ABSTRACT

HERRERA, OLGA LUCIA. Investigation of the role of pre- and post-admission variables in undergraduate institutional persistence, using a Markov student flow model. (Under the direction of Frank J. Smith).

This study used selected student record data to investigate the effect of students' characteristics prior to university admission (pre-admission variables), and academic actions and educational achievement indicators (post-admission variables) on retention in higher education. The analysis followed first-year undergraduate students at a large Midwestern university through four academic levels (freshman-senior).

A Markov student-flow model was employed to estimate the probabilities of stopping out, staying at the same academic level, or advancing to a higher academic level up to graduation. Logistic regression was used to calculate fourteen transition probabilities of specific flow-model events given a profile of independent variable scores. Based on the yearly transitions, predicted probabilities of graduating after 4, 5 and 6 years were also computed.

The key results are (a) The Markov student flow model and its use as a predictive tool, which allow calculation of a persistence risk value using institutional data. (b) The finding that many variables vary in predicting persistence depending on the academic level, which corroborates the need to organize the model by academic levels and indicates that it is incorrect to conclude that variables that affect persistence at one academic level do so at all levels.

Relevant to the specific institution studied are the findings that variables such as Age at Entrance, and Pell Grant Indicator consistently predict lower probabilities of progressing

towards graduation for all academic levels, holding other variables in the model constant. Cumulative GPA and Not Changing Majors also predict higher transition probabilities, with the strongest effect at the sophomore level. Target Minority, ACT score and High School Percentile predict higher probabilities of persisting at the Freshman level, but the effect becomes negative at the Senior level.

If tested and implemented in an institution, the proposed simulation tool would allow decision-makers to examine potential effects of policies by altering variable profiles and analyzing the predicted changes in the institutional persistence of students. The probabilities obtained can be interpreted as an empirical persistence risk value.

INVESTIGATION OF THE ROLE OF PRE- AND POST-ADMISSION VARIABLES IN UNDERGRADUATE INSTITUTIONAL PERSISTENCE, USING A MARKOV STUDENT FLOW MODEL

by

OLGA LUCIA HERRERA

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy

PSYCHOLOGY

Raleigh, NC
2006

Approved by:

Lynne E. Baker-Ward	Craig C. Brookins
Aaron M. Brower	Donald W. Drewes

Frank J. Smith
Chair of Advisory Committee

This dissertation is dedicated to my family

Olga Velasco, my mother whom instilled in me the virtue of determination through making learning fun.

Federico Herrera, my father who always reminded me of the value of education for women.

Daniel Herrera, my brother through whom I always see the bright side of everything.

…and finally, to my exchange student family, the Satterfields, who are my anchor family in the U.S., and who made me a Wolfpack fan to the point where a Ph. D. was inevitable.

Agradecimientos

Muchas gracias a todos los que hicieron este proyecto posible, desde mis padres que me inculcaron la importancia de la educación, hasta las personas que hacían día a día el proceso de acabar mi tesis viable (mi esposo, los profesores, colegas, amigos, vecinos, niñeras y familia).

[*translation:* Thank you to everyone who made this project possible, from my parents who instilled in me the importance of education, to those who made the daily process of finishing my thesis possible (my husband, professors, colleagues, friends, neighbors, babysitters and family)]

BIOGRAPHY

As I describe my academic life at length I remind myself that I have lived passionately doing what I love. I was born in New York city, but grew up in Popayán, Cauca, a lovely colonial city in the southwest of Colombia. From elementary to high school I attended “El Colegio San Jose de Tarbes.” Doing community work with the nuns in impoverished areas of the city, I became interested in how children with very few resources were ingenious and thrived in non-formal settings (i.e., selling in the streets). I knew then that I would study something related to education and community. Before college in 1985-86, I was an exchange student at Broughton High School in Raleigh NC and lived with the Satterfields, who became my family in the US. With lots of love and lots of teasing they made sure I learned good English. I returned to Colombia for my undergraduate studies, and in 1991 completed a five year B.S. program in Psychology from “La Pontificia Universidad Javeriana” in Bogotá. In this program I did practica in organizational and educational psychology. In the latter, I helped to implement innovative reading, writing and math curriculums for first graders of Bogotá’s economically disadvantaged public schools. My undergraduate thesis was a comparison of the effectiveness of traditional ways of teaching math in elementary school vs. a more hands-on activities approach.

As school-ish as it might sound, having seen the libraries and the access to technology in the US, I knew where I would go for graduate school. While figuring out financial aid and graduate schools, I spent a year with my host sister Carson Satterfield doing seasonal work in Breckenridge, CO and in Martha’s Vineyard, MA. In 1993 I entered the Psychology in the Public Interest Ph.D. program which was called Human Resource Development. I met Frank J Smith, my advisor, whose fluency in Spanish and empathy for my interest in bridging theories of psychology and day-to-day challenging social problems encouraged me to further my education.

I spent the summer of 1994 working with two ethnic rural communities in Colombia collecting data on peoples' views about the quality, equity and relevance of the education in elementary school. This became the topic of my Master's thesis, degree acquired in 1997. In search of a dissertation topic, I worked during the summer of 1998 for the United Nation Population Fund (UNFPA) in Tegucigalpa, Honduras. During the three month consulting job I did qualitative research to understand attitude, knowledge, and behavior towards pregnancy of underprivileged adolescents from metropolitan Tegucigalpa.

While camping in the North Carolina Mountains in 1994, I met Nigel Orton who was finishing a Ph. D. in Nuclear Engineering at NCSU. In 1999 we married in my home town and soon after he accepted a residency in Medical Physics at the University of Wisconsin-Madison. Somewhat comfortable with my ABD (i.e., All But Dissertation) status we moved to Madison and I worked for nearly three years at the Learning through Evaluation, Adaptation, and Dissemination (LEAD) Center at UW-Madison. I was part of a research team in charge of providing high-quality formative and summative evaluation for programs in education, technology, health, and social services. LEAD director Susan Millar's high expectations for my career, and (oddly enough) the birth of my first child, Sebastian Felipe Orton, incited an interest in pursuing my dissertation. In 2002, Dr. Steve Kosciuk and Dr. Aaron Brower helped me negotiate access to student record data with the UW Registrar's office. I was offered a limited-term position which funded part of my study. With the data in hand and the knowledge that there was a need to systematically track problems of student retention, this dissertation got started. I enjoyed the process and the restriction of working part-time in order to keep up with my other responsibilities. These include taking care of Sebastian and my second son Benjamin, plus spending time with my husband, playing Frisbee, as well as keeping up with our organic cooking, gardening, camping and traveling in the country and overseas.

ACKNOWLEDGEMENTS

No “individual” work of this length is really done alone. Therefore I would like to start by first thanking my committee for their expertise and suggestions. Profound appreciation goes to my advisor Dr. Frank Smith, for his guidance, patience and ongoing support through out my career. To Dr. Donald Drewes for making this project a rich learning experience and helping to convert my ideas into a productive & realistic project. Dr. Aaron Brower, for mentoring me and treating me as if I was really a student at his university. Thanks to my colleagues at the LEAD center, especially Susan Millar, for always expecting that I would earn a doctoral degree. My gratitude extends to: Steve Kosciuk, for the long hours spent teaching me how to dive into the cubes and data views and to handle the huge amount of data; Marilyn McIntire for her support and encouragement; Doug Hemken for quickly responding to all my SAS consultations; and Jeff Gear and Ben Payne for solving all the hardware problems.

I thank my babysitters (Olga Neira, Olga Tavarez, Marta Muñoz, Alejandra Garate, Antonieta Temosique y Maritza Duran) for respecting my work schedule, taking good care of my children and assuring me they would speak Spanish to them in my absence.

Finally, special thanks to my husband Nigel-ito for being supportive, valuing my career and being my personal editor and back-up babysitter. Thanks for making financial sacrifices as well as sacrificing much of our time that might have otherwise been spent camping and playing Frisbee. I only hope that your statement many years ago comes true “…by the time you finish your Ph. D. you’ll know how to write in English.”

TABLE OF CONTENTS

LIST OF TABLES .... **ix**

LIST OF GRAPHS .... **x**

LIST OF FIGURES .... **xi**

TITLE **1**

CHAPTER 1: INTRODUCTION **1**

Background **1**
Area of Concern **4**
Resilience in the Context of Education 7
Problem Statement **9**
Purpose **10**

CHAPTER 2: REVIEW OF RELEVANT LITERATURE **13**

PART 1: RETENTION IN HIGHER EDUCATION **13**

Definition of Terms **13**
Theoretical Models of Persistence **16**
Student Integration Model 16
Input-Environment-Output (IEO) Model 18
Student Attrition Model 19
Integrated Model of Student Retention 20
Student Faculty Interaction Model 21
Student Social and Academic Integration Community Model 22
Summary of Theoretical Models of Persistence 23

PART II: RESEARCH EVIDENCE **25**

Traditional Predictors of College Persistence **25**
Summary and Discussion of the Literature on Persistence 36
At-Risk Students 38
A Review of the College Choice Literature 39
Summary and Discussion: Understanding who are at-risk students 43

PART III: EDUCATIONAL RESILIENCE **45**

The Concept of "Resilience" **45**
Resilience: A trait within the individual or a product of his /her environment? 46
Protective Factors 48
Summary and Discussion of the Literature Review on Risk and Resilience 51
Why does the resilience concept matter in higher education and how would a study from this perspective be different than any study in the persistence literature? 52

CHAPTER 3: METHODS **54**

Goal and Objectives **54**

Delimitations ... **56**
Research Design ... **56**
Quantitative approach to undertake the theoretical model proposed in this study ... 57
Population ... 61
Sampling ... 62
Description of Retention Data Views (RDV) or student record data source ... 63
Measures ... **64**
Variable Selection Process ... 64
Independent Variables ... 68
Administrative variables ... 73
Dependent Variable ... 74
Data Analysis ... **75**
Research hypotheses ... 78
Hypothesis testing ... 79
Informed Consent ... 79

CHAPTER 4: RESULTS ... **80**

Data Management ... **80**
Recoding variables and dealing with missing values ... 82
Descriptive Analysis ... **85**
Descriptive results for the dependent variables and transition probabilities ... 89
Actual graduation rates ... 93
Test of the Hypothesized Model ... **94**
Test of hypothesis 1: results of beta coefficients for pre-admission variables ... 95
Test of hypothesis 1.1: beta coefficient values and trends ... 96
Confirmation of hypothesis 1 and hypothesis 1.1: test of the significance for the beta coefficients ... 108
Test of hypothesis 2: results of beta coefficients for post-admission variables ... 111
Test of hypothesis 2.1: beta coefficient values and trends ... 113
Confirmation of hypothesis 2 and hypothesis 2.1: test of the significance for the beta coefficients ... 127
Simulation Application ... **129**
Example: how to calculate probabilities by hand ... 131
Four year and six year transition probabilities ... 134

CHAPTER 5: SUMMARY & DISCUSSION OF RESULTS ... **145**

Validation of the Persistence Probability Model for the Pre-Admission Variables ... **147**
Expected results ... 147
Unexpected results ... 152
Validation of the Persistence Probability Model for the Post-Admission Variables ... **158**
Expected results ... 158
Unexpected Results ... 161
Limitations ... **165**

CHAPTER 6: IMPLICATIONS OF THE STUDY ... **168**

Implications for Theory ... **168**
Implications for Practice ... **169**

Implications for Research 171

REFERENCES 175

APPENDICES 182

Appendix A: Glossary of Terms 183
Appendix B: Supplementary Tables 184
Appendix C: Hypothetical Profile Scores for Policy Options 1,2, and 3 188
Appendix D: Routine to Estimate Transition Probabilities 191
Appendix E: Simulations Output TP, $(TP)^4$, $(TP)^5$, $(TP)^6$ 198

LIST OF TABLES

Table 1 Effect of Environmental Characteristics or Experiences on Students' Outcomes 30
Table 2 A Summary of Factors that Affect Academic Persistence 35
Table 3 Academic groups available to undergraduate students............ 72
Table 4 Independent variables............ 73
Table 5 Administrative Variables............ 74
Table 6 Data organization summary............ 81
Table 7 Description by cohort and matriculation academic level 86
Table 8a Descriptive statistics for the independent variables............ 87
Table 9 Descriptive results for the DV............ 92
Table 10 Observed transition probabilities 92
Table 11 Guide to Interpreting Beta Coefficients............ 97
Table 12 Pre-Admissions Predictors by Academic Level............ 101
Table 13 Post-Admissions Predictors by Academic Level 116
Table 14 Transition probability matrix for 18 year old student 130
Table 15 Transition probability matrix for 25 year old student 130
Table 16 Transition probability matrix for student A 137
Table 17 Transition probability matrix for student B............ 137
Table 18 Probabilities of stopping out and graduating from $(TP)^4$, $(TP)^5$, $(TP)^6$............ 138
Table 19 Transition probability matrix for student X 140
Table 20 Transition probability matrix for student Y 141
Table 21 Probabilities of graduating for student X and Y, from $(TP)^4$, $(TP)^5$ & $(TP)^6$ 143

LIST OF GRAPHS

Graph 1 Number of students who do and do not graduate, by years to degree..............93
Graph 2 Beta Coefficients for Age at Entrance Across Academic Levels........................98
Graph 3 Beta Coefficients for all Ethnic Groups Across Academic Levels 100
Graph 4 Beta Coefficients for Gender Across Academic Levels....................................102
Graph 5 Beta Coefficients for Target Minority Status Across Academic Levels 104
Graph 6 Beta Values for Pell Grant Indicator across Academic Levels 105
Graph 7 Beta Values for High School Percentile Across Academic Levels...................106
Graph 8 Beta Values for ACT scores Across Academic Levels......................................108
Graph 9 Beta Values for Cumulative GPA Across Academic Levels............................115
Graph 10 Beta Values for Change Academic Group across Academic Levels.............118
Graph 11 Beta values for Changing Majors across Academic Levels...........................119
Graph 12 Beta Values for M112 Course and Grade across Academic Levels..............121
Graph 13 Beta Values for M221 Course and Grade across Academic Levels.............122
Graph 14 Beta Values for CH103 Course and Grade across Academic Levels............124
Graph 15 Beta Values for Econ 101 Course and Grade across Academic Levels........125
Graph 16 Beta Values for Psy 202 Course and Grade across Academic Levels..........126

LIST OF FIGURES

Figure 1 Pipeline of Institutional Persistence in Undergraduate Education...................58

TITLE

Investigation of the role of pre- and post-admission variables in undergraduate institutional persistence, using a Markov student flow model

CHAPTER 1: INTRODUCTION

Background

The importance of postsecondary education has never been greater. Education beyond high school is increasingly crucial to young people's ability to compete and prosper in the nation's job market. Agricultural and factory-related jobs that once sought less educated individuals now require operation of advanced technological equipment, which demands workers to have skills typically attained at the college level. The pressure to develop human talent and extend educational levels to higher education is of concern to all nations. The National Center for Public Policy and Higher Education warns that the nations most successful in developing human talent through the postsecondary levels will have the greatest competitive advantage over those that do not (Callan & Finney, 2003). The need for a workforce with postsecondary skills and education is therefore imperative for public policy, the government, and higher educational institutions, where institutional and public interests are bridged.

Higher education is becoming more accessible, and students are placing a greater value on acquiring more education. This is reflected in the number of students who enroll in college immediately after high school graduation. The college enrollment rates of high school graduates rose from 49% in 1972 to 66% in 1998 (U.S. Department of Education, 2000a). Enrollment in four-year institutions increased from 7.7 million in 1984 to an estimated 8.8 million in 1997, an average annual growth rate of 1.0%. For the nation, college enrollment is projected to increase from 15.3 million in 2000 to 17.7 million by 2012, an increase of 15 percent (U.S. Department of Education, 2002).

A large portion of students entering higher education and later the workforce will come from low-income families and demographic groups that are least well served by American education. Targeted minorities, first generation college goers, students from low-income families and English language learners often have the lowest rates of completing high school and enrolling in college (Callan & Finney, 2003). Once in college, individuals from these groups also have difficulty persisting in a four-year college and attaining a degree (Cabrera, La Nasa, & Burkum, 2001).

Despite the motivation of students from all backgrounds to pursue postsecondary education, persistence (i.e. degree completion) rates lag far behind college enrollment. According to the National Center for Education Statistics (U.S. Department of Education, 2000a), 3.3 million students enrolled in postsecondary education in 1995–96 for the first time. Their outcomes three years later varied with their initial goal, the type of institution in which they first enrolled, and whether they transferred. By 1998, 37% had left postsecondary education without an award.

Concern regarding degree completion is growing, based on the amount of funds invested by federal and state government, as well as institutional dollars that are spent to identify factors related to students' persistence. Institutional persistence, keeping students in the university where they started, is important to universities, students, their families and legislators. Universities are realizing that from a financial standpoint they are better off retaining their students than attracting new ones. Departure from the institution may reflect the lack of coherent efforts to accomplish the goal of making sure students earn a college degree. Graduating their students and reducing time to graduation is of great concern to institutions, and the only viable solution to accommodate enrollment increases (Stancill, 2006; U.S. Department of Education, 2002)

For students, the cost of attending postsecondary education continues to increase. In the 2003–04 academic year, the average total cost for full-time undergraduates to attend 4-year institutions—including tuition, fees, room, board—was estimated to be $10,636 at public institutions and almost $26,854 at private institutions (The College Board, 2003). Terminating school without a degree results in income loss and high levels of frustration.

A college degree brings a variety of benefits to society and to the individual. College graduates are more empowered in society. They are more likely to assume civic leadership positions, more likely to vote, and support advanced education for their children (Pascarella & Terenzini, 1991). Finally, educational attainment and income almost always show a positive linear relation (U.S. Department of Education, 2001a).

There is no doubt that need to improve degree completion in American higher education is imperative for those providing and receiving the education. Thus, the ability to identify potential institutional stop-outs from student record data can lead to the enhancement of retention rates through planned interventions for those students who may be at-risk. This study proposes and tests a predictive model that will help to understand the institutional persistence of students as they move from one academic level to the next (e.g., advancing from freshman to sophomore status) up to graduation.

Area of Concern

If administrators and admissions offices want their students to attain degrees, parents are invested in their children's education, and students are choosing to attend college with obvious reasons to attain a degree, why is student institutional retention still a problem? Some authors have tried to understand the problem of retention by understanding behaviors and factors affecting students' decisions and their conditions before attending college (Hossler, Schmit, & Vesper, 1999).

Educational researchers have also sought to understand the effect of college attendance on student growth and development (Pascarella & Terenzini, 1991). Research has brought conceptual understanding of student academic persistence and college success. The phenomenon of student dropouts has been thoroughly analyzed from a psychological, economical, societal, and institutional point of view (Tinto, 1993). A specific set of the literature concentrates on understanding the adjustment process and reasons for departure from different types of institutions (i.e., private vs. public; four or two year colleges) and grade levels (e.g., first year of college (Brower, 1990; Green, 1998). In addition, researchers have looked specifically at students with differing attributes, such as race, academic ability, and socio-economic status (Cabrera & La Nasa, 2000c; Jones, 2001; Tinto, 1993).

The literature indicates that certain conditions and factors are associated with an increased probability of leaving college and not persisting (Pascarella & Terenzini, 1991; Tinto, 1993). Some of these factors are highly correlated to students' demographic characteristics: for example, belonging to the lowest socioeconomic status (SES), gender (Leppel, 2002) age, ethnicity, academic performance (Tinto, 1993), having parents with a low level of education or being first generation in college (i.e., having parents who have not gone to college)(Choy, Horn, Nuñez, & Chen, 2000); to the individuals personality, level of motivation, self-efficacy, interest in academics, study skills, and type of employment (Green, 1998). Some conditions depend on the institution: academic advising, mentoring programs, teaching practices, role models, availability of financial resources, type of academic enrollment (Bean, 1980; Cabrera, Stampen, & Hansen, 1990; Colbeck, Cabrera, & Terenzini, 2001); and to the community: encouragement and family support (Nora, 2001).

Researchers have identified groups of students who have a tougher time securing their path to college or are at risk of not finishing their degree. These groups include low-income students, targeted minorities, first-generation college students and academically under-prepared students. Once in college, low-income students are more likely to follow attendance behaviors that reduce the likelihood of completing their degree (e.g., working outside of school or full time, not choosing a major, and starting at a community college) (Choy, 2002; Pascarella & Terenzini, 1991). Targeted minorities face difficulties in postsecondary institutions, especially when they attend predominately white institutions (PWI) (Newman, 1997). Cross-cultural and communications barriers which limit involvement in academic and social activities are two common factors affecting minorities' persistence (Newman, 1997; Tinto, 1993). First-generation students, those whose parents received no education beyond high school, are less likely to enroll in college, and if they do they remain at a disadvantage with respect to staying enrolled and attaining a degree (U.S. Department of Education, 2001a). This is true even after controlling for variables such as income, educational expectations, academic preparation and family support (Choy, 2002; U.S. Department of Education, 2001a). Belonging to the lowest SES and coming from families where parents have less education correlate with poor high school academic preparation (U.S. Department of Education, 2001a). Finally, achievement test scores and high school GPA have long been known to be predictors of success in college (Pike & Saupe, 2002).

In summary, the literature is rich in identifying factors that lead to school failure, and pointing at groups who due to their characteristics are at a higher risk of not persisting (Bernal, Cabrera, & Terenzini, 2000; Choy 2002; Doolittle, 1996). However, less is known about what leads to success. What are quantitative ways for institutions to determine who is in fact at risk at the time of enrollment and after enrollment? And why do some students admitted to schools with the so called at-risk characteristics manage to be academically successful and persist? Despite the overwhelming number of studies on the topic of persistence, college student departure still poses a puzzle to college and university administrators (Braxton & Mundy, 2001).

Therefore, there is a need to broaden the understanding of the problem by taking a different approach. In response to this need, the present study investigates what factors contribute to students' institutional persistence in higher education by proposing and testing a predictive model using institutional data. Of special interest is to be able to quantify risk in order to later identify students who succeed despite their at-risk characteristics. In this sense academic success or persistence is defined as successfully advancing from one academic level to the next and graduating in a timely manner, despite having risk factors that according to the literature hinder probabilities of graduation. The educational resilience literature will be used as lens to develop and test an early warning system for institutions that will predict a four academic level progression towards graduation of both at-risk and non-at-risk students.

Resilience in the Context of Education

In the field of psychology and K-12 education, resilience is a term used to describe a set of qualities that foster a process of successful adaptation and transformation despite risk and adversity (Benard, 1995; Masten, 1994). Risk factors can be biological or psychosocial hazards that increase negative outcomes (Worrel, 1996). Resilience is often used to describe people from high-risk groups who "beat the odds" or do better in life than expected considering their stressors and unfavorable conditions (Thielemann, 1999). Doll and Lyon (1998) warned that neither risk nor resilience is a characteristic of the individual in isolation. Risk and resilience are characteristics of an individual living within the context of the family, neighborhood, community, and school. Thus, the individual's environment can amplify or mitigate the impact of risk factors. In other words, the environment can offer protective factors that counteract the at-risk condition.

According to Benard (1995) there are three critical protective factors that, when present in an individual's environment (i.e., family, school, and community), can help him or her persevere despite adversity. These are:

(a) Caring environment: the presence of a caring adult who knows the individual well and cares about his or her well-being.

(b) Positive expectations: high, clearly-articulated expectations, and the purposeful support necessary to meet those expectations.

(c) Opportunities for participation/contribution: responsibilities and opportunities for meaningful involvement with others.

