PREDICTION OF RETENTION AND PROBATION STATUS OF FIRST-YEAR COLLEGE STUDENTS IN LEARNING COMMUNITIES USING BINARY LOGISTIC REGRESSION MODELS

A Dissertation

by

RITA A. SPERRY

BS, Texas A&M University-Corpus Christi, 2004
MS, Texas A&M University-Corpus Christi, 2007

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR of PHILOSOPHY

in

CURRICULUM AND INSTRUCTION

Texas A&M University-Corpus Christi
Corpus Christi, Texas

August 2014

© Rita Annette Sperry

All Rights Reserved

August 2014

PREDICTION OF RETENTION AND PROBATION STATUS OF FIRST-YEAR COLLEGE STUDENTS IN LEARNING COMMUNITIES USING BINARY LOGISTIC REGRESSION MODELS

A Dissertation

by

RITA A. SPERRY

This dissertation meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

Bryant Griffith, PhD
Chair

Kamiar Kouzekanani, PhD
Committee Member

Dan Pearce, PhD
Committee Member

Susan Wolff Murphy, PhD
Graduate Faculty Representative

JoAnn Canales, PhD
Dean, College of Graduate Studies

August 2014

ABSTRACT

The first year of college is a critical period of transition for incoming college students. Learning communities have been identified as an approach to link students together in courses that are designed with first-year students' needs in mind. Yet, learning community teaching teams are often not provided with data prior to the start of the semester about their students in order to target interventions. One question then becomes, what variables known on or before the first day of classes are predictive of first-year student success, in terms of retention and probation status, for first-year college students in learning communities?

The correlational study employed univariate and multivariate analyses on pre-college data for three consecutive cohorts of first-year students (n = 4,215) in learning communities at a regional public university in South Texas. Logistic regression models were developed – for all students as well as for individual learning community categories – to predict retention and probation status using the variables of first-semester hours, developmental status, high school percentile, transferred hours, SAT score, age, gender, first-generation status, ethnicity, Pell Grant eligibility, admission date, admission status, and orientation date.

Results indicated that group differences were statistically significant for retention based on all pre-college variables excluding first-generation status or age, while group differences were statistically significant for probation status on the basis of all pre-college variables except age. The model to predict retention for all students included five variables (high school percentile, SAT score, Pell Grant eligibility, days since admission, and days since orientation), and the model to predict probation status included three additional variables (transferred hours, gender, and ethnicity). The models for individual learning communities contained different sets of

predictor variables; the most common predictors of retention or probation status were high school percentile and orientation date.

The study has practical implications for admissions officers, orientation planners, and learning community practitioners based on the pre-college variables, such as orientation date, that were found to be predictive of retention or probation status. Topics for further research include exploring the pre-college variables that did not predict either outcome, such as first-generation status, for first-year students in learning communities.

DEDICATION

I dedicate this dissertation to my husband, Mark Sperry, without whom I would never have had the courage to pursue graduate work, let alone a PhD. His support makes me want to be a better student, teacher, wife, mother, friend, and human being. I know it was not easy on him as a passenger on this dissertation journey, and I will be ever thankful for his encouragement and love. I also dedicate this dissertation to our son, Andrew Jacob, who was not even a glimmer in our eyes when I entered the doctoral program. He continues to amaze me each day, and I am more proud of being his mother than of any academic distinction.

ACKNOWLEDGEMENTS

I acknowledge my husband, Mark, for his understanding and patience throughout my lengthy career as a student. I also acknowledge our son, Andrew, for his smiles, babbles, and kisses that served as constant reminders of my goal to finish the degree. I look forward to post-graduate life with my family.

I acknowledge Dr. Bryant Griffith, my dissertation chair, for his support for the past five years. I was extremely nervous about entering the doctoral program, but those fears were immediately put at ease after the first day of his Curriculum Theory course. I was fortunate to take several courses with him over the next few years, all of which forced me to critically analyze who I was as an educator and an individual. His guidance throughout the dissertation process – even after it admittedly stalled a bit after the birth of my son – was tremendous.

I also acknowledge Dr. Kamiar Kouzekanani, who served as the methodologist for the dissertation study. I took Dr. Kouzekanani's Applied Statistics II class in Spring 2011, and it was probably the most challenging course in my doctoral studies, yet it was also the one I enjoyed the most. Although I did well in statistics as an undergraduate mathematics minor, I did not really understand its real-world application until explained to me in Dr. Kouzekanani's lectures. Each week brought with it new types of problems to solve, and I made it my personal mission to learn as much as possible about the material, especially when it involved interpreting and reporting results. I am proud to have the distinction of being the first student at TAMU-CC to ace all three exams in the course. When it came time for me to select someone for my dissertation committee to assist with quantitative research, there was really no alternative. Fortunately, Dr. Kouzekanani agreed and I was able to work closely with him this past semester. Although no causal inferences can be drawn, his work ethic, responsiveness, and guidance were all directly related to my completion of the dissertation.

I would like to thank my committee members, Dr. Dan Pearce and Dr. Susan Wolff Murphy. I could not have asked for a more supportive group of individuals to guide me through this process. Your questions and encouragement challenged me while at the same time making me feel like my goal of obtaining a PhD was achievable.

To my friends, Chelsey Patterson and Luke Jackson, I thank you for your support as the three of us took distinctively different paths in pursuing doctorate degrees. I look forward to getting back in touch as graduates.

To my colleague and friend, Michelle Riley, I thank you for the lengthy conversations about my study and its implications. Our brainstorming sessions were invaluable to the writing process. I promise to be there for you when you need the same.

To Dr. Juan Carlos Huerta, my supervisor, mentor, and friend, I thank you for your words of wisdom and encouragement throughout the course of my studies.

To Dr. John D. Fernandez, thank you for recognizing my potential as a timid college sophomore, and serving as my academic father-figure and life coach over the past twelve years. I wish you the best of happiness in your retirement.

To my students, past and present, thank you for your enthusiastic support and interest about my research over the past few years. I have the greatest job in the world.

Finally, I acknowledge my family, especially my father, John Hodges, and mother, Annette Hodges, who instilled in me a sense of personal responsibility and commitment to excellence that made this journey possible. Thank you to my siblings, Joanna Grubaugh and Jack

Hodges, for not only being my best friends, but also serving as my greatest intrinsic motivators throughout my academic career. I hope your big sister has made you proud!

TABLE OF CONTENTS

CONTENTS PAGE

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

Background and Setting

In the current global and economic climate, student persistence and graduation rates from institutions of higher education in the United States are more scrutinized now than ever before. Not surprisingly, it is a student's academic performance that is the greatest predictor of retention and ultimate graduation (Astin, 1975; Hall, 2007). Student success in the first semester of college has been shown to have a significant impact on persistence (Hosch, 2008; Tharp, 1998). In response, colleges and universities have implemented a wide variety of programs in an attempt to positively affect two well-documented predictors of future academic success: first-semester grade point average (GPA) and first-year retention.

The Association for American Colleges and Universities (AAC&U) launched the Liberal Education and America's Promise (LEAP) initiative in 2005 to address ongoing issues in higher education, including a study of which educational practices have the greatest impact on the success of college students at all levels. Kuh (2008) outlined ten specific teaching and learning practices that the LEAP initiative found to be most effective at increasing student retention and engagement, each of which are linked to retention and ultimate graduation. Two of these practices – first-year seminars and learning communities – are particularly salient to an examination of first-year student success.

First-year seminar courses, although widely varied in implementation, have been repeatedly supported as a successful initiative to increase first-year student success rates (Engstrom & Tinto, 2008; Griffin & Romm, 2008; Tobolowsky, Cox, & Wagner, 2005). Additionally, learning communities, which are formed by the linking of two or more courses for a shared cohort of students, have demonstrated significant rewards for students and faculty alike (Hill, 1985; Huerta, 2004; Lardner & Malnarich, 2008; Smith, MacGregor, Matthews, & Gabelnick, 2004; Tinto, 2000). Along with the Washington Center for Improving the Quality of Undergraduate Education, the National Resource Center for the First-Year Experience and Students in Transition supports the embedding of first-year seminar courses within learning communities as an effective practice for integrating the entire first-year experience (Henscheid, 2004). However, it is still unclear which students benefit the most from these initiatives and which ones continue to struggle to succeed and persist. Thus, further exploration that includes the specific characteristics of students in learning communities is necessary in order to appreciate this phenomenon and continue the effort to improve the quality of higher education.

Statement of the Problem

A regional public four-year university in South Texas has had a required learning community experience that includes a first-year seminar course since it began admitting first-year students in 1994. Several published studies have demonstrated the achievement of the program in helping students successfully make the transition from high school to college (Araiza, 2006; Huerta, 2004; Sterba-Boatwright, 2000). The program has also gained national recognition as a leader in the learning community movement (Kutil & Sperry, 2012; Smith, MacGregor, Matthews, & Gabelnick, 2004). Yet, it is still unclear as to the particular characteristics of the learning community program that contribute most to student success in terms of retention and probation status. In addition, little is known on the first day of class about which students are most at risk of landing on probation or not returning for their sophomore year. In order to gain this valuable insight so that interventions can be targeted within the learning communities from the onset of the academic year, a thorough analysis of the characteristics of recent first-year students, as well as their ultimate successes or failures in the first year, is warranted.

Theoretical Framework

Tinto's (1975) Student Integration Model postulated that the students who persist and succeed in college are those who are able to successfully integrate into an institution's social and academic environment. Alternatively, the students who are more likely to struggle and fail to persist are those who do not attempt or achieve social and academic integration. The Student Integration Model, depicted in Figure 1, identified a variety of external or pre-college factors that play a role in college student integration, including past academic performance (prior qualifications), family background (family attributes), and personal goals (individual attributes), as well as experiences at the institution (inside and outside of the classroom). Tinto's model and these external factors provided the theoretical framework for the study.

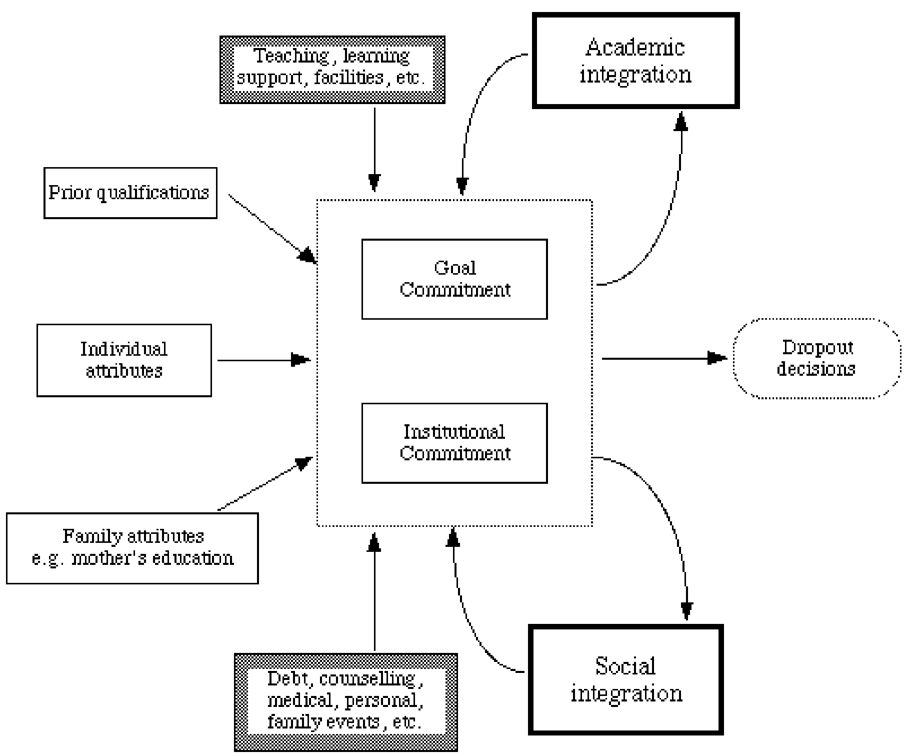

Figure 1. Tinto's Student Integration Model (Draper, 2008)

Borrowing from Van Gennep's (1960) anthropological concept of rites of passage, Tinto (1988) defined three stages of student departure: separation, transition, and incorporation. At each of these three abstract stages, students decide whether or not to remain in college. Tinto argued that the first semester of college is particularly crucial to helping students make the successful social and academic transition that leads to persistence and ultimate graduation. Tinto (1997) later updated his Student Integration Model to include the significance of classroom experience and faculty interactions on student success and persistence, arguing for the implementation of learning communities to assist in this critical academic period.

Tinto's (1975; 1997) SIM is particularly salient to the prediction of outcomes such as retention and probation status because it provides a framework for identifying and categorizing the types of incoming variables that are related to student success. According to the SIM, student persistence is a function of various factors – past academic performance, personal and family background, personality and goals, and college experiences – each of which plays a role in explaining the end result. If these factors could each be measured, then it would be feasible to develop statistical models to predict whether or not a given student would be successful.

Over the past several decades, Tinto's emphasis on the first semester of college has been answered by numerous studies attempting to predict first-semester GPA and retention using a multitude of variables. Tinto's model and personal experience has also helped to bring learning communities to the forefront of research in higher education, most recently in the creation of assignments in learning communities that require students to integrate content and skills across disciplines (Huerta & Sperry, 2010; Lardner & Malnarich, 2009), social network analysis studies to examine peer relationships within learning communities (Chamberlain, 2011; Smith, 2010; Stuart, 2008), and the opportunities that learning communities provide for developmental education (Hansen, Meshulam, & Parker, 2013; Heany & Fisher, 2011; Synder, Hakett, Stewart, & Smith, 2002). The intersection between the research on the prediction of first-year student success and the implementation of learning communities, however, presents a research void that has yet to be filled.

Purpose of the Study

The purpose of the study was to determine which pre-college variables, that is, independent variables that can be collected on or before the first day of classes, were predictors of retention or probation status for first-year students in a learning communities program, with the goal of developing models to predict the probability of success (in terms of retention or probation status) for future students based on these variables. The following questions informed the research study:

1. What pre-college variables are predictors of the retention of first-year students in learning communities?
2. What pre-college variables are predictors of the probation status of first-year students in learning communities?

Operational Definitions

For the purpose of the study, the following operational definitions were employed:

Retention was measured based on whether or not a student in the Fall 2010, Fall 2011, or Fall 2012 cohort was registered for at least one course in the Fall 2011, Fall 2012, or Fall 2013 semester, respectively (0 = no, 1 = yes).

Probation Status was measured by determining whether a student's grade point average (GPA) after his or her first semester was at or above a 2.0 on a 4-point scale (1 = below 2.0, on probation; 0 = at or above 2.0, not on probation).

First-Year Students were defined as college students between the ages of 18 and 24 (Bean & Metzner, 1985; Clark, 2012; Wyatt, 2011) who matriculated with less than 30 hours, were admitted less than a year before the start of classes, and were enrolled in a learning community at the university for the first time during the Fall 2010, Fall 2011, or Fall 2012 semester.

The term *Learning Community* signified a class schedule with at least two linked courses in which the same cohort of students were co-enrolled, one of which was First-Year Seminar. Variations of learning communities included triads (three linked courses) and tetrads (four linked courses).

Pre-College Variables described data that were available about students from the registrar on or before the first day of classes.

Delimitations, Limitations, and Assumptions

The study was delimited to (a) first-year students at a single South Texas public university; (b) 13 pre-college variables which served as potential predictors; and (c) the outcome measures of retention and probation status. Due to the non-experimental nature of the study, no causal inferences were drawn. Some of the pre-college variables were self-reported; the underlying assumption was that students were truthful in reporting details such as high school rank and size, birthdate, and first-generation status on their applications for admission. It was also assumed that the data collected from the registrar were accurate and complete.

Significance of the Study

According to the National Center for Higher Education Management Systems (2013), the national retention rate for first-time college freshmen at all four-year institutions in 2010 was 77.10%, while the Texas state retention rate came in a bit lower at 73.30%. Research has indicated that student performance in the first year is predictive of cumulative undergraduate

GPA and subsequent graduation (Terenzini, Springer, Yaeger, Pascarella, & Nora, 1996). The aim of the study was to identify the pre-college student characteristics that are useful in prediction of the retention and probation status of first-year students in learning communities. Clearly, any empirical evidence to support a particular intervention (such as learning communities) to increase first-year student success would be worthy of note to any institution concerned about student persistence and graduation rates.

In addition, the learning community program under review is determined to combat complacency and actively seeks out ways to increase the probability of student success. The results of the study shed light on particular characteristics of incoming students so interventions can be targeted to the students who might profit from them the most. Learning community teaching team members can identify which students are most at risk at being on probation and not returning the following year. With increasingly limited resources, being able to predict where attention is needed is invaluable. Because the results are based on former students who participated in the program, the learning communities that are shown to have been successful in the past in helping students stay off probation and remain at the institution could be further explored. Ideally, the traits of the successful learning communities could then be extended to the program as a whole in order to help the entire first-year class. Although the particular predictors and models created in the study may not be generalizable to outside institutions because of differing student populations and situations, the process of analyzing pre-college traits of first-year students in learning communities could easily be replicated in other settings. National supporters of the learning communities' movement, such as the Washington Center or the LEAP initiative, would undoubtedly be interested in results that contribute to the growing body of literature about how learning communities contribute to student success.

