-college factors on the educational trajectory of African American students. Doctoral dissertation. Miami University, Miami, FL.

UPCEA & InsideTrack. (2013). Measuring post-traditional student success: Institutions making progress, but challenges remain. [Whitepaper]. Retrieved from Insidetrack Website: http://www.insidetrack.com/wp-content/uploads/2013/06/upcea_whitepaper_2013.pdf

Vandamme, J.-P., Meskens, N., & Superby, J.-F. (2007). Predicting academic performance by data mining methods. *Education Economics, 15* (4), 405-419.

Walsh, T. A. (1997). *Developing a postsecondary education taxonomy for inter-institutional graduation rate comparisons*. PhD dissertation, University of New York, Buffalo, NY.

Woosley, S. & Jones, D. (2011). The Foundation of Map-Works: Research and theoretical underpinnings of Map-Works. [Whitepaper]. Retrieved from http://indstate.edu/studentsuccess/pdf/The%20Foundation%20of%20MAP-Works.pdf

Zernike, K. (2010, January 21). Annual Poll of Freshmen Shows Effect of Recession. *New York Times*. Retrieved from http://www.nytimes.com/2010/01/21/education/21college.html?_r=0

BIOGRAPHY

Bolorchimeg Boldbaatar was born in 1979 in Ulaanbaatar, Mongolia. She graduated with honors from the lyceum at the Mongolian University of Science and Technology (MUST) in 1996 and received an invitation to enroll at MUST without any examinations. She decided to pursue her higher education in the U.S. and received her Bachelor of Science in Marketing from Minnesota State University Moorhead in 2005. After graduating, she went back to Mongolia to work as an internet marketing specialist in a bank and as a project manager for the Mongolian National Amusement Park. After five years of working in the corporate world, she wanted to follow her childhood dream of making a difference in the field of education and started her graduate degree in higher education in the U.S. While studying at George Mason University, she became fascinated with big data analytics and realized the need for data analysts in the higher education field. Within two and half years, she had completed her Masters in Higher Education and also obtained a Graduate Certificate in Data Analytics. Upon completion of this degree, she was recognized by the Phi Kappa Phi honor society as being in the top 10 percent of her graduating class and became an active member. She also joined the American Educational Research Association and participated its 2014 annual meeting in Philadelphia as a graduate student. Her dedication to educational research made her become a lifetime learner of learning, teaching, and data analysis. Her professional interests include retention and completion of college students, instructional design, learning science, learning analytics, data visualization, and data mining.