Following up on the effective schools research for urban elementary and secondary schools, and in line with other researchers (Benard, 1991; LePage-Lees, 1997), Wang

(1998b) looked at caring attitudes and positive expectations as protective factors in the school environment. A caring attitude from a teacher can be demonstrated directly through academic and social interactions, and indirectly through their classroom structure, curriculum and instructional practices. In the context of higher education, forms of advising, class structure, class size and curriculum may serve as measures of a caring environment.

The assumptions teachers have about their students' capabilities affect how they relate to students and how they conduct their classes. It is easier to hold positive expectations for students who perform better, are well-prepared academically or do not belong to a disadvantaged group. Disadvantaged groups include the economically deprived, low socioeconomic status families, and ethnic minorities (Hossler et al., 1999). In higher education, possible measures of positive expectations may be the type of courses a student takes, the number of credits, the type of major, the continuity in a selected major and high academic achievement.

Positive academic, social and moral expectations are more common among parents who are involved in their children's education (Wang, Haertel, & Walberg, 1998a). High expectations and involvement are closely related. Most important, however, is students' own opportunity for participation and responsibility in their educational activities. The knowledge that involvement of students in school academic and social activities has long been supported by theoretical models of persistence in higher education (Astin, 1993; Tinto, 1993). According to Harper (2003), involvement in academic and out-of-class activities is essential for school success, especially for African-American men.

Opportunities for participation in higher education are often offered through federal and institutional programs. One well-known group of academic programs is TRIO. TRIO refers to three original programs, Upward Bound, Talent Search and Student Support Services, created with the first reauthorization of the Higher Education Act in 1968. These are outreach and support programs designed to assist at-risk students (e.g., low-income, first-generation college and students with disabilities) to begin and complete post-secondary education.

In the context of education, Wang (1994) referred specifically to *educational resilience*, which is defined as “success in school and in other life accomplishments, despite environmental adversities brought about by early traits, conditions and experiences" (p. 46). In a school setting, educational resilience may be reflected in two ways: at-risk students’ ability to identify protective factors in their environment, and the school’s capacity to offer protective opportunities to build academic resilience. For example, students who are able to identify and use the support offered to them, and schools that make students aware of programs available to them and offer to students of all backgrounds a recommend path towards attaining a degree.

Problem Statement

The review of the literature on persistence leads to the conclusion that less is known about how to identify successful paths towards institutional persistence and degree completion; more specifically, how to quantify risk in order to help students replicate positive experiences of staying in school despite the presence of risk factors, as determined by the literature. In this sense institutional persistence is defined as successfully advancing from one academic level to the next and graduating in a timely manner, despite having risk factors that according to the literature hinder probability of graduation.

The present study aims to identify demographic characteristics and enrollment behavior patterns that can positively or negatively affect the probabilities of successfully attaining a degree in the institution where the student first enrolled. The study uses a multivariate quantitative analysis approach to investigate the effects of selected factors on institutional persistence. The ultimate goal is to understand what in the setting of a higher education institution serves as protective factors and facilitates the success of students, allowing them to advance from one academic level to the next until graduation within the expected time frame of 4 to 5 years.

Purpose

The present study contributes to the literature on institutional persistence in higher education by augmenting the knowledge on how higher education communities can implement a systems-wide prediction tool to differentiate their at-risk populations from others. An alternate purpose is to determine whether in a university with strict admission policies (i.e. the majority of the students have a high ACT and school rank), the so called at-risk students are in fact experiencing problems in their educational continuity when compared to the rest of the population. This was accomplished in this study by designing and testing a Markov student flow model, which is expected to help identify factors that protect students and promote institutional persistence.

The educational resilience theory offers a focus on positive phenomena like "success" and "strength" and the opportunity to explore what causes or supports these phenomena. The understanding of these concepts might ultimately help to find out whether educational resilience is something that can be deliberately reproduced or encouraged, and if so, how?

This paradigm shift, concentrating on success and not failure, could also improve the understanding of what factors in the university environment foster educational resilience among students, especially those who are thought to be at-risk because of their characteristics. Identifying an academic path of successful at-risk students will give a new insight into the problem of student institutional persistence.

The main guiding questions in this study are:

1. What role can educational resilience play in understanding institutional persistence? Educational resilience is defined as advancement of "at-risk students" to graduation. Risk implies that a designated group of students demonstrates a greater tendency to not persist in comparison to the rest of the population. In the context of this study, early indication of failure to persist is shown by failure to advance in a timely fashion (e.g., stopping out at freshman, sophomore, junior, or senior levels, not advancing from one level to next and undue retention time at each academic level).

2. What variables are significant in predicting whether a student will advance to the next academic level, remain in the current level, or stop-out? In the context of this study, what constitutes a risk factor? Does the effect of these factors on persistence vary depending on the students' academic level?

3. Are there variables that serve as protective factors and moderate against the likelihood of academic failure? Do these variables function as "protective" for all students or only for those students who the literature considers at-risk due to their particular risk characteristics?

Understanding what fosters educational resilience can help university administrators to identify and promote protective mechanisms for non-resilient students, and to provide the foundation for building an academic environment that can retain its students. A Markov student flow model will be used in this study to investigate the empirical value of factors influencing institutional persistence. The model has the potential to enable administrators to track students' progress from year to year throughout their undergraduate tenure. Finally, by identifying the factors that promote persistence, this study will facilitate a systematization of policies that benefit a broader range of students.

CHAPTER 2: REVIEW OF RELEVANT LITERATURE

This chapter is organized in three parts. The first part presents all the literature related to institutional persistence in higher education. It includes definitions of terms[1] used in this section, such as retention, persistence, and academic success, and ends with a summary of the theoretical models that have attempted to explain institutional persistence. The second part complements the theoretical models of persistence by identifying supporting research that investigates predictors of persistence. The college choice literature is used to define the at-risk population and to understand the possible outcomes in the present study. The third part of the chapter aims at explaining educational resilience. The construct of resilience is explained, placing emphasis on understanding environmental characteristics that foster educational resilience. The last part of this literature review provides a definition of educational resilience and some examples of studies in educational resilience. These examples are expected to facilitate the understanding of how the theory of resilience can be used to understand college institutional persistence.

PART 1: RETENTION IN HIGHER EDUCATION

Definition of Terms

In this study, the term *retention* in higher education refers to the institutional efforts to keep students in college so that they successfully attain a degree. Retention is appropriate when used to refer to what institutions do to promote educational attainment and degree

---

[1] Some other terms used may appear in the Glossary of Terms, after the References.

completion. Retention is not to be confused with common secondary school usage which refers to retaining a student in his/her present grade/course (Afolayan, 1996).

Persistence, a term often used in parallel with retention, has been introduced in the educational literature with reference to undergraduate degree and educational attainment (Pascarella & Terenzini, 1991). The term refers more to the efforts of the individual, not the institution, to continue his or her education and attain a degree. Sometimes persistence is specified as academic persistence, referring more specifically to the academic success of the individual (e.g., GPA progress). In other situations, persistence refers to students who have completed or are enrolled (full-time or part-time) in their initial institution within a 5 to 6 year period. Most of the literature and well-known theoretical models explaining 'persistence' refer to this last definition, known as institutional persistence. Finally, in other instances persistence is used to refer to students who have either completed their higher education or are still enrolled, regardless of the institution. In these cases, a more specific term is system persistence. The present study does not track the students who leave the university or transfer in, therefore the terms retention and institutional persistence are more appropriate and are used interchangeably.

Persistence research points toward the importance of both social and psychological factors in students' decisions to complete a degree. A host of factors have been shown to have a direct or indirect influence on persistence, including parental educational levels and income, college selection, high school grades, academic ability, motivation, and study skills. Persistence is often measured in terms of educational achievement (e.g., GPA, number of courses taken, grades achieved) and degree attainment (e.g., degree completion, and time elapsed to earn a degree: persistence after the first year, 3rd year or 6th year in school).

Academic Success and/or *Academic Achievement* are also considered outcome measures of student retention (Astin, 1993; Ybarra, 2000). Traditionally, these are measured through student's grade point average (GPA), graduating with honors and other ways of measuring student performance. College grades represent an important index of student accomplishment in college and embody a number of casual factors including academic ability, motivation, perseverance and study skills (Astin, 1993; Pascarella & Terenzini, 1991).

Student Departure, as opposed to student "dropout", is a less controversial term to describe what happens when a student leaves an institution of higher education. From a broad perspective, all students who leave are dropouts, meaning they withdraw and their absence creates a vacancy. However, Tinto (1993) notes three reasons why labeling all departures as dropouts can be misleading. First, it obscures the differing forms of departure. Without distinguishing the alternative reasons why students leave, institutions may apply a single policy action to treat the problem. Different forms of departure are *stop outs* (students who leave school temporarily) *transfers* (students who leave one institution and enroll in another one), *voluntary withdrawals* (students who leave college, but have good academic standing) and *involuntary withdrawals* or academic dismissals. Second, using the term 'dropout' to describe all forms of departure may lead institutions to believe that the problem is within the institution and thus treat the problem only with institutional actions. Research has shown that students leave school for reasons other than those dependent on the institution (Bean & Metzner, 1985; Cabrera, Nora, & Castañeda, 1993; Tinto, 1993). Third, the term 'dropout' usually connotes individual failure, indicating that an individual can not measure up to the college demands. However, in many circumstances the institution and the environment fail to meet the needs of the students. The term chosen in this study to refer to students who are not longer enrolled is *stop-outs.* It is important to clarify that in the present study students who *transfer* out are part of the group labeled stop-outs. Yet, the differences between these two groups are theoretically acknowledged.

Theoretical Models of Persistence

Student Integration Model

What keeps students in school? To answer this question several theoretical models have been proposed, to explain how individual, environmental and institutional characteristics, and their interactions predispose students to stay in or leave college. A common assumption is that students who leave college (i.e. desert the institution) do so because of poor performance. Although academic failure is a common reason for departure, "only 15 to 25 percent of all institutional departures arise because of academic failure" (Tinto, 1993 p. 81-82).

According to Tinto's (1993) model, the decision to leave college is determined by the match between students and institutions. Tinto uses Durkheim's theoretical explanation of "egotistical suicide" to explain school departure. According to the sociologist Durkheim, egotistical suicide arises when individuals are unable to become integrated into their social communities in two respects, socially (i.e., the day-to day interactions) and intellectually (i.e., the sharing of values). This notion applied to higher education emphasizes the importance of social and academic integration of the student into the institution. The more a student's collegiate experience helps him/her integrate socially and intellectually into the life of the institution, the more likely the student is to commit to the institution and to the goal of education. This commitment enhances persistence. Poorly integrated students are more likely to leave college before finishing their degree (Tinto, 1993).

Tinto's model takes in to account the influence of the attributes, background, skills and pre-college experiences each individual brings (family background, gender, SES, high school grades, motivation and interests, etc.) as well as the characteristics of the institution where the student attends (small, large, private, public, etc.). Students' academic integration results from his/her interactions with the formal and informal academic system. Indicators are academic performance and the quality of interactions with faculty, staff and other students. Social integration results from students' interactions with the formal and informal social system. Indicators are students' participation in extracurricular activities and peer-group interactions. In his model, Tinto also takes in to account the student's community external to the college (e.g. family, work and community). This leaves open the possibility that experiences outside the institution can have a positive or negative influence on a student's decision to stay in college. For example, in a study by Cabrera, Stampen and Hansen (1990), it was found that ability to pay moderates the effect of educational aspirations by affecting goal and institutional commitment. These finding indicate that an external variable like ability to pay can indirectly influence college persistence. However, the student integration model emphasizes the explanatory strength of understanding factors internal to the college. According to Tinto (1993), voluntary departure reflects more what goes on inside the institution than what goes on before entering college or what goes on outside college.

Input-Environment-Output (IEO) Model

The IEO model developed by Alexander Astin (1993) is a conceptual guide for studying student development in college. Understanding the impact that college has on students can help to achieve desired educational outcomes such as degree attainment/persistence (Astin, 1993). In Astin's model, the term *input* refers to the incoming characteristics of the students; *environment* refers to the many aspects of the campus environment (educational experiences, programs, policies, faculty, living arrangements, other students, etc.); and *outcomes* refer to the characteristics of the students following their encounters with the campus environment.

Similar to Tinto, Astin takes into account the characteristics students come with when they enter college, the environment to which they are exposed and the outcomes resulting from the college experience. This model is more linear, perhaps because it was not developed to understand persistence alone, but rather to understand the impact of various environmental conditions on students' growth or change over time in school. The key in this model is the diversity of students' conditions in their environment and the outcomes in their development. Astin's IEO model is often used by institutions to help relate a university's environment and student's outcomes.

Astin classifies outcomes using three dimensions: 1) type of outcome data, which can be affective (e.g., non-cognitive outcomes: values, attitudes, aspirations) or cognitive (e.g., intellective: knowledge, critical thinking, basic skills, academic achievement); 2) type of data, which refers to the manner in which outcome is actually measured: "psychological" data, relating to the internal states or traits of the individual; and "behavioral" data relating to directly observable activities; and 3) time, which involves the short-term and the long term effects of college. An example of a short-term effect of college could be dropping out of school. A long-term effect, which can only be observed after many years, would be having low income as a result of having left the educational system without a degree.

Astin's longitudinal research resulted in institutions paying more attention to recruitment. Institutions realized that students' varying characteristics as they entered school had different effect on outcomes. The finding that outcomes vary among different kinds of students also drew attention to research in persistence for specific populations (e.g., male/female; minority/non-minority)

Student Attrition Model

Bean's Student Attrition Model posits that students' beliefs about their experiences in school affect their intention to stay and subsequently their persistence (Bean 1980). Bean suggests that student attrition is similar to employee turnover in work organizations. Similar to turnover theory, he incorporated psychological variables such as satisfaction and institutional commitment, the organizations' characteristics, and the students' background to predict dropout. Bean expected organizational determinants to have a causal effect on employee satisfaction with the resulting effect being the decision to leave. Similar to psychological models, intent to leave as an attitude affects the behavior to dropout. In his model, intent to leave is the best predictor of dropping out. Bean's theory added to the previous two models by introducing the effect that a psychological variable, like satisfaction, can have in the decision to leave. Similar to Astin's and Tinto's models, Bean also emphasized the role of the environment and the effect this has on the student 'fit' in the institution. Not 'fitting' results in dissatisfaction, which triggers the intent to leave.

Integrated Model of Student Retention

Cabrera, Castañeda, Nora, and Hengstler (1992), combined Tinto's integration model with Bean's attrition model for a better understanding of what affects students' decisions to stay in college. The convergence of these two theories resulted in the finding that both models "are correct in presuming that college persistence is the product of a complex set of interactions among personal and institutional factors as well as in presuming that intent to persist is the outcome of the successful match between the student and the institution." (p.158).

The convergence study validated several hypotheses of Tinto's model (student integration), making this model a strong predictive model. This study did not validate as many hypotheses in Bean's model (student attrition model) as it did for Tinto's model. However, Bean's model accounted for more variance in intent to persist (60.3% vs 36%) and persistence (44% vs 38%). The authors attribute this finding to the significant effect of external factors in the form of parental encouragement and support from friends and finances, which validated Bean's previous findings. Cabrera's et al. (1992) convergence of two models resulted in a more comprehensive understanding of persistence. Nevertheless, the design of a new model that integrates the "leading factors in each theory may contribute to explain this process better (Cabrera et al., 1992, p.160).

In a follow up study, Cabrera, Nora, and Castañeda (1993), re-tested the above findings by "simultaneously testing all non-overlapping propositions underlying both conceptual frameworks" (p.124). Once again, the conclusion of their study confirmed the importance of student commitment on persistence and the influence that environmental factors exert in the socialization and academic experiences of the students.

Cabrera's early findings with regards to the importance of environmental factors has influenced his current research on external variables that characterize the student population before they enter college (e.g., high school ability and preparation, access and information about college cost, and the actual graduation application process to college) and their effect on persistence (Cabrera & La Nasa, 2000c). The review of this literature is used further in this study to define the at-risk populations.

Student Faculty Interaction Model

Pascarella and Terenzini (1980) were some of the first researchers to test Tinto's model. Using an instrument that was developed based on Tinto's constructs of academic and social integration, they found that goal commitment and institutional commitment are consistent predictors of persistence. They also found particularly strong contributions of student-faculty relationships to college persistence. This was measured by looking at interactions with faculty and faculty concern for student development. Students who scored high in these scales were more likely to be in school by the end of their freshman year.

Pascarella and Terenzini's studies confirmed that Tinto's model was a conceptually useful framework for the understanding of student attrition. Their findings contributed to broadening the perspective of Tinto's model. Quality and frequency of student-faculty informal contacts is viewed as an important measure of the students' institutional integration. Their research also pointed out differences in college experiences between men and women. However, they did not specify which features of college life lead to attrition among genders. In their conclusions, they urged researchers to consider a more comprehensive understanding of the process of student attrition (Balistreri-Clarke, 1996).

Student Social and Academic Integration Community Model

In 1998, Tinto wrote: "one thing we know about persistence is that involvement matters." (Tinto, 1998; p.168). All models of persistence support the assertion that academically and socially involved individuals who interact with their peers and faculty are more likely to persist. Tinto's latest model is a call for reform in higher education that aims at converting institutions into learning communities. This means promoting student involvement in and outside the classroom, organizing curriculum based on connected learning and organizing classroom experiences around collaborative learning. He promotes block scheduling of classes and course continuity. The former allows students to stay together in a class. The later relates curriculum themes from one course level to the next.

Tinto (1998) points out some important benefits of learning communities. First, students tend to form their own supportive peer groups. Sharing a curriculum encourages them to spend more time together. Second, he argues that students become more actively involved in classroom learning, which contributes to a third benefit: enhanced quality of learning. By learning together, students get to share processes for understanding knowledge. Therefore learning communities help bridge students' academic and social life, which helps them become more integrated in both aspects. Finally, the nature of learning communities also encourages students' services, staff and faculty to work more closely together.

Tinto also advocates the reorganization of the Freshman Year as a stand-alone academic and administrative unit. He proposes creation of a type of organization within an organization, with its own faculty, academic organization and pedagogical orientation. Tinto believes that this model allows faculty and staff to gear their services towards students in the freshman year without the pressure of other institutional responsibilities. Consistent with his community model for learning, the curriculum would emphasize interdisciplinary teams and collaborative learning.

Finally, he recommends reorganizing faculty in learning communities as well. Faculty would become connected learners by seeking collaboration of other faculty from different disciplines and departments, developing supportive peer groups with other faculty, and involving student affairs personnel. The expectation is that a connected organization and a connected teaching body will facilitate student integration and thus promote persistence.

Summary of Theoretical Models of Persistence

There are similarities among the models of persistence. All of the models include background variables, individual abilities and goals as important determinants of persistence in higher education. In addition, to some extent they all consider students' interactions or involvement with the academic and social aspects of the college to be important.

Vincent Tinto was one of the first researchers to explain the problem of college departure (1975). According to his model, student departure is a consequence of the interaction between the individual student and the college or institution. Commitment to the goal of education and commitment to the institution are shaped by the social and academic 'match' or 'fit' between the student and the institution. Pascarella and Terenzini's (1980) research confirmed that Tinto's model was a conceptually useful framework for the understanding of student attrition. In testing Tinto's model, they found that quality and frequency of student-faculty informal contacts were important to the students' institutional integration.

Bean (1980) proposed a model that emphasized the student's intention to leave or stay. In his model, behavioral intentions are shaped by attitudes and vice versa. Environmental, personal and organizational variables are said to have an effect on attitude and intent. Cabrera, et al. (1992) brought convergence of Tinto's and Bean's models and concluded that the two complemented each other. They found that environmental variables also exert an influence in the social and academic life of the students.

Alexander Astin, proposed a model that relates student recruitment to retention. His IEO model consists of assessing outcome variables (e.g. retention) based on the students' incoming characteristics (i.e., input variables) and how are they are affected by their experience (environmental variables) in college. Finally, Tinto's latest model of learning communities seeks to transform the experiences of the students by integrating academic assistance into the curriculum so that the students get academic support while progressing toward a degree (Tinto, 2002).

The most important theoretical models of persistence have been reviewed. Models are important because they orient the direction of inquiry. This permits systematic evaluation and interventions that target problem areas. Theoretical models have made feasible the generation of testable hypotheses and predictive equations, which have been the focus of much of the research over the past decade.

PART II: RESEARCH EVIDENCE

Traditional Predictors of College Persistence

Theoretical models of persistence have helped to effectively organize factors previously found to be related to retention and attrition into three broad interactive categories:

(1) The individual;

(2) The academic system of the university;

(3) The social system inside and outside the university.

Each of these categories has subcategories that involve, among others, the following variables:

(1) The individual: family background (e.g., socioeconomic status, parental education, family income, quality of family relationships), individual personal attributes, psychological variables, pre-college schooling (e.g., high school GPA, ACT, SAT), and the effect of these variables on grade performance (GPA), especially in the first year of college.

(2) The academic system: curriculum, classroom environment and institutional organizations.

(3) The social system: peer-group interaction and faculty interactions.

Traditionally, studies have focused on finding ways to measure the constructs proposed in the theoretical models of persistence. Researchers approach this from different perspectives, psychological, sociological, environmental, as well as organizational. Tinto's model is one that has been widely tested (Pascarella & Terenzini, 1991) and validated across institutions and populations (Braxton, Sullivan, & Johnson, 1997; Cabrera et al., 1993).

Academic Integration is a core construct in Tinto's interactionalist theory, and many studies support its influence on persistence. This construct is usually measured by assessing student satisfaction or 'fit' with the academic aspect of the institution, which often involves combining two or more perspectives (e.g., the individual within a social and academic system.)

Braxton and Lien (2000) assessed the extent to which there was empirical support for the influence of academic integration on subsequent institutional commitment and on student departure decisions. Their assessment included several studies that used multi-institutional and single institutional samples. Their findings indicate that multi-institutional studies provide a strong empirical backing for the positive effect of academic integration on persistence and subsequent institutional commitment. On the other hand, single-institutional studies provide weak empirical support. Tinto's model was developed to understand the longitudinal process of student departure within a particular institution, and it cannot be considered a systems model to explain departure. Therefore, the authors suggest finding different ways of measuring academic integration.

However, academic integration remains a viable construct because it is a strong predictor of persistence in single institutional tests that involve commuter universities (Braxton, Johh & Lien, 2000). The lack of social interactions necessary for social integration in non-residential institutions leaves academic integration as the only way students can become connected to the university. Braxton and Lien's assessment of the construct suggests that academic integration should be measured and tested in different ways. A viable way to tackle this problem could be to introduce non-traditional forms of measuring and testing academic integration.

In another study, Braxton et al. (1997) appraised Tinto's theory of college departure by reviewing all the studies that tested or extended Tinto's theoretical model. The authors wanted to assess the extent to which there was empirical evidence supporting the 15 testable propositions in the model. Two primary propositions were strongly supported by both multi-institutional and single institutional tests: initial institutional commitment affects subsequent levels of institutional commitment; and initial level of commitment to the goal of graduation from college affects the subsequent levels on commitment to the goal of college graduation.

Braxton et al. (1997) concluded that multi-institutional assessment provided robust support for two primary propositions: 1) student entry characteristics affect the level of initial commitment to the goal of graduation from college; and 2) the greater the level of subsequent commitment to the goal of college graduation, the greater the likelihood of student persistence in college. Single-institutional tests strongly supported two primary propositions: 1) the greater the level of social integration, the greater the level of subsequent commitment to the institution; 2) the greater the level of subsequent commitment to the institution, the greater the likelihood of student persistence in college.