CHAPTER II

LITERATURE REVIEW

Introduction

The first year of college is a critical period in a student's academic career. Many students struggle to successfully make the transition. The factors that contribute to first-year student success and persistence are of obvious interest to anyone in the field of higher education and the body of relevant literature is impressive in its depth and scope. The majority of research on first-year student success and persistence is based on the work of one of four major theoretical founding fathers. This chapter outlines their respective models and describes relevant research regarding existing models of first-year student retention and predictors of first-year success. The discussion concludes with a brief exploration of learning communities, particularly in relation to first-year students.

Theoretical Models of Student Success and Persistence

Much of the research on first-year student success is based on the theoretical models of Spady (1970; 1971), Astin (1975; 1991; 1993), Bean (1980; 1983), and Tinto (1975; 1988; 1997). Each model has been extensively researched, supported, and challenged in the literature on college student achievement over the past four decades. While each of the theorists examined the success and persistence of college students through a distinctive theoretical lens, all four collectively demonstrate that the puzzle of first-year student departure contains many pieces.

Spady's Model for Student Attrition

Spady (1970; 1971) was the first to propose a model for student attrition and his was based on Durkheim's Theory of Suicide. According to Spady (1970), when students do not achieve social integration (ie, fail to make friends at college), then they are at risk of academic "suicide" and more prone to drop out. Spady (1971) tested his own theory with 683 first-year students at the University of Chicago. Through various questionnaires and interviews over four years, he discovered that friendship was a significant factor in determining student persistence, as well as GPA, social integration, and overall satisfaction with the college experience for the male participants.

Astin's Input-Environment-Outcomes Model

Based on his longitudinal study of student persistence, Astin (1975; 1984) postulated that student success in college was directly related to personal involvement in college life. Thus, students who were not involved were more likely to drop out. Astin (1991) identified 146 precollege variables (including high school grades, ethnicity, and parental level of education), as well as 192 environmental variables (such as institution type and size, financial aid, and peer group characteristics), that contribute to student success. In his seminal work, *What Matters in College? Four Critical Years Revisited,* Astin (1993) outlined the key components of his Input-Environment-Outcomes Model that divided the student success puzzle into three mutually exclusive areas, the third of which focused on 82 possible "output" characteristics of students once they entered college (such as satisfaction, achievement, and retention).

Bean's Causal Model of Student Attrition

In the early 1980s, John Bean joined the conversation on student success in college. Bean (1980) developed a causal model for student attrition that borrowed from the concept of employee turnover in the work place. He defined several background characteristics – such as past academic achievement and distance from home to school – that contributed to student success and retention when combined with features of the institution's organizational structure. Using regression results from surveys of 1,171 first-year students, Bean (1980) found that the intervening variable of institutional commitment was statistically significant in predicting dropout for both men and women. Bean (1983) later refined his theory to emphasize student satisfaction. Based on his questionnaire study of 876 women, Bean (1983) found that intent to leave was the greatest predictor of attrition.

Tinto's Student Integration Model

Perhaps the most prolific research on student attrition and retention over the past 40 years has been led by Vincent Tinto of Syracuse University. Tinto's (1975) Student Integration Model furthered Spady's (1970; 1971) work, combined it with the sociocultural notion of "rites of passage," and theorized that the students who persist and succeed in college are those who are able to successfully integrate into an institution's social and academic environment. The Student Integration Model also identified a variety of external factors that play a role in college student achievement, including past academic performance, family background, and personal goals.

Tinto (1988) defined four distinct phases of student departure, beginning with recruitment and admission, and argued that the first semester of college was particularly crucial to persistence and ultimate graduation. A few years later, Tinto's (1993) work developed into an interactionalist model focused on a student's interactions with faculty and peers both inside and outside of the classroom. Tinto (1997) later updated his Student Integration Model to include the significance of classroom experiences on student success and persistence, specifically noting that learning communities were uniquely situated to bridge the gap between social and academic integration.

Tinto's Student Integration Model (SIM) has not been without its critics. Some have argued against Tinto's (1975) "rite of passage" claims regarding the transition to the social and academic setting of the university, citing that true rites of passage occur within a single culture and not in the transition from one environment to another (Kuh & Love, 2000; Tierney, 1992). In her illuminating ethnographic piece *My Freshman Year: What a Professor Learned by Becoming a Student*, Nathan (2005) contended that college represents a “liminal state” in which reality is suspended for a period until the individual reintegrates into society, and not a traditional “rite of passage” as argued by Tinto. According to Nathan and others, Tinto’s sociological foundation is flawed.

Other critics of Tinto’s SIM have argued that it is oriented only towards traditional full-time and residential college students (ages 18-24) at four-year institutions and cannot be generalized to the entire population of students in higher education (Braxton, 2000; McCubbin, 2003; Pascarella & Terenzini, 1991). Researchers discovered inconsistencies with the SIM when it was applied to particular ethnic minorities (Torres & Solberg, 2001) or students with disabilities (Duquette, 2000). Voorhees (1987) defined a gap in the SIM’s ability to model community college student persistence, which was echoed in the results of a study of public community college students by Borglum and Kubala (2000).

Many have also argued that Tinto’s SIM is incomplete (Braxton, Sullivan, & Johnson, 1997; McCubbin, 2003; Tinto, 1982). In a test of convergent validity between the two theories, Cabrera, Castañeda, Nora, and Hengstler (1992) recommended the integration of selected elements of Tinto’s SIM with others from Bean’s (1980) model of student attrition to gain a more comprehensive understanding of student departure. Alternatively, Milem and Berger (1997) suggested the combination of Astin’s (1984) theories about involvement with Tinto’s SIM to shed light on first-year student retention. More recent attempts to model student retention, such as Nora’s (2002) student engagement model, have extended Tinto’s SIM to include the significant role that culture and ethnicity play in student retention for today’s students. Over the years, Tinto (1993; 2000) has incorporated elements from both Astin's and Bean's models in an attempt to make the SIM more comprehensive.

In the past two decades, much of the research on college student departure has taken a qualitative turn and has moved away from the development of statistical models (Braxton, 2000; Seidman, 2012). Regardless, student success remains to be a puzzle with many unknown pieces, many of which are beyond the control of the institution. Although Tinto (2006-2007) himself acknowledged the complexity of student behaviors, the limits of current models, and a "gap between research and practice" (p. 4), the tenets of his SIM continue to provide the theoretical foundation for many retention programs and services across the nation, as well as much of the literature on student success and persistence (Seidman, 2012). Some have gone as far to argue that Tinto's model has achieved paradigmatic status in the study of student retention (Braxton & Hirschy, 2005). The SIM has been cited in no less than 775 research articles and three of its underlying propositions have been accepted as "reliable knowledge" in the field based on decades of research and empirical examinations (Braxton & Lee, 2005). Despite its debated shortcomings, the SIM defines a straightforward empirical process for categorizing student variables that can work together to explain, or predict, complex outcomes such as retention and probation status.

Models for Predicting Student Retention

The first documented studies of "student mortality" took place in the 1930s when researchers conducted academic autopsies in order to discover when and why students left college (Berger & Lyon, 2005). Decades later, Spady (1971) examined the literature in the field of retention in the 1950s and 1960s and grouped the studies into six categories: philosophical, census, autopsy, case, descriptive, and predictive. The final category, predictive studies, involved the use of admissions criteria to forecast student potential. While much research has been conducted in this area, there currently is no standard statistical model for predicting whether or not a given student will succeed in the first year of college.

Berger and Lyon (2005) posited that any research involving retention is inherently complex. The student population in higher education has changed dramatically over time and continues to change. In addition, as Tinto's (1975; 1993) SIM suggests, retention challenges are particular to the context of the campus, the types of students it attracts, the roles and interactions of faculty and other professionals on the campus, the interventions available at the institution, and the relationship with the outside community at large. The creation of a single statistical model to predict student success or retention on a large scale regardless of institutional context is a considerable – if not impossible – challenge.

Yet this challenge is exactly what Kuh, Cruce, Shoup, Kinzie, and Gonyea (2008) tackled when they attempted to model the factors that contribute to first-year college grades and persistence based on data from 18 institutions across the country. Their work combined National Survey of Student Engagement (NSSE) data with ACT, College Board, and institutional records for 6,193 students in an attempt to discover what variables were related to higher GPAs and persistence regardless of institutional context. Kuh et al. (2008) found that prior academic achievement had the largest influence on first-year GPA, but that the magnitude of the effect was decreased when student engagement measures were added to the model. However, Kuh et al. (2008) also noted that the results of their study might not be the same in another year or with other institutions.

Along a similar vein, Crissman Ishler and Upcraft (2004) attempted to catalogue all of the variables that have been found to play a role in first-year student retention. Their review of the literature included student input variables (prior academic achievement, socioeconomic status, gender, age, race/ethnicity, parents and other family, and degree commitment); institutional variables (selectivity, type, size, private vs. public, gender composition, race composition); and an array of environmental variables (first-year GPA, major, enrollment status, quality of student effort, interactions with faculty, interpersonal interactions, extracurricular activities, work, satisfaction, alcohol abuse, Greek life, campus climate, financial aid, athletics, campus interventions, classroom experiences, first-year seminars, orientation, living environments, learning communities, academic advising, service-learning, supplemental instruction, developmental education, and other support services). Instead of attesting to the universality of their findings, however, Crissman Ishler and Upcraft (2004) noted that persistence is “very much institution specific” (p. 32

Because retention is so dependent on context, Bean (2005) encouraged institutions to confront the student departure problem by diving deeper into the student data and experiences at their individual campuses. The lists of variables provided by Kuh et al. (2008) and Crissman Ishler and Upcraft (2004) provide a starting point for this type of undertaking, but often the variables used in prediction models are limited to the predictors that are readily available to campus registrars (Snyder, Hackett, Stewart, & Smith, 2002). Not surprisingly, big business has found a market in providing consultation services to campuses that combine enrollment and survey data to create prediction models. Noel-Levitz (2014), for example, now offers a Student Retention Predictor (SRP) that determines the probability of attrition and risk factors for incoming students based on the results of their College Student Inventory combined with campus data. Another similar product created by Educational Benchmarking called MAP-Works has recently been used by Indiana University in the creation of a campus-wide prediction model for first-year student retention (Drake, 2011a).

Institutions not interested in paying a premium for this type of statistical modeling service often rely on the expertise of their existing employees to create useful prediction models based on available student data, as evidenced by the work of Drake (2011b) at Purdue University and the Office of Institutional Research at the Community College of Philadelphia. *College and University* published a series of articles on the research done at the University of South Florida to establish a predictive formula to determine attrition risk and implement targeted interventions based on the results (Herreid & Miller, 2009; Miller, 2007; Miller & Herreid, 2008; Miller & Tyree, 2009; Miller, Tyree, Riegler, & Herreid, 2010). A cursory search of academic databases reveals an abundance of recent dissertations and articles regarding the prediction of first-year student retention at individual institutions (Heaney & Fisher, 2011; McKenzie, Gow, & Schweitzer, 2004; Perry, 2010; Pizzo, 2010; Synder, Hackett, Stewart, & Smith, 2002).

Various methods of statistical analysis have been used to examine first-year student GPA and persistence in the literature. The most basic approaches involve the utilization of one-way analysis of variance (ANOVA) procedures to compare the GPA or retention of various groupings of students (Borglum & Kubala, 2010; Heaney & Fisher, 2011). By far the most widespread statistical technique regarding prediction of first-year student success or retention is regression analysis, typically in the form of logistic regression using retention as a binary outcome variable (Glynn, Sauer, & Miller, 2003; Herreid & Miller, 2009; Kahn & Nauta, 2001; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; Nora, Cabrera, Hagedorn, & Pascarella, 1996; Porter, 1999; Snyder, Hackett, Stewart, & Smith, 2002; Voorhees, 1987; Wetzel, O'Toole, & Peterson, 1999). A more advanced form of regression modeling, survival analysis, was proposed by Murtaugh, Burns, and Schuster (1999) to predict not only *if* a student would leave college but also *when* the withdrawal would occur.

Other student persistence models have incorporated the use of discriminant analysis and path analysis (Elkins, Braxton, & James, 2000; Pascarella & Chapman, 1983). First-year student persistence and achievement have also been investigated using structural equation modeling and exploratory factor analysis (Allen, 1999; Cabrera, Nora, & Castañeda, 1993; McKenzie, Gow, & Schweitzer, 2004). Thomas (2000) and Stuart (2008) used social network analysis to study the relationship between student integration and persistence. Finally, Bogard, Helbig, Huff, and James (2011) suggested the superiority of data mining approaches, such as neural networks and decision trees, over more traditional statistical methods to predict retention.

Predictors of First-Year Student Success

Research on the unique variables that contribute to first-year student success – which is usually defined using GPA or retention as the outcome measure – is extensive and varied, but the predictors can be roughly grouped into four categories defined by Tinto's (1975; 1997) Student Integration Model. The first is academic achievement and background, including variables such as high school GPA, advanced placement participation, and standardized test scores. The second grouping of factors includes student personal and family characteristics such as parent education levels, gender, and financial aid status. Various instruments and inventories have been created to examine the relationship between student personality traits and first-year student success. Finally, college-related experiences have been viewed as potential contributing factors to overall academic achievement in the first semester of college.

Past Academic Achievement

Perhaps the most widely demonstrated research on first-semester GPA prediction involves the use of high school GPA as an independent variable (Astin, 1971; Goldman & Widawski, 1976; Noble & Sawyer, 2004; Stiggins, Frisbie, & Griswald, 1989; Zheng, Saunders, Shelley, & Whalen, 2002). Maggard (2007) found high school GPA to be a significant predictor of first-semester GPA for student athletes, even more than class rank or standardized test scores, accounting for 21% of the variance alone. Scott, Tolson, and Lee (2010) determined that participation in Advanced Placement (AP) courses led to higher first semester GPAs. In a study by Aleamoni and Oboler (1977) on the use of ACT and SAT standardized test scores to predict first-semester GPA, both were found to be equally predictive but less powerful than high school class percentile. The predictive power of high school GPA and standardized SAT/ACT scores varies by study and accounts for anywhere between 12 and 29 percent of total variance in retention (Hanover Research, 2011).

Personal and Family Background

There have been several studies on the predictive nature of gender on first-semester GPA (Chase, 1981; Mattson, 2007). Parent education levels have also been related to GPA at all undergraduate levels (Nelson, 2009). Race has also been shown to play a small role (Boyer & Hickman, 2007), as well as financial aid status (Cabrera & Nora, 1992; Coperthwaite & Klimczak, 2008), in first-year student GPA. Cirillo and Smith (2008) constructed an inventory of 37 pieces of biographical information – self-reported figures regarding the frequency of engaging in experiences related to the theme areas of expectations, academic preparedness, technological savvy, independence, distractions, ease of adjustment, personal habits, and success orientation – for prediction that was more closely correlated to first-semester GPA than standardized test scores.

Student Personality and Goals

Dispenzieri and Giniger (1971) found that a student's readiness to work hard, as well as insight into his or her own abilities, was positively correlated to first-semester GPA. Tross, Harper, Osher, and Kneidinger (2000) conducted a similar study and found that student conscientiousness was predictive of first-semester GPA, explaining nearly 7% of the variance and more than SAT standardized test scores. Strayhorn (2011) found that a combination of precollege performance and positive beliefs about academic skills could explain 30% of the variance in first-semester GPA.

College Experiences

A variety of college-related factors have also been linked to first-semester performance and retention into the second year. The Community College of Philadelphia (2011) found that the number of credit hours attempted in the first semester was the strongest predictor of persistence for its Latino students, and significant differences in retention were found between college-ready students and those who had to enroll in at least one developmental course. Hoyt (1999) conducted a study of community college students which found that the number of remedial courses was inversely related to first-semester GPA, and directly related to student dropout rates. Time factors – such as application, orientation, and admission dates that may serve as proxies for harder-to-quantify student personality traits like motivation and commitment – have also been connected to the first-semester performance of conditionally admitted students at Indiana University (Drake, 2011b). Finally, participation in a learning community has been linked to first-semester GPA and retention (Tinto, 2000; Waldron & Yungbluth, 2007).

Learning Communities

Learning communities have become common practice in higher education, especially as first-year student retention and success interventions. Loosely defined as “clusters of courses organized around a curricular theme that students take as a group” (Laufgraben, 2005b, p. 371), learning communities can be historically traced to the early 20th century pedagogical foundations of John Dewey and the experimentation of Alexander Meiklejohn at the University of Wisconsin (Smith, MacGregor, Matthews, & Gabelnick, 2004). Although the learning community model is rarely identical from one institution to the next, the growing base of literature on the positive impact of learning communities on first-year student success and persistence is hard to deny.

Influenced by Dewey’s challenge for educators to promote “habits of mind” while avoiding the fragmentation common in undergraduate education curricula, Meiklejohn created Experimental College and the University of Wisconsin in 1927 (Shapiro & Levine, 1999). This two-year program involved an intense interdisciplinary study of culture from Athens to the Americas. Seminars, one-on-one conferences, and team teaching were common elements of this unique residential program that attracted students on the margins, challenged traditional college notions about grading policies and student-faculty interactions, and struggled to gain momentum in its short five years of existence (Smith, MacGregor, Matthews, & Gabelnick, 2004).

Although decreased Depression-era funding, negative press, and campus politics ended Meiklejohn’s experiment prematurely, its legacy lived on to inspire the work of Joseph Tussman at Berkeley in the 1960s (Shapiro & Levine, 1999). Tussman’s Experiment included the creation of a program of interrelated courses for students to take during their lower division undergraduate years to integrate the entire learning experience. This deceptively simple model for faculty collaboration inspired similar reforms at institutions across the country, sparking a movement that has gained significant traction over the decades (MacGregor & Smith, 2005).

In a study of 365 four-year institutions, Zhao and Kuh (2004) found that roughly 30% of all first-year students had participated (or planned to participate) in a learning community. Smith, McGregor, Matthews, and Gabelnick (2004) estimated that over 500 institutions had learning communities programs across the nation and that the majority fell into one of three general categories: learning communities with courses that were unmodified, learning communities of linked or clustered classes, and team-taught learning communities. Shapiro and Levine (1999) specified an additional model, residential learning communities. The differences between the models are based on the size and overlap of the linked courses, as well as the amount of integration of course materials that occurs in the semester.

The simplest learning community model, paired or clustered courses, requires the enrollment of a cohort of students in at least two classes together (Shapiro & Levine, 1999). The faculty members may elect to integrate their courses, but the degree of integration frequently depends on the amount of time the faculty members choose to invest in creating those links. Paired or clustered learning communities can also be theme-based and frequently include a first-year experience course. The second and most cost efficient learning community structures involve small cohorts from large courses that enroll in an attached weekly seminar course; these are often referred to as Freshman Interest Groups (FIGs). Coordinated studies or team-taught learning communities are perhaps the most complex because the involved faculty members are expected to completely integrate their disciplines, course requirements, and lesson plans. Finally, residential learning communities require that students not only enroll in the same courses based on some common interest or major, but also live in student housing together. Cocurricular activities are designed to encourage students in residential learning communities to integrate experiences inside and outside of the classroom.

The structure of learning communities inherently creates classroom environments that inspire good practice in teaching – as defined by Chickering and Gamson (1987) – which include increased contact between first-year students and faculty, more collaboration among first-year peers in smaller classes, and active and cooperative learning strategies (Laufgraben, 2005a). In addition, learning communities allow for the type of interactive, supportive, and time-on-task environment that Astin (1993) describes as positively contributing to first-year student success (Engstrom, 2008).

The current center for learning community pedagogy, practice, and research can be found at The Evergreen State College in Washington, which was founded in the 1970s as one of the first colleges with a coordinated interdisciplinary curriculum (Shapiro & Levine, 1999). The Washington Center for Improving the Quality of Undergraduate Education at The Evergreen State College serves as the National Resource Center for Learning Communities and hosts an annual Summer Institute on Learning Communities to assist institutions across the nation in building their own learning communities programs (Washington Center, 2014). Led by Emily Lardner and Gillies Malnarich, the Washington Center is also home to *Learning Communities Research and Practice*, an online peer-reviewed journal for learning community practitioners.

Under the leadership of Lardner and Malnarich, the Washington Center has made learning community work a research priority. Lardner and Malnarich (2008) developed a heuristic for the creation of assignments within learning communities to assess students' abilities to integrate their learning experiences, allowing for so-called "intentional integration". In conjunction with Boix-Monsilla from Harvard's Project Zero, the Washington Center led the National Project on Assessing Learning in Learning Communities from 2006 to 2008. This voluntary project included 22 institutions that came together to assess integrated assignments using a protocol developed by Boix-Monsilla. The results of this study increased awareness of the interactive relationship between interdisciplinary understanding and integration that exists in learning communities, and led to the development of a consistent protocol for responding to students' integrative work (Lardner & Malnarich, 2009).

Despite the growth of learning communities as a movement in higher education, there is limited published research, especially in relation to first-year student success and persistence at four-year universities (Andrade, 2008; MacGregor & Smith, 2005). Taylor, Moore, MacGregor, and Lindbland (2003) identified 32 formal research studies and 119 institutional reports on learning communities programs. In a quasi-meta-analysis of first-year learning community programs, Andrade (2008) found 17 published articles on the impact of learning communities on first-year students, only 12 of which measured persistence. Fifteen of the studies addressed first-semester GPA. Andrade's (2008) results indicated that the research on learning communities appears to demonstrate their positive contribution to student outcomes – such as increased GPAs and persistence – but that it remains unclear as to what specific aspects of learning communities contribute the most to their success. The heterogeneity of programs across the country, as well as the self-selection effect common to most learning community programs, makes interpretation of the data difficult (Andrade, 2008; Habley & Bloom, 2012).

Roccini (2011) was also interested in the impact of learning communities on first-year students. His study explored the literature and identified more than 40 studies that supported the positive outcomes commonly attributed to learning communities (Habley & Bloom, 2012). Roccini's (2011) study used path analysis to examine the responses to the College Student Experience Questionnaire, and the results indicated that although learning communities do contribute to the success of first-year students, the impact is indirect via student engagement. In other words, learning community participation is related to increased student engagement, which is in turn related to educational gains. These findings echo the results of Zhao and Kuh's (2004) examination of the experiences of learning community participants who responded to the National Survey of Student Engagement.

A small number of published studies have attempted to aggregate findings about learning communities across multiple institutions. In 1993, Vincent Tinto directed a project for the National Center on Postsecondary Teaching, Learning and Assessment that examined three learning community programs for first-year students: The University of Washington's Freshman Interest Groups (FIGs), LaGuardia Community College's learning community clusters, and Seattle Central Community College's Coordinated Studies Program. Tinto, Love, and Russo (1993) found through quantitative and qualitative methods that not only do students in learning communities report positive perceptions of classes, peers, faculty, and themselves at higher rates than non-learning community participants, but also that these students persisted at significantly higher rates.

The Manpower Demonstration Research Corporation (MDRC) recently partnered with the National Center for Postsecondary Research on a six-year grant to explore the variations of learning communities and their impact on student success for community college students. A study of six community colleges found that learning communities had small positive effects on overall academic progress, but no impact on persistence, for developmental students (Visher, Weiss, Weissman, Rudd, & Wathington, 2012). The concerning MDRC findings have spurred recent entreaties by the Washington Center for learning community programs across the nation to conduct self-assessments in order to contest the results (E. Lardner, personal communication, April 25, 2014). Although the impact is often hard to isolate, learning communities are considered one of ten High-Impact Practices (HIPs) endorsed by the Association of American Colleges and Universities' LEAP initiative because of their relationship to deep learning, effective educational practices, and self-reported personal and practical gains (Kuh, 2013).

Summary

All first-year students who choose to attend college bring with them a multitude of experiences that can ultimately impact their future achievement (Astin, 1993; Tinto, 1975). Experiences in the higher education environment also play a role in student adjustment, performance, and retention (Kuh, 2008; Engstrom & Tinto, 2008). Kuh (2013) reported that 18% of all first-year students who participated in the 2012 administration of the National Survey of Student Engagement enrolled in a learning community during their first year. Yet, little to no research exists on the prediction of first-year student success and retention in the context of learning communities.

The goal of the study, therefore, was to answer the call of the National Learning Communities Project by bridging the research gap between the literature on predictors of first-year student success and the recent discussions about the impact of learning communities. Specifically, this venture focused on predicting the retention and probation status of first-year students enrolled in a single learning communities program. Tinto (1975; 1997; 2000) suggests that success in the first semester of college is critical to student persistence and that learning communities can make a difference. The study took both of these propositions as a given; the purpose, therefore, was not to determine if, but rather *to what extent*, pre-college variables play a role in the first-year transition to college for students in learning communities. On the basis of an extensive review of the literature, it was hypothesized that pre-college variables would be useful in predicting both retention and probation status for students in learning communities.

CHAPTER III

METHOD

The purpose of the study was to test the hypothesis that pre-college variables are useful in predicting first-year retention and probation status among students in learning communities. The setting was a public university in South Texas. The study was guided by the following research questions:

1. What pre-college variables are predictors of the retention of first-year students in learning communities?
2. What pre-college variables are predictors of the probation status of first-year students in learning communities?

Research Design

The study employed a correlational design. Correlational research is commonly used in applied behavioral sciences, especially when the manipulation of variables is difficult or impossible (Kamil, Langer, & Shanahan, 1985; Vogt, 2007). The purpose of the correlation is to determine whether or not relationships exist between or among variables (Triola, 2002). In the study, archival data were collected for 13 pre-college independent variables to explore how they were related – either individually or in combination – to each of the dependent, or outcome, variables. Specifically, the study was predictive in nature, in which the pre-college variables were used to explain variation in retention and probation status. Additionally, results were used to formulate binary logistic regression prediction models. Practical significance of the findings were investigated. No causal inferences were drawn due to the non-experimental nature of the study.

Independent/Predictor Variables

Variables were selected for the study based on the availability and accessibility of data from campus records related to each of the major categories of Tinto's (1975; 1997) SIM, as well as the literature on predictors of first-year student performance. The 13 pre-college variables selected fell into three of the categories of the SIM: past academic performance (high school percentile, transferred hours, and SAT score), personal and family background (gender, first-generation status, ethnicity, Pell Grant eligibility, and age), and college experiences (first-semester hours, developmental status, admission date, admission status, and orientation date). The fourth category of the SIM, personality and goals, was not included in the study due to its inherent reliance on measurements not readily available on or before the first day of classes.

High School Percentile

High school class rank and size were self-reported in the application for admission to the university and were confirmed by the admissions office when final transcripts were submitted upon high school graduation. Percentiles were then created to standardize class ranks across the student population. For example, a class percentile of .97 indicated that the student was in the 97^{th} percentile of his or her class and ranked higher than 97% of all other graduating students. High school percentile scores ranged from zero (0) to one (1).

Transferred Hours

Some students entered college with dual credit, transfer, and/or AP credits. The variable ranged from 0 to 29 hours for the students in the learning communities program.

SAT Score

The Scholastic Assessment Test (SAT) or American College Test (ACT) scores were required for admission and were included in the student's academic record. For the purpose of the study, the ACT scores were converted into the SAT scores, using the ACT's (2008) concordance tables. If a student had scores for both tests, the higher score was selected. The SAT scores ranged from 510 to 1600.

Gender

The student's gender, male or female, was included in the application. A dichotomous variable represented gender in the study (1 = female, 0 = male).

First-Generation Status

As part of admission to the university, students were asked whether or not at least one of their parents graduated from college. The student's self-reported data were coded as a one (1) for those whose one or both parents were college graduates, and a zero (0) if both parents had not graduated from college.

Ethnicity

Student ethnicity was self-reported as part of the application process to the university. The institution utilized the following categories for reporting purposes: Hispanic/Latino, Black or African American, White, American Indian or Alaskan Native, Asian, Native Hawaiian or Pacific Islander, and Nonresident aliens for international students. For the purposes of the study, the variable was recoded into either Hispanic or non-Hispanic.

Pell Grant Eligibility

Pell Grants were distributed based on the determined financial need of the student. Data were coded as either eligible or non-eligible for the grant.

Age

The student's age, in years, at the start of the semester was determined based on student date-of-birth, which was included in the application for admission.

First-Semester Hours

Students enrolled in their first year can take anywhere between 1 to 21 hours; most take between 12 and 18 hours. The variable represented the number of hours attempted at the start of the semester, not the number of hours earned.

Developmental Status

There were two developmental courses at the university during the course of the study. State mandates required students not meeting certain criteria in the areas of reading and mathematics to enroll in a corresponding developmental course during their first semester of study. A dummy-coded variable represented the status (1 if the student had enrolled in at least one of the two developmental courses, 0 if the students had been classified as “college-ready” and did not have to take either of the two courses).

Days since Admission

Each student’s record contained an admission date. The number of days between the admission date and the start of the appropriate fall semester was calculated to determine the days since admission.

Admission Status

The institution used three levels of admission, which included one level for student accepted by normal admission standards and two levels for alternatively-admitted students. For the purpose of the study, data were recoded into either accepted or alternatively admitted.

Days since Orientation

Each incoming first-year student had attended a summer orientation. The number of days between the orientation date and the start of the appropriate fall semester was calculated and served as the operational definition for this predictor variable.

Dependent Variables/Outcome Measures

Retention Status

The first dependent variable, retention into the subsequent fall semester, was included in each record. A value of one (1) for this variable represented a student who had returned to the university in the following fall semester and zero (0) represented a student who had not been retained.

Probation Status

The second dependent variable, probation status after the first semester, was based on the overall first-semester GPA for the each of the fall cohorts of first-year students in the learning communities program. First-semester GPA ranged from 0 to 4.0 and students who earned below a 2.0 were placed on academic probation. A dummy-coded value of zero (0) represented students not on academic probation, while a value of one (1) signified the on-probation status of the study participants.

Subjects and Setting

The learning communities program for the study was located at a public four-year institution in South Texas. At the onset of the study, the undergraduate student population was primarily composed of Hispanic and White students, representing 46.02% and 40.07% of the 9,152 students, respectively, and the university was designated as a Hispanic-Serving Institution. All traditional incoming first-year students were required to enroll in a learning community during their first and second semesters. In Fall 2010, 1533 students enrolled in the learning communities, 1503 students in Fall 2011, and 1806 in Fall 2012.

Most of the learning communities in the program were triads, meaning that they contained three courses in which students co-enrolled in cohorts of 25 students. There were other learning communities ranging from two to five linked classes with varied cohort sizes. Every learning community contained a section of UCCP 1101 (First-Year Seminar) that supported the other courses in the learning community. The UCCP 1101 was a requirement for graduation from the institution. Most of the learning communities were also linked to ENGL 1301 (Composition I), a core curriculum first-year writing course. There were learning community options for students who had entered the program with credit for ENGL 1301.

Each learning community in the program was centered on one or two large core curriculum courses. For example, the Sociology learning community in Fall 2010 (Triad B) had 200 seats. All students in the learning community enrolled in the sociology course, Human Societies (SOCI 1301) on Tuesdays and Thursdays at 9:30am. The students were divided into eight groups of 25 for their UCCP 1101 course. Six of the sections were also linked to two sections of ENGL 1301, so the same 25 students who were in UCCP 1101 also attended their First-Year Composition course together. The instructors for the Sociology, First-Year Seminar, and First-Year Composition courses met weekly to plan assignments and activities. Students completed several assignments based around themes from the Sociology course and grades were often shared in more than one of the linked classes.

Table 1 shows a summary of the learning communities offered in the Fall 2010, Fall 2011, and Fall 2012 semesters. For the purpose of the study, the learning communities were grouped by subject. The six learning community categories were as follows: Sociology (Triad B), History (Triads C, E, K, and M), Political Science (Triads F and L), Science (Triad S and Tetrads V and W), Developmental History (Tetrad N), and Other (Triads G and T).

Table 1

Learning Communities Offered in Fall 2010, Fall 2011, and Fall 2012

Learning Community	Fall 2010	Fall 2011	Fall 2012
Triad B – Sociology	Yes	Yes	Yes
Triad C – History	---	---	Yes
Triad E – History	Yes	Yes	Yes
Triad F – Political Science	Yes	Yes	Yes
Triad G – Geology	---	---	Yes
Triad K – History	Yes	Yes	Yes
Triad L – Political Science	Yes	Yes	Yes
Triad M – History	Yes	Yes	---
Tetrad N – Developmental History	Yes	Yes	Yes
Triad S – Biology/Chemistry	---	Yes	Yes
Triad T – Chemistry	---	Yes	Yes
Tetrad V – Biology/Chemistry	Yes	Yes	Yes
Tetrad W – Biology/Chemistry	Yes	Yes	Yes

Collection and Preparation of Data

The data used for the study were collected from the university's registrar's office. The researcher requested all the demographic information and student records from the department directly, a process that required both Institutional Review Board (IRB) and registrar approval (Appendix). The following data were obtained for the Fall 2010, Fall 2011, and Fall 2012 cohorts of students who were enrolled in UCCP 1101: (a) retention into the subsequent fall semester, (b) first-semester GPA, (c) UCCP 1101 section, (d) first-semester credit hours attempted, (e) first-semester developmental hours attempted, (f) high school class rank, (g) transferred hours, (h) SAT score and/or ACT score, (i) date of birth, (j) gender, (k) first-generation status, (l) ethnicity, (m) Pell Grant eligibility, (n) admission date, (o) admission status, and (p) orientation date.

The data were provided in three comma-delimited spreadsheet files, one for each fall cohort of students, and contained a total of 4,673 student records. Each file was first opened individually in Microsoft Excel. Students in the online section of UCCP 1101, not intended for first-year students, were removed from the study, as well as students who were not in their first semester of college or who were classified as sophomores, juniors, or seniors. Additionally, any students who did not have both an admission date and an orientation date were not included in the study. Finally, early college high school students who were enrolled in the learning communities were removed from the date file. After these deletions, the Fall 2010, 2011, and 2012 data files contained 1,376, 1,401 records, and 1,698 records, respectively.

The next task was to prepare the data, which required the use of both Microsoft Excel and IBM Statistical Package for the Social Sciences (SPSS). In Excel, the date-of-birth, admission date, and orientation date fields were converted to shorter date forms which did not include a timestamp. The high school class rank was provided in the original spreadsheet as a fraction; for example, a student who graduated second in a class of 340 students had “2/340” listed in the column for class percentile. In order to work with the numbers mathematically, the characters had to be separated into two columns using an Excel function that found the “/” delimiter. Then, an Excel formula was created to calculate the percentile for the student by subtracting the class rank quotient from one.