In efforts to expand the understanding of persistence, researchers have examined environmental, social, and psychological variables in conjunction with Tinto's model. For example, Brower (1992) combined concepts of cognitive social psychology with Tinto's constructs to explore how students can affect their environment and how this affects persistence. According to Brower, students' perception of their environment is affected by their college goals, plans, values and expectations (i.e., life tasks), and these in turn affect their academic and social integration. Brower argues that students' commitment to different life task domains will facilitate or hinder their integration into the academic and social life of the institution. "Research has identified up to seven life task domains that are important to college life: academic achievement, social interaction, future goal development, autonomy, identity formation, time management, and physical maintenance/well-being." (p. 446; Brower, 1992).

Using a sample of 311 freshman students at the University of Wisconsin-Madison, Brower used linear regression to compare constructs in Tinto's model (background characteristics, initial commitment, academic integration, social integration and later commitment) to the life task persistence model. The dependent variable was persistence, measured as the number of semesters students remained enrolled in the university. The results of the study indicated that by adding variables of life task predominance to Tinto's model, the ability to predict persistence (as defined in his study) improves significantly. Nevertheless, much of the variance in persistence remains unexplained and therefore the author suggested that research be done to explore differences in how students shape their environment.

In summary, "Tinto's theoretical perspective possesses logical internal consistency, it lacks, in the aggregate, empirical internal consistency" (Braxton, et al. 1997; p. 156). Thus Braxton et al. (1997) suggest revising Tinto's theory, integrating it with psychological, social, organizational and especially environmental perspectives or creating new theory with greater internal validity. What it is clear about the revision of major constructs and the variables used to predict persistence, is that the research needs to expand its definitions, the variables used and the different effects they exert on the outcome variable.

In various theoretical models, certain variables have been found to be key in affecting persistence. Alexander Astin and his research team have identified several variables that are associated or have an effect on student persistence. Through the Cooperative Institutional Research Program (CIRP), they have collected data since 1966. Table 1 is a summary of the results of a longitudinal study that assessed the effect of environmental characteristics, and experiences on students' outcomes. The results are based on a nationwide college freshman sample of nearly 25,000 students from 217 institutions. Multiple regression was applied to predict different student outcomes (e.g., whether or not students finish school, attain a degree, or are retained in the institution where they started).

Table 1 shows some relevant variables from Astin's study. Among the predictor variables were SES, parent's education, academic majors, and academic records, shown to the left of Table 1. The second column (from left to right) shows the variables that have a negative direct effect on outcomes such as retention and bachelor's degree completion. The third column shows variables that are positively or negatively correlated with the outcome variables presented in the fourth column (for more details see Astin, 1993)

Table 1 Effect of Environmental Characteristics or Experiences on Students' Outcomes

Variables with positive direct effect on students' outcomes	**Variables with negative direct effect on students' outcomes**	**Variables that present (+) or (-) associations with students' outcomes**	**Two types of Student Outcomes Retention/Persistence**
		Environmental Variables	
Attending a private university			Staying in school and continuing to graduate school
	Lack of student community		Persistence to degree completion in four year
	4 Year public university		Degree attainment
	Majoring in Engineering		Bachelor's degree completion in four year
	Majoring in health professions		Bachelor's degree completion in four year
		(-) Majoring in Engineering	Retention
		(-) Attending a large institution	Retention
Majoring in Physical sciences			Bachelor's degree completion in four year
		(+) Enrollment in honors or advance placement courses	Degree attainment Retention
		(-) Receiving tutoring	Bachelor's degree completion in four year
		(+) Receiving vocational or career counseling	Retention
		(-) Receiving personal or Psychological counseling	Retention
		Economic Variables	
Financial aid (based on need)			Bachelor's degree completion in four year
Parental income			Degree attainment
		Academic Variables	
		(+) Hours spent studying	Retention of student in school
College internship program			Bachelor's degree completion in four year
		(+) Working on an independent research project	Degree attainment
Giving class presentations			Degree attainment
		(+) Taking essay exams	Bachelor's degree completion in four year

Table 1. (Continuation)

Variables with positive direct effect on students' outcomes	**Variables with negative direct effect on students' outcomes**	**Variables that present (+) or (-) associations with students' outcomes**	**Two types of Student Outcomes Retention/Persistence**
Background variables			
Mother's education			Degree attainment
Father's education			Degree attainment
Socio-Economic Status			Bachelor's degree completion in four years
Involvement Variables			
"to become more knowledgeable about things that interest me"			Attending College
Student-student and faculty student interaction			Retention
Hours per week spent socializing with friends			Retention
Talking with faculty outside the class			Retention
Being a guest in a professor's home			Retention
		(-) Working off campus at a part-time job	Retention
		(-) Working full time as a student	Retention
		(-) Number of science courses	Retention
		(-) Number of math courses	Retention
		(-) hours per week spent reading for pleasure	Retention

* From Astin (1993) What Matter's in College.

Table 1 shows undergraduate major as a variable that affects persistence. Seymour and Hewitt (1997) have done extensive research on factors that lead to switching majors. Their focus is on the move from Science, Math, Engineering, and Technology (SMET) majors to non-SMET majors, and the effect of switching on Bachelor's degree completion. According to a 1993 CIRP report cited by Seymour and Hewitt (1997), the relative loss of students from SMET majors between the freshman and senior year is 40%. Persistence among SMET majors remains below the average of other disciplines, and these rates are lower for women and students of color (Seymour, 2001).

Seymour and Hewitt (1997) looked at various factors that have contributed to high rates of switching, especially among women and students of color. Factors related to school curriculum, advising or high school preparation were identified. These included: (1) lack/loss of interest in science, mathematics and engineering due to the poor teaching experience created by the weed-out system; 2) inadequate advising or help with academic problems; 3) curriculum overload, fast pace overwhelming; 4) discouraged/lost confidence due to low grades in early years; 5) discouraged by professors in their field to follow a teaching career path in SMET; 6) inadequate high school preparation in subjects/study skills; 7) unexpected length of SMET degree; and 8) morale undermined by competitive SMET culture (i.e., faculty and male peers).

Seymour's ethnographic research does not follow the traditional theoretical framework or quantitative approach used to explain retention in higher education, yet her findings support the importance of academic and social fit in understanding persistence. Her research focuses on the reasons for switching to non-SMET fields and suggests that initial discouragement impacts students' decisions to pursue their college degrees.

Newman (1997), in her dissertation on retention of students of color at UW-Madison, looked at factors that affect academic success in the College of Engineering (COE). She defined academic success as being retained to complete the degree in the COE. Her study consisted of identifying factors that encourage and discourage students of underrepresented populations to complete their degree.

Using theories in ergonomics (human factors and systems engineering, person-environment fit, and community-environment fit) and social theories (cultural deprivation, cultural differences and cultural fit), she developed a model to identify students' needs to be academically successful in the COE. Her model identified 5 factors that affect persistence, which she categorized as follows: 1) personal characteristics; 2) resources; 3) ownership; 4) exposure to new opportunities; and 5) social support services. These constituted what she called the institutional climate and the absence or presence of these posed the "barriers" or forms of "access" to academic success. To understand students' perspectives on these barriers or opportunities for success, she did group interviews. One group was formed of current engineering students, another consisted of students who had changed from the COE to another field and a third group was composed of high school students applying to college.

Newman's findings corroborated her theoretical model, but also pointed to other complementary factors. To the personal characteristics, students added the importance of pre-college preparation, career purpose and reasons for choosing the discipline. Under the broad institutional climate, students pointed out the importance of classroom environment, course scheduling and interaction with faculty. She added a "feedback" factor, which was needed between the institution and the students. Feedback includes improving the communication with administrators, faculty and other peers. She also identified the need for a “rescue loop”, which refers to the university support programs for students (e.g., mentoring, improved advising, connection to the university environment, role models, programs to improve academic achievement). Finally, she pointed to the need to establish connections between high schools and colleges, the idea being that students needed to become more familiar with the university environment, the qualifications needed to enter school and the meaning and purpose of academic research.

In the next page, Table 2 shows a summary of factors that affect academic persistence. The table presents a summary of all the factors that have been named in studies and known to have a positive or negative effect on persistence. The column to the right contains only some of the authors or studies where the selected factors where found.

Table 2 A Summary of Factors that Affect Academic Persistence

Categories	**Factors**	**Authors/Study**
The Individual Students' personal characteristics	Demographics: High school academic preparation Freshman GPA Race/Ethnicity Gender Religious background Family Background Socio-Economic Status/parents' income Number of Family members in college Parents' level of education History of college education in family/siblings Motivators -Intrinsic: self-efficacy, goal to get meaningful job, intention to attain degree, interest in field/academics -Extrinsic: Scholarship, family expectations, job expectations, purpose for acquiring a degree	Tinto, V (1993) Newman, L (2000) Doolittle, M (1996) Leppel, K (2002) Astin, A (1993) Astin, A (1971) Brower, A (1992) Mashburn, (2000)
Academic System Inside the Institution - Curriculum -Availability and access to resources	Social Support Services Academic advising Mentoring programs Role models Career development resources Social activities Financial Aid Exposure to new opportunities Internships Study abroad programs Guided research opportunities -papers/publications -conferences Curriculum Cooperative learning Academic Major	Newman, L (2000) Swail, W (2000) Astin, A (1993) Cabrera, Stampen, & Hansen (1990) Seymour and Hewitt (1997)
Institutional fit	Level of ownership Institutional familiarity Feeling of belonging Peer support Academic and social experiences	Newman, L (2000) Doolittle, M (1996) Harper, S (2003)
Academic System Outside the University Institutional characteristics	Type of school (Public vs Private) 2year institutions vs. 4year institutions Large vs Small	Astin, A (1993) Astin, A (1971) Tinto, V (1993)

Summary and Discussion of the Literature on Persistence

College retention is contingent on the individual, academics and the social system surrounding the individual and university. Researchers have used psychological, sociological, environmental and organizational perspectives to understand what factors better explain attrition. Tinto's academic and institutional integration constructs have been one of the most researched. Braxton et al. (1997) report moderate to strong support for the effect of academic and institutional integration on student persistence. However, a limitation of Tinto's theory is the emphasis on person-institution fit. Cabrera et al. (1993) state that "a major gap in Tinto's theory and allied research has been the role of external factors in shaping perceptions, commitments and preferences" (p.124).

Other variables related to individual characteristics, the environment, academics and the institution have received consistent support in explaining college persistence. Tables 1 and 2 present a summary of variables that have been found to affect student persistence, obtained from quantitative and qualitative research.

After doing an extensive review of the literature on persistence, two gaps are evident. First, there is a need to broaden the existing theoretical models used to explain persistence and to increase their validity or explicative scope (Braxton, et al., 1997). Although the literature on persistence is rich and has contributed to the development of successful retention programs, it has yet to provide a thorough understanding of the reasons students leave college. Greater understanding of the "departure puzzle" is now possible using advanced technology and improved institutional data that facilitate the analysis of student records in greater detail (e.g., achievement patterns before entering college and during their career, educational milestones, course history, and demographics). Historical student records can help to understand the role of other variables in explaining persistence. For example, a profile can be generated from historical data for those students who have completed a degree despite being at-risk. These profiles can be used to answer questions such as: Is educational resilience different among students from different colleges? Does choosing a major early in the career affect the success of students in college? and Is 'risk' a simple construct or a multidimensional one? A profile of successful students alone may not be enough to explain the complexity of the departure puzzle, but it can facilitate the early identification of common characteristics and help determine levels of risk.

The second gap in the literature on persistence is that few researchers have employed longitudinal designs with cohorts of students who have persisted through successive years of college. With the exception of the research done by Astin (1975) and studies that have analyzed large National surveys and databases, such as the National Educational Longitudinal Study of 1988 (U.S. Department of Education, National Center for Education Statistics (NCES), 2001c) the Beginning Postsecondary Students [BPS] (U.S. Department of Education, National Center for Education Statistics (NCES), 2000b, 2001d)Baccalaureate and Beyond Longitudinal Study [B&B] (U.S. Department of Education, National Center for Education Statistics (NCES), 2003), and the High School and Beyond Study [HS&B] (U.S. Department of Education, National Center for Education Statistics (NCES), 2001b). In contrast, most researchers have focused on specific schools using institutional data, and have investigated persistence only over a short period of time (Bean, 1980; Braxton et al., 1997; Green, 1998; Nonis & Wright, 2003; Pascarella & Terenzini, 1980).

Nevertheless, universities continue to struggle to retain their students, and although some students leave one institution to attend another, approximately 20% of college students drop out entirely from the postsecondary system (Choy, 2002).

At-Risk Students

What does being an "at-risk" student mean? The terms at-risk, disadvantaged or underprivileged generally refer to students who are equal to others in terms of their capabilities, but whose academic background, prior performance and/or socioeconomic characteristics put them at a high probability for academic failure and limited educational attainment (Yeh, 2002; O'Connor, 2002). At-risk students have different profiles. Most current research states that these students are more likely to be from low-income families, belong to a minority group, be from their families' first-generation in college, be academically under-prepared, or have several or all of these characteristics (Cabrera & La Nasa, 2000.; Choy 2002; Gandara, 1995; Gloria, A, Kurpius, Hamilton, & Wilson, 1999; Horn, L. J. & Chen, 1998; LePage-Lees, 1997; Riehl, 1994). Risk factors are defined as the characteristics that students have that put them at risk of academic failure and school departure.

In the context of this dissertation, the term at-risk implies that the group of students with one or more risk characteristics (i.e., lower scores, target minority, financial need, etc.) as described in the literature has a greater expected probability of experiencing delays in graduation or not graduating at all.

A Review of the College Choice Literature

The college choice literature is important for the understanding of risk and persistence because it incorporates the conditions necessary for students to consider, enroll, and attend postsecondary education. College choice models have helped to identify pre-college factors and background variables that put students at risk of leaving school before finishing. Pre-college characteristics affect how well students adjust to their institutions (Cabrera & La Nasa, 2000a; Hossler et al., 1999; Tinto, 1993)

Models have been proposed to explain how traditional-aged students go about planning their educational career. According to Hossler, Schmit and Vesper (1999), there are economic models that assume that a student's decision process to attend college is based on a cost-benefit rationale. Students are thought to weigh the perceived benefits of school and work against the cost and make decisions to choose and attend an institution based on this analysis. In a typical model, school-related factors (e.g., college size) and student characteristics (e.g., ability) are associated with income, wealth and commodities. The outcome of these relations constitutes the college choice decision.

Other models, such as the status-attainment models, derive from sociological theories. These models attempt to identify the variables and circumstances that may narrow students' possibilities of attending college. These models assume that behavioral variables such as academic performance and family educational interest interact with background variables such as socioeconomic status and parents' occupation to determine educational attainment (Hossler et al., 1999).

A more concise model proposed by Hossler and Gallagher (cited by Hosller et. al, 1999; pg 149) combines sociological and economic theories. The model theorizes that there are three stages in the college choice process: predisposition, search and choice. The predisposition stage refers to the educational or labor plans the student has after high school. The choice to attend college is thought to occur during this stage. Background variables are correlated with the predisposition stage and are cumulative in terms of their effect on the whole choice process. The search stage involves the discovery and evaluation of possible institutions. This stage is easily affected by social and educational conditions and is considered the most important and the most open to intervention. The choice stage refers to the actual selection of the schools. This stage assumes that students have made application decisions consistent with the search stage.

Hossler, Schmit and Vesper (1999) applied the three-stage model of college choice in a longitudinal study, surveying a sample of 4,923 ninth-grade students and their parents. They found that parents play the most significant role in shaping the educational aspirations of their children. More important than parents' education and income was parent encouragement and support (i.e., parents talking to their children about educational expectations, hopes and dreams). The second best predictor of postsecondary aspirations was student achievement (i.e., reported GPA). The higher the students' grades the more likely the students were to be consistent with their ninth-grade plans to attend college. The authors explain that college attendance is influenced by achievement because good grades may augment the expectations parents have about their children, which results in more encouragement and support.

Although the college choice theory points to specific variables that are modifiable such as students' information about colleges, financial aid availability, parents' knowledge of college requirements, and parental support and encouragement, there are predetermined societal factors that make the path to college more difficult for some students to negotiate. These factors include parental income, ethnicity, parents' education, and high school preparation (Hossler, et. al., 1999; Cabrera & La Nasa, 2000).

Income affects all three stages of the college choice process (i.e., predisposition, search and choice). Families of lower SES are more likely to rely on financial aid and less likely to save for the total cost of college education (Cabrera & La Nasa, 2000c). Not saving for college may indicate lower *predisposition* to attend college.

During the *search* stage, students visit colleges, learn about institutions and talk to their parents and peers about places to go for their postsecondary education. Students of lower SES have less chances of making visits to colleges and have parents who are less knowledgeable about what schools their children can attend (Cabrera & La Nasa, 2000c).

Finally the *choice* stage involving applying and enrolling in college is likewise different for students with a lower SES. Although the availability of federal financial assistance facilitates the enrollment of low-income students in college, it does so for institutions that are less expensive and for community colleges.

As logical as it might sound, in order to attain a postsecondary degree, it is first necessary to apply and enroll in college. Cabrera and La Nasa (2000a) state that "enrolling in a four year college requires the completion of at least three critical tasks: meeting minimal college qualifications, graduating from high school and actually applying to a 4 year college." (p. 31). These 3 tasks are more difficult to attain for students of lower SES. In their study, Cabrera and La Nasa analyzed the college choice processes of 1,000 eighth-grade students from a 1988 cohort, and found that by the time these student's were in high school

71% of the lowest-SES students did not have the academic qualifications necessary to enroll in college.

Moreover, low SES is associated with poor academic preparation and low parental education (Cabrera & La Nasa, 2000b). Academic preparation and more importantly parents' education, are two variables that are strongly related to college enrollment and persistence (Choy 2002; U.S. Department of Education, 2001a)

Based on data from the National Education Longitudinal Study of 1988, third follow up [NELS: 1988-94] and cited in (U.S. Department of Education, 2001a) graduates whose parents did not go to college were much less likely than those with parents who had some education, to complete various steps to college enrollment (e.g., aspirations to college, academic preparation, taking SAT and ACT, applying to a 4-year college and enrolling in a 4-year college). For example, 46% of graduates whose parents had a high school education or less aspired to a Bachelor's degree, compared to 86% whose parents held a Bachelor's degree or higher. Having completed all other steps in the pipeline, only 21% of these students were likely to enroll in a 4-year institution, compared to 65% of those who had parents with a college degree.

Parents' education is also associated with remaining enrolled in college. After 3 years in college, first generation students with Bachelor's degree goals are less likely to remain enrolled in 4-year institutions than their peers whose parents have Bachelor's or advanced degrees (U.S. Department of Education, 2001a, p.25).

Additionally, first generation college students tend to be less prepared academically. In a study of 2,190 incoming freshman at Indiana State University, Riehl (1994) found that first generation students had lower SAT scores and high school grade point averages than their peers who came from homes where at least one parent had attended college.

In summary, the literature on college choice helps to identify relevant pre-admission factors, and serves to illuminate their impact on the students' paths to college. Although the focus of this study is to understand what factors help at-risk students succeed at the college level, it is important to understand what incoming variables affect future performance and put students at risk. This, along with identifying what populations are at-risk, is the contribution of the college choice literature.

Summary and Discussion: Understanding who are at-risk students

The higher education literature is saturated with studies that focus on determining what variables affect dropout and retention rates of the students enrolled. Students have been found to be at risk of leaving school for a variety of reasons that may apply to those labeled disadvantaged, those who belong to the mainstream and even the academically elite students. The great majority of these studies focus on identifying those students who, given their background, educational patterns and decisions, are more likely to fail somewhere in their educational path. An alternative approach is to focus on understanding what helps some at-risk students succeed.

Few researchers have addressed the issue of at-risk students who have succeeded academically, and most of the existing research is qualitative. Padilla (1999), for example, is a pioneer in developing student retention strategies based on the experience of successful students. Utilizing a qualitative approach and a framework derived from systems thinking (i.e., input-process-output), he demonstrates how successful students' knowledge and behavior can be identified systematically and used to help similar groups succeed in college.

The work of O'Connor (2002) on generational opportunities and constraints, LePage-Lees (1997) on patterns of achievement and intellectual development, and Zeldin (2000) on self-efficacy and beliefs towards science and math, has been more specific in exploring the path of women (mostly minority) who beat the odds and become successful in their academic careers.

Closer to the quantitative nature of the present study is the work by Horn (1997), who analyzed the data on 1992 high school graduates from the National Educational Longitudinal Study of 1988 [Survey: NELS:88/94]. This survey began with eighth graders in 1988 and followed them every two years. She studied at-risk high school students who passed through the pipeline to college. Pipeline is a term used to define the stages students go through in their path to college enrollment. At-risk students were defined as 1992 high school graduates who had risk characteristics (e.g., being from a single parent household, having an older sibling who dropped out of high school, changing schools more than once, having C or lower grades between 6^{th} and 8^{th} grades, repeating an early grade and being of low SES) that increased their chances of dropping out.

Horn found that students with zero risk factors were more likely to navigate the college pipeline to enrollment. Among the at-risk students, those who succeeded had all completed at least one advanced math course and reported receiving help from school personnel in filling out their applications. Successful students were also more likely to have participated in extracurricular activities, have parents who discussed school-related matters and have more friends who were also planning to attend college.

It is important to note that Horn's study looked at high school students and defined success as enrollment in college. Similar analyses could be conducted to determine why some at-risk students succeed in college. Little is known about the cumulative effect on college graduation of having several risk characteristics, and what protective processes aid students in the course of their careers. Finally, more needs to be understood about how different risk factors interplay with the protective processes and allow individual success.

PART III: EDUCATIONAL RESILIENCE

The Concept of "Resilience"

The term resilience is used in the fields of mental health, education, and psychology, with varying definitions. The Resilience Net web site (*Resilience NET*, 2003) offers a general definition of resilience as the "human capacity and ability to face, overcome, be strengthened by, and even be transformed by experiences of adversity" (Resilience Net section, Para. 2).

The fundamental interpretation of this concept consists of believing that people vary in their capacity to successfully change despite life's negative circumstances. The term resilience is almost always used to describe a set of qualities that foster a process of successful adaptation, development and transformation despite risk and adversity (Cyrulnik, 2000; Masten, 1994).

Rutter (1993) points out that resilience does not reside in the avoidance of risk experiences. Resilience results from facing a problem at a time and in a way that the individual can successfully cope in the face of the negative challenges posed. Rutter uses an analogy from the medical field: "for example, immunity to infections whether natural or therapeutically induced through immunization derives from controlled exposure to the relevant pathogen and not through its avoidance." (Rutter, 1993, p. 627). This clarification is important for understanding the possible outcomes derived from resilience research. It is often thought that resilience means that only a few individuals have unique capabilities to adapt and transform which help them succeed. In the context of a school, for example, if we know that only a few of the young adults who are academically under-prepared will succeed, then a reasonable policy would be to raise standards so that the school has fewer academically under-prepared students. However, taking into account Rutter's clarification, a sound option would be to aid those few under-prepared students and offer protective measures so that prior academic preparation does not pose such a negative force against them graduating.

Resilience: A trait within the individual or a product of his /her environment?

Whether the capacity to adapt and transform depends on the individual, the context surrounding the individual, or both is something that is viewed differently by different experts in the field. In addition, 'barriers' or 'stressors,' and what constitutes successful adaptation have been defined and measured in many different ways. The following paragraphs present the different theories and research on how resilience develops, especially as it relates to education.