The remainder of data preparation was completed in SPSS. The UCCP 1101 section number was recoded into a new column representing the learning community letter (see Table 1) to which the First-Year Seminar course was linked. Ultimately, this column was recoded into a third column that combined the learning communities together by category. The number of developmental hours taken in the first semester was recoded to a new column for developmental status that became a one (1) for students with at least one developmental course and a zero (0) for students with no developmental courses. The transferred hours column was filled with a zero (0) for students who did not have any credits coming into their first semester. Gender was recoded with one (1) representing female and zero (0) representing male. The Ethnicity field contained text which was recoded with one (1) for Hispanic students and zero (0) for all other students. Similarly, the Admission Status field was recoded with a one (1) for "Accepted" students and a zero (0) for all other statuses.

The student's age was calculated using the Datediff function in SPSS to calculate the number of days between the start of class and the student's date-of-birth. The result was divided by 365 to determine the age in years. A comparable method was used to calculate the days between admission and the first day of classes, and the days between orientation and the first day of classes. Two columns in the data file contained the SAT and ACT composite scores for each student, but not all records contained scores for one or both of the fields. A concordance table was used to recode the ACT scores to SAT scores (ACT, 2008). The higher of the two scores (SAT or ACT-to-SAT) was recorded in a new column to be used in the analysis.

Once the preparation process was applied to all three data files, they were merged into one SPSS file. First-semester GPA was used to determine whether or not students landed on academic probation by means of a SPSS syntax statement that tested whether the GPA was below 2. The last step was to exclude any student under the age of 18 or over the age of 24. The final date file consisted of 4,215 student records.

Data Analysis

Data analysis involved descriptive statistics, univariate analyses, and multivariate analyses. Descriptive statistics were used to summarize and organize the variables into a profile of subjects.

Univariate analyses were conducted to explore the relationship between each independent variable and each of the two dependent variables. Group comparisons for categorical variables were performed, using a series of Chi-Square Test of Independence and the odds ratios were calculated to examine the practical significance of the findings (Field, 2013). Group comparisons for continuous variables were conducted, using a series of t-test for Independent Samples. When the homogeneity of variances assumption was not met, the Welch approximate t was used as the test statistic (Stevens, 1999). Effect sizes were calculated to determine the practical significance of the findings (Cohen, 1988).

A correlation matrix, using Pearson Product-Moment Correlation Coefficient (Field, 2013), was created to examine the independent variables for evidence of multicollinearity prior to multivariate analyses. Correlations were also calculated between each independent variable and each of the two dependent variables, using the Pearson Product-Moment Correlation Coefficients for continuous variables and Phi Coefficients (Field, 2013) for categorical variables. Multivariate analyses employed binary logistic regression. Logistic regression is used to estimate the probability of an event occurring – in the study, either retention or probation – based on a set of predictor variables (Field, 2013). In binary logistic regression, a dichotomous outcome is transformed into a linear model by comparing each independent variable to the log odds of the event taking place. In an exploratory study, each independent variable is tested to determine its unique contribution to the prediction of the outcome, that is, its relationship to the log odds of the event to determine if it meets the inclusion criteria to be included in the final model. The model can then be used to estimate the probability of the event occurring as p(event) = $1/(1 + e^{-z})$, where $z = \text{Constant} + B_1X_1 + B_2X_2 + \ldots + B_nX_n$. The Constant and Bs are coefficients obtained from the logistic regression. The Wald statistic tests whether the coefficient for each of the independent variables in the model is zero (0); it has a chi-square distribution (Field, 2013; Hosmer & Lemeshow, 2000). The Nagelkerke R^2 and classification table were examined to evaluate the practical significance and power of the logistic regression models. Similar to other coefficients of determination, the Nagelkerke R^2 represents the amount of variance in the outcome that is explained by the model's variables (Nagelkerke, 1991). The Hosmer and Lemeshow Test was used to examine the goodness-of-fit of logistic regression models (Hosmer & Lemeshow, 2000).

Odds ratios were calculated and examined to interpret the variables which defined various models. An example of the calculation of odds ratios follows (Howell, 1992):

In the study, 1,969 students who were college-ready in mathematics and reading were retained into the second fall semester, while 1,183 of the college-ready students were not retained. The number of college-students who were retained (1,969) was divided by the number of college-ready students who were not retained (1,183), resulting in odds of 1.66 to 1 for the retention of college-ready students. Similarly, 588 students who were not college-ready were retained and 475 who were not college-ready were not retained; the odds of being retained among non-college-ready students were equal to 588 divided by 475, or 1.24 to 1. The odds ratio which compared the odds of being retained for college-ready students to the odds of being retained for non-college-ready students was calculated as 1.66 (retention odds of college-ready students) divided by 1.24 (retention odds of non-college-ready students), which resulted in an odds ratio of 1.34 to 1. The practical interpretation of this finding is that students who were college-ready were 1.34 as likely to be retained as were students who were not college-ready.

CHAPTER IV

RESULTS

The purpose of the study was to determine the pre-college variables that predict the retention and probation status of first-year students in learning communities at a regional university in South Texas. The study used existing data for three consecutive cohorts of first-year students and employed both univariate and multivariate statistical techniques to analyze a large number of categorical and continuous variables. Several binary logistic regression analyses were conducted, using two dichotomous outcome measures, namely, one-year retention and probation status following the first semester. The Likelihood-ratio Chi-square was used to test the statistical significance of the prediction models. The Wald statistic was employed to examine the statistical significance of the individual predictor variables. The Hosmer-Lemeshow Chi-square Test was performed to examine the goodness-of-fit of the models. The Nagelkerke R^2 and classification tables were used to examine the practical significance and power of the models. The exponential regression coefficients (odds ratios) were examined to interpret the individual predictors which defined the models; for the purpose of the study, odds ratios ≥ 1 indicated the likelihood of the event happening. The level of significance was set, a priori, at .01.

Profile of Subjects

The data for the study consisted of 4215 first-year student records for the Fall 2010, Fall 2011, and Fall 2012 semesters. The subjects were between the ages of 18 and 24, matriculated with less than 30 transferred hours, and enrolled in a First-Year Seminar (UCCP 1101) course in a learning community. The majority (62.20%) were enrolled in either History (35.40%) or Political Science (26.80%) learning communities, were college-ready in reading and mathematics (74.80%), were female (58.60%), were not first-generation (68.70%), and were alternatively admitted (54.10%). While no ethnicity was in the majority, 45.50% and 40.60% of the subjects had been identified as Hispanic and White, respectively. Results are summarized in Table 2.

Table 2

Profile of Subjects, Categorical Variables, n = 4215

Variable	F	%
Learning Community		
Sociology	514	12.20
History	1493	35.40
Political Science	1131	26.80
Science	765	18.10
Developmental History	168	4.00
Other	144	3.40
Developmental Status		
College-Ready	3152	74.80
Not College-Ready	1063	25.20
Gender		
Male	1745	41.40
Female	2470	58.60
First-Generation Status		
Non-First-Generation	2896	68.70
First-Generation	1319	31.30
Ethnicity		
Hispanic	1919	45.50
Non-Hispanic	2296	54.50
Pell Grant Eligibility		
Eligible	2094	49.70
Ineligible	2121	50.30
Admission Status		
Accepted	1934	45.90
Alternative Admit	2281	54.10

The high school percentile data were treated as ordinal with the median of .68. The average SAT score was 966.22 (SD = 140.72). The distribution of age at the start of the fall semester was positively skewed; the median age was 18.59 years. A typical first-year student enrolled in 13.77 hours (SD = 1.25) during the first semester, was admitted 175.52 days (SD = 78.47) prior to the start of the semester, and attended orientation 37.06 days (SD = 24.21) before the first day of classes. The distribution of transferred hours brought in by first-year students was positively skewed with a median of 0.00 hours. Results are summarized in Table 3.

Table 3

Profile of Subjects, Continuous Variables, n = 4215

Variable	Mean	Median	SD
High School Percentile	NA	.68	NA
SAT Score	966.22	950.00	140.72
Age	18.68	18.59	.63
First-Semester Hours	13.77	14.00	1.25
Days since Admission	175.52	182.00	78.47
Days since Orientation	37.06	34.00	24.21
Transferred Hours	3.96	.00	6.27

Univariate Analyses

Retention Status

Students who returned after their first year of college were compared to students who did not return after their first year of college on the basis of categorical variables of learning community, developmental status, gender, first-generation status, ethnicity, Pell Grant eligibility, and admission status. With the exception of learning community, $\chi^2(5, N = 4215) = 11.10$, $p =$

.05, and first-generation status, $\chi^2(5, N = 4215) = 2.07$, $p = .15$, all group differences were statistically significant. Results are summarized in Table 4.

Table 4

Group Comparisons by Retention Status after First Year, Categorical Variables

	Retained Yes F/%	Retained No F/%	χ^2
Learning Community			
Sociology	287/55.80	227/44.20	
History	903/60.50	590/39.50	
Political Science	679/60.00	452/40.00	
Science	488/63.80	277/36.20	
Developmental History	103/61.30	65/38.70	
Other	97/67.40	47/32.60	11.10
Developmental Status			
College-Ready	1969/62.50	1183/37.50	
Not College-Ready	588/55.30	475/44.70	17.05*
Gender			
Male	1015/58.20	730/41.80	
Female	1542/62.40	928/37.60	7.79*
First-Generation Status			
Non-First-Generation	1778/61.40	1118/38.60	
First-Generation	779/59.10	540/40.90	2.07
Ethnicity			
Hispanic	1106/57.60	813/42.40	
Non-Hispanic	1451/60.70	845/39.30	13.56*
Pell Grant Eligibility			
Eligible	1158/55.30	936/44.70	
Ineligible	1399/60.70	722/39.30	50.17*
Admission Status			
Accepted	1300/67.20	634/32.80	
Alternative Admit	1257/55.10	1024/44.90	64.33*

* $p < .01$

Calculation and interpretation of the odds ratios indicated that students who were college-ready in reading and mathematics were 1.34 times as likely to be retained as were students who were not college-ready. Females were 1.20 times as likely to be retained as were males, while non-Hispanics were 1.26 times as likely to be retained as were Hispanics. Students not eligible for Pell Grants were 1.57 times as likely to return for their second year as were students who were eligible for Pell Grants. Students who were accepted based on established admission criteria were 1.67 times as likely to be retained as were students who were alternatively admitted.

Students who were retained into the fall semester of their second year were compared to students who were not retained into the fall semester of their second year on the basis of the continuous variables of SAT score, age, hours enrolled during the first semester, days since admission, days since orientation, transferred hours, and the ordinal variable of high school percentile. The homogeneity of variance assumption was met for SAT score (*Levene's* $F = 4.45$, $p = .04$), age (*Levene's* $F = .00$, $p = .99$), hours enrolled during the first semester (*Levene's* $F = .44$, $p = .51$), and days since admission (*Levene's* $F = .28$, $p = .60$). The homogeneity of variance assumption was not met for days since orientation (*Levene's* $F = 22.82$, $p < .01$) and transferred hours (*Levene's* $F = 51.13$, $p < .01$).

A series of t-test for Independent Samples showed that retained students had higher SAT scores, more enrolled hours during the first semester, more days between admission and the start of the semester, more days between orientation and the first day of class, and a greater number of transferred hours than did the students who were not retained; all group differences were statistically significant at the .01 level. The age differences were not statistically significant. Mean difference effect sizes were computed to examine the practical significance of the findings. High school percentile rankings were treated as ordinal data and the Mann-Whitney-Wilcoxon U

Test showed that group differences were statistically significant, favoring the retained group, and the effect size was small (*r* = .15). Results are summarized in Table 5.

Table 5

Group Comparisons by Retention Status after First Year, Continuous Variables

	Retention Status					
	Retained		Not Retained			
	M	SD	M	SD	t[a]	ES
SAT Score	978.06	141.99	948.01	136.80	6.77*	.21
Age	18.68	.63	18.68	.64	.07	<.01
First-Semester Hours	13.84	1.24	13.66	1.25	4.64*	.14
Days since Admission	183.59	77.77	163.07	77.94	8.36*	.26
Days since Orientation	39.52	24.59	33.28	23.11	8.35*	.27
Transfer Hours	4.41	6.53	3.26	5.77	6.02*	.19
High School Percentile[b]	.67	NA	.61	NA	9.43*	.15

ES = Effect Size, .20 = small, .50 = medium, >.80 = large
[a] Welch approximate t when the homogeneity of variances assumption was not met
[b] Percentile scores were treated as ordinal. Means are reported for the ease of interpretation. The test statistic is z and the effect size is r (.10 = small, .30 = medium, .50 = large)
* $p < .01$

<u>Probation Status</u>

Students on probation after first semester were compared to students not on probation on the basis of learning community, developmental status, gender, first-generation status, ethnicity, Pell Grant eligibility, and admission status. With the exception of learning community, $\chi^2(5, N = 4215) = 4.73$, $p = .45$, all group differences were statistically significant (Table 6).

Table 6

Group Comparisons by Probation Status after First Semester, Categorical Variables

	Probation		
	Yes F/%	No F/%	χ^2
Learning Community			
Sociology	138/26.80	376/73.20	
History	398/26.70	1095/73.30	
Political Science	330/29.20	801/70.80	
Science	215/28.10	550/71.90	
Developmental History	53/31.50	115/68.50	
Other	34/23.60	110/76.40	4.73
Developmental Status			
College-Ready	799/25.30	2353/74.70	
Not College-Ready	369/34.70	694/65.30	34.70*
Gender			
Male	559/32.00	1186/68.00	
Female	609/24.70	1861/75.30	27.79*
First-Generation Status			
Non-First-Generation	753/26.00	2143/74.00	
First-Generation	415/31.50	904/68.50	13.50*
Ethnicity			
Hispanic	608/31.70	1311/68.30	
Non-Hispanic	560/24.40	1736/75.60	27.75*
Pell Grant Eligibility			
Eligible	690/33.00	1404/67.00	
Ineligible	478/22.50	1643/77.50	57.06*
Admission Status			
Accepted	352/18.20	1582/81.80	
Alternative Admit	816/35.80	1465/64.20	161.35*

* $p < .01$

The odds ratio for developmental status showed that students who were not college-ready in mathematics and reading were 1.57 times more likely to be on probation after the first semester as were students who were college-ready. Males were 1.44 times as likely to be on probation as were females, non-first-generation students were l.31 times as likely to be on probation as were first-generation students, and Hispanics were 1.44 times as likely to be on probation as were non-Hispanics. Students who were eligible for Pell Grants were 1.69 times more likely to land on probation as were students who were not eligible for Pell Grants. Finally, students who were alternatively admitted were 2.50 times more likely to be on probation as were students who were accepted via standard admission criteria.

Students who were on probation after their first semester were compared to students who were not on probation after their first semester on the basis SAT score, age, hours enrolled during the first semester, days since admission, days since orientation, and transferred hours. The homogeneity of variance assumption was met age (*Levene's* $F = 6.09$, $p = .014$), hours enrolled during the first semester (*Levene's* $F = 1.33$, $p = .25$), and days since admission (*Levene's* $F = 1.90$, $p = .17$). The homogeneity of variance assumption was not met for SAT score (*Levene's* $F = 20.84$, $p < .01$), days since orientation (*Levene's* $F = 16.18$, $p < .01$), and transferred hours (*Levene's* $F = 158.38$, $p < .01$).

A series of t-test for Independent Samples showed that students on probation had lower SAT scores, fewer enrolled hours during the first semester, fewer days between admission and the start of the semester, fewer days between orientation and the first day of class, and a smaller number of transferred hours than did students who were not on probation; all group differences were statistically significant at the .01 level. Age differences were not statistically significant. Mean difference effect sizes were computed to examine the practical significance of the findings.

High school percentile rankings were treated as ordinal data and the Mann-Whitney-Wilcoxon U Test showed that group differences were statistically significant, favoring the non-probation group, and the effect size was small (r = .26). Results are summarized in Table 7.

Table 7

Group Comparisons by Probation Status after First Semester, Continuous Variables

	Probation Status					
	Yes		No			
	M	SD	M	SD	t[a]	ES
SAT Score	933.53	129.99	978.77	142.67	9.78*	.41
Age	18.70	.68	18.67	.61	1.09	.03
First-Semester Hours	13.62	1.23	13.82	1.25	4.58*	.14
Days since Admission	158.16	76.12	182.17	78.35	8.97*	.28
Days since Orientation	30.84	22.73	39.45	24.34	10.79*	.45
Transfer Hours	2.55	5.20	4.49	6.55	10.06*	.39
High School Percentile[b]	.57	NA	.68	NA	16.06*	.26

ES = Effect Size, .20 = small, .50 = medium, >.80 = large
[a] Welch approximate t when the homogeneity of variances assumption was not met
[b] Percentile scores were treated as ordinal. Means are reported for the ease of interpretation. The test statistic is z and the effect size is r (.10 = small, .30 = medium, .50 = large)
* $p < .01$

Multivariate Analyses

Table 8 shows the correlation matrix for the study's 13 pre-college variables. Out of the resulting 78 correlations, 62 were statistically significant at the .01 level. The noteworthy correlations were between SAT score and admission status ($r = .62$, $n = 4163$, $p < .01$), high school percentile and admission status ($r = .54$, $n = 3905$, $p < .01$), days since admission and days since orientation ($r = .48$, $n = 4215$, $p < .01$), and SAT score and the subjects' developmental status ($r = -.51$, $n = 4163$, $p < .01$). The remaining 58 statistically significant correlations indicated negligible to moderate relationships. For example, the relationship between Pell Grant eligibility and transferred hours was found to be statistically significant, but the magnitude of the relationship ($r = -.04$, $n = 4215$, $p < .01$) indicated a practically insignificant correlation. For the purpose of regression analysis, the assumption of no multicollinearity was met.

Table 9 contains the correlations between the 13 independent variables with the dependent variables of one-year retention and probation status after the first semester. Nearly all of the relationships were statistically significant at the .01 level, but the degree of the correlations ranged from negligible to weak for the majority of the variables. The strongest relationships were between high school percentile and probation status ($r = -.25$, $n = 3905$, $p < .01$) and admission status and probation status ($\varphi = -.20$, $n = 4215$, $p < .01$). A statistically significant relationship was not found between age and retention ($r = < .01$, $n = 4215$, $p = .94$), age and probation ($r = .02$, $n = 4215$, $p = .28$), and first-generation status and retention ($\varphi = -.02$, $n = 4215$, $p = .15$). Thus, age was not related to either retention or probation status. Relatively speaking, high school percentile showed the strongest relationship with both outcome variables.

Table 8

Correlation Matrix for Independent Variables (N = 4215)

	2	3	4	5	6	7	8	9	10	11	12	13
1	-.07*	.04	-.02	.18*	.01	<.01	-.06*	-.10*	-.05*	.06*	.11*	.09*
2		-.21*	-.19*	-.51*	.11*	.07*	.07*	.11*	.12*	-.17*	-.38*	-.24*
3			.30*	.20*	-.07*	.17*	.09*	.10*	.03	.27*	.54*	.25*
4				.21*	.10	.09*	<.01	.01	-.04*	.14*	.23*	.21*
5					-.09*	-.13*	-.13*	-.25*	-.21*	.17*	.62*	.25*
6						-.10*	-.04	-.07*	-.03	-.24*	-.01	-.17*
7							.05*	.05*	.10*	.12*	-.02	.05*
8								.21*	.29*	.05*	-.06*	-.01
9									.24*	.01	-.14*	-.03
10										.01	-.15*	-.07*
11											.17*	.48*
12												.25*

1 = First-Semester Hours, 2 = Developmental Status, 3 = High School Percentile, 4 = Transferred Hours, 5 = SAT Score, 6 = Age, 7 = Gender, 8 = First-Generation Status, 9 = Ethnicity, 10 = Pell Grant Eligibility, 11 = Days since Admission, 12 = Admission Status, 13 = Days since Orientation

* $p < .01$

Table 9

Correlations of the Independent Variables with Retention and Probation Status (N = 4215)

	Retention	Probation Status
First-Semester Hours[a]	.07*	-.07*
Developmental Status[b]	-.06*	.09*
High School Percentile[a]	.14*	-.25*
Transferred Hours[a]	.09*	-.14*
SAT Score[a]	.10*	-.14*
Age[a]	<.01	.02
Gender[b]	.04*	-.08*
First-Generation Status[b]	-.02	.06*
Ethnicity[b]	-.06*	.08*
Pell Grant Eligibility[b]	-.11*	.12*
Days since Admission[a]	.13*	-.14*
Admission Status[b]	.12*	-.20*
Days since Orientation[a]	.13*	-.16*

* $p < .01$
[a] Pearson Product Moment Correlation Coefficient
[b] Phi Coefficient

Two logistic regression models were developed to identify the best predictors of retention and probation status regardless of learning community membership. The dependent variable for the first model was retention (1 = retained, 0 = not retained) into the second academic year. Probation status (1 = on probation, 0 = not on probation) following the first semester of college served as the outcome measure for the second model. Models 1 and 2 were then repeated for each learning community in order to develop prediction models for students in each of the six learning community categories.

Model 1: Predicting Retention Independent of Learning Community

Out of the 13 independent variables, five (high school percentile, SAT score, Pell Grant eligibility, days since admission, and days since orientation) met the criteria to be included in the model to predict retention. The model was statistically significant, $\chi^2(5) = 215.44$, $p < .01$, correctly classified 62.20% of the students, and accounted for 7.30% of the variance in retention. The goodness-of-fit test was not statistically significant, $\chi^2(8) = 9.69$, $p = .29$, indicating that the model fit the data. Inspection of the odds ratios revealed that retention was likely for students with higher high school percentiles, higher SAT scores, more days since admission, and more days since orientation. Students who were eligible for Pell Grants were less likely to be retained. Results are summarized in Table 10.

Table 10

Logistic Regression Model for Retention Independent of Learning Community (N = 3886)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	.96	.17	30.56*	2.62
SAT Score	.01	<.01	8.99*	1.00
Pell Grant Eligibility	-.34	.01	23.74*	0.71
Days since Admission	.01	<.01	28.08*	1.00
Days since Orientation	.01	<.01	15.95*	1.00
CONSTANT	-1.51	.26	33.13*	0.22

* $p < .01$

Model 2: Predicting Probation Status Independent of Learning Community

As shown in Table 10, eight variables were included in the final model to predict probation status. The model was statistically significant, $\chi^2(8) = 425.88$, $p < .01$, correctly classified 72.10% of the students, and accounted for 14.90% of the variance in retention. The goodness-of-fit test was not statistically significant, $\chi^2(8) = 13.83$, $p = .09$, indicating that the model fit the data. The odds ratios showed that probation was likely for students who were Hispanic and eligible for Pell Grants. Probation was less likely for students with higher high school percentiles and females, as well as for students with more transferred hours, higher SAT scores, more days since admission, and more days since orientation. Results are summarized in Table 11.

Table 11

Logistic Regression Model for Probation Independent of Learning Community (N = 3886)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	-2.06	.20	106.08*	0.13
Transferred Hours	-.03	.01	18.46*	0.97
SAT Score	-.01	<.01	12.38*	0.99
Gender	-.24	.08	9.47*	0.79
Ethnicity	.24	.08	9.02*	1.27
Pell Grant Eligibility	.34	.08	18.97*	1.41
Days since Admission	-.01	<.01	13.65*	0.99
Days since Orientation	-.01	<.01	19.01*	0.99
CONSTANT	2.01	.32	40.13*	7.48

* $p < .01$

Learning Community Membership

Descriptive statistics were obtained for each of the 13 predictor variables, sorted by the learning community, to explore the role that learning community membership might have played in the retention and probation status of students. Group differences on the basis of means (for continuous variables) and frequencies (for categorical variables) were examined, using One-Way ANOVA and Chi-Square Test of Independence, respectively. Statistically significant group differences were found for all 13 predictor variables. For example, the average number of hours taken during the first semester by Science learning community students was 14.04 (SD = 1.46), which was significantly higher than the average number of first-semester hours for students in all other learning communities except those placed in the "Other" category, *Welch's F*(5, 793.16) = 18.55, $p < .01$. Students in the "Other" category of learning communities came in with the highest SAT scores, and group differences were statistically significant when compared to every other learning community category excluding the Science learning community, *Welch's F*(5, 746.71) = 102.23, $p < .01$.

Another example of notable group differences was developmental status, which ranged from 9.00% of the students in the Science learning community to 99.00% of the Developmental History learning community students; group differences were statistically significant, $\chi^2(5, N = 4215) = 607.32$, $p < .01$. Group difference based on admission status were also statistically significant, $\chi^2(5, N = 4215) = 223.42$, $p < .01$; while 65% of the students in the "Other" category of learning communities had been accepted based on standard admission criteria, 95% of the students in the Developmental History learning community had been alternatively admitted. Results are summarized in Table 12.

Table 12

Independent Variable Means by Learning Community (N = 4215)

	SOCI	HIST	POLS	SCI	DHIST	OTHER
1[a]	13.50	13.78	13.71	14.04	13.39	13.97
2[a]	.63	.66	.64	.69	.54	.60
3[a]	4.05	3.56	4.07	5.48	0.73	2.51
4[a]	953.46	965.89	947.84	1021.66	827.52	1027.10
5[a]	18.66	18.67	18.73	18.60	18.71	18.87
6[a]	174.21	174.70	167.05	197.13	166.60	150.71
7[a]	31.61	37.42	33.24	47.07	33.32	33.93
8[b]	.29	.23	.27	.09	.99	.17
9[b]	.56	.59	.59	.63	.70	.24
10[b]	.34	.32	.33	.27	.41	.18
11[b]	.47	.46	.48	.42	.54	.29
12[b]	.55	.50	.51	.44	.61	.31
13[b]	.40	.46	.42	.62	.05	.65

SOCI = Sociology, HIST = History, POLS = Political Science, SCI = Science, DHIST = Developmental History, OTHER = All Other Learning Communities

1 = First-Semester Hours, 2 = High School Percentile, 3 = Transferred Hours, 4 = SAT Score, 5 = Age, 6 = Days since Admission, 7 = Days since Orientation, 8 = Developmental Status (1 = Not College-Ready, 0 = College-Ready), 9 = Gender (1 = Female, 0 = Male), 10 = First-Generation Status (1 = First-Generation, 0 = Non-First-Generation), 11 = Ethnicity (1 = Hispanic, 0 = Non-Hispanic), 12 = Pell Grant Eligibility (1 = Eligible, 0 = Ineligible), 13 = Admission Status (1 = Accepted, 0 = Alternatively Admitted)

[a] One-Way ANOVA indicated significant differences between group means

[b] Chi-square Test of Independence indicated significant differences between group frequencies

Means for categorical data are reported for ease of interpretation.

Predicting Retention for each Learning Community

Sociology

Out of the 13 independent variables, only one (days since orientation) met the criteria to be included in the model to predict retention for students who had enrolled in a Sociology learning community. The model was statistically significant, $\chi^2(1) = 15.24$, $p < .01$, correctly classified 60.10% of the students, and accounted for 4.20% of the variance in retention. The goodness-of-fit test, $\chi^2(6) = 4.66$, $p = .59$, showed that the model fit the data. The odds ratios suggested that retention was likely for students with more days since orientation. Results are summarized in Table 13.

Table 13

Logistic Regression Model for Retention, Sociology Learning Community (N = 481)

Predictor	B	SE	Wald	Odds Ratio
Days since Orientation	.02	<.01	14.63*	1.02
CONSTANT	-.29	.16	3.18	0.75

* $p < .01$

History

High school percentile, days since orientation, Pell Grant eligibility, and days since admission formed the model to predict retention for students who had enrolled in a History learning community. The model was statistically significant, $\chi^2(4) = 85.11$, $p < .01$, correctly classified 62.60% of the students, and accounted for 8.00% of the variation in the outcome measure. The goodness-of-fit test, $\chi^2(8) = 4.74$, $p = .79$, showed that the model fit the data. The odds ratios showed that retention was likely for students with higher high school percentiles, more days since admission, and more transferred hours. Students who were eligible for Pell Grants were less likely to be retained. Results are summarized in Table 14.

Table 14

Logistic Regression Model for Retention, History Learning Community (N = 1398)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	1.35	.31	19.51*	3.84
Transferred Hours	.03	.01	7.53*	1.03
Pell Grant Eligibility	-.30	.11	6.95*	0.74
Days since Admission	.01	<.01	22.70*	1.00
CONSTANT	-1.07	.22	23.17*	0.34

* $p < .01$

Political Science

There were three variables (high school percentile, days since admission, and Pell Grant eligibility) which defined the model to predict retention for students who had enrolled in a Political Science learning community. The model was statistically significant, $\chi^2(3) = 46.73$, $p < .01$, correctly classified 61.00% of the students, and accounted for 6.10% of the variance in retention. The goodness-of-fit test showed that the model fit the data, $\chi^2(8) = 2.27$, $p = .97$. An examination of the odds ratios showed that retention was likely for students with higher high school percentiles and more days between admission and the start of the semester. Students who were eligible for Pell Grants were less likely to be retained. Results are summarized in Table 15.

Table 15

Logistic Regression Model for Retention, Political Science Learning Community (N = 1017)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	.88	.33	7.36*	2.42
Days since Admission	.01	<.01	13.97*	1.00
Pell Grant Eligibility	-.56	.13	17.83*	0.57
CONSTANT	-.46	.24	3.71	0.63

* $p < .01$

Science

Only two independent variables, high school percentile and days since orientation, met the inclusion criteria and formed the model to predict retention for students who had enrolled in a Science learning community during their first semester. The model was statistically significant, $\chi^2(2) = 54.13$, $p < .01$, correctly classified 65.40% of the students, and accounted for 9.80% of the variance in retention. The goodness-of-fit test, $\chi^2(8) = 3.07$, $p = .93$, showed that the model fit the data. On the basis of the odds ratios, it was concluded that retention was likely for students with higher high school percentiles and more days between orientation and the start of the semester. Results are summarized in Table 16.

Table 16

Logistic Regression Model for Retention, Science Learning Community (N = 728)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	1.57	.41	14.77*	4.79
Days since Orientation	.02	<.01	29.09*	1.02
CONSTANT	-1.44	.31	21.01*	0.24

* $p < .01$

Developmental History

None of the 13 independent variables met the criteria to be included in the model to predict retention for students who had enrolled in the Developmental History learning community.

Other

Two of the 13 independent variables, days since admission and SAT score, met the inclusion criteria and were useful in predicting retention for students who had registered in the "Other" category of learning communities during their first semester of college. The model was statistically significant, $\chi^2(2) = 15.74$, $p < .01$, correctly classified 67.00% of the students, and accounted for 18.50% of the variance in retention. The goodness-of-fit test showed that the model fit the data, $\chi^2(8) = 16.83$, $p = .03$. The odds ratios showed that retention was likely for students who were admitted earlier and had higher SAT scores. Results are summarized in Table 17.

Table 17

Logistic Regression Model for Retention, Other Learning Communities (N = 109)

Predictor	B	SE	Wald	Odds Ratio
SAT Score	.01	<.01	6.77*	1.01
Days since Admission	.01	<.01	6.39	1.01
CONSTANT	-6.38	2.22	8.28*	< .01

* $p < .01$

<u>Predicting Probation Status for each Learning Community</u>

<u>Sociology</u>

Out of the 13 independent variables, only one (days since orientation) met the criteria to be included in the model to predict probation status for students who had enrolled in a Sociology learning community. The model was statistically significant, $\chi^2(1) = 22.29$, $p < .01$, correctly classified 72.60% of the students, and accounted for 6.60% of the variance in probation status. The goodness-of-fit test, $\chi^2(6) = 3.98$, $p = .68$, indicated that the model fit the data. Inspection of the odds ratios showed that probation was less likely for students with more days since orientation. Results are summarized in Table 18.

Table 18

Logistic Regression Model for Probation Status, Sociology Learning Community (N = 481)

Predictor	B	SE	Wald	Odds Ratio
Days since Orientation	-.02	.01	20.07*	0.98
CONSTANT	-.31	.17	3.21	0.74

* $p < .01$

History

High school percentile, transferred hours, Pell Grant eligibility, and days since orientation formed the model to predict probation status for students who had enrolled in a History learning community. The model was statistically significant, $\chi^2(4) = 170.46$, $p < .01$, correctly classified 73.70% of the students, and accounted for 16.60% of the variance in probation status. The goodness-of-fit test, $\chi^2(8) = 4.09$, $p = .85$, showed that the model fit the data. The odds ratios showed that probation was likely for students who were eligible for Pell Grants. Students with higher high school percentile scores, more transferred hours, and more days since orientation were less likely to land on probation. Results are summarized in Table 19.

Table 19

Logistic Regression Model for Probation Status, History Learning Community (N = 1398)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	-2.67	.34	61.74*	0.07
Transferred Hours	-.05	.01	13.68*	0.95
Pell Grant Eligibility	.42	.13	10.16*	1.52
Days since Orientation	-.01	<.01	23.29*	0.99
CONSTANT	1.13	.23	24.95*	3.09

* $p < .01$

Political Science

There were three variables (high school percentile, SAT score, and days since orientation) which defined the model to predict probation status for students who had enrolled in a Political Science learning community during their first semester. The model was statistically significant, $\chi^2(3) = 106.08$, $p < .01$, correctly classified 70.50% of the students, and accounted for

14.00% of the variance in probation status. The goodness-of-fit test, $\chi^2(8) = 13.35$, $p = .10$, showed that the model fit the data. On the basis of the odds ratios, it was concluded that probation was less likely for students with higher high school percentiles, higher SAT scores, and more days since orientation. Results are summarized in Table 20.

Table 20

Logistic Regression Model for Probation, Political Science Learning Community (N = 1017)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	-2.10	.36	34.31*	0.12
SAT Score	-.01	<.01	20.40*	0.99
Days since Orientation	-.01	<.01	16.68	0.99
CONSTANT	3.37	.57	35.48*	28.99

* $p < .01$

Science

Four variables (high school percentile, transferred hours, SAT score, and days since orientation) formed the model to predict probation status for students who had enrolled in a Science learning community. The prediction model was statistically significant, $\chi^2(4) = 112.77$, $p < .01$, correctly classified 72.40% of the students, and accounted for 20.50% of the variance in probation status. The goodness-of-fit test, $\chi^2(8) = 13.15$, $p = .11$, indicated that the model fit the data. The odds ratios showed that probation was less likely for students with higher high school percentiles, more transferred hours, higher SAT scores, and more days since orientation. Results are summarized in Table 21.

Table 21

Logistic Regression Model for Probation Status, Science Learning Community (N = 728)

Predictor	B	SE	Wald	Odds Ratio
High School Percentile	-2.94	.47	38.82*	0.05
Transferred Hours	-.04	.02	7.80*	0.96
SAT Score	-.01	<.01	10.31*	0.99
Days since Orientation	-.01	<.01	10.22*	0.99
CONSTANT	4.18	.74	31.83*	65.18

* $p < .01$

Developmental History

None of the original 13 independent variables met the criteria to be included in the model to predict probation status for students who enrolled in the Developmental History learning community.