Resilience is sometimes understood as a psychological trait. A definition offered by Linquanti (1992) (cited in Finley, 1999), states that resilience describes the quality in children who have been exposed to significant stress and adversity in their lives but do not succumb to the school failure, substance abuse, and mental health problems that have been predicted for them. This definition suggests that resilience is a trait of a child's personality and temperament. Garmezy (1983) (cited in Winfield, 1994) has identified individual characteristics among high school students who come from poor areas but have succeeded in school despite their disadvantaged circumstances. These personal characteristics include a wide range of social skills, positive peer and adult interactions, and teachers' ratings of the youths' low degrees of defensiveness and aggressiveness and high degrees of cooperation, participation and emotional stability. Other traits identified by Garmezy are a positive sense of self and power, an internal locus of control, empathy, a sense of humor, and intelligence (measured by IQ).

Even researchers who assert that the environment plays an important role in building resilience have found personal characteristics unique to resilient individuals (LePage-Lees, 1997). Some of these traits are summarized as having high emotional intelligence, which includes attributes like persistence, self awareness, self-motivation, self-esteem, even temper, openness to experience and adaptability (Wang, Haertel, & Walberg, 1994).

Although many researchers place resilience within the individual, others see the influence of the environment, which presents conditions where individuals can display their resilient behavior. Bernard and Marshall (1997) describe resilience as a process that starts within the individual and propagates to his/her environment. The authors clarify that the environment is critical to permit a healthy human development. By this, they mean an environment that permits individuals to meet their basic human needs "for caring and connectedness, for respect, challenge, and structure, and for meaningful involvement, belonging and power." The environment offers a system of support that fosters resilience.

Rhodes and Hoey (1994) state that the systematic occurrence of favorable factors in an individual's life is what leads to the development of the individual's success. In their study, academic achievement was a factor used to define success. Similarly, Bernard and Marshall (1997) assert that individuals beat the odds not just because of luck, but because of the combination of individual and environmental factors that produce a positive outcome. According to these authors, factors influencing an individual's success include parental support and encouragement, encouraging student-teacher relationships, neighborhood influence and personal assertion.

A supportive environment may explain why some children with what seems like an inappropriate background nevertheless show resilience and succeed. In a more recent analysis of the resilience literature, Masten (2001) concluded that resilience is made of ordinary individual qualities and resources representing basic protective systems in their families and communities, rather than anything extraordinary. In other words, institutions like schools, families and the community can provide a support system and play an important role in the process of creating an environment that can offer what Bernard (1991) refers to as 'protective factors.'

Protective Factors

Protective factors are situations or characteristics of the environment that can alter or even reverse expected negative outcomes, and enable individuals to elude life stressors and manifest resilience despite the risk situation they may be encountering (Benard, 1995). In simpler terms, protective factors enable the individual to counter the effects of risk factors or stressors (Miller & MacIntosh, 1999). Protective factors can be thought of as the building blocks of resilience. Research consistently reveals the presence of individual and environmental factors associated with resilience (Benard, 1991; Rutter, 1993; Wang et al., 1994). Among the most widely cited protective factors are: caring relationships or the presence of someone who conveys understanding, compassion and respect; high expectation messages, and opportunities for participation/contribution.

Similarly, Doll and Lyon (1998) who have done an extensive review of the research on risk and resilience, state that three factors that protect against risk and therefore promote resilience are: (a) close affectionate relationships with at least one parent or caregiver, effective parenting (characterized by warmth, structure, and high expectations); (b) access to consistent, warm care giving and positive adult models in a variety of extra-familial contexts; and (c) strong connections with other pro-social organizations or institutions, including schools.

A caring person can be someone from the family, and a caring environment can be that of an educational setting, a teacher or a program. LePage-Lees (1997), in a qualitative study exploring the experience of women who achieved highly in academics despite growing up disadvantaged, found that all women in her study wanted to be known by their teachers and be cared for by them. Ultimately the women in this study recalled having a teacher or a mentor adult who had made a difference in their lives, because they cared and got to know their individual circumstances.

Benard (1991) defines positive expectations as high, clearly articulated expectations, and the purposeful support necessary to meet those expectations. Expectations can come from the individual, their parents or their teachers. The resilience research suggests that all of these three sources are important in serving as protective factors (Wang & Gordon, 1994).

Zeldin and Pajares (2000) explain that according to the social-cognitive theory, “people are more likely to perform tasks that they believe they are capable of accomplishing and are less likely to engage in tasks in which they feel less competent.” (p.216). In other words, when students are convinced that they are influential in their performance they work harder to overcome difficulties. This is to say that they expect highly of themselves or have self-efficacy. Parents, teachers and the students’ environment are sources from which students build their self-efficacy.

Opportunities for meaningful participation, means providing individuals with responsibilities and opportunities to be involved with others (Benard, 1991). In an educational setting this would be providing students with opportunities to connect with their school (Wang et al., 1998a). LePage-Lees (1997), in a qualitative study of resilient women, found that their activities in school and closeness to the academic work and teachers kept their mind off of problems at home.

Education research has demonstrated that features of homes, schools and communities promote the development of beliefs and behaviors that can result in positive educational outcomes and educational resilience (Wang et al., 1994). It is therefore important to identify and implement those features that act as protective factors, in order to design environments that promote resilience. Less has been researched concerning environments, such as school settings for older children and adults. For example, an important question that arises from promoting protective factors in higher education could be: what in a school environment signals individual characteristics that can serve as protective factors? To what extent is being malleable reflected in the student’s or the school's ability to realize that the best decision for persistence is to change academic programs?

Summary and Discussion of the Literature Review on Risk and Resilience

Another way of broadening the explicative scope of existing theoretical models in persistence is to introduce new paradigms such as the concept of educational resilience. The absence in the literature of studies that institute a resilience paradigm in higher education reveals a need for a study of this kind. Interestingly, the correlates of resilience appear to be similar variables involved in the educational persistence of all individuals, but the approach to studying them is different. Another difference is that when at-risk populations are studied, the persistence studies tend to focus on one specific risk characteristic (e.g., studies about the persistence of first generation college students or minorities or low SES students). Studies on risk and resilience look at what the environment had to offer for individuals and this in itself is what determines the level of risk.

Rodriguez (2003), without relating her work to theories of resilience, did a qualitative study with first generation college students and tried to isolate the factors that aided in their academic success and determine whether these can be replicated in other people's lives. She indicated that the usual success promoting factors such as financial aid availability, parental support, academic preparedness and college counseling played an important role in the lives of these students. But she also identified very similar factors to those proposed by researchers studying resilience. She came up with her own labels and called them "positive naming", which refers to having someone who would help observe a positive way of life or opportunity. Her participants experienced this when someone who cared about them or knew them well helped them develop to their potential. Another of Rodriguez' factors is called "the special status phenomenon", which refers to participants being singled out in a positive way. Other themes in her participants' explanations were a "sense of belonging" and "high expectations." Rodriguez, states that the participants in her study experienced these influences by chance, but she believes that these can be replicated to influence students' outcomes in college.

An objective of the present study is to follow a similar approach to that used by Rogriguez (2003) but using quantitative data. Perhaps certain academic actions that occur while the student is in school have a protective effect and influence their retention.

Why does the resilience concept matter in higher education and how would a study from this perspective be different than any study in the persistence literature?

It is not the purpose of the study to confront two constructs like resilience and persistence and conclude that one approach is better than the other. Rather, the purpose is to use the lenses of the research in resilience to view the problem of institutional persistence. What differentiates the literature on resilience from that on persistence is the focus on success. The persistence research is good at pointing out why individuals fail, but less is said about the successful steps or processes needed for a successful outcome to occur and how to determine these.

The challenge is to identify the source of resilience in the processes that occur in an institution. What the literature does tell us is what is required from the individual, as well as the general characteristics of the environment, but what we do not know are the specific steps that need to be taken to build these supportive processes. Where resilience is low, the question is how to develop resilience or replicate the environment that the successful students experienced. In the higher educational environment, what are those factors that become protective of the individuals and prevail over risk? Any insight gained by addressing these questions will help to design interventions at the college level. Most importantly, this study addresses these questions using data that is available to the institution, such as student record data.

CHAPTER 3: METHODS

This chapter describes the methods used to model and predict the transition probabilities of undergraduate students at each stage of the college pipeline. There are three possible educational outcomes: to advance from one academic level to the next one, including graduation; to be enrolled but not advance to the next academic level; and to stopout. The chapter is organized as follows: the delimitation of the study; the goal and objectives of the research; and the research design, including an introduction of the measures chosen for the study, and a description of the site, the data source, the population, the sampling technique, measures, research hypotheses and data analysis procedure

Goal and Objectives

The main goal is to develop and test progression model of institutional persistence in undergraduate higher education, which will facilitate understanding educational resilience, especially for those considered to be at risk of failure.

The enabling objectives are:

1. To predict and quantify the probabilities of retention outcomes over the course of a four-year college career.
2. To investigate whether certain pre-admission variables that, according to the literature, are factors that put students at risk of not succeeding in college do in fact decrease the predicted retention probabilities.

3. To investigate whether there are post-admission variables that serve as protective factors and contribute to the retention of students, and how this determines and at risk population. Some academic actions are expected to serve as protective factors. The protective factors considered are limited by availability in the student record data and are derived from the theory of resilience and persistence. The findings from this objective are expected to help identify what factors contribute to the institutional persistence and how these may affect students traditionally not retained due to their incoming at-risk characteristics.

4. To develop and parameterize a Markov simulation model using the retention data views (RDV) technology. The simulation model is expected to help faculty and administrators study institutional persistence of undergraduate students.

The RDV are an On-Line Analytical Processing[2] (OLAP) system, newly developed at the host university, which contains archival student record data dating from 1985. The product of this model is a composite of predictor equations that provide estimated outcome probabilities for students who are enrolled and advance from one academic level to the next including graduation, who are no longer enrolled (i.e., stop-out), and who are matriculated but have not advanced to the next educational level. The variables used in the model have been drawn from a historical student record database maintained by a large mid western university. This historical database includes fields such as major, program, award, GPA, placement and standardized test score, and more detailed data such as 3rd, 5th, and 7th semester GPA, High school rank, 1st term credits, academic group and major, academic actions, and degree completion.

---

[2] OLAP: "On-Line Analytical Processing" a database technology that has been optimized for querying and reporting instead of processing transactions. OLAP data is organized hierarchically and stored in cubes instead of tables. The OLAP organization allows fast analysis of multidimensional data.

Delimitations

The following delimitations are applicable to this study:

1. The study sample is restricted to those students who began their postsecondary education during the 1996-97, 1997-98 and 1998-99 academic years. These cohorts were chosen to allow enough time for students to graduate. Three cohorts are also used to increase the N of the study.
2. The study involves only undergraduate students who enrolled for the first time in college (i.e., coming directly from high school or First Year Students, FYR). A few students included in the model entered with a sophomore and junior standing because they came in with a higher number of credits (typically through having completed Advanced Placement courses and examinations while in high school). However, these students were still considered FYR.
3. The variables utilized in this study are limited to those available through the University Retention Views.

Research Design

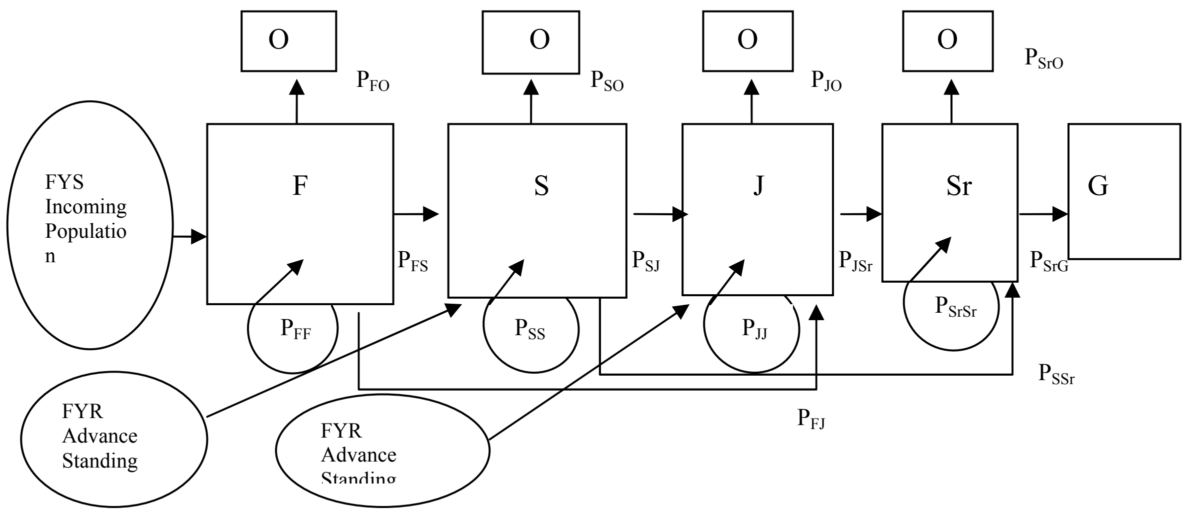
Most studies on retention and attrition are quantitative in design and have viewed undergraduate student retention from a 'deficiency' perspective, concentrating on what students are lacking or what institutional variables cause students to leave (Bean & Metzner, 1985; Cabrera et al., 1992; Tinto, 1993). A newer trend views the problem from the perspective of successful students and uses mostly qualitative methods (Ford-Edwards, 2002; Newman, 1997; Padilla, 1999). Although qualitative methods provide the opportunity for rich description through narrative and the clarification of individual experiences (Harper, 2003), they do not allow for inferential hypothesis testing and comparative analyses. Also, qualitative studies have questionable generalizability, make no claim as to sampling from a larger population, and often focus on unique groups.

The main strengths of the present study concerning the data are the size of the sample, relative consistency of data definitions, and availability of RDV database technology. Also, once the model is implemented it can be used to assess the effect of university policy changes on retention. In that data from a single institution has been utilized, the results will not generalize to the same extent had a national dataset been the source. However, student record data drawn from a single university permits greater control over contextual variables.

Quantitative approach to undertake the theoretical model proposed in this study

In quantitative terms, this study uses student record data to investigate the separate and combined effects of pre-admission and post-admission variables on institutional persistence. Profiles of selected pre-admission variables are used to identify students at risk of not being retained. In this study, the term “risk variables” refers to the pre-admission variables. Some pre-selected, post-admission variables have the potential of protecting students from attrition. Thus, the term “protective factors” refers to these post-admission variables.

The Markov model used in this study views university undergraduate education as a pipeline with a collection of flow regulators. Each flow rate is a function of a number of causal variables, some of which are subject to administrative control. Figure 1 provides a visual representation of the model:



Legend for model:

FYR=First year students	F = Freshman status	J = Junior status
	S = Sophomore status	Sr = Senior status

O = Stopout (transfer, temporarily not enrolled, or dropout)
G = Graduated
P_{FO} = probability of stopping in freshman year
P_{FF} = probability of remaining in freshman status
P_{FS} = probability of advancing to sophomore status
P_{FJ} = probability of advancing to junior status
P_{SO} = probability of stopping in sophomore year
P_{SS} = probability of remaining in sophomore status
P_{SJ} = probability of advancing to junior status
P_{SSr} = probability of advancing to senior status
P_{JO} = probability of stopping in junior year
P_{JJ} = probability of remaining in junior status
P_{JSr} = probability of advancing to senior status
P_{SrO} = probability of stopping in senior year
P_{SrSr} = probability of remaining in senior status
P_{SrG} = probability of graduating

Figure 1 Pipeline of Institutional Persistence in Undergraduate Education

In the above model the transition is at academic level intervals. At the end of each academic level, undergraduate students are in one of three progressive states: 1) they have stopped out; 2) they are enrolled but have not advanced to the next academic level; 3) they are enrolled and have advanced to the next academic level including graduation (e.g., advanced from sophomore to junior or from senior to graduation). The last academic level, seniors, has an option of graduating instead of advancing to higher academic levels. As these categories are complete and mutually exclusive, the proportions in each category for each status level will sum to 1.0 and can be regarded as transition probabilities. The original freshman cohort will decline, as no provision is made for incoming transfers. In this manner, the model operationally defines institutional persistence, since only those undergraduate students who persist will ultimately advance their academic status. To adapt the model to the reality of the university being studied, students who enroll at an advanced academic level are allowed in the model (e.g., students who start as sophomores). Another group allowed in the model is students who move to the next academic level in less than a full academic year. These are students who complete enough credits to advance two academic levels in one school year.

The pipeline transition model serves to structure the research design. In the data analysis, the transition probabilities have been estimated by logistic regression, and the predictor variables are classified as pre-admission or post-admission variables. Pre-admission variables are those that characterize the student prior to university admission. They serve to determine which students were at-risk prior to starting their academic life. Post-admission variables are those that capture the fit between the student and the university. They are indicators of the intellectual development, personal characteristics, educational achievement, and academic actions that take place while the student is in school. The variables selected from this category are those expected to serve as protective factors against risk factors. Since the transition options (remain, advance, or stop) are ordinal categories, the use of logistic regression with categorical dependent data is appropriate (Menard, 1995). The predictor variables are a combination of continuous and categorical variables.

The predicted transition probabilities can be arranged into a matrix that reflects the best statistical estimate of the probabilities of moving from one educational state to another. An exemplary state transition matrix for a cohort group is shown below:

	F	**S**	**J**	**Sr**	**O**	**G**
F	P_{FF}	P_{FS}	P_{FJ}	0	P_{FO}	0
S	0	P_{SS}	P_{SJ}	P_{SSr}	P_{SO}	0
J	0	0	P_{JJ}	P_{JS}	P_{JO}	0
Sr	0	0	0	P_{SrSr}	P_{SrO}	P_{SrG}
O	0	0	0	0	1	0
G	0	0	0	0	0	1

The 0s in the above matrix indicate that transition from the row state to the column state in one year is impossible. The 1s indicate absorbing states, where once in no escape is possible. Once a student leaves the university (0 state), there is no return as incoming transfers are not allowed in the model. Similarly, once a student enters the graduated state, reentry into the model as an undergraduate is impossible. As expected, there are 14 non-zero probabilities to be estimated for each cohort. The probability that a student moves from state i to state j in N steps (years) is $\mathbf{P}^N$, where $\mathbf{P}$ is the 6 x 6 matrix of estimated transition probabilities. Thus, given that the transition probabilities have been estimated, raising the matrix to the N^{th} power will give the probability of attaining the j state in N steps given that it starts in state i at time 0. For example, if state i is entering as a freshman at time 0, then the entry in cell (1,6) of the $\mathbf{P}^N$ matrix is the probability of graduating in N years.

Using a Markov model, the proposed study computes prediction equations for each of the four academic levels in the pipeline model, with a total of fourteen transition probabilities. These probabilities are computed using the derived logistic equations. Finally, the predicted probability of graduating in four years or more years is computed using the transition probabilities and recorded for each subject observation that persisted four, five and six years.

The results of the present study and future applications of this model are expected to have immense implications for decision support. For example, if it is found that the protective factors have a positive effect on persistence policy changes can be implemented to alter the effect upon transition probabilities. Also, the model will yield a clearer picture of what constitutes ‘risk’ in this particular university and what route entering students need to follow to have a smoother path towards successfully finishing their degree.

Population

The target population for this study is comprised of the aggregate of about 16,500 records of entering undergraduate students, beginning with the fall term of 1996, 1997 and 1998 through Spring of 2004. For the purpose of this study, entering classes of students are conceptualized as demographic cohorts. A cohort is understood to be individuals entering a system at the same time. Each cohort has roughly 5,500 new freshmen.

Sampling

Three cohorts have been chosen in order to increase total sample size. Each entering cohort has been selected for this research to allow an adequate amount of time for graduation (i.e., up to 6 years). The analysis of the data focuses on all undergraduate freshman students who enrolled in college for the first time (i.e., coming directly from high school or First Year Students [FYR]). The analysis includes only enrolled students, and excludes all foreign students. Foreign students would augment the number of missing values for many variables such as minority status, ACT scores, high school rank, etc. Matriculated students who take a leave of absence during a four-year period have been coded stop-outs. The analysis is limited to incoming first year college students who enter as freshmen or in advanced standing as sophomores, and to those who enter in the summer or fall semester of the academic year. The purpose of this selection is to have a more homogeneous group, and is deemed appropriate considering the demographic characteristics of the school's entering class (see statistics below).

The institution from which the sample has been drawn is located in a midsized (pop. 200,000) capital of a Midwestern state in the US, with a predominantly white student body. In the fall of 2002, the total undergraduate population enrolled was 28,677, of whom 24,653 were Caucasian, 2,768 were ethnic minorities, 347 unknown and 909 were foreign. The total freshman population enrolled was 5,479, 95% of whom were new freshmen. Of the 5,180 new freshmen, 99% were full-time (12 semester credit hours or higher). The ethnic breakdown for the total freshmen enrolled in the fall of 2002 was, for the minorities: 164 African-American, 321 Asian-American, 29 Native American, 160 Hispanic; and for the non-minorities: 127 International, 4,625 Caucasian and 53 Unknown (Office of the Registrar, 2002).

Description of Retention Data Views (RDV) or student record data source

Technological advances now permit universities to keep student records in computer databases. However, these data management systems do not lend themselves readily to computation for the purpose of statistical and trend analysis. Rather, these production systems were designed for the consistent entry/updating of individual records. To address this limitation of production systems, information managers have constructed “data warehouses.” A data warehouse extracts vast amounts of related data from a production system and organizes it into sets of logically related files for end-user analysis. Authorized faculty, staff, administrators, researchers, and associated personnel with a business need to know can access this collection of integrated institutional data. These data views can be accessed through desktop software such as Microsoft Access-Office and BRIO and other commercial packages that have Open Data Base Connectivity (ODBC).

During the last several years, a campus-wide committee has constructed a set of ‘data views’ referred to as the "Retention Data Views" (RDV) in the university data warehouse. The RDV, designed to provide users with longitudinal student records, contain the vast majority of student record data relevant to retention issues (e.g., every major, program, award, course, grade, placement and standardized test). A university staff member has facilitated access to the retention data views by extracting and integrating the RDV onto a type of data retrieval/analysis system referred to as an On-Line Analytical Processing (OLAP) system, which uses Microsoft Excel software. It is important to understand that these Excel files contain the same retention views available through the data warehouse, yet they are organized into Excel Pivot Tables to facilitate analysis. These Excel tables are often referred to as the Retention Data Cubes. The cubes allow analysis of data in a way not possible when directly querying the RDV's and can be directly read by any statistical package. The data is organized so that there is one row per student/ID and several columns with the fields, instead of several rows per student ID.

Measures

This section includes the process used to select the variables, a description of all the variables used in this study and an explanation of why they are used in the particular grouping where they have been classified.

Variable Selection Process

For the present study, the researcher was given access to administrative student record data pertinent to retention and longitudinal studies. These consist of a set of at least 300 fields such as student history, enrollment, courses, grades, program and plan history. Based on the research questions proposed in this study it was clear that some of the fields available would entail student information from prior to enrolling to the university (e.g., student background, test scores, high school information) and student information recorded once the student is already enrolled (e.g., GPA, course and credit history). A logical main split to select the data were then determined by pre-admission variables, a set of variables that refer to the characteristics or experiences that students have prior to entering school; and post-admission variables, a set of variables that represent school actions, accomplishments and requirements which may influence degree completion.