Other

SAT score was the only variable which met the inclusion criteria to form the model to predict probation status for students who registered in the “Other” category of learning communities. The model was statistically significant, $\chi^2(1) = 8.23$, $p < .01$, correctly classified 70.60% of the students, and accounted for 10.60% of the variance in probation status. The goodness-of-fit test, $\chi^2(8) = 13.78$, $p = .09$, indicated that the model fit the data. Inspection of the odds ratios showed that probation was less likely for students with higher SAT scores. Results are summarized in Table 22.

Table 22

Logistic Regression Model for Probation Status, Other Learning Communities (N = 109)

Predictor	B	SE	Wald	Odds Ratio
SAT Score	-.01	<.01	7.05*	0.99
CONSTANT	5.13	2.29	5.01	169.15

* $p < .01$

Summary

The study explored the role of 13 independent variables in predicting the one-year retention and first-semester probation status of first-year students in learning communities. Univariate analyses were performed to explore the individual relationships between each of the predictors and the two outcomes. Multivariate analyses resulted in several binary logistic regression models. The first two models explored the significance of the 13 independent variables in predicting retention and probation status, independent of learning community membership. The models were:

Retention = -1.51 + .96(High School Percentile) + .01(SAT Score) - .34(Pell Grant Eligibility) + .01 (Days since Admission) + .01(Days since Orientation)

Probation Status = 2.01 – 2.06(High School Percentile) - .03(Transferred Hours) - .01(SAT Score) - .24(Gender) + .24(Ethnicity) + .34(Pell Grant Eligibility) - .01(Days since Admission) - .01(Days since Orientation)

Additionally, prediction models were formulated for the outcome variables in each of the learning communities, as follows:

Sociology

Retention = -.29 + .02(Days since Orientation)

Probation Status = -.31 - .02(Days since Orientation)

History

Retention = -1.07 + 1.35(High School Percentile) + .03(Transferred Hours) - .30(Pell Grant Eligibility) + .01(Days since Admission)

Probation Status = 1.13 - 2.67(High School Percentile) - .05(Transferred Hours) + .42(Pell Grant Eligibility) - .01(Days since Orientation)

Political Science

Retention = -.46 + .88(High School Percentile) + .01 (Days since Admission) - .56(Pell Grant Eligibility)

Probation Status = 3.37 – 2.10(High School Percentile) - .01(SAT Score) - .01(Days since Orientation)

Science

Retention = -1.44 + 1.57(High School Percentile) + .02(Days since Orientation)

Probation Status = 4.18 – 2.94(High School Percentile) -.04(Transferred Hours) - .01(SAT Score) - .01(Days since Orientation)

Other

Retention = -6.38 + .01(SAT Score) + .01(Days since Admission)

Probation Status = 5.13 - .01(SAT Score)

Table 23 shows a summary of the statistically significant predictors of retention in various models.

Table 23

Predictors of Retention in Various Logistic Regression Models

Predictor	M1	M2	M3	M4	M5	M6	M7
First-Semester Hours							
Developmental Status							
High School Percentile	✓		✓	✓	✓		
Transferred Hours			✓				
SAT Score	✓						✓
Age							
Gender							
First-Generation Status							
Ethnicity							
Pell Grant Eligibility	✓		✓	✓			
Days since Admission	✓		✓	✓			✓
Admission Status							
Days since Orientation	✓	✓			✓		

M1 = Model for predicting retention independent of learning community, M2 = Sociology Learning Community, M3 = History Learning Community, M4 = Political Science Learning Community, M5 = Science Learning Community, M6 = Developmental History Learning Community, M7 = Other Learning Communities

Table 24 shows a summary of the statistically significant predictors of probation status in various models.

Table 24

Predictors of Probation Status in Various Logistic Regression Models

Predictor	M1	M2	M3	M4	M5	M6	M7
First-Semester Hours							
Developmental Status							
High School Percentile	✓		✓	✓	✓		
Transferred Hours	✓		✓		✓		
SAT Score	✓			✓	✓		✓
Age							
Gender	✓						
First-Generation Status							
Ethnicity	✓						
Pell Grant Eligibility	✓		✓				
Days since Admission	✓						
Admission Status							
Days since Orientation	✓	✓	✓	✓	✓		

M1 = Model for predicting probation status independent of learning community, M2 = Sociology Learning Community, M3 = History Learning Community, M4 = Political Science Learning Community, M5 = Science Learning Community, M6 = Developmental History Learning Community, M7 = Other Learning Communities

CHAPTER V

SUMMARY OF RESULTS AND CONCLUSIONS, DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH

Summary

It is an unfortunate reality that students often struggle in their first year of college. Many are in academic trouble after only one semester of coursework and a large portion do not return after the first year. Although programs have been developed at institutions across the nation to assist first-year students in making a successful transition from high school to higher education, the problem persists. Learning communities have been identified as a high-impact practice to engage students early in their academic careers. In addition, many colleges have used institutional data to predict which students have the most risk of being on probation or not being retained based on past student performance. Yet, little research exists on the prediction of first-year student success for students in learning communities using variables that are known on the first day of classes.

The study was designed to identify the pre-college variables that are useful predictors of one-year retention and probation status after the first semester for learning community students. Archival data available through the registrar were collected on three cohorts (Fall 2010, Fall 2011, and Fall 2012) of first-year students at a public South Texas university. The study was guided by the following research questions:

1. What pre-college variables are predictors of the retention of first-year students in learning communities?
2. What pre-college variables are predictors of the probation status of first-year students in learning communities?

The correlational study was conducted between Fall 2013 and Spring 2014. The data were collected from the registrar in Fall 2013, after retention information was determined for the Fall 2012 cohort of students, while the preparation and analysis of the data were conducted in Spring 2014. The initial database of student records contained 4,673 entries that was whittled down to 4,215 students based on the study's definition of traditional first-year students. The data were analyzed using the SPSS. Univariate and multivariate analyses were employed to determine the relationship between the 13 pre-college variables and each of the two outcome measures, namely, retention and probation status. Binary logistic regression models were developed to predict retention and probation status for students in each learning community category, as well as for all students regardless of learning community membership.

Summary of Results and Conclusions

Group differences on the basis of retention into the second fall semester were statistically significant for all variables except learning community, admission status, and age. Group differences on the basis of probation status after the first semester were statistically significant for all variables excluding learning community and age. Participation in a particular learning community, therefore, did not statistically indicate that students would be more or less likely to be retained or on probation.

When data from all 4,215 first-year students were included in formulation of the prediction models to address the two research questions, five variables were identified as predictors of retention (high school percentile, SAT score, Pell Grant eligibility, days since admission, and days since orientation). Together they accounted for 7.30% of the variance in retention. Three additional variables (transferred hours, gender, and ethnicity) met the criteria to be included in the model to predict probation status and accounted for 14.90% of the variance in probation status. It is concluded that these eight pre-college variables are useful in predicting retention and probation status for first-year students in the learning communities program at the university.

The binary logistic regression models included different sets of predictor variables when the data were analyzed by learning community (see Tables 23 and 24). High school percentile and the number of days since admission were included in four of the seven models to predict retention, and Pell Grant eligibility and days since orientation were each included in three of the retention models. In terms of probation status, the number of days since orientation was included in five of the seven models. High school percentile and SAT score were each included in four models, while transferred hours met the criteria to be included in three of the predictive models. First semester hours, developmental status, age, first-generation status, and admission status did not make it into any of the models to predict retention or probation status. Although this does not mean that these variables are not important, it does imply that perhaps, due to multicollinearity, these variables shared variance with one or more of the other pre-college variables used in the study.

High school percentile and the number of days since orientation were the most common pre-college variables used in the prediction of retention and probation status within individual learning communities. Both variables met the criteria to be included in eight of the study's 14 models. The SAT score also appeared in six of the 14 models, while Pell Grant eligibility and the number of days since admission appeared in five. These results are consistent with previous research on the relationship between high school and college academic performance (Astin, 1971; Goldman & Widawski, 1976; Maggard, 2007; Noble & Sawyer, 2004; Stiggins, Frisbie, & Griswald, 1989; Zheng, Saunders, Shelley, & Whalen, 2002), as well as the link between standardized test scores and first-semester GPA (Aleamoni & Oboler, 1977). The findings also support past research regarding the impact of financial aid status on first-year success (Coperthwaite & Klimczak, 2008). Additionally, the results align with the self-study conducted by Indiana University concerning the significance of time factors such as admission and orientation dates (Drake, 2011b).

Discussion

The purpose of the study was to identify pre-college variables that could serve as predictors of retention and probation status of first-year students in learning communities. The results indicated that several of the 13 variables used in the study were useful in predicting the retention and probation status of first-year students, but also that the predictor variables changed based on the learning community under scrutiny. Additionally, although the students in each learning community were markedly different from one another when the groups were compared on each of the 13 pre-college variables (see Table 12), there was not a statistically significant difference in retention or probation status between any of the learning communities. This seems to indicate that some factor within the learning community experience or program – or some inexplicable outside factor – somehow mitigated these incoming differences so students landed on probation and were retained at similar rates across the learning community categories.

Comparing the prediction models for each learning community revealed some notable patterns. The number of days since orientation was the only variable included in the models to predict retention and probation status for the Sociology learning community. The Sociology learning community is often the last learning community to fill during summer orientations. It is hypothesized that this is due to student’s unfamiliarity with the subject matter, as well as the fact that students can choose to take either sociology or psychology to meet core curriculum requirements. Thus, the learning community is frequently comprised of a substantial number of students from the final few orientations. This finding has significant implications for both the Sociology learning community and orientation schedule.

The History and Political Science learning communities differed on only one predictor (transferred hours) when retention was the outcome measure, but only shared two predictors in common (high school percentile and days since orientation) when probation status was the dependent variable. History and Political Science learning communities typically consist of a mixture of students of all majors and the courses are taught with the understanding that students are not history or political science majors. The number of pre-college transferred hours – or college-level courses completed prior to enrolling at the university – was statistically significant in the prediction of the retention of students in the History learning community but not the Political Science learning community. Transferred hours were also included in the model to predict probation status of students in the History learning community, indicating that pre-college coursework was linked to success for these students. The History learning community was the only learning community in which Pell Grant eligibility was a predictor of probation status.

The predictor variables for students in the Science learning community for both retention and probation status were all related to previous academic performance (high school percentile, transferred hours, SAT score) or time factors (days since orientation). Most of the Science learning communities are tetrads, meaning that the students enroll in two challenging science courses, biology and chemistry, which have lab components, in addition to first-year seminar and first-year composition. These students not only enrolled in 14.04 hours on average during their first semester, more than any other learning community group, but they also entered the program with the highest mean high school percentile (.69). Although it might be hypothesized that this learning community would have fewer students on probation and would retain students at higher rates, the results were not significantly different for either outcome for the Science learning community students. It could be that the number of hours in challenging courses served as a counterweight to the higher incoming pre-college measures.

None of the study's 13 variables were useful in predicting retention or probation status for students in the Developmental History learning community. This finding, although initially perplexing, was unsurprising when the pre-college variables were analyzed for this group of students. All students in the learning community were in developmental mathematics, which contributed to the homogeneity of the group on the majority of the 13 pre-college variables. Their similarity made it difficult to identify even one variable that was useful in predicting either retention or probation status. One significant finding about the Developmental History learning community was that it did not have a significantly different retention rate or higher number of students on probation than any of the other learning communities despite the fact that its students arrived with the lowest high school percentiles and SAT scores, and the highest percentage of first-generation, Pell Grant eligible, and alternatively admitted students.

Finally, the SAT score was the only variable included in the model to predict probation status for students in the "Other" category of learning communities, while the SAT score and the number of days since admission were found to be predictive of retention. The learning communities in the "Other" category are major-specific and are reserved for engineering, geology, and environmental science majors. The students in these learning communities had the highest SAT scores when compared to any other learning community category. The majority of students in these learning communities were academically successful in the past, making for a rather homogenous and competent cohort. It is worthy to note that the SAT score met the criteria to be included in the model while high school percentile did not. Like the Science learning community, this group had similar outcomes in terms of retention and probation status to all of the other learning communities, despite its higher scores on each of the 13 pre-college variables.

Binary logistic regression models use existing data about past events to create equations to predict future outcomes. Thus, the study's models can be used to determine the probability and odds of landing on probation or being retained for new incoming students with similar characteristics based on their pre-college variables. For example, let us take a typical profile for a hypothetical incoming student named Jane Doe. Jane is college-ready in reading and mathematics, not first-generation, female, 18.68 years old, non-Hispanic, and ineligible for Pell Grants. She was ranked in the 68th percentile of her high school class, earned the SAT score of 960, is enrolled in 13 hours during her first semester, and has 3 transferred hours. She was alternatively admitted 176 days before the first day of class and attended the orientation 37 days before the start of the semester. Table 25 includes the calculations for determining the probability and odds of Jane being retained based on her pre-college variables. Table 26 details the probability and odds of Jane being on probation after her first semester using various models.

Table 25

Prediction of Retention using Binary Logistic Regression Models

Model 1 (Independent of Learning Community Membership)
Retention = -1.51 + .96(High School Percentile) + .01(SAT Score) - .34(Pell Grant Eligibility) + .01 (Days since Admission) + .01(Days since Orientation) Retention = -1.51 + .96(.68) + .01(960) - .34(0) + .01(176) + .01(37) = 10.87 p(Retention) = 1 / (1 + $e^{-10.87}$) = **.99** odds(Retention) = .99/(1-.99) = **99.00 (in favor of retention)**

Sociology Learning Community

$$\text{Retention} = -.29 + .02(\text{Days since Orientation})$$
$$\text{Retention} = -.29 + .02(37) = .45$$

p(Retention) = $1 / (1 + e^{-.45}) =$ **.61**
odds(Retention) = .61/(1-.61) = **1.57 (in favor of retention)**

History Learning Community

$$\text{Retention} = -1.07 + 1.35(\text{High School Percentile}) + .03(\text{Transferred Hours}) - .30(\text{Pell Grant Eligibility}) + .01(\text{Days since Admission})$$

$$\text{Retention} = -1.07 + 1.35(.68) + .03(3) - .30(0) + .01(176) = 1.70$$

p(Retention) = $1 / (1 + e^{-1.73}) =$ **.85**
odds(Retention) = .85/(1-.85) = **5.46 (in favor of retention)**

Political Science Learning Community

$$\text{Retention} = -.46 + .88(\text{High School Percentile}) + .01(\text{Days since Admission}) - .56(\text{Pell Grant Eligibility})$$

$$\text{Retention} = -.46 + .88(.68) + .01(176) - .56(0) = 1.90$$

p(Retention) = $1 / (1 + e^{-1.90}) =$ **.87**
odds(Retention) = .87/(1-.87) = **6.68 (in favor of retention)**

Science Learning Community

$$\text{Retention} = -1.44 + 1.57(\text{High School Percentile}) + .02(\text{Days since Orientation})$$
$$\text{Retention} = -1.44 + 1.57(.68) + .02(37) = .37$$

p(Retention) = $1 / (1 + e^{-.37}) =$ **.59**
odds(Retention) = .59/(1-.59) = **1.44 (in favor of retention)**

Other Learning Communities

$$\text{Retention} = -6.38 + .01(\text{SAT Score}) + .01(\text{Days since Admission})$$
$$\text{Retention} = -6.38 + .01(960) + .01(176) = 4.98$$

p(Retention) = $1 / (1 + e^{-4.98}) =$ **.99**
odds(Retention) = .99/(1-.99) = **99.00 (in favor of retention)**

Table 26

Prediction of Probation Status using Binary Logistic Regression Models

Model 2 (Independent of Learning Community Membership)

Probation = 2.01 – 2.06(High School Percentile) - .03(Transferred Hours) - .01(SAT Score) - .24(Gender) + .24(Ethnicity) + .34(Pell Grant Eligibility) - .01(Days since Admission) - .01(Days since Orientation)

Probation = 2.01 – 2.06(.68) - .03(3) - .01(960) - .24(1) + .24(0) + .34(0) - .01(176) - .01(37) = -11.45

p(Probation) = 1 / (1 + $e^{11.45}$) = **1.06 * 10^{-5}**
odds(Probation) = 1.06 * 10^{-5}/ (1 - 1.06 * 10^{-5}) = **1.06 * 10^{-5} (not in favor of probation)**

Sociology Learning Community

Probation = -.31 - .02(Days since Orientation)

Probation = -.31 - .02(37) = -1.05

p(Probation) = 1 / (1 + $e^{1.05}$) = **.26**
odds(Probation) = .26/ (1 - .26) = **.35 (not in favor of probation)**

History Learning Community

Probation = 1.13 - 2.67(High School Percentile) - .05(Transferred Hours) + .42(Pell Grant Eligibility) - .01(Days since Orientation)

Probation = 1.13 – 2.67(.68) - .05(3) + .42(0) - .01(37) = -1.21

p(Probation) = 1 / (1 + $e^{1.21}$) = **.23**
odds(Probation) = .23/ (1 - .23) = **.30 (not in favor of probation)**

Political Science Learning Community

Probation = 3.37 – 2.10(High School Percentile) - .01(SAT Score) - .01(Days since Orientation)

Probation = 3.37 – 2.10(.68) - .01(960) - .01(37) = -8.03

p(Probation) = 1 / (1 + $e^{8.03}$) = **3.26 x 10^{-4}**
odds(Probation) = 3.26 x 10^{-4}/ (1 - 3.26 x 10^{-4}) = **3.26 x 10^{-4} (not in favor of probation)**

Science Learning Community

Probation = 4.18 – 2.94(High School Percentile) - .04(Transferred Hours) - .01(SAT Score) - .01(Days since Orientation)

Probation = 4.18 – 2.94(.68) - .04(3) - .01(960) - .01(37) = -7.91

p(Probation) = 1 / (1 + $e^{7.91}$) = **3.67 x 10^{-4}**
odds(Probation) = 3.67 x 10^{-4}/ (1 - 3.67 x 10^{-4}) = **3.67 x 10^{-4} (not in favor of probation)**

Other Learning Communities

Probation = 5.13 - .01(SAT Score)
Probation = 5.13 - .01(960) = -4.47

p(Probation) = 1 / (1 + $e^{4.47}$) = **.01**
odds(Probation) = .01/ (1 - .01) = **.01 (not in favor of probation)**

According to all of the retention models, Jane Doe would be likely to be retained and not be on probation. In fact, her odds for both measures, regardless of learning community membership, indicate a high probability that she would return for the second year of college and that she would have a 2.0 GPA or higher after the first semester. Within the learning communities, the odds of Jane being retained are the greatest in the “Other” category of learning communities, but Jane would need to be an engineering, geology, or environmental science major to register for those courses. The probabilities of Jane being retained after enrolling in the History or Political Science learning community are 85% and 87%, respectively. Jane’s odds of returning after one year are smaller in the Science and Sociology learning communities, but are still in favor of a positive result. The odds of Jane landing on probation after her first semester are small to negligible in all of the learning communities. The highest probability that Jane would land on probation is in the Sociology learning community at 26%, but the odds are nearly three to one (3:1) against that outcome.

Implications

The results of the study have implications for admissions and orientation officers, as well as student support services and learning community practitioners at the university level and scholars across the country. Perhaps the most obvious implication implied by the results of the study is that pre-college variables can be used as predictors of both retention and probation status, as suggested by Tinto's (1975; 1993) SIM which classified student characteristics into four categories that work together to explain, or predict, outcomes such as retention and probation status. Although additional surveys and more student data could undoubtedly contribute to an increased understanding of student success in the first year, it is possible to formulate prediction models based on student data that are available on the first day of classes. This implication applies to any institution interested in first-year student success.

Admission status did not meet the criteria to be included in any of the prediction models. This seems to indicate that the admissions criteria at the university level are somehow otherwise accounted for in the models because they are related to another variable, such as high school percentile or the SAT score. Since the admissions officers rely on high school performance and standardized test scores as part of the admissions process, this was not surprising. However, there was a statistically significant difference both in the retention rates and the probation status of students who were admitted by the standard criteria and those who were admitted alternatively. The results showed that the traditionally admitted students were more likely to be retained and not on probation, and that students who were alternatively admitted were more at risk of not being retained and landing on probation. Thus, the current admission criteria are performing their function. The alarming figures for admissions officers to consider, however, are the number of students who were alternatively admitted (54.10% of the 4,215 in the study, or

2,281 students) and their impact on the retention and probation rates of the first-year student population as a whole. In fact, more than any other variable, admission status had largest odds ratios when grouped by retention status and probation status, indicating that the institution might want to reconsider its policies for alternatively admitting students.

Another pre-college variable that was not included in any of the models to predict retention or probation status was first-generation status. Although first-generation students have been described as distinctively different from their non-first-generation classmates in ways that disadvantage them with respect to college knowledge, family support, and preparation (Pascarella, Pierson, Wolniak, & Terenzini, 2004; Terenzini, Springer, Yaeger, Pascarella, & Nora, 1996; Warburton, Bugarin, & Nunez, 2001), self-identifying as a first-generation student was not predictive of either outcome in the study. One explanation is that the predictors that actually made it into the models – Pell Grant eligibility or SAT score, for example – washed out and accounted for any unique effect that first-generation status might have on the outcomes. However, another interpretation of this omission is that the learning communities themselves somehow neutralize any inherent disadvantages that first-generation students bring in with them. This intriguing discovery opens up an avenue in the research about the intersection of learning communities and first-generation students that has yet to be explored.

The results of the study also have practical implications for the summer orientation coordinators, advisors, and student support services. The number of days between orientation and the start of the semester was a unique predictor of both retention and probation status in several of the models, including both of the models that were independent of learning community membership. In fact, the "Days since Orientation" variable was the sole predictor of both retention and probation status in the Sociology learning community. Essentially, the more days between orientation and the start of the semester, the greater the odds that students would be retained and not end up on probation. This finding validates the commonly-held belief of many faculty members in the learning community program that the students who attend the final few summer orientations are usually the most at risk of not succeeding in their first year. These faculty members, like Drake (2011b), have argued that orientation date can serve as a proxy for student motivation. One implication of this finding is that universities could consider modifying the orientation schedule so that the later orientation sessions include more intentional programming about academic and financial assistance available on campus, and perhaps even pilot more intrusive practices that include regular mandatory check-ins with advisors or peer mentors for students who are predicted to be on probation or not retained. Miller, Tyree, Riegler, and Herreid (2010) proposed a similar intervention at the University of South Florida based on a model they developed to predict retention and were able to report positive results. Another modification based on the significance of orientation date would be to stagger and monitor enrollments across all of the learning communities so that the sections fill evenly throughout the summer orientations, which would disperse the students from the final few orientations across the program.

There were no statistically significant differences in the retention and probation rates among the different learning communities despite their significantly different incoming student populations. This is perhaps most noteworthy when the Developmental History learning community is considered; this group contained students with the most incoming risk factors including, but not limited to, the lowest SAT scores, the greatest financial need, and the lowest high school percentiles, all of which were predictors of retention or probation status. The fact that these students were retained and kept off probation at similar rates to the rest of the first-year students in learning communities is not trivial. The learning community program should investigate the Developmental History learning community in an attempt to replicate its practices and impact student performance program-wide.

Because statistically significant differences were found in the retention and probation rates of students when divided into groups based on the pre-college variables, it would seem that information about these variables should be shared with the learning community teaching teams as soon as it is available, preferably before the start of the semester. The faculty members in each learning community usually meet at least once during the weeks leading up to the fall semester. After students are registered in the various courses, a profile of students in each learning community could be created with aggregated information about the students who would be in their linked classes. This profile could be shared at the planning meeting and would contain information such as the mean SAT score of the incoming students for the learning community, the number of first-generation students, and the number of students who were alternatively admitted. Knowing this information about their incoming students could assist the teams in the creation of targeted interventions and activities for the upcoming semester.

Learning community teaching teams could also request to have reports created for each student that include the pre-college variables, as well as the prediction models for retention and probation status (with and without respect to learning community membership). The learning community teams could elect to target specific interventions for students in a particular range of risk for being on probation or not returning the next fall. These suggestions would require the learning community leadership team to run the models with incoming student data and create the reports to distribute to each of the teaching teams. However, the ability to target particular students for interventions might outweigh the amount of effort required on the front end if successful.

Another implication of the findings for learning community leadership is that the differences among the learning community prediction models could be used for self-assessment. For example, in some learning communities, high school percentile was a predictor of retention or probation, but it was not a predictor in others. Although there were no statistically significant differences in the ultimate retention and probation rates among the different learning communities, the odds of landing on probation or being retained went up or down depending on the learning community when the different models were used with pre-college data for a typical student. These differences seem to indicate inherent differences between the learning communities themselves that beseeches further exploration by the program administrators. It seems that there might also be a way to use the prediction models to assess the effectiveness of individual learning communities, such as the Developmental History learning community, by comparing the predicted number of students on probation (or retained) at the start of the semester with the actual number of students on probation (or retained) at the end.

Ultimately, this study contributes to a niche in the literature about first-year students in learning communities that had been lacking up to this point. The body of research on learning communities argues that they contribute to first-year student success and are a high-impact practice, especially if they include first-year seminar courses, but no research had previously been published about the use of pre-college variables to predict first-year student success in various learning communities at an individual institution. The implication, therefore, for learning community scholars is the potential for this type of analysis to be conducted in programs across the country as a way to explore the relationship between the information available about students on the first day of class and first-year success rates. Resulting prediction models can be employed on campuses in order to help target interventions led by learning community teaching teams. Learning community leaders at the Washington Center could subsequently compile national data regarding the use of prediction models to share with the field; many learning community programs often have to justify their existence and campus administrators often ask for quantitative results as evidence of added value. Success rates after participating in learning communities with targeted interventions would be beneficial in both emphasizing and clarifying the role that learning communities play in solving the puzzle of student departure.

Perhaps most importantly, the results of the study invite further analysis – of both qualitative and quantitative natures – to dive even deeper into the data available at individual institutions about the first-year students who engage in learning communities. Now that variables have been identified as unique predictors (such as orientation date) and non-predictors (such as admission or first-generation status) of retention or probation status, differences regarding their impact within the context of the learning communities program can be further investigated either through additional quantitative analysis or by qualitative means to shed more light on the phenomenon. Because of the relative dearth in the literature regarding the impact of learning communities on first-year students (Andrade, 2008; MacGregor & Smith, 2005), it would seem that new avenues for research would be welcomed by anyone interested in the fate of the learning community movement and first-year college students as a whole. The MDRC's recent findings (Visher, Weiss, Weissman, Rudd, & Wathington, 2012) question the impact of learning communities on college student success and are a real threat to the future of learning communities; the study is only the first step in the formation of a complete, undeniable, and empirically-based repudiation.

Recommendations for Further Research

It is recommended that the study be replicated to include a larger set of more recent student data, particularly because the learning community program has expanded its offerings since the Fall 2012 semester and the first-year student population has increased significantly in the past two years. It is also recommended that the models be analyzed for predictive validity, using data from cohorts after Fall 2012. The study used pre-college variables that were readily available from the registrar; perhaps a future study could attempt to include other data that are known on the first day of classes that were previously not available.

It is recommended that the impact of orientation date be further explored at the university level, as well as targeted interventions based on the data and prediction models shared with learning community teams prior to the fall semester. Several studies could be conducted based on interventions developed at the university as a result of increased data and information about incoming students. In addition, it is recommended that the learning communities program explore the differences among the learning communities that led to similar results despite dissimilar student populations. Finally, it is recommended that other institutions – with or without learning communities – replicate the study with their own readily available student data to determine which pre-college variables can serve as predictors of retention or probation status for their students. An institution that allows first-year students to opt into learning communities, rather than requiring them like the university in the study, would provide a valuable setting for the comparison of prediction models for students inside and outside of learning communities in order to explore their relationship to student success.

REFERENCES

ACT. (2008). ACT-SAT concordance. Retrieved from http://www.act.org/aap/concordance/

Aleamoni, L. M., & Oboler, L. A. (1977, April). *ACT versus SAT in predicting first semester GPA.* Paper presented at the Annual Meeting of the American Educational Research Association, New York, NY.

Allen, D. (1999). Desire to finish college: An empirical link between motivation and persistence. *Research in Higher Education, 40*(4), 461-485.

Andrade, M. S., (2008). Learning communities: Examining positive outcomes. *Journal of College Student Retention, 9*(1), 1-20.

Araiza, I. (2006). An assessment of the First-Year Learning Community Program: A report on three focus groups with Texas A&M University – Corpus Christi students enrolled in the First-Year Learning Communities Program. Unpublished manuscript, First-Year Experience Council, Texas A&M University – Corpus Christi, Corpus Christi, TX.

Astin, A. W. (1971). *Predicting academic performance in college: Selectivity data for 2300 American colleges.* New York, NY: The Free Press.

Astin, A. W. (1975). *Preventing students from dropping out.* San Francisco, CA: Jossey-Bass.

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

Astin, A. W. (1991). *Assessment for excellence.* Old Tappan, NJ: Macmillan.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco, CA: Jossey-Bass.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education, 12*(2), 155-187.

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

Bean, J. P. (2005). Nine themes of college student retention. In A. Seidman (Ed.), *College student retention: Formula for student success.* Westport, CT: ACE/Praeger.

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

Berger, J. B., & Lyon, S. C. (2005). Past to present: A historical look at retention. In A. Seidman (Ed.), *College student retention: Formula for student success.* Westport, CT: ACE/Praeger.

Bogard, M., Helbig, T., Huff, G., & James, C. (2011, August). *A comparison of empirical models for predicting student retention* [White paper]. Retrieved from http://www.wku.edu/instres/documents/comparison_of_empirical_models.pdf

Borglum, K., & Kubala, T. (2000). Academic and social integration of community college students: A case study. *Community College Journal of Research and Practice, 24,* 567-576.

Boyer, T. A. & Hickman, K. R. (2007). Predicting first-semester agricultural college students' college performance using various measures of academic preparation. *NACTA Journal, 51*(3), 50-55.

Braxton, J. M. (2000). *Reworking the student departure puzzle*. Nashville, TN; Vanderbilt University Press.

Braxton, J. M., & Hirschy, A. M. (2005). Theoretical developments in the study of college student departure. In A. Seidman (Ed.), *College student retention: Formula for student success.* Westport, CT: ACE/Praeger.

Braxton, J. M., & Lee, S. D. (2005). Toward reliable knowledge about college student departure. In A. Seidman (Ed.), *College student retention: Formula for student success.* Westport, CT: ACE/Praeger.

Braxton, J. M., Sullivan, A. V., and Johnson, R. M. (1997). Appraising Tinto's theory of college student departure. In J. C. Smart (Ed.), *Higher education: Handbook of theory and* research, Vol. 12. New York, NY: Agathon Press.

Cabrera, A., Nora, A., & Castaneda, M. (1992). The role of finances in the persistence process: A structural model. *Research in Higher Education, 33*(5), 571–593.

Cabrera, A. F., Castañeda, M. B., Nora, A., & Hengstler, D. (1991). The convergence between two theories of college persistence. *The Journal of Higher Education, 63*(2), 143-164.

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

Chamberlain, T. A. (2011). *Branching out or breaking off? An exploration of first-year student social networks.* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3488197).

Chase, C. I. (1981). *GPA prediction procedures and normative data for freshmen (Indiana Studies in Higher Education No. 44.)* Bloomington, IN: Bureau of Evaluative Studies and Testing, Division of Research and Development, Indiana University.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 40*(7), 3-7.

Cirillo, M., & Smith, T. (2008). The development of a biographical inventory for use in the prediction of first-year college success. *The College of New Jersey Journal of Student Scholarship, 10*. Retrieved from http://www.tcnj.edu/~joss/archive/2008.html

Clark, L. (2012). When nontraditional is traditional: A faculty dialogue with graduating community college students about persistence. *Community College Journal of Research and Practice, 36*(7), 511-519.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Community College of Philadelphia (2011, April). Predicting student persistence: Factors that make a difference. Retrieved from http://www.ccp.edu/vpfin-pl/ir/ir_reports/ir_report_214.pdf

Coperthwaite, C.A., & Klimczak, B. (2008). *Does financial aid status affect student performance, retention, persistence, and academic success?* Hartford, CT: Connecticut Community-Technical Colleges.

Crissman Ishler, J. L, & Upcraft, M. L. (2004). The keys to first-year student persistence. In M. L. Upcraft, J. N. Gardner, & B. O. Barefoot (Eds.), *Challenging and supporting the first-year student: A handbook for improving the first year of college*. San Fransisco, CA: Jossey-Bass.

Dispenzieri, A., & Giniger, S. (1971, February). *Student expectations and experiences as predictors of college performance.* Paper presented at the 55th Annual Meeting of the American Educational Research Association, New York, NY.

Drake, B. M. (2011a, Summer). Predictive modeling of first-year student success factors. Retrieved from http://irt2.indstate.edu/ir/assets/sem/Modeling1.pdf

Drake, B. M. (2011b, Summer). Predictors of academic success for conditional admits. Retrieved from http://irt2.indstate.edu/ir/assets/sem/Modeling3.pdf

Draper, S. W. (2008, February 25). *Tinto's model of student retention*. Retrieved from http://www.psy.gla.ac.uk/~steve/localed/tinto.html

Duquette, C. (2000). Experiences at university: Perceptions of students with disabilities. *The Canadian Journal of Higher Education, 30*(2), 123-142.

Elkins, S. A., Braxton, J. M., & James, G. W. (2000). Tinto's separation stage and its influenc of first-semester college student persistence. *Research in Higher Education, 41*(2), 251-268.

Engstrom, C. M. (2008). Curricular learning communities and unprepared students: How faculty can provide a foundation for success. *New Directions for Teaching and Learning, 115*, 5-19.

Engstrom, C. M. & Tinto, V. (2008). Access without support is not opportunity. *Change, 40*(1), 46-50.

Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: And sex and drugs and rock 'n' roll* (4th ed.). London: Sage.

Glynn, J. G., Sauer, P. L., & Miller, T. E. (2003). Signaling student retention with prematriculation data. *NASPA Journal, 41*(1), 41-67.

Goldman, R. D., & Widawski, M. H. (1976). A within-subjects technique for comparing college grading standards: Implications in the validity of the evaluation of college achievement. *Educational and Psychological Measurement, 36*(2), 381-90.

Griffin, A. M & Romm, J. (Eds.). (2008). *Exploring the evidence: Reporting research on first-year seminars, Volume IV*. Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Habley, W. R., & Bloom, J. L. (2012). *Research-based strategies for college student success*. San Francisco, CA: Jossey-Bass.

Hall, R. A. (2007). *Freshman experience at a community college: Its relationship to academic performance and retention.* (Doctoral dissertation). Retrieved from Digital Commons at Liberty University.