This major split of the variables is important because it will allow users of the proposed model to clearly differentiate how much of the effect on institutional persistence is dependent on the individual characteristics of the students before they are admitted. As explained by Astin (1993), the assessment of how outcomes are affected by environments is biased unless one measures and controls for as many input characteristics as possible. This split also adds significance to the post-admission variables because they occur once the student is in the university system, thus they can be affected by policy changes.

The next major split of the variables involved determining which variables were those that would put students at risk of not being retained and which served as protective factors against risk. The independent variables where then categorized as variables that put students "at risk" of failure and variables that "protect" against failure.

For the purpose of this study, risk factors are defined as those independent variables that are expected to have a positive beta weight in the logistic regression model of retention (i.e., predict stop out) Protective factors are those independent variables that are positively related to the prediction of retention. In the logistic regression model they are expected to carry a negative beta weight (i.e., predict persistence).

Based on the literature and voluntary informal interviews with students, local admissions counselors, administrators and staff working in university based programs, the following constructs were chosen to help determine which variables identify risk among students and which may serve as a protective factors.

To select the risk variables, the following constructs were considered: minority status (Astin, 1993; Gloria et al., 1999; Newman, 1997; Tierney, 1992; Tinto, 1993; U.S. Department of Education National Center for Education Statistics, 2001; Ybarra, 2000), socioeconomic status (Astin, 1993; Bean & Metzner, 1985; Choy et al., 2000; Tinto, 1993), high school preparation and first semester GPA (Astin, 1993; Mc Grath & Braunstein, 1997). As has been presented in the literature review, students who belong to a minority group, students of low socio-economic status and students with less than successful high school records are more likely to drop out of school. Also, students with the weakest academic preparation, often measured through performance on standardized tests, take longer than 4 years to complete their bachelor's degree (Astin, Tsui, & Avalos, 1996).

To select the protective factors, the following constructs were considered: Presence of critical courses and performance in these courses, change in programs and academic majors, participation in school programs, college GPA (Tinto, 1993) and yearly change in GPA (Fletcher, Halpin, & Halpin, 1999).

The idea that some courses may help keep students in school or hinder their overall performance, especially courses required in the first two years of college, has been suggested by a few researchers (Levin & Wyckoff, 1990; Zhu, 2002). In summary, and taking into account school administrators' perceptions, the grades and the timing of certain basic courses may have an effect on persistence. In an informal conversation with a student, when asked about critical core courses he named a math course and said: "you know you are in trouble if you can't handle math 221 or if you take too many of the high credit courses right away."

The timing of college major selection and whether the student changes the initial program have also been considered in the literature as factors affecting persistence. Hagedorn and colleagues (2001) found that early identification of a college major was salient in predicting retention among African-American male students attending a community college. In another study, Fredda (2000) found that among freshman students advancing to the next term, there were no statistically significant differences in the drop out rates between those who changed their academic major and those who did not. There are not enough studies that can help clarify the effects on persistence of the timing of selecting or changing a major. However, research does suggest that there are some correlations, and thus this construct seems appropriate to investigate.

In informal interviews, university administrators also identified changing majors and changing colleges as important milestones suspected of having an effect on retention. The theory of resilience suggests that academically successful students have a sense of future (Wang, Haertel, & Walberg, 1997), so changing a major early may correlate with persistence. On the other hand, leaving a flexible schedule and switching to another program may indicate the student's intent to adapt and find a better fit, which may contribute to persistence.

Theoretically, it is known that the student's academic integration and social relations influence attitudes such as college completion goals, which in turn affect retention (Cabrera et al., 1992; Pascarella & Terenzini, 1991; Tinto, 1993). Based on this knowledge, special programs have been created in universities with the goal of helping students adapt to the academic and social environment of the university. The central dimension of the support that students receive through these programs is often based on staff mentorship, tutoring, academic advising and providing a caring environment and opportunity for social integration in to the campus life. Based on the persistence and resilience literature, which indicates that opportunity for participation and a caring environment are protective factors, it was decided to analyze the influence that two different programs offered to targeted groups and underrepresented groups had on institutional persistence.

Finally, academic achievement, often measured using college GPA and most precisely first year GPA, is highly associated with continuing in college (Bean, & Metzner, 1985; Fredda, 2000; Mc Grath & Braunstein, 1997; Nora, Cabrera, Hagedorn, & Pascarella, 1996; Pascarella & Terenzini, 1980; Tinto, 1993). In a study by McGrath & Braunstein (1997), first semester GPA was the best predictor of persistence to the sophomore year. Nora and colleagues (1996), studying differential impacts of academic and social experiences among minorities and non-minorities and males and females, found that college GPA was a strong predictor of persistence for all groups.

Independent Variables

The variables used in this research have been judged to manifest or indicate constructs previously identified in the literature, with the purpose of testing and assessing the validity of models of student persistence (Astin, 1975; Bean, 1980; Bernal et al., 2000; Cabrera & La Nasa, 2000a; Cabrera et al., 1990; Horn, L. J. & Chen, 1998; Pascarella & Terenzini, 1980; Pascarella & Terenzini, 1991; Riehl, 1994; Tinto, 1993). However, these variables are unique to the host university and have not been used by other studies reported in the literature. The following is a description of the variables chosen for the study:

Minority Status is a variable that indicates whether a student belongs to a targeted minority group. This category refers to minority students who are underrepresented at the university and are targeted for campus diversity initiatives and local university programs. A student is an underrepresented minority if he or she belongs to a minority group that does not compare in terms of retention, degree attainment, or number of students to the population in general. In this particular university, a student who is a US citizen, permanent resident or refugee, and who is Black, Hispanic, American Indian or Southeast Asian (of Laotian, Cambodian, or Vietnamese heritage and came to the US prior to the end of the Vietnam war or is descended from such a person) is considered a targeted minority. The variable is operationalized as Not Targeted Minority=0; Targeted Minority =1

Socioeconomic Status (SES): The variable chosen identifies students of low socioeconomic status. These are students receiving financial aid in the form of Federal Pell Grants. Pell Grants are funded by the federal government and range from about $400 to $4,050 per academic year. Pell grant recipients are considered by the federal government and the university to be the students most in need of financial aid, due to their parents' low income. The variables used to assess any information related to student SES come from the financial aid office. The financial aid data were operationalized as those students who either received a Pell grant or did not receive a Pell Grant: Not Grant recipient= 0; Grant recipient=1.

High School Academic Preparation: uses the following pre-admission variables, high school rank (H%) and test scores (ACT Composite). The ACT Composite is the average of the four ACT test scores (i.e., English, Math, Reading and Science), rounded to the nearest whole number. ACT scores have been chosen over SAT scores because the former are more populated in the database. A concordance table will be use to translate SAT scores to ACT for students who only have the SAT math and verbal scores. A table and details about the use of concordance tables can be found in College Board, 2004. In previous studies (Astin, 1975; Pascarella & Terenzini, 1980), high class rank and high ACT and SAT scores have been shown to positively affect student institutional persistence.

Other background variables such as Gender, Age and Race are also included in the analysis. In previous studies, these variables have been found to be important in explaining the effects on persistence (Doolittle, 1996; Leppel, 2002; Tinto, 1993). For example, these studies indicate that students who enter at an older age are more likely to drop out. The argument behind this is that older students are more likely to be married, have children, live off campus, and be working, so they have many other obligations in addition to going to school. However, in this study the differences in age are expected to be minimal because of the range of ages of the participants, which has been limited to FYR students.

The race variable, which is different than minority status, will be used to specify the race group when necessary to better interpret results (e.g., differences among Hispanic, Whites, African –American, Native, etc). Gender has also been considered, to broaden the understanding of the data. Studies have reported that females have lower attrition rates (Astin, 1993; Tinto, 1993). However, results must be carefully interpreted because important variables such as marital status and number of children are not available in the database. Leppel (2002) found lower persistence rates among women, because in her study women were more likely to be older, married and have children, which had negative effects on persistence.

Further investigation is needed into what constitute protective factors for at-risk individuals in higher education and how these can be operationalized. Similarly to the way risk factors were selected (i.e., based on the literature and voluntary informal interviews with students, local admissions counselors, administrator and staff working in university based programs), the following variables have potential as protective factors against risk factors and academic failure.

Critical Courses: This construct has been measured using derived variables based on five critical courses, using the course history and drop history file in the data views. Most of these courses are taken in the freshman or sophomore years. The five courses were also selected based on the large number of students who take these courses. At the local university these are considered critical because many programs require taking them and many students do fail the courses. Among some administrators it is believed that failure to succeed in these courses may trigger problems with retention. Five courses seemed an appropriate number that would allow a variety of courses without an overwhelming amount of data. Delaying taking these courses or making sure that few of these are taken at a time may constitute a protective factor. This data were coded: 0=course not taken, 1= took course in X academic year 1. This type of coding was done for each of the five courses selected: Chemistry 103, Math 112, Math 221, Psychology 201, Economics 101. Letter grades are also available for these courses.

Academic disciplines: The variable measuring this construct was derived from the academic plan history file in the data views and operationalized as change of major. However, this variable only applies to those students who started with one major and eventually added or changed to another major. This measure excludes those who started with more than one major. Coded (0=did not change major, 1= changed major; 9= more than one major (eg., dual majors)

Academic groups: This variable indicates change in student academic group (e.g., Education to Engineering, etc). Table 3 shows the 8 colleges and schools available to undergraduate students. Coded (0=did not change academic group, 1= changed academic groups)

Table 3 Academic groups available to undergraduate students

SCHOOLS and COLLEGES	**Abbreviations**
Agricultural and Life Sciences	ALS
School of Business	BUS
School of Education	EDU
College Engineering	EGR
School of Human Ecology	HEC
College of Letters& Sciences	L&S
School of Medicine	MED
School of Nursing	NUR
School of Pharmacy	PHM

College Grades: To measure this construct, the cumulative grade point average reported in the data records for each student is used. GPA is one of the best predictors of retention (Astin, 1993; Bean, 1980; Bean, J & Metzner, 1985). In this particular study, first semester GPA is a post-admission variable that is critical to determine retention. It is important to note that the GPA used is an accurate number as reported to the school administration, not one reported by the student. In many studies of retention, the variable GPA is obtained from student surveys, which makes it a less accurate measure. The GPA is recorded from fall to fall term, following the students' academic level. A field was available for academic level status (e.g., freshman, sophomore, junior, senior, not enrolled). Current policies require that undergraduate students have a total of 120 credits to graduate. They need 24 credits to become a sophomore; 54 credits to become a junior, and 86 credits to become a senior. Full-time students have 12 credits or more, and half-time students have fewer than 12 credits. Table 4 below shows all the predictive variables used in the study and their categories.

Table 4 Independent variables

Variable Category	**Underlying Construct**	**Data Base Label**	**Variable Type**
Demographic	Gender	Gender	categorical
Demographic	Age	Age at entrance	numeric
Demographic	Race	Ethnic Group	categorical
Pre-admission	Minority status	Targeted minority	categorical
Pre-admission	Socio Economic status	Pell Indicator	categorical
Pre-admission	Socio Economic status	AidTrmCount	numeric
Pre-admission	High school preparation	HS Percentile	numeric
Pre-admission	High school preparation	ACTcomposite	numeric
Post-admission	College grades	First semester GPA	numeric
Post-admission	College grades	CumGPA	numeric
Post-admission	Critical course 1	Chemistry 103 count	categorical
Post-admission	Critical course 1 grade	Grade Chem103	categorical
Post-admission	Critical course 2	Math 112m count	categorical
Post-admission	Critical course 2 grade	Grade Math112	categorical
Post-admission	Critical course 3	Math 221 count	categorical
Post-admission	Critical course 3 grade	Grade Math 221	categorical
Post-admission	Critical course 4	Psychology 201 count	categorical
Post-admission	Critical course 4 grade	Grade Psy 201	categorical
Post-admission	Critical course 5	Economics 101 count	categorical
Post-admission	Critical course 5 grade	Grade Econ 101	categorical
Post-admission	Academic discipline	ChangMaj	categorical
Post-admission	Academic Program	ChgAcdGrpInd	categorical

Administrative variables

The following variables have been named *administrative variables* because they are used for administrative and classification purposes. In this study they serve to identify individual rows and grouping categories (e.g., undergraduate, cohort, academic year, academic level and ID).

Table 5 Administrative Variables

Variable Category	Data Base Label	Example
Cohort definition	Type of career	UNDGR
Cohort definition	Description of academic term	Fall 1996-97
Cohort definition	Description of level	FYR
Cohort definition	Description of Academic level	10=freshman, 20=Sophomore.
Cohort definition	Relative Term	Spring. Summer, fall.
Row ID	Assigned random code	#00000 ID replaced with random code.
Milestone	Degree Conferred Indicator	Y = Student graduated with a degree; otherwise not populated
Milestone	Years to degree	The number of elapsed calendar years between the student's first term and their degree conferral date
Milestone	Discrete years to Degree.	The value of Years_To_Degree rounded down to the nearest 0.5 years.
Milestone	Degree completion Term	The term in which the student completed their first degree

The administrative variables (See Table 5 above) years to degree, degree conferred and degree completion term will help determine if and when a student graduates.

Dependent Variable

The dependent variable in this study is institutional persistence. This variable is treated as ordinal. The assumption of order is based on academic persistence and the logical argument that moving on to the next academic year manifests more persistence than stopping out. The persistence time frame in this study will be continuous enrollment for up to four years including graduation.

At each of Freshman, Sophomore, Junior and Senior status, the dependent variable institutional retention is as follows: 1= no longer enrolled/ no data available. These are the stopouts. 2= enrolled and not advancing to the next academic level; 3= enrolled and advancing to the next academic level, including graduation. Whenever graduation is present there may be a fourth level. This study treats transfers and stopouts (students no longer enrolled) as dropouts, because major persistence models are based on institutional dropout (Bean, 1980; Bean & Metzner, 1985; Cabrera et al., 1992; Tinto, 1975, 1993). For more details on the dependent variable and the coding for each particular academic status, see the fourth paragraph of the data analysis section.

Data Analysis

This study applies a Markov model and logistic regression to predict probabilities of institutional persistence up to graduation for each of four academic statuses (freshman, sophomore, junior and senior). The model carries pre- and post- admission variables drawn from a student database. An integrated database has been created for the analysis by pooling from the university's retention data views the variables selected for the study.

Microsoft Access and Excel were used to query the main database and arrange it so that each record corresponds to a unique student ID. Each record has a set of data fields, which for this study have been organized according to the independent and dependent variables selected. Administrative fields are used to identify and assign students to the various analysis categories (e.g., corresponding cohort, risk categorization, graduation, etc). Independent variable fields, which are arranged by academic status, will include all predictive variables shown in Table 4. Dependent variable fields include fourteen dummy variables corresponding to each of the three outcome states at each of four academic levels. There are 14 of these dummy dependent variables because the sophomore status and the junior status have an additional level for those students who advance to the next academic level in less than one academic year (e.g., Freshman to Junior and Sophomore to Senior).

To facilitate the manipulation of the data and to predict the transition probabilities at each of four academic levels, the data has been divided into four data sets, each with a dependent variable (i.e., Data Set I, II, III, IV). The classification criteria refer to the academic classification of the students. In this way Data Set I comprises all freshman students, data set II all sophomores, data set III all juniors and data set IV all seniors. Data set I start with all FYR undergraduate students from the three cohorts who matriculated in the summer and fall. Data Set II allows FYR incoming students classified as sophomores; Data set III comprises all juniors, and Data Set IV comprises all students who became seniors after their second, third and fourth years in school.

The dependent variable indicates the academic status of a student after completing one academic year. For example: Data Set I has a dependent variable (DVI) with the following values: 1=Freshmen whom are no longer enrolled at the beginning of the second academic year (these are the stopouts); 2=Freshmen who stayed freshman at the beginning of the second academic year; 3= Freshmen who became Sophomores at the beginning of the second academic year; 4= Freshmen who became Juniors at the beginning of the second academic year. Predicted probabilities have been computed and recorded for each student.

In order to perform the statistical analyses, the data cubes were imported into a research database created using SAS/STAT (Version 8) and descriptive statistics were computed. Descriptive statistics include frequency distributions for discrete demographic variables and means for continuous variables. The PROC CONTENTS, PROC FREQ, and PROC UNIVARIATE procedures were used to obtain the descriptive data. Predictions of the outcome probabilities were made using logistic regression. The PROC LOGISTIC procedure with ordered responses was used because the transition Markov states can be considered ordinal with respect to institutional persistence.

Logistic regression is a form of multivariate analysis that was specifically designed for use with J categorical dependent measures. Logistic regression provides more flexibility and is more appropriate to use with nominal as well as ordinal categorical dependent variables than other multivariate techniques. The ordinal response allows prediction of the probability of falling into one category rather than the others. Unlike other types of multiple regression, logistic regression with ordinal response categories uses a score statistic which is asymptotically distributed as chi-square.

Based on the general ordinal response model of parallel slopes across j categories for *k* explanatory variables and *i*= 1,2,...,n individuals, the model is:

$$g\ (\Pr(Y \leq j \mid \mathbf{x}) = \alpha_j + \beta \cdot \mathbf{x_i}$$

where β is a kx1 vector of slope parameters, $\mathbf{x_i}$ is a 1xk vector of explanatory variable scores for the i^{th} individual, α_j is the intercept for the j^{th} response category, and the expression on the left-hand side is referred to as the link function.

For a three-category response model, the probability that the i^{th} observation has response j is given by:

$$\Pr(Y_i = 1 \mid \mathbf{x_i}) = F(\alpha_1 + \beta \cdot \mathbf{x_i})$$

$$\Pr(Y_i = 2 \mid \mathbf{x_i}) = F(\alpha_2 + \beta \cdot \mathbf{x_i})$$

$$\Pr(Y_i = 3 \mid \mathbf{x_i}) = 1 - F(\alpha_2 + \beta \cdot \mathbf{x_i})$$

Where F(x) is the cumulative logit function:

$$\frac{1}{1+e^{\left(\alpha_i+\sum\beta.\chi\right)}}$$

If there are more than three transition states j= 1,…,J +1, the probability that the ith observation has response j is given by:

$$\Pr(Y_i = 1 \mid \mathbf{x_i}) = F(\alpha_1 + \beta\cdot\mathbf{x_i}) \qquad j=1$$

$$\Pr(Y_i = j \mid \mathbf{x_i}) = F(\alpha_j + \beta\cdot\mathbf{x_i}) - F(\alpha_{j-1} + \beta\cdot\mathbf{x_i}) \qquad 1 < j \leq J$$

$$\Pr(Y_i = 3 \mid \mathbf{x_i}) = 1 - F(\alpha_2 + \beta\cdot\mathbf{x_i}) \qquad j = J + 1$$

Research hypotheses

1. *Pre-admission variable*s will predict outcome probabilities at a greater than chance level for students in each academic level.

1.1. <u>Several risks vs. specific risk hypothesis</u>: The predictability of pre-admission risk variables will diminish across academic level, with fewer risk variables contributing to prediction of retention outcomes at the junior and senior level.

2. *Post-admission variable*s will predict outcome probabilities at a greater than chance level for students in each academic level.

2.1. <u>The protective factors hypothesis</u>: The predictability of post-admission variables will increase across academic levels, with more post-admission variables exhibiting significant predictive contribution at junior and senior levels.

Hypothesis testing

Hypothesis 1: will be regarded as confirmed if the number of significant beta coefficients for pre-admission independent variables exceeds the number expected by chance at each academic level.

Hypothesis 1.1.: will be tested by plotting beta coefficients over the four academic levels for comparable pre-admission variables. The hypothesis will be regarded as confirmed if the number of significant variables exceeds that expected by chance.

Hypothesis 2: will be regarded as confirmed if the number of significant beta coefficients for post-admission independent variables exceeds the number expected by chance for each academic level.

Hypothesis 2.1: will be tested by plotting beta coefficients over the four academic levels for comparable post-admission variables. The hypothesis will be regarded as confirmed if the number of significant variables exceeds that expected by chance.

Informed Consent

The research proposal for this study was submitted to the institutional review board for research with human subject at North Carolina State University, and in April 29, 2004 it was approved as exempt from the policy as outlined in the Code of Federal Regulations (Exemption: 46.101.b.4). For NCSU projects, the Assurance Number is: FWA00003429; the IRB Number is: IRB00000330. No IRB renewal was necessary as long as no significant changes were made to the study based on the proposal.

CHAPTER 4: RESULTS

This chapter comprises the results from the data analysis, which are presented in the following order: First, data management details are described. This is important for the purpose of replication and understanding of data classification in accordance with the model. Second, the descriptive statistics for the entire sample are presented. This includes descriptive statistics for all pre- and post-admission variables, and a table reporting the descriptive data for the dependent variable, representing the transition probabilities from one academic level to the next. These transition probabilities illustrate the persistence patterns in the sample and are used as a benchmark for comparison with predicted probabilities resulting from hypothetical score profiles. Third, the results from the logistic regression test for each of the four data sets are shown. The results from this section and plots for the parameter estimates are used to answer the proposed hypotheses. Fourth, three data simulations with hypothetical score profiles for pre-admission and post-admission variables are presented. The simulation provides an example of college persistence probabilities given a set of possible combinations of student characteristics and policy options.

Data Management

As explained in the Methods section, the data for three cohorts 1996-97, 1997-98, and 1998-99 summer and fall, First Year Students (FYR) totaling 16,507 students were divided into four data sets. (i.e., Data Set I, II, III, IV). The classification criterion for the sets was Academic Level (e.g., Data I= all freshman; Data II= all sophomore). Data were recorded at the beginning of each academic year, every odd term (i.e., 3rd academic term, 5th term, 7th

term, to graduation). Following the stated logic and the predetermined rule that once a student stopped s/he could not enter into the model again, the data sets were formed (refer to Table 6).

Table 6 <u>Data organization summary</u>

Data I Freshman status =10	**Data II** Sophomore status=20	**Data III** Junior status=30	**Data IV** Senior status=40
Data I (n=15,682)	Data II a. (n=12,892)	Data III a. (n=10,677)	Data IV a. (8,788)
Matriculation=**10**	Matriculation=10 3rd term=**20**	Matriculation=10 3rd term=20 5th term=**30**	Matriculation=10 3rd term=20 5th term=30 7th term=**40**
	Data II b. (n=643)	Data III b. (n=475)	Data IV b. (n=364)
	Matriculation=**20**	Matriculation=20 3rd term=**30**	Matriculation=20 3rd term=30 5th term=**40**
		Data III c (n=881)	Data IV c (n=611)
		Matriculation=10 3rd term=**30**	Matriculation=10 3rd term=30 5th term=**40**
		Data III d. (n=20)	Data IV d (n=9)
		Matriculation=**30**	Matriculation=30 3rd term=**40**
			Data IV e. (n=20)
			Matriculation=20 3rd term=**40**
			Data IV f. (n=132)
			Matriculation=10 3rd term=20 5th term=**40**
Data appending			
Data I = Data I	Data II a + Data II b. = Data II ab	Data III a + Data III b + Data III c + Data III d =Data III abcd	Data IV a + Data IV b + Data IV c + Data IV d + Data IV e + Data IV f =Data IV abcdef
Test performed			
Logistic Regression PROC LOG (SAS version 8)	Logistic Regression PROC LOG (SAS version 8)	Logistic Regression PROC LOG (SAS version 8)	Logistic Regression PROC LOG (SAS version 8)
Simulation: SAS/IML program used to calculate predicted probabilities from a set of hypothetical values			

Table 6 above provides a visual representation of how the data were organized in to different datasets. So, for example: Data II (second column in the table) with students classified as sophomores were formed of two data subsets: Data II (a): contains students who matriculated as freshman, and at the beginning of their 3rd term had enough credits to be classified as sophomores. Data II (b) contains only students who had matriculated as sophomores. Notice in the table that the majority of the students are classified as freshman at entrance. A smaller number of students come with advance credits and are classified as sophomores, and even fewer enter as juniors.