Hanover Research. (2011, May.) *Predicting student retention* [White paper]. Retrieved from http://www3.algonquincollege.com/student-success-committee/files/2013/11/Predicting-College-Student-Retention-Literature-Review-1.pdf?file=2013/11/Predicting-College-Student-Retention-Literature-Review-1.pdf

Hansen, M., Meshulam, S., & Parker, B. (2013). Assessing the effectiveness of a learning community course design to improve the math performance of first-year students. *Learning Communities Research and Practice, 1*(1), 1-24.

Heaney, A., & Fisher, R. (2011). Supporting conditionally-admitted students: A case study of assessing persistence in a learning community. *Journal of the Scholarship of Teahcing and Learning, 11*(1), 62-78.

Henscheid, J. M. (Ed.). (2004). *Integrating the first-year experience: The role of first-year seminars in learning communities* (Monograph No. 39). Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Herreid, C. H., & Miller, T. E. (2009). Analysis of variables to predict first year persistence using logistic regression analysis at the University of South Florida: Model v2.0. *College and University, 84*(4), 12-19.

Hill, P. (1985, October). *The rationale for learning communities*. Speech presented at the Inaugural Conference on Learning Communities, Olympia, WA.

Hosch, B. J. (2008, May). *The tension between student persistence and institutional retention: An examination of the relationship between first-semester GPA and student progression rates of first-time students*. Paper presented at the Association for Institutional Research Annual Forum, Seattle, WA.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*. New York: Wiley.

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Boston, MA: PWS-Kent Publishing Co.

Hoyt, J. E., (1999, Fall). Remedial education and student attrition. *Community College Review, 27,* 51-73.

Huerta, J. C. (2004). Do learning communities make a difference? *PS: Political Science and Politics, 37*(2), 291-296.

Huerta, J. C. & Sperry, R. A. (2010). Pulling it together: Using integrative assignments as empirical direct measures of student learning for learning community program assessment. *Journal of Learning Communities Research, 5*(1), 85-92.

Kahn, J. H., & Nauta, M. M. (2001). Social-cognitive predictors of first-year college persistence: The importance of proximal assessment. *Research in Higher Education, 42*(6), 633-652.

Kamil, M. L., Langer, J. A., & Shanahan, T. (1985). *Understanding reading and writing research*. Boston, MA: Allyn & Bacon.

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter.* Washington, DC: Association of American Colleges and Universities.

Kuh, G. D. (2013). Taking HIPs to the next level. In G. D. Kuh & K. O'Donnell (Eds.), *Ensuring quality and taking high-impact practices to scale*. Washington, DC: Association of American Colleges and Universities.

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79*(5), 540-563.

Kuh, G. D., & Love, P. G. (2000). A cultural perspective on student departure. In J. Braxton (Ed.), *Rethinking the departure puzzle: New theory and research on college student retention* (pp. 196-212). Nashville, TN: Vanderbilt University Press.

Kutil, B., & Sperry, R. (2012). The science of first-year learning communities. In M. Soven, D. Lehr, S. Naynaha, & W. Olsen (Eds.), *Linked courses for general education and integrative learning: A guide for faculty and administrators*. Sterling, VA: Stylus Publishing.

Lardner, E. & Malnarich, G. (2008). A new era in learning community work: Why the pedagogy of intentional integration matters. *Change, 40*(4), 30-37.

Lardner, E., & Malnarich, G. (2009). Assessing integrative learning: Insights from the Washington Center's National Project on Assessing Learning in Learning Communities. *Journal of Learning Communities Research, 3*(3), 1-20.

Laufgraben, J. L. (2005a). Building learning communities for first-year students. In R. S. Feldman (Ed.), *Improving the first year of college* (pp. 195-218). Mahwah, NJ: Lawrence Erlbaum Associates.

Laufgraben, J. L. (2005b). Learning communities. In M. L. Upcraft, J. N. Gardner, B. O. Barefoot, & Associates (Eds.), *Challenging and supporting the first-year student* (pp. 371-387). San Francisco, CA: Jossey-Bass.

MacGregor, J., & Smith, B. L. (2005). Where are learning communities now? National leaders take stock. *About Campus, 10*(2), 2-8.

Maggard, B. S. (2007). *Selected academic variables as predictors of first semester academic success of at-risk football student-athletes at the University of Missouri.* (Doctoral dissertation). Retrieved from MOspace.

Mattson, C. E. (2007). Beyond admission: Understanding pre-college variables and the success of at-risk students. *Journal of College Admission,* (196), 8-13.

McCubbin, I. (2003). An examination of criticisms made of Tinto's 1975 students integration model of attrition. Retrieved from http://www.psy.gla.ac.uk/~steve/localed/icubb.pdf

McKenzie, K., Gow, K., & Schweitzer, R. (2004). Exploring first-year academic achievement through structural equation modelling. *Higher Education Research and Development, 23*(1), 95-112.

Milem, J. F., & Berger, J. B. (Jul/Aug 1997). A modified model of college student persistence: Exploring the relationship between Astin's theory of involvement and Tinto's theory of student departure. *Journal of College Student Development*, *38*(4), 387-400.

Miller, T. E. (2007). Will they stay or will they go?: Predicting the risk of attrition at a large public university. *College and University*, *83*(2), 2-3 & 6-7.

Miller, T. E., & Herreid, C. H. (2008). Analysis of variables to predict first year persistence using logistic regression analysis at the University of South Florida. *College and University, 83*(3), 2-11.

Miller, T. E., & Tyree, T. M. (2009). Using a model that predicts individual student attrition to intervene with those who are most at risk. *College and University*, *84*(3), 12-19.

Miller, T. E., Tyree, T. M., Rieger, K. K., & Herreid, C. H. (2010). Results of the use of a model that predicts individual student attrition to intervene with those who are most at risk. *College and University, 85*(3), 12-19.

Murtaugh, P. A., & Burns, L. D., & Schuster, J. (1999). Predicting the retention of university students. *Research in Higher Education*, *40*(3), 355-371.

Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*: 691-692.

Nathan, R. (2005). *My freshman year: What a professor learned by becoming a student*. Ithaca, NY: Cornell University Press.

National Center for Higher Education Management Systems. (2013). *Retention rates – First-time college freshmen returning their second year*. Retrieved from http://www.higheredinfo.org/dbrowser/index.php?submeasure=223&year=2010

Nelson, J. K. (2009). *Impact of parent education on student success.* Orem, UT: Utah Valley University.

Noble, J. P., & Sawyer, R. L. (2004). Is high school GPA better than admission test scores for predicting academic success in college? *College and University*, *79*(4), 17-22.

Noel-Levitz. (2014). *Student Retention Predictor*. Retrieved from https://www.noellevitz.com/student-retention-solutions/retention-management-system-plus/student-retention-predictor

Nora, A. (2002). A theoretical and practical view of student adjustment and academic achievement. In W.G. Tierney & L.S. Hagedorn (Eds.), *Increasing access to college: Extending possibilities for all students*. Albany: State University of New York Press.

Nora, A., Cabrera, A., Hagedorn, L. S., & Pascarella, E. (1996). Differential impacts of academic and social experiences on college-related behavior outcomes across different ethnic and gender groups at four-year institutions. *Research in Higher Education, 37*(4), 427-451.

Pascarella, E. T., & Chapman, D. W. (1983). A multiinstitituional path analytic validation of Tinto's model of college withdrawal. *American Educational Research Journal, 20*(1), 87-102.

Pascarella, E. T., Pierson, C. T., Wolniak, G. C., & Terenzini, P. T. (2004). First-generation college students: Additional evidence on college experiences and outcomes. *The Journal of Higher Education, 75*(3), 249-284.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students.* San Francisco, CA: Jossey-Bass.

Perry, J. (2010). *Critical elements affecting retention of first-time, full-time freshmen at Tennessee State University* (Doctoral dissertation). Retrieved from ProQuest database. (UMI No. 3433418)

Pizzo, C. C. (2010). *A study of retention and graduation rates as related to student profiles of freshmen, sophomores, and seniors at a large, urban, Catholic university*. (Doctoral dissertation). Retrieved from ProQuest database. (UMI No. 3451006)

Porter, S. R. (1999, June). *Viewing one—year retention as a continuum: The use of dichotomous logistic regression, ordered logit and multinomal logit.* Paper presented at the Annual Meeting of the Association of Institutional Research, Seattle, WA.

Rocconi, L. M. (2011). The impact of learning communities on first year students' growth and development in college. *Research in Higher Education, 52*, 178-193.

Scott, T. P., Tolson, H., & Lee, Y. H. (2010). Assessment of advanced placement participation and university academic success in the first semester: Controlling for selected high school academic abilities. *Journal of College Admission*, (208), 26-30.

Seidman, A. (Ed.). (2012). *College student retention: Formula for student success* (2nd ed.). Lanham, MD: ACE/Roman & Littlefield.

Shapiro, N. S., & Levine, J. J. (1999). *Creating learning communities: A practical guide to winning support, organizing for change, and implementing programs*. San Francisco, CA: Jossey-Bass.

Smith, B., MacGregor, J., Matthews, R., & Gabelnick, F. (2004). *Learning communities: Reforming undergraduate education*. San Francisco, CA: Jossey-Bass.

Smith, R. A. (2010). *"Only connect": A mixed methods study of how first-year students create residential academic and social networks.* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3437586).

Snyder, V., Hackett, R. K., Stewart, M., & Smith, D. (2002, April). *Predicting academic performance and retention of private university freshmen in need of developmental education*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1*(1), 64-85.

Spady, W. G. (1971). Dropouts from higher education: Toward an empirical model. *Interchange, 2*(3), 38-62.

Sterba-Boatwright, B. (2000). The effects of mandatory freshman learning communities: A statistical report. *Assessment Update*, *12*(2), 4-5.

Stevens, J. (1999). *Intermediate statistics: A modern approach* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice, 8*(2), 5-14.

Strayhorn, T. L. (2011). Bridging the pipeline: Increasing underrepresented students' preparation for college through a summer bridge program. *American Behavioral Scientist, 55*(2), 142-159.

Stuart, G. S. (2008). *Making connections: The role of social networks of first-year students enrolled in learning communities in predicting success in college*. (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3343351).

Taylor, K., Moore, J. S., MacGregor, J., & Lindbland, J. (2003). *Learning communities research and assessment: What we know now.* National Learning Communities Project Monograph Series, Olympia, WA: The Evergreen State College, Washington Center for Improving the Quality of Undergraduate Education and the American Association for Higher Education.

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences and cognitive development. *Research in Higher Education, 37*(1), 1-22.

Tharp, J. (1998). Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data. *Community College Journal of Research and Practice*, *22*(3), 279-292.

Thomas, S. L. (2000). Ties that bind: A social network approach to understanding student integration and persistence. *The Journal of Higher Education, 71*(5), 591-615.

Tierney, W. (1992). An anthropological analysis of student participation in college. *Journal of Higher Education*, *63*, 603-618.

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

Tinto, V. (1982). Limits of theory and practice in student attrition. *Journal of Higher Education, 53*(6), 687-700.

Tinto, V. (1988). Stages of student departure: Reflections on the longitudinal character of student leaving. *The Journal of Higher Education*, *59*(4), 438-455.

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

Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *The Journal of Higher Education, 68*(6), 599-623.

Tinto, V. (2000). What have we learned about the impact of learning communities on students? *Assessment Update*, *12*(2), 1-2.

Tinto, V. (2006-2007). Research and practice of student retention: What next? *Journal of College Student Retention*, *8*(1), 1-19.

Tinto, V., Love, A. G., & Russo, P. (1993). *Building learning communities for new college students: A summary of research findings of the collaborative learning project*. University Park, PA: National Center on Postsecondary Teaching, Learning, and Assessment.

Tobolowsky, B. F., Cox, B. E., & Wagner, M. T. (Eds.). (2005). *Exploring the evidence: Reporting research on first-year seminars, Volume III* (Monograph No. 42). Columbia, SC: University of South Carolina, National Resource Center for the First-Year Experience and Students in Transition.

Torres, J., & Solberg, S. (2001). Role of self-efficacy, stress, social integration, and family support in Latino college student persistence and health. *Journal of Vocational Behavior, 59*, 53-63.

Triola, M. F. (2002). *Essentials of Statistics.* Boston, MA: Pearson Education, Inc.

Tross, S. A., Harper, J. P., Osher, L. W., & Kneidinger, L. M. (2000). Not just the usual cast of characteristics: Using personality to predict college performance and retention. *Journal of College Student Development*, *41*(3), 323-34.

Van Gennep, A. (1960). *The rites of passage*. (M. B. Vizedom & G. L. Caffee, Trans.). Chicago, IL: University of Chicago Press (Original work published 1908).

Visher, M. G., Weiss, M. J., Weissman, E., Rudd, T., & Wathington, H. D. (2012). *The effects of learning communities for students in developmental education: A synthesis of findings from six community colleges* [Executive summary]. New York, NA: National Center for Postsecondary Research.

Vogt, W. P. (2007). *Quantitative research methods.* Boston, MA: Pearson Education, Inc.

Voorhees, R. A. (1987). Toward building models of community college persistence: A logit analysis. *Research in Higher Education*, *26*(2), 115-129.

Waldron, V. R., & Yungbluth, S. C. (2007). Assessing student outcomes in communication-intensive learning communities: A two-year longitudinal study of academic performance and retention. *Southern Communication Journal, 72*(3), 285-302.

Warburton, E., Bugarin, R., & Nunez, A. (2001). *Bridging the gap: Academic preparation and postsecondary success of first-generation students* (NCES 2001–153). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Washington Center. (2014). *The National Resource Center for Learning Communities*. Retrieved from http://www.evergreen.edu/washingtoncenter/

Wetzel, J. N., O'Toole, D., & Peterson, S. (1999). Factors affecting student retention probabilities: A case study. *Journal of Economics and Finance*, *23*(1), 45-55.

Wyatt, L. G. (2011). Nontraditional student engagement: Increasing adult student success and retention. *The Journal of Continuing Higher Education*, *59*, 10-11.

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education*, *45*(2), 115-138.

Zheng, J. L., Saunders, K. P., Shelley, M. C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development*, *3*(2), 267-83.

APPENDIX
IRB Approval
Records Request

ERIN L. SHERMAN, MAcc, CRA, CIP, CPIA
Research Compliance Officer
Division of Research, Commercialization and Outreach

6300 Ocean Drive, Unit 5844
Corpus Christi, Texas 78412
O 361.825.2497 • F 361.825.2755

Human Subjects Protection Program	Institutional Review Board

APPROVAL DATE:	April 29, 2013
TO:	Ms. Rita Sperry
CC:	
FROM:	Office of Research Compliance Institutional Review Board
SUBJECT:	Initial Approval

Protocol Number:	40-13
Title:	Predicting the Retention and Probation Status of First-Year Students in Learning Communities
Review Category:	Exempt

Approval determination was based on the following Code of Federal Regulations:

Eligible for Exempt Review (45 CFR 46.101)

Criteria for Approval has been met (45 CFR 46.101) - The criteria for approval listed in 45 CFR 46.101 have been met (or if previously met, have not changed).

(4)Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

Provisions:

Comments:

This research project has been approved. As principal investigator, you assume the following responsibilities:

1. Informed Consent: Information must be presented to enable persons to voluntarily decide whether or not to participate in the research project unless otherwise waived.
2. Amendments: Changes to the protocol must be requested by submitting an Amendment Application to the Research Compliance Office for review. The Amendment must be approved before being implemented.
3. Completion Report: Upon completion of the research project (including data analysis and final written papers), a Completion Report must be submitted to the Research Compliance Office.
4. Records Retention: Records must be retained for three years beyond the completion date of the study.
5. Adverse Events: Adverse events must be reported to the Research Compliance Office immediately.

Texas A&M University – Corpus Christi
University Core Curriculum Programs
MEMORANDUM

To: Dr. Michael Rendon, University Registrar

From: Dr. David Billeaux, Associate Vice President for Academic Affairs

CC: Ms. Barbara Hand, Manager, Academic Programming
Dr. Carlos Huerta, Director, Core Curriculum Programs
Ms. Rita Sperry, Seminar Coordinator

Subject: Student Records Request

Based on recent communications with the Manager of Academic Programming, approval must be obtained from the University Registrar when requesting student information from Banner. In order to assess the First-Year Learning Communities Program and construct a model for predicting the retention and probation status of incoming first-year students, the following data is requested for all TAMU-CC students who were enrolled in **UCCP 1101** in the **Fall 2010**, **Fall 2011**, and **Fall 2012** semesters:

(a) List of courses taken in the first fall semester
(b) Number of total credit hours attempted in the first fall semester
(c) Number of developmental credit hours attempted in the first fall semester
(d) Grades for courses taken in the first fall semester
(e) First fall semester GPA
(f) Retention into the following spring semester (1 for retained, 0 for not retained)
(g) Retention into the following fall semester (1 for retained, 0 for not retained)
(h) High school class rank
(i) Number of credit hours earned prior to enrollment at TAMU-CC
(j) SAT scores
(k) ACT scores
(l) Date of birth
(m) Gender
(n) First-generation status (1 for first-generation, 0 for not)
(o) Ethnicity
(p) Pell-grant eligibility status (1 for eligible, 0 for not)
(q) Admission date
(r) Admission status
(s) Orientation date
(t) Major
(u) College

The data is requested in electronic format and should be directed to Rita Sperry, who will store it on a secured campus computer.