Recoding variables and dealing with missing values

Fields in the original data source for which a 'blank' (the absence of data) had a value were recoded in SAS. For example: in the source data a [blank] in the field Target _Minority means that the student is not a target minority. Since in SAS missing character data are represented by blanks and missing numerical data are represented by a single [.] period, Target _Minority and Change _Major were recoded to [0], [1] (Delwiche & Slaughter, 2003).

Post-admission fields such as change of major and courses taken (e.g. Math 112) were recoded in each data sub-set and conditions assigned so that data corresponded to the correct relative term. For example: changing major during academic year 1 was matched with relative terms 1 and 2 (i.e., first and second semesters). A 1 was assigned when the student took the course in the corresponding level, and a 0 when they did not. The same logic was applied to all courses. For example: taking the course Math112 in academic year 2 was matched with relative terms of 3 and 4 (i.e., third and fourth semesters). A 1 was assigned when the student took the course in the given academic level, and a 0 when the student did not take during that level. A zero could also mean that the student did not take the course at all.

Missing values are a problem is SAS when running statistical tests such as PROC LOGISTIC because SAS deletes the entire row, which means that the whole field is excluded from the model. To enhance the analysis and include all possible cases, some variables were recoded. As explained in the Methods section, missing ACT scores were replaced with SAT scores when available by using a concordance table. Refer to (College Board, 2004) for a table and details about the use of concordance tables. Grades for all courses were recoded in order to have a code for students not taking the course at all or not taking it during a specific academic year. Thus, letter grades were recoded to numeric grades.

The scale used is:

A =4
AB =3.5
B =3
BC =2.5
C =2
D =1.5
*F,NR, NW, IN or U=1 Did not take the course=0

In the present study it was possible to track Change of Major per academic term only for students who started with one major. For students who started with more than one major it was impossible to differentiate whether a different major indicated picking up a new major

* F=0.0
NR+ Grade list received from department, but no grade for this student. NR becomes an F, if no grade is forthcoming by the final grade run (i.e., early next term).
NW = No work
IN = Incomplete (for Credit/No Credit grading basis only)
U = Unsatisfactory (for Pass/Fail grading basis and some others)

or changing majors. Thus, students with field 1^{st}.Major FTE Percent[*] not equal to 1 had a missing value. To included these students in the study the variable was recoded so that when 1^{st}Major FTE Percent is not equal to 1, Change of Major _Academic _Year=9. In other words, a 9 signifies that the student had more than one major or never had a major, which is rare but possible because some students leave before they ever declare a major.

When assessing the field CUM_GPA (Cumulative GPA) it was found that 162 students had a GPA=0 in their first term and taken credits =0. This data were interpreted as meaning that these students attempted going to school, but never did. Based on this logic these students where excluded from the analysis. The initial total of 16,507 students changed to 16,345.

Another case concerning missing values in CUM_GPA was fixed in the following manner. Data were recorded at the beginning of each academic year, yet, CUM_GPA was recorded at the end of the academic year (e.g. the even semester). For example: for those starting their fifth semester CUM_GPA 4^{th} Term was used. However, in a few cases students stop out during the even semester creating a missing record for this semester. Since the proposed probability model predicts the probability of stopping out at least until the last semester a student is enrolled, a statement was created in the SAS program indicating to replace the missing record with their previous cumulative GPA (e.g., if CUM_GPA $4^{nd.}$Term =. then CUM_GPA $4^{nd.}$Term=CUM_GPA 3^{rd}Term). Finally, four variables were created

---

[*] The Major FTE Percent is equal to 1 divided by the number of Plans of type MAJ that the student has declared as of the given term. Allows one to sum over a record set consisting of plans of type MAJ to get the FTE (Full-Time-Equivalent) number of majors.

when adding all subsets, for the purpose of unifying the different term cumulative GPA into one. In other words, a GPA for all freshmen, a GPA all sophomore, a GPA for all junior and a GPA for all seniors.

The field High School Percentile contributed to the majority of missing values. Out of 16,345 students 1,917 (12%) did not have a record in this field, and for the majority of these students this constituted the only missing record. There was no reason or systematization for the missing values. Thus, in order to keep these students in the analysis HS _Percentile rank scores were estimated using multiple regression. The purpose was to use other pre-admission variables to come up with a New Predicted High School Percentile for records missing a value in the original field. PROC GLM was run using all pre-admission variables (e.g., Gender, Age at entrance, Ethnic group, Pell _grant indicator, and Target minority) including other fields such as Fall _cohort _ term, and Residency _at _entrance. The overall *F* statistic for the equation was significant ($P<0.001$) the R^2 of (0.1881) accounted for 19% of the variance on high school percentile. The Type I test for all the variables (except Pell _Ind.) were significant ($P<0.001$). The mean for the New predicted _HS% is 85.21 with a median of 86 similar to the original HS% mean of 85.25 with a median of 87. A statement was written in SAS to assigned predicted HS% scores to High School Percentile missing records.

Descriptive Analysis

This section presents the descriptive statistics of the data used in this study. The total population analyzed in the study was 16,345 students. Table 7 shows the sample distributed by student's fall cohort, and the matriculation level assigned to the student. Students entering in the spring were excluded from the analysis. Notice in the table that the number of students for all three cohorts is similar. As the table indicates, the majority of the students come in classified as freshmen, and fewer receive advanced placement.

Table 7 Description by cohort and matriculation academic level

Count of ID	**Fall Cohort Term**			
Matriculation Academic Level	**1996-97**	**1997-98**	**1998-99**	**Total**
Freshman	5,069	5,468	5,145	15,682
Sophomore	193	194	256	643
Junior	8	4	8	20
Grand Total	**5,270**	**5,666**	**5,409**	**16,345**

Tables 8a and 8b show descriptive statistics (counts and frequencies, and mean, median, standard deviation and range, when applicable) for seven pre-admission and thirteen post-admission variables respectively. As seen in Table 8a, the percentage of females in the sample is slightly higher (55.1%) than the percentage of males (45.9%). The great majority of the population is white (89.2%). The percentage of non-target minority students increases in relation to ethnicity, because Asian Americans who are not "pacific islanders" are not a target minority. The age range is from 13 years to 52 years. . Most students in the sample are 18 years old and close to 15% are Pell-grant recipients. The data on academic preparation shows that the standards for this closed-admission type university are high. The average ACT score is 26.6 points. The average high school rank with predicted values for missing scores is 85.3th.percentile.

Table 8b illustrates how the post-admission variables change across academic level. The data indicates that close to half of the students in the sample (7,894; 48.3%) did not change majors. Nearly one-fifth of the students chose more than one major or did not choose a major. Logically, everyone who attains a degree chooses a major at some point, so all of those who did not chose a major eventually left. Not all students are required to take the critical courses chosen, and the data indicates that the majority of students taking these courses do so while they are freshmen or sophomores. As the academic level advances, the number of students taking these courses decreases. At the bottom of Table 8b is the cumulative GPA. This comes from the recoded GPA for all those in data I (freshman), GPA _sophomore, junior and senior. The cum _GPA increases by 0.1 point per academic level, starting with a 2.9 Cum _GPA freshman and finishing with a 3.3 Cum _GPA _senior. The mean for each course was calculated based on the number of students who took the course (took course=1). For example: 3,700 students took Econ101 while freshmen. The mean for those taking this course was 2.7 with a 0.7 standard deviation.

Table 8a Descriptive statistics for the independent variables

Independent (pre-admission)						Range	
Variables N=16,345	Count	Cell%	Mean	Median	S.D.	Min.	Max.
Gender							
1=Female	9,001	55.1	-	-	-	-	-
0=Male	7,344	45.9	-	-	-	-	-
*Ethnicity							
1=White	14,574	89.2	-	-	-	-	-
2=African American	316	1.9	-	-	-	-	-
3=Hispanic	388	2.4		-			
4=Asian /Pacific Islander	787	4.8	-	-	-	-	-
5=Native American- Alaskan	79	0.5	-	-	-	-	-
6=Unknown	200	1.2	-	-	-	-	-
*Target minority							
1=Target Minority	885	5.4	-	-	-	-	-
0=Not a Target Minority	15,460	94.6	-	-	-	-	-
Pell_Grant Receiver							
1=Yes	2,388	14.6	-	-	-	-	-
0=No	13,957	85.4	-	-	-	-	-
Age	-	-	18	18	0.6	13	52
*New_ACT	-	-	26.6	27	3.4	12	36
*New_HighSchool%	-	-	85.3	87	10.9	8	99

* # Missing values: Ethnicity=1; New_HighSchool%=12; New_ACT=15.

Table 8b Continuation of descriptive statistics for independent variables

Independent (post-admission) Variables N=16,345	All levels N=16,345		Freshman status n=15,682		Sophomore status n=13,535		Junior status n=12,053		Senior status n=9,924	
	count	**cell%**	**count**	**cell%**	**count**	**cell%**	**count**	**cell%**	**count**	**cell%**
Changed Academic Group										
1=changed	5,020	30.7	-	-	-	-	-	-	-	-
0= did not change	11,325	69.3	-	-	-	-	-	-	-	-
Changed Major_indicator										
1=changed	5,378	32.9	328	2.1	1,025	7.6	1,722	14.3	1,047	10.6
0=did not change/during acad.year	7,894	48.3	12,409	79.1	10,618	78.5	9,078	75.3	7,894	79.5
9=+1 major/ never declared a major	3,073	18.8	2,945	18.8	1,892	13.9	1,253	10.40	983	9.9
M112Course										
1=took course	3,593	21.9	3,267	20.8	197	1.5	32	0.3	14	0.1
0=no course at all/ during academic .year	12,752	78.1	12,415	79.2	13,338	98.5	12,021	99.7	9,910	99.9
M221Course										
1=took course	4,246	26.0	3,676	23.4	376	2.8	62	0.5	18	0.2
0= no course at all/ during academic .year	12,099	74.0	12,006	76.6	13,159	97.2	11,991	99.5	9,906	99.8
CHM 103 Course										
1=took course	6,320	38.7	5,644	36.0	501	3.7	67	0.6	16	0.2
0= no course at all/ during academic .year	10,025	61.3	10,038	64.0	13,034	96.3	11,986	99.4	9,908	99.8
ECON 101 Course										
1=took course	7,020	42.9	3,700	23.6	2,176	16.1	558	4.6	124	1.3
0= no course at all/ during academic .year										
current academic level	9,325	57.1	11,982	76.4	11,359	83.9	11,495	95.4	9,800	98.8
PSY 202 Course										
1=took course	8,412	51.5	5,493	35.0	1,891	13.0	430	3.6	126	1.3
0= no course at all/ during academic .year	7,933	48.5	10,189	65.0	11,644	86.0	11,623	96.4	9,798	98.7
			Mean	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
Cumulative GPA	-	-	2.9	0.6	3.1	0.5	3.2	0.4	3.3	0.4
*M112Grade	-	-	2.5	0.8	2.4	0.9	2.7	1.0	2.4	1.0
*M221Grade	-	-	2.8	0.8	2.4	0.9	2.3	0.9	2.7	1.0
*CHM 103 Grade	-	-	2.9	0.7	3.0	0.8	3.1	0.8	3.0	1.0
*ECON 101 Grade	-	-	2.7	0.7	2.9	0.7	2.9	0.8	3.0	0.8
*PSY 202 Grade	-	-	2.7	0.8	2.9	0.8	3.0	0.8	2.9	0.9

*The mean for the course is calculated based on the number of students who took the course (took course=1). For example: for M112grade in freshman year , n=3,267

Descriptive results for the dependent variables and transition probabilities

The dependent variable in this study is institutional persistence. Each of the four data sets had a dependent variable (DV), and the frequencies and percentages of these four DVs are presented in Table 9. The data illustrates the transition from one academic level to the next. The categories of the DVs were ordered according to the extent of persistence. In this sense, stop-outs exhibit the least persistence (SO), retention in the same academic level the next lowest level [i.e, freshman to freshman (FF); sophomore to sophomore (SS); junior to junior (JJ); senior to senior (SrSr)], followed by transition to the next higher level [i.e., FS, SJ, SrG] followed by two-level transition as the highest level [i.e., FJ, SSr].

As seen in Table 9, DV 1 illustrates the path of the students who matriculate as freshmen. The percentage column shows that 7.6% of the students stop out by the beginning of the second year. A smaller number (4.5%) stay at the freshman level, the great majority (82.2%) advance to the next academic level, and 5.6% advance to the highest level in that group, junior. DV 2_AB illustrates the path of the students who become sophomores by the beginning of the second year (n=12,892), plus those who matriculate as sophomores (n=643). Refer to Table 6 to view the origin of the N for each data set. The path of the sophomores also shows that only a small proportion of students stop out (7.8%), although the percentage of stop-outs is highest for sophomore status. Compared to the freshmen, more sophomores stay at the sophomore level (8.6%), but the majority (82.4%) move to the next level and a few (1.2%) become seniors at the beginning of the second year (3rd term in school). The earliest a student graduates in the sample is the 4th term (see Graph 1).

The data for the juniors indicates that a smaller percentage stops out (3.7%), a higher percentage than that of the two previous groups stays in the same level (JJ=15.1%). A total of 81.1% advance to the next level, and a few (0.2%) graduate. There are students who stop a the beginning of their senior year (2.6%), a little over one third of the students (39.9%) stay at the senior level, which means most of these graduate after more than four years of college, and (57.6%) graduate.

The dark shadows in Table 9 illustrate states forbidden by the model. First, once a student stops out, a path can not be modeled. Second, no student goes backwards in the system (i.e., from junior to sophomore). Finally, students only graduate after their junior and senior year. Logically all students need to attain a senior status in order to graduate, yet since the data in this study was collected at the beginning of each odd term, some students become seniors during the even term (e.g., 6^{th} term), and graduate by the time the data is collected.

The percentages of the descriptive data in Table 9 can be summarized as the 14 non-zero probabilities, presented in Table 10. The results in these two tables are the same, one expressed in terms of percentages and one expressed in terms of probabilities (Table 10). The transition probabilities have been arranged into a matrix that reflects the movement from one educational state to another (See Methods section, pg.60 for details on the matrix). Transitions forbidden by the model are again indicated with dark shading.

In Table 10, the probability of stopping out in the freshman year is 0.076, which appears in Table 9 as 7.6% of the students who are no longer enrolled at the beginning of their third term. A few of these students who leave come back to school, but the majority do not attain a degree, even if they re-enroll later on. The literature (Astin, 1993; Astin et al., 1996; Pascarella & Terenzini, 1991) supports the finding that the freshman year yields a higher number of students who stop out or leave the school.

Somewhat higher is the probability of stopping out as sophomore (0.078). This means that 7.8% (see Table 9) of the students stop out after completing their sophomore credits. It is unknown if this difference means that at the sophomore level more students come and go, or if these are students who stop out for good. As expected, the number of students stopping out gets smaller as the student achieves a higher academic level. At the senior level this probability is only (0.012), and it is suspected that the majority of these students seek to attain a degree in another university. Of interest to this study is the number of students who stay at the senior level (39.9%). Later on when presenting the results it will be explained how the results for the senior level can be interpreted as a separate study in itself. Once a student completes enough credits to be classified as senior, the student has achieved the highest level and there are no other levels to advance to other than graduation. Since the proportion of those who stay as seniors is large, and the proportion of those who stop is small, it is likely that the predictors that result with significant betas in this last level are explaining what contributes to students' graduating later than the usual four years vs. more than four years.

Table 9 Descriptive results for the DV

DV levels & transitions N=16,345	Stop out		Freshman		Sophomore		Junior		Senior		Graduate	
	f	%	f	%	f	%	f	%	f	%	f	%
DV1 n=15,682												
S _O	1,195	7.6										
FF			713	4.6								
FS					12,892	82.2						
FJ							882	5.6				
DV 2_AB n=13,535												
S_O	1,060	7.8										
SS					1,171	8.7						
SJ							11,152	82.4				
SSr									152	1.2		
DV 3_ABCD n=12,053												
S_O	444	3.7										
JJ							1,815	15.1				
JSr									9,771	81.1		
Gr											23	0.2
DV 4_ABCDEF n=9,924												
S_O	254	2.6										
SrSr									3,956	39.9		
SrGr											5,714	57.6

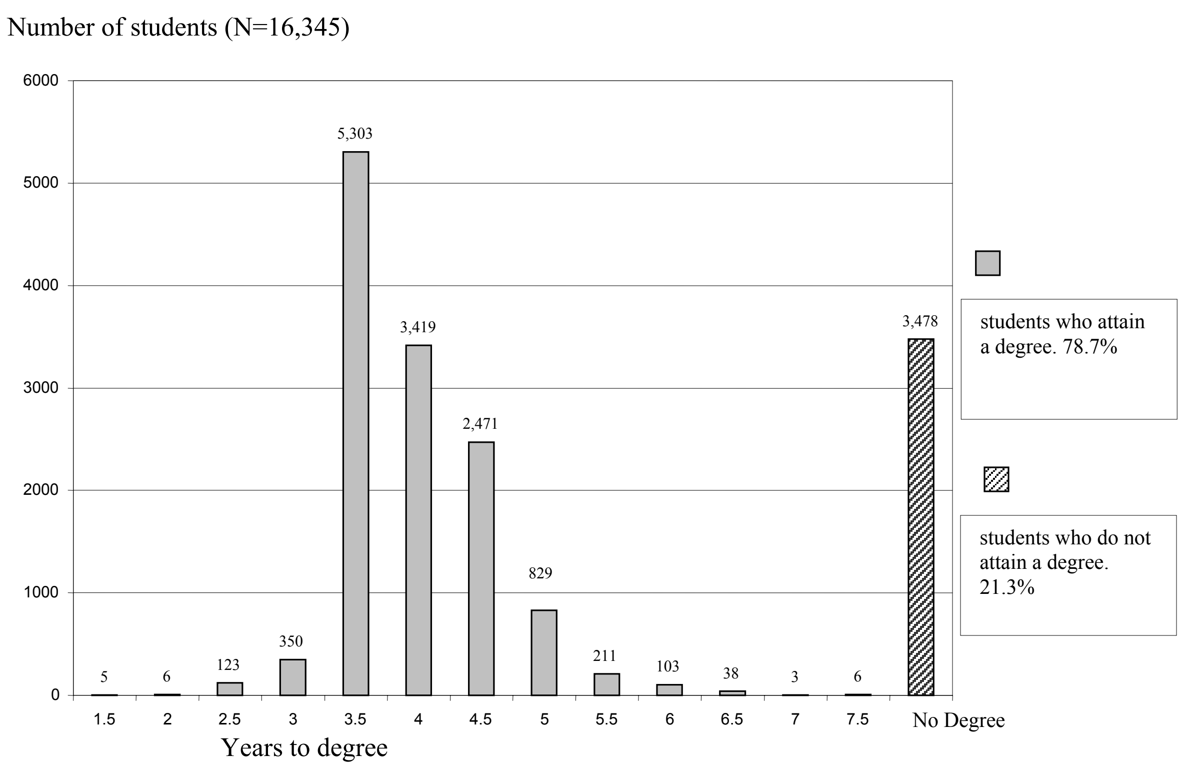
Table 10 Observed transition probabilities

	F	S	J	Sr.	S_O	Grad	Probability
F	0.046	0.822	0.056	0.000	0.076	0.000	1.000
S	0.000	0.087	0.824	0.012	0.078	0.000	1.000
J	0.000	0.000	0.151	0.811	0.037	0.002	1.000
Sr.	0.000	0.000	0.000	0.399	0.026	0.576	1.000
S_O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
Grad	0.000	0.000	0.000	0.000	0.000	1.000	1.000

Actual graduation rates

Graph 1 provides a visual description of the actual graduation rates in the whole sample. This graph uses all data from the three cohorts before is divided in to the four data sets, so it includes students who come and go, but eventually graduate. This is why the numbers in this graph do not match the numbers in Table 9. However, the numbers are still very similar. This confirms that despite the fact that the model used in this study does not carry students who stop out in between levels, this condition does not impose a different effect on the data. In the three cohorts there are some students who stop out for one semester and come back, but most of those who stop out end up not enrolling again and thus not completing a degree in this university.

Graph 1 Number of students who do and do not graduate, by years to degree.



The following section of this chapter presents the results from the logistic regression test used to compute prediction equations for each of the four academic stages in the pipeline model. The results have been organized in a manner in which they can help respond to the proposed research hypotheses.

Test of the Hypothesized Model

A 25-predictor logistic model was fitted to each of the four data sets to test the research hypotheses. Eleven pre-admission variables are relevant to hypotheses 1 and 1.1, and fourteen post-admission variables are relevant to hypotheses 2 and 2.1 (see Methods p.79).

The overall fit for all four models showed an improvement over the intercept-only model (the null model). The three inferential statistical tests: the Likelihood ratio, Score test and Wald tests, with a significant chi-square p<0.0001, yield the conclusion that model 1 (i.e., all 25 variables for the freshman group), model 2, model 3 and model 4 were more effective than the null model. (See Testing Global Null Hypothesis in Tables B1-B4 in the Appendix B. These tables show the logistic regression results for all four models).

Since the probability of a transition event is the focus of this study, and since the model developed is expected to serve as a predictive tool to identify categories of individuals at risk of not persisting, it was considered more appropriate to interpret the beta coefficients instead of the usual odds ratios, a common approach in logit models (see DeMaris, 1993) for a discussion on interpreting odds ratios versus probabilities.

In the following paragraphs, data for the significant predictors in the four data sets will be presented, as well as the interpretation of their symbol. The level of confidence for the number of significant betas and plots of parameter estimates are also presented, to facilitate the understanding of trends across academic levels. Establishing the validity of the parameter estimates will justify the use of the resulting equations to compute the estimated transition probabilities.

Test of hypothesis 1: results of beta coefficients for pre-admission variables

To test the first hypothesis, this study examined the effects of Age (Age _at _entrance), the class variable Ethnic Group (American Indian/Alaskan native, Asia /Pacific Islander, African American, Hispanic, unknown, vs. White), the class variable Gender (F vs. M), Target Minority, Pell Grant Indicator (Pell _IND.), High School Percentile (New_ HSP) and ACT Scores (New _ACT) on transitions to the next academic level.

As seen in Table B-1, there are 6 significant pre-admission predictors for the freshman level:

Age	(β =.1273, $p<$ .0047)
American –Indian vs. white	(β=.9159, $p<$ .0007)
Asian/pacific islander vs. white	(β= - .5455, $p<$ .0002)
Target Minority	(β = - .6856, $p<$ .0114)
New _High School Percentile	(β= -.0106, $p<$.0001)
New _ACT	(β= -.1032, $p<$.0001).

Table B-2 shows 4 significant pre-admission predictors for the sophomore level:

Age	(β=.2012, $p<$.0001)
Asian/pacific islander vs. white	(β= -.3925, $p<$.021)
New _Pell grant indicator	(β=.1970, $p<$.0036)
New _ACT	(β=-.0368, $p<$.0001).

Table B-3 shows 3 significant pre-admission predictors for the junior level. These are:

Age	(β= .1452, p<.0058)
Gender (female vs. male)	(β= -.1962, p<.0001)
Pell grant indicator	(β= .2913, p<.0001).

Table B-4 shows 7 significant pre-admission predictors for the senior level. These are:

Age	(β= .1169, p<.0184)
Gender (female vs. male)	(β= -.1664, p<.0001)
Ethnic group Hispanic vs. white	(β= -.5278, p<0.025)
Target Minority	(β= 1.059, p<.0032)
Pell grant indicator	(β= .3674, p<.0001)
New _High School Percentile	(β=.0192, p<.0001)
New _ACT	(β=.0365, p<.0001).

As expected, the predictability of pre-admission variables diminished across academic level until the junior level. However, the increase in the number of significant predictors at the senior level and the sign reversal for some of the predictors were unexpected. Please refer to the Discussion under unexpected findings for further comments. Table 12 shows a summary of all coefficients, 6 significant predictors at the freshman level, 4 at the sophomore level, 3 at the junior level and 7 at the senior level.

Test of hypothesis 1.1: beta coefficient values and trends

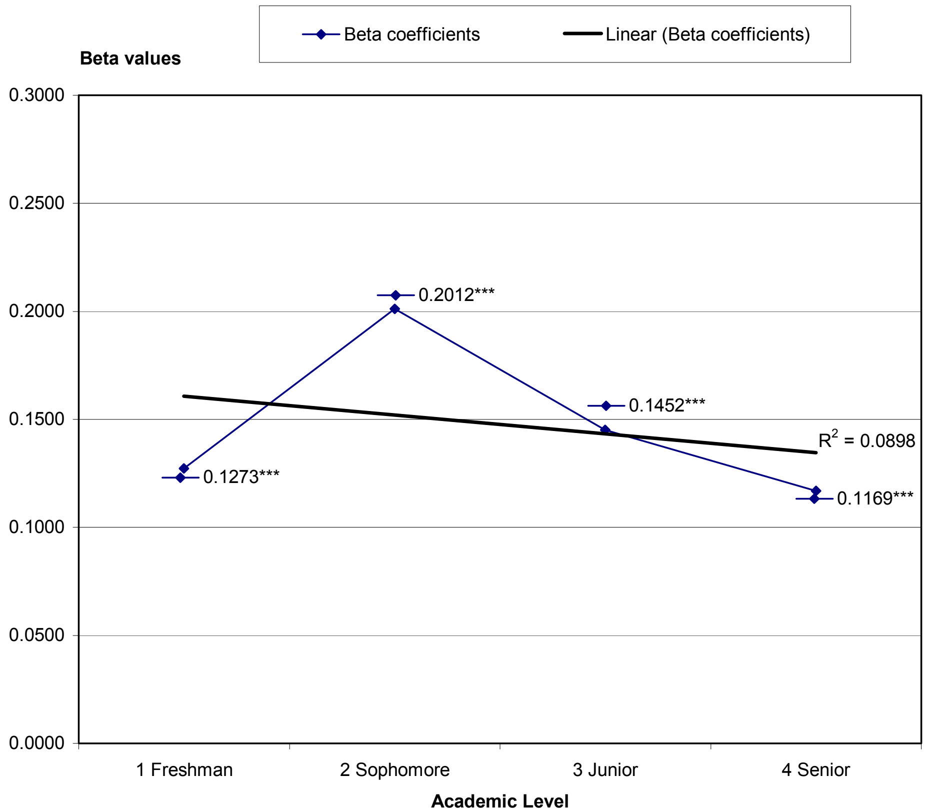
Table 11 is a guide to facilitate the interpretation of the beta coefficients. Coefficient size, sign and trend changes will be presented to respond to hypothesis 1.1.

Table 11 Guide to Interpreting Beta Coefficients

Coefficient characteristics	**Interpretation example**
Coefficient value (e.g., β= 0.1273)	A logit coefficient of 0.1273 tells us that the log-odds increases by 0.1273 for every 1 unit increase in the explanatory variable. For this study: it refers to the probability of stopping out vs. staying in an academic level vs. advancing to the next academic level vs. graduating.
Coefficient value increase (e.g., 0.1273 to 0.2012) /decrease (e.g., 0.1452 to 0.1169)	An increase in the coefficient value can be interpreted as an augmentation of the effect of the predictor on the dependent variable. A decrease can be interpreted as a reduction on the effect of the predictor on the dependent variable.
Coefficient sign (+ / -)	Similar to OLS, the sign associated to the beta weights indicates the direction of the effect that the independent variable has on the DV. This study used the default in SAS (Ascending), which models the response coded with the lowest value. The values in the DV are ordered from 1 Stop-Out to 4 the highest academic level/graduation: Thus, the effect of *positive* beta coefficients is to *increase* the probability of stopping out and of remaining at grade level while reducing the probability of advancing one or two academic levels. The effect of a negative beta is just the reverse. (i.e., *negative* beta= *decrease* the probability of stopping out). + sign = negative contribution to transition probabilities – sign = positive contribution to transition probabilities

Age at entrance: is a significant predictor for all four academic levels. The (+) sign of the beta coefficients indicates a negative contribution to probability prediction for each state of the dependent variable. The beta values preserve the order of the + sign across all four levels. Thus, lower transition probabilities to the next academic level are predicted as a function of increasing age, other variables in the model held constant. The effect of age across academic levels is shown in Graph 2. A linear fit is shown, and explains only 9% of the variation in the estimated beta coefficients. As shown, in the graph, being older has the greatest negative effect on persistence for sophomores, compared to other academic levels.

Graph 2 Beta Coefficients for Age at Entrance Across Academic Levels



Ethnicity: was entered as a class variable in the model. Logistic regression modeled the effect of each ethnic group vs. Whites as the base class. The results indicate that being American Indian has a significant positive contribution to the probabilities of stopping out or failing to advance, at the freshman level. Thus, lower transition probabilities to the next academic level are predicted as a function of being a Native American in the freshman year, compared to that of white students and holding other predictors in the model constant. This result is interesting in light that Native Americans in this study account for only 0.5% of the students.

Results also show that being a member of an Asian/Pacific islander ethnic group has a significant negative contribution to the probabilities of stopping out or failing to advance, at the freshman and sophomore levels when compared to that of white students. Thus, higher transition probabilities are predicted as a function of being Asian/Pacific Islander compared to white and holding all other variables constant at the freshman and sophomore levels. The effect of this characteristic is negative, but it is not statistically significant at the junior and senior level (See Graph 3). The change in the signs in the last two levels may result from the fact that this group is formed by Asians, a group often thought of as high academic achievers, and Pacific Islanders, whom for the most part are an at-risk group (Yeh, 2002). Among the 787 Asian Americans in this study, 106 are target minority (13%), possibly all pacific islanders. For the African American ethnic group, the effect on transition probabilities when compared to white students is not significant at the freshman, sophomore, junior or senior level. For the Hispanic ethnic group, the effect on transition probability is not statistically significant at the freshman, sophomore or junior level, and is slightly significant at the senior level. The negative (-) sign indicates a positive contribution to probability prediction, resulting in higher transition probabilities towards graduation as a function of being Hispanic in the last academic level when compared to white students, with other predictors being held constant.

As seen in Graph 3 the beta values are very different for each one of the ethnic groups and they vary across academic levels. The effect on transition probabilities is very much alike for all groups at the sophomore level when compared to whites. The exception is Asian-Americans' where the beta is negative, resulting in higher transition probabilities towards advancing to the junior and senior levels. At the junior level most betas cluster within a similar range. The exception here is for American-Indian. Something similar happens at the senior level and here the exception is for Hispanics, where the effect, as mentioned before favors transition probabilities towards graduation and it is statistically significant. More about these differences will be commented in the conclusions chapter.

Graph 3 Beta Coefficients for all Ethnic Groups Across Academic Levels

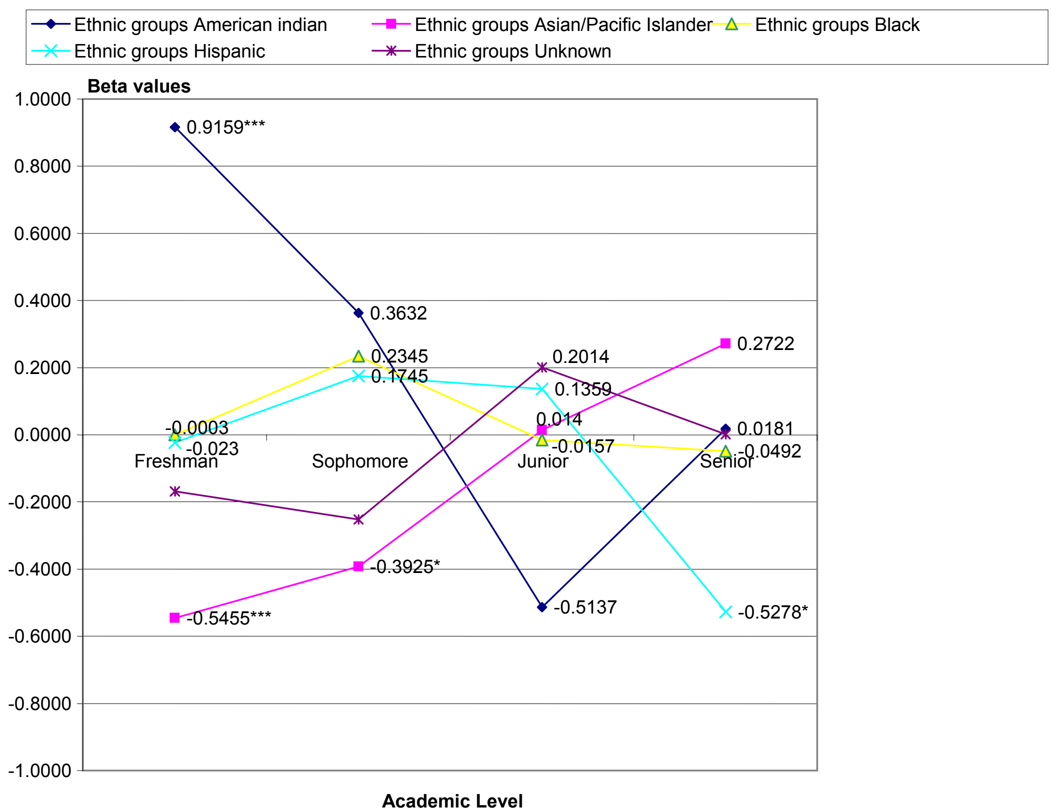


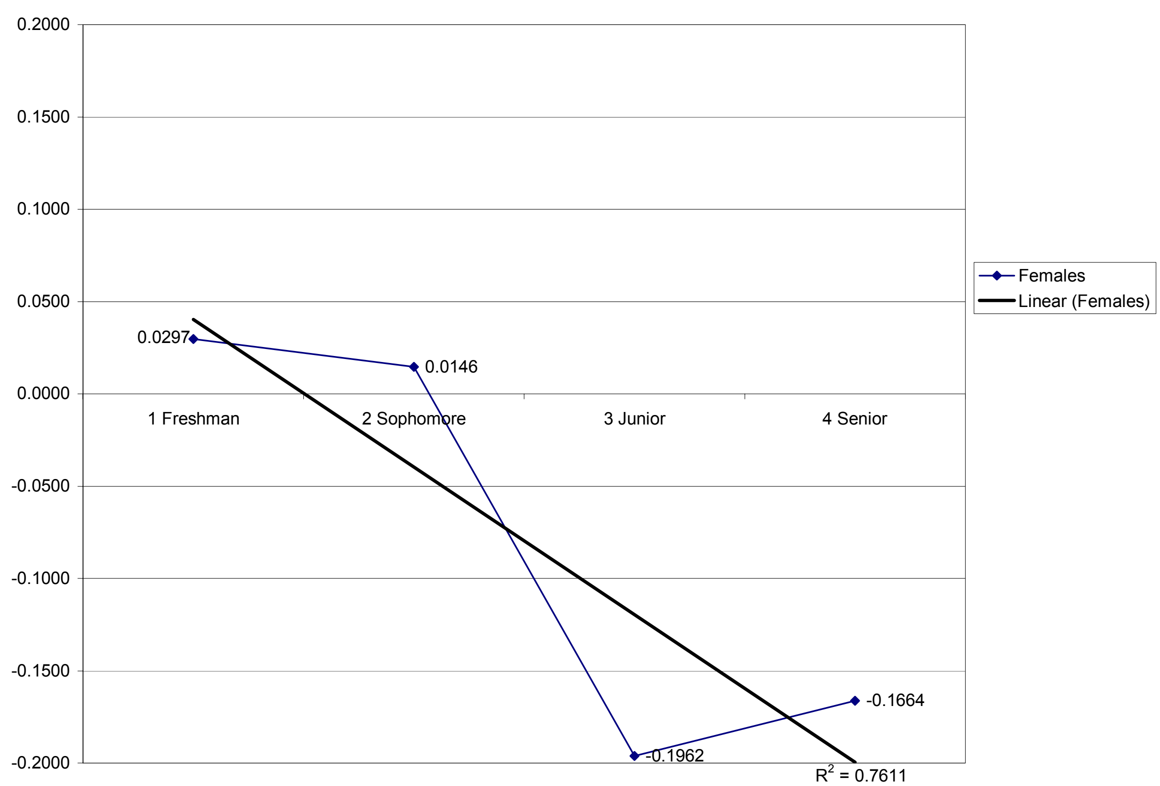
Table 12 Pre-Admissions Predictors by Academic Level

Predictor Variables Variables N=16,345	Freshman status n=15,661			Sophomore status n=13,523			Junior status n=12,049			Senior status n=9,920		
	+ / −	β	*p*	+ / −	β	*p*	+ / −	β	*p*	+ / −	β	*p*
Pre-admission predictors												
1-Age	+	**.1273**	**.0047****	+	**.2012**	**.0001*****	+	**.1452**	**.0058****	+	**.1169**	**.0184***
2-Eth. grp. American Indian	+	**.9159**	**.0007*****	+	.3632	.2748	−	.5137	.2245	+	.0181	.9615
3-Eth. grp. Asian/Pacific Islander	−	**.5455**	**.0002*****	−	**.3925**	**.021***	+	.0140	.9381	+	.2722	.1619
4-Eth grp. African American	−	.00025	.9989	+	.2345	.2672	−	.0157	.9475	−	.0492	.08471
5-Eth grp. Hispanic	−	.0230	.8994	+	.1745	.3992	+	.1359	.5409	−	**.5278**	**.0254***
6-Eth grp. Unknown	−	.1686	.4709	−	.2527	.3318	+	.2014	.4376	+	.00103	.9968
7-Gender (F)	+	.0297	.2270	+	.0146	.5808	−	**.1962**	**.0001*****	−	**.1664**	**.0001*****
8-Target minority ind.	−	**.6858**	**.0114****	−	.3190	.3108	+	.2582	.4197	+	**1.059**	**.0032****
9-Pell_ind (yes)	+	.1216	.0594	+	**.1970**	**.0036****	+	**.2913**	**.0001*****	+	**.3674**	**.0001*****
10-High School %	−	**.0106**	**.0001*****	−	.0032	.2173	+	.00401	.1409	+	**.0192**	**.0001*****
11-ACT	−	**.1032**	**.0001*****	−	**.0368**	**.0001*****	−	.00945	.2755	+	**.0365**	**.0001*****
Total No. of significant predictors	6			4			3			7		

Note: Freshman (n): 21 observations were deleted in the logistic procedure due to missing values
Sophomore (n): 12 observations were deleted in the logistic procedure due to missing values
Junior (n): 4 observations were deleted in the logistic procedure due to missing values
Senior (n): 4 observations were deleted in the logistic procedure due to missing values

Gender: This variable was entered in the model as a class variable. Logistic regression modeled the effect of females vs. males on the DV. The (-) sign of the beta value indicates that being a female contributes positively to favorable transition probabilities at the junior and senior level in comparison with males at the same academic levels, holding all other predictors in the model constant. The effect of being a female vs. male is not statistically significant on transition probabilities at the freshman and sophomore levels. As seen in Graph 4 below, the sign changes from positive (+) in the first two academic levels to negative (–) in the last two academic levels. A linear fit accounts for two thirds of the variance in the beta coefficients (R^2 = .76). This finding is consistent with studies are based on national data, which found that women, as compared with men, are more likely to complete their bachelor's degree in four years (Astin, 1975, 1993).

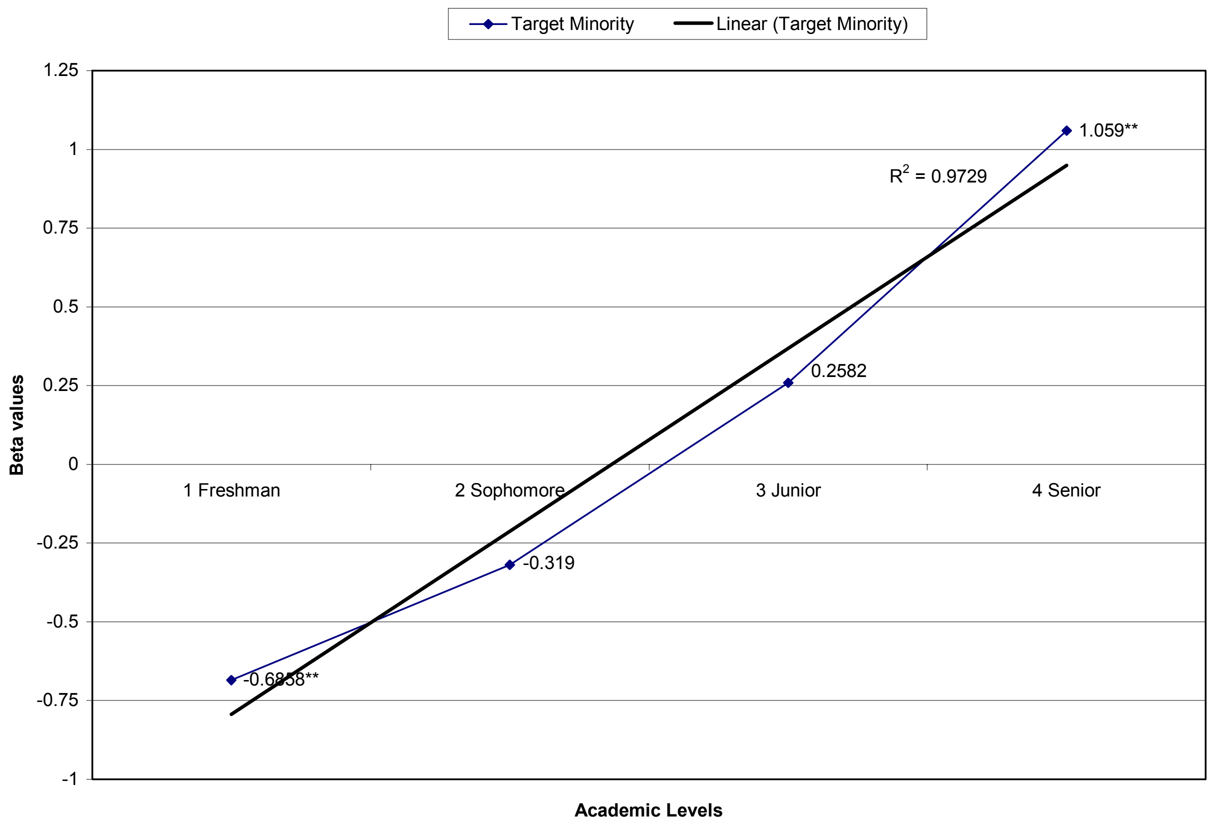
Graph 4 Beta Coefficients for Gender Across Academic Levels



Targeted Minority: being a target minority has a positive significant effect on transition probabilities for freshmen, and a positive but insignificant effect for sophomores. At the junior level the beta sign shifts from - to + although insignificant, and at the senior level value of the beta is (+) and statistically significant. Thus, lower transition probabilities are predicted as a function of being a target minority vs. not being a target minority, in the last academic level while holding all other variables constant. As is evident in Graph 5 below, at the senior level the weight of the beta is larger than in any of the other 3 academic levels and the linear fit accounts for almost all the variance in the estimated beta coefficients. ($R^2 = .97$)

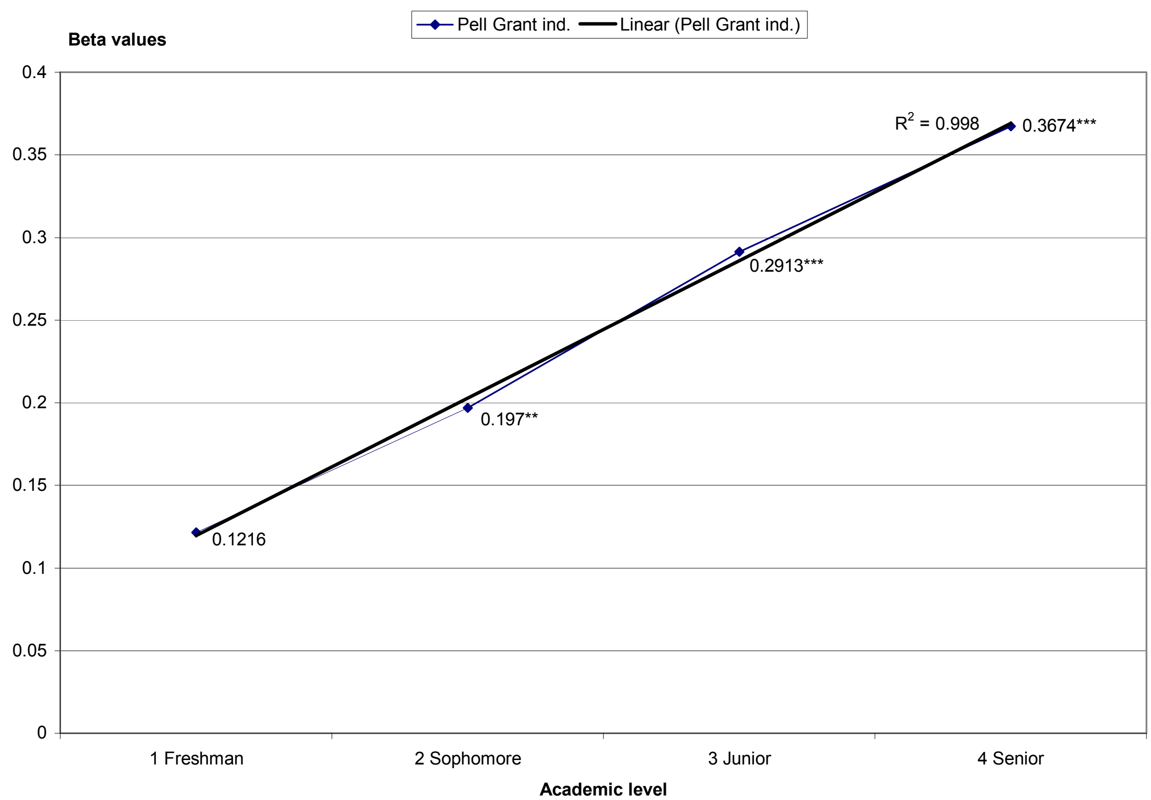
The interesting aspect of the shift the predictor Target Minority has towards inhibiting persistence in the last two academic levels is that it is apparently contradictory to the results obtained for individual ethnic group categories (see Graph 3, ethnic groups). The (+) beta sign at senior level for Target Minority is not being enhanced by the (+) sign for Asian Americans, because 87% of the students in this group are not Target Minority. The disparity in results may be a result of insufficient number of students for each of the ethnic group categories. A more reliable predictor to understand transition probabilities towards graduation appears to be Target minority, in general terms this predictor is the sum of all underrepresented groups. Other studies have associated being part of an underrepresented minority group with lower graduation rates (Alexander, Foertsch, & Bowcock, 1998; Astin, 1993; Gloria et al., 1999; Tinto, 1993; U.S. Department of Education, National Center for Education Statistics (NCES), 2001c)

Graph 5 Beta Coefficients for Target Minority Status Across Academic Levels



Pell Grant indicator: the (+) sign of the beta values indicate that receiving a grant has a negative effect on transition probabilities over all academic levels. The effect is not significant at the freshman level but it is significant in all other levels. Notice that the sign of the beta (+) values is the same across the four academic levels (see Graph 6). This finding is apparently counterintuitive, since one would expect for grants to be helping student move forward. Despite this stated logic, it is important to remember that Pell Grants are just one form of student aid, and it is a need-based program targeted to those who are economically disadvantaged. According to the literature, economically disadvantaged students are identified as a group at risk of not persisting in college (Astin, 1975; U.S. Department of Education, 2000a). Nevertheless, an interesting finding for this predictor is that the linear fit accounts for almost all the variance in the estimated beta coefficients. (R^2 = .99). Why does the effect of Pell Grant gets worse in a linear fashion over academic levels? There is no simple answer, but this will be discussed under summary and conclusions.

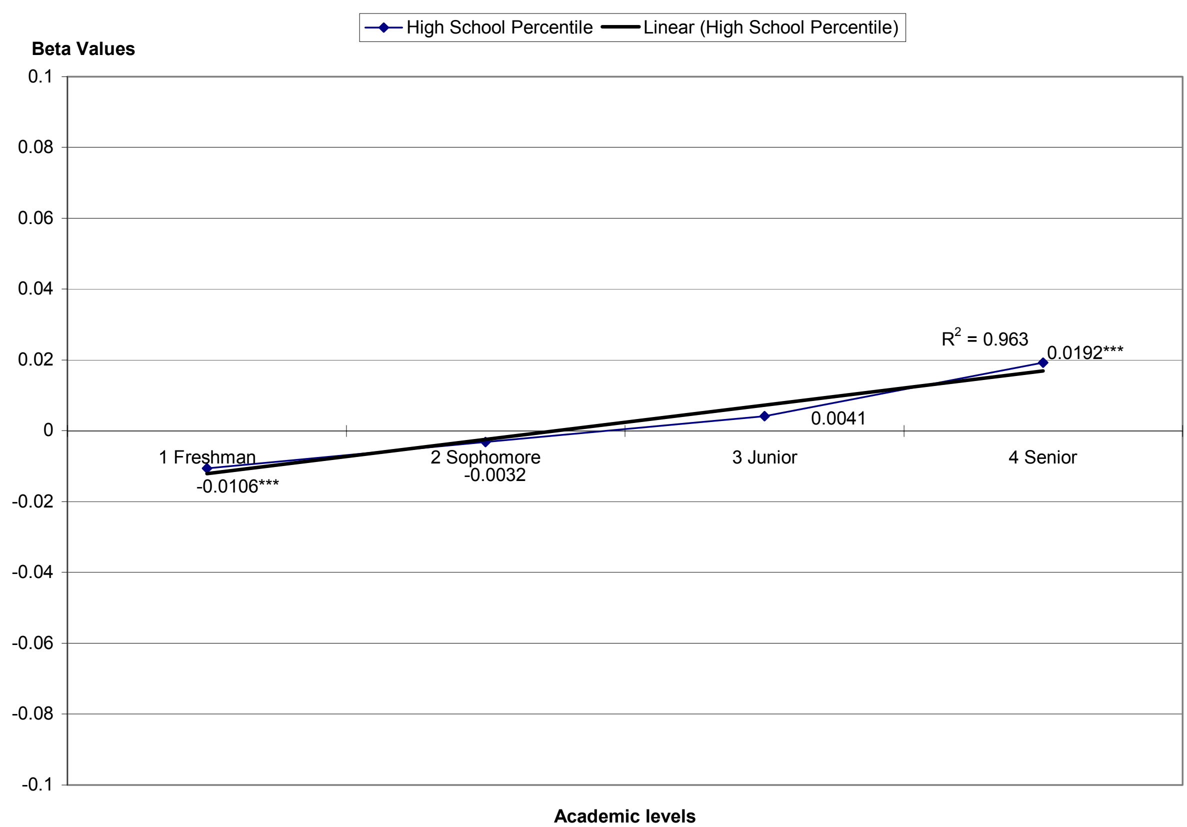
Graph 6 <u>Beta Values for Pell Grant Indicator across Academic Levels</u>



<u>High school percentile</u>: This is a significant predictor of persistence at the freshman and senior levels. As expected, the negative (-) sign in the freshman level indicates a positive contribution to probability prediction towards advancing academic levels. In this sense, higher transition probabilities are expected as a function of higher percentile scores. Nevertheless, the value of the beta at the freshman level is small, indicating little weight, although significant. As seen in Graph 7 a reverse of the sign occurs at the junior level, and

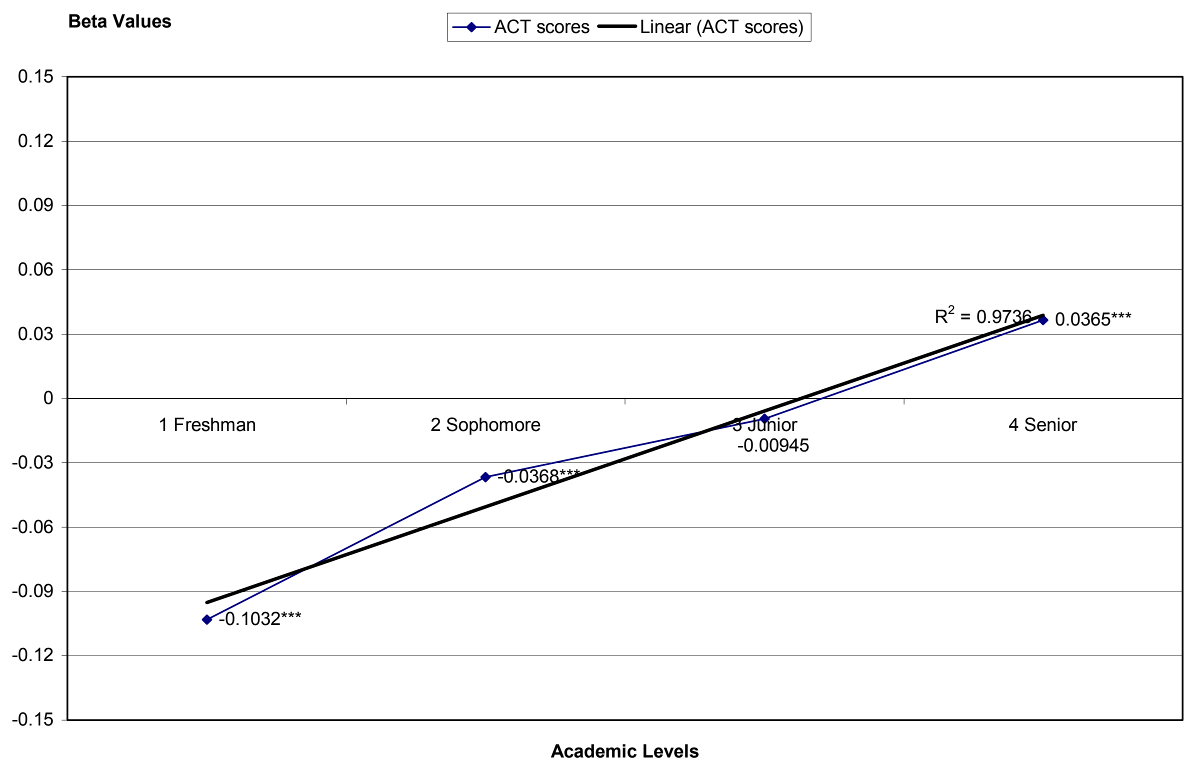
by the senior level the beta values is (+), the effect on transition probabilities is therefore negative, has a slightly higher weight, and is significant. The sign reversal in the last academic level is considered an unexpected result. According to the hypothesis, the expected results would have produced a graph with values sloping downward in the negative values and never becoming positive ($R^2 = .96$). The literature supports the finding for the significant positive effect on persistence at the freshman level (Pascarella & Terenzini, 1980), the sign reversal at the senior level is something that will be further discussed in the conclusions.

Graph 7 Beta Values for High School Percentile Across Academic Levels



<u>ACT scores</u>: ACT score is a significant predictor of persistence at the freshman and sophomore level. The (-) sign at the freshman and sophomore level indicate a positive contribution to academic level transition probabilities. Thus, higher transition probabilities are expected as a function of increasing scores, other predictor variables held constant. The beta value is much higher at the freshman level than at the sophomore level, indicating that the positive effect on academic level transition probabilities is larger for freshmen. The weight of the beta at the junior level is very close to that for sophomore level, but no longer significant. However, at the senior level the order of the sign for the beta value reverses and the predictor, although not large, becomes significant again. A (+) sign here indicates lower transition probabilities associated with graduation (see Graph 8). Similar to most of the pre-admission variables the linear fit accounts for almost all the variance in the estimated beta coefficients. (R^2 = .97).

Graph 8 Beta Values for ACT scores Across Academic Levels



Confirmation of hypothesis 1 and hypothesis 1.1: test of the significance for the beta coefficients

A binomial probability distribution was used to compute the probability of observing k significant betas from a population of n binary variables. In this study, a binary event is acceptance or rejection of the hypothesis that a given beta weight is equal to 0 in the population. In other words, this test was used to find out the probability of obtaining significant betas by chance, and thus decide whether hypothesis 1 could be regarded as confirmed. For example: for the 11 pre-admission variables, using the 0.05 level of significance, the probability of rejecting the null hypothesis is p = 0.05 and the probability of acceptance of the null hypothesis is q = 0.95 for each of the n = 11 pre-admission variables. Thus, the probability of observing k significant pre-admission variables by chance is found using the following probability function:

$$P(k) = \binom{n}{k}(p^k)(q^{n-k}) \tag{4.1}$$

Where,

$\binom{n}{k}$ is the binomial coefficient, $\dfrac{n!}{k!(n-k)!}$

p is the probability of success (i.e., rejecting the null hypothesis)
q is the probability of failure (i.e., accepting the null hypothesis)
k is the value of the random variable success (i.e., significant betas)
n is the number of trials (i.e., 11 pre-admission variables; 14 post-admission variables)

Writing out the binomial coefficient and using p=0.05 and q=0.95, this becomes:

$$\mathrm{P(k)} = \frac{11!}{k!(11-k)!}(.05^k)(.95)^{11-k} \tag{4.2}$$

The probability of observing more that *j* significant betas by pure chance can be computed as the sum of the probabilities of $k = j+1$, $k = j+2$, ... , $k = 11$, which can be expressed symbolically as:

$$\sum_{i=j+1}^{11} P(k=i) = \sum_{i=j+1}^{11} \frac{11!}{i!(11-i)}(.05^i)(.95^{11-i}) \tag{4.3}$$

or since the total probability must equal 1:

$$\sum_{i=j+1}^{11} P(k=i) = 1 - \sum_{i=0}^{j} P(k=i) = 1 - \sum_{i=0}^{j} \frac{11!}{i!(11-i)}(.05^i)(.95^{11-i}) \tag{4.4}$$

Therefore, to determine the probability that more than 2 significant betas occur by chance, the three probabilities P(k=0), P(k = 1) and P(k=2) need to be computed.

$$P(k=0) = \frac{11!}{0!(11-0)!}(.05^{0})(.95^{11-0}) = 0.5688$$

And similarly:

$P(k{=}1)$ = 0.3293
$P(k{=}2)$ = 0.0867

Which when summed equal 0.5688 + 0.3293 + 0.0867 = 0.9848. From this, it can be concluded that the probability of observing three or more significant betas by pure chance from a total of 11 variables is:

$$\sum_{i=4}^{11} P(k=i) = 1 - 0.9848 = 0.0152$$

Since the analyses of pre-admission variables in all academic levels have three or more significant betas, we can conclude that the probability of that happening by chance is 0.0152 or less. This is the same as saying that if we have three betas significant at the 0.05 level, we can be 98.48% confident that they did not occur by chance. This allows us to reject the null hypothesis and conclude that pre-admission variables have an effect on transition probabilities (i.e., something other than chance is operating).

Because the binomial distribution assumes independence and many of the variables in this study may be somewhat correlated, the above probability estimates must be considered as approximations. Beyond the extent necessary to either confirm or reject the hypotheses, an in-depth investigation of the correlations between the pre-admission variables is not within the scope of this work. Since nearly half (20 out of 44) of the beta coefficients computed were found to be significant -- many at a level much better than $p<0.05$ -- it is reasonable to reject the null hypothesis. One further confirmation of this can be obtained by looking at the p=0.005 significance level. At the freshman level, 5 of the 11 betas were found to be significant with p<0.005. Using the values p=0.005 and q=0.995 in equation 4.4 yields P (k=0) = 0.9435 and P(k=1) = 0.0523. These sum to 0.9987, meaning that the probability of obtaining more than 1 beta significant to the p<0.005 level by pure chance is 0.0013%.

Test of hypothesis 2: results of beta coefficients for post-admission variables

All post-admission variables with the exception of Change _Academic _Group are time-dependent variables. This means that their values are specific to academic level. For example: GPA at the freshman level is the cumulative GPA up to the second term. GPA at the sophomore level is the cumulative GPA up to the 4^{th} term, for those who entered as freshman and the cumulative GPA up to the second term for those who entered as sophomores.

To test the second hypothesis, this study examined the effects of 14 post-admissions variables: cumulative GPA , the class variable Change _Academic _Group _indicator (ChgAcdGrp_Ind; no change vs. change), the class variable Change Major (Changemajor_Acadyear; more than one major vs. change; no change vs. change), M112course (M112course_Acadyear), M112Grade, M221course (M221course_Acadyear); Ch103course (Ch103course_Acadyear), Ch103 Grade, Econ101course (Econ101course_Acadyear), Econ101Grade, Psy202Course (Psy202Course_Acadyear), Psy202Grade.

As seen in Table B-1, there are 8 significant post-admission predictors for the freshman level:

GPA-F	(β = -1.5489, p< 0.0001)
Change Academic Group Ind. (no change vs. change)	(β =0.0772, p< 0.0052)
Chg _Major _Acadyear_Ind. (+1 major vs. change)	(β = 1.078, p<0.0001)
Chg _Major _Acadyear _Ind (no change vs. change)	(β = - 0.38650, p<0.0001)
M112course _Acadyear	(β = -0.311, p<0.0091)
M112Grade	(β = 0.1205, p<0.0079)
M221Grade	(β = 0.1230, p<0.0019)
Econ101Grade	(β = -0.0801, p<0.0001)

Table B-2 shows 7 significant post-admission predictors for the sophomore level:

Cum _GPA So.	(β =1.6661, p<0.0001)
Change _Academic _Group Ind (no change vs. change)	(β =0.1301, p<0.0001)
Chg _Major _Acadyear Ind. (+1 major vs. change)	(β = 1.0112, p<0.0001)
Chg _Major _Acadyear Ind (no change vs. change)	(β = -0.5267, p< 0.0001)
M221course_Acadyear	(β = 0.3002, p<0.0262)
Ch103course_Acadyear	(β = 0.3864, p<0.0018)
Econ101Grade	(β = -0.0661, p<0.0015)

Table B-3 shows 9 significant post-admission predictors for the junior level. These are:

GPA _jun	(β= -1.448, p<0.0001)
Change Academic Group Ind (no change vs. change)	(β=0.0983, p<0.0003)
Chg _Major _Acadyear_Ind. (+1 major vs. change)	(β = 0.4826, p<0.0001)
Chg _Major _Acadyear_Ind (no change vs. change)	(β = -0.0805, p<0.0343)
M112Course Acadyear	(β =0.8828, p<0.0195)
M221Grade	(β =0.0599, p<0.0032)
Ch103course_Acadyear	(β =0.5697, p<0.0418)

Ch103Grade	(β =0.0450, p<0.0077)
Psy202course_Acadyear	(β=0.2672, p<0.0274)

Table B-4 shows 6 significant post-admission predictors for the senior level. These are:

GPA_Sen	(β = -0.7086, p<0.0001)
Change Academic Group Ind (no change vs. change)	(β= -0.2346 p<0.0001)
Chg _Major _Acadyear_Ind (no change vs. change)	(β = -0.3739, p<0.0001)
M221Grade	(β =0.0714 p<0.0001)
Ch103Grade	(β=0.1220, p<0.0001)
Psy202Grade	(β= -0.0758, p<0.0001)

An unexpected result was that the predictive power of post-admission variables does not increase steadily across academic levels. This is seen in the decrease number of significant variables at the sophomore and senior levels. Refer to the Discussion chapter for comments on unexpected findings for post-admission variables. Table 13 presents a summary of all coefficients, showing that there are 8 significant predictors at the freshman level, 7 at the sophomore level, 9 at the junior level and 6 at the senior level. To not overwhelm the reader the values of the parameter estimates will not be presented, but they are visible in the graphs.

Test of hypothesis 2.1: beta coefficient values and trends

Original beta values and significance levels for post-admission variables can be found in Table 8 and Appendix B, Tables 1-B to 4-B

Cum_GPA is a significant predictor for all four academic levels. The (-) sign of the beta coefficients indicates a positive contribution to probability prediction for each state of the dependent variable. The beta values preserve the order of the (-) sign across all four levels, and the weight of the beta values is higher at the freshman and sophomore level, diminishing a little at the junior level and much more at the senior level. Thus, higher transition probabilities associated with persistence are predicted as a function of increasing GPA, other variables in the model held constant. The effect of GPA across academic levels is seen in Graph 9. A linear trend was fitted to the data, to show a visual representation of the direction of the effect of GPA across the four academic levels. The R^2 for the linear model is 0.67, which indicates that a linear fit explains about two thirds of the variation in the estimated beta coefficients. The results for this predictor are well supported by the literature as strong predictor degree attainment and persistence (Arredondo & Knight, 2005; Choy 2002; Fredda, 2000; Mc Grath & Braunstein, 1997; Nora et al., 1996; Pascarella & Terenzini, 1991; Tinto, 1993) An additional finding in this study is the fact that GPA is large and statistically significant at the sophomore level. Most studies emphasize the importance of the freshman GPA, but this has to do more with not having access to information about GPA after the first year, or having to rely on students' reported GPA.

Graph 9 Beta Values for Cumulative GPA Across Academic Levels

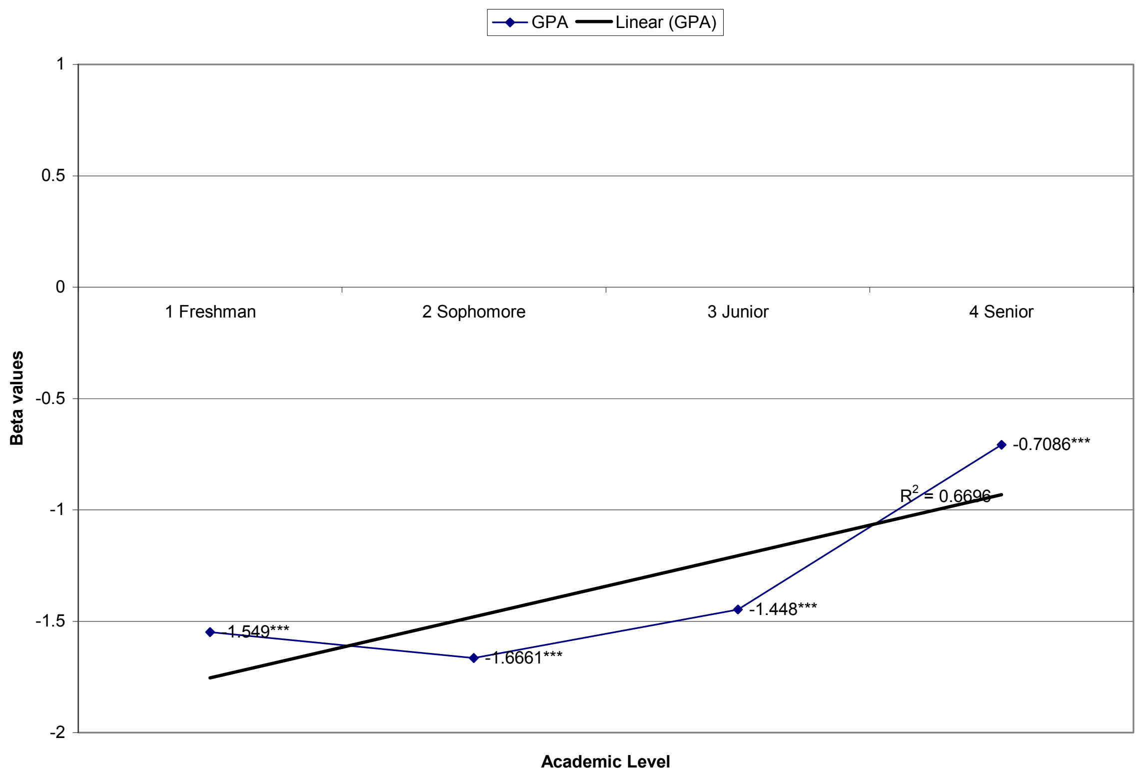


Table 13 Post-Admissions Predictors by Academic Level

Predictor Variables Variables N=16,345	Freshman status n=15,661			Sophomore status n=13,523			Junior status n=12,049			Senior status n=9,920		
	+ / −	β	p	+ / −	β	p	+ / −	β	p	+ / −	β	p
Post-admission predictors												
12-Cum_GPA	−	1.549	**.0001*****	−	1.6661	**.0001*****	−	1.448	**.0001*****	−	.7086	**.0001*****
13-Change Academic group ind. (no)	+	.0772	**.0052****	+	.1301	**.0001*****	+	.0983	**.0003*****	−	.2346	**.0001*****
14-Chg Major Acadyear ind.(dual major+)	+	1.078	**.0001*****	+	1.0112	**.0001*****	+	.4826	**.0001*****	−	.0581	.2363
15-Chg Major Acadyear_ind (no change)	−	.3865	**.0001*****	−	.5267	**.0001*****	−	.0805	**.0343****	−	.3739	**.0001*****
16-M112course_Acadyear	−	.3111	**.0091****	+	.3296	.0619	+	.8828	**.0195****	+	.5548	.3075
17-M112Grade	+	.1205	**.0079****	+	.00345	.8817	+	.0215	.3462	−	.0127	.5277
18-M221course_Acadyear	−	.0190	.8698	+	.3002	**.0262***	−	.1736	.5714	+	.2061	.6648
19-M221Grade	+	.1230	**.0019****	+	.0369	.1095	+	.0599	**.0032****	+	.0714	**.0001*****
20-Ch103course_Acadyear	+	.0991	.3487	+	.3864	**.0018****	+	.5697	**.0418***	+	.7414	.1472
21-Ch103Grade	+	.0109	.7542	+	.0298	.1078	+	.0450	**.0077****	+	.1220	**.0001*****
22-Econ101course_Acadyear	−	.1150	.0860	−	.0604	.4225	+	.1842	.0971	+	.2256	.2250
23-Econ101Grade	−	.0801	**.0001*****	−	.0661	**.0015****	−	.0318	.0779	−	.0134	.3586
24-Psy202course_Acadyear	+	.0174	.7806	+	.0452	.5450	+	.2672	**.0274***	−	.1210	.5214
25-Psy202Grade	−	.0233	.2695	+	.00985	.6165	+	.0264	.1364	−	.0758	**.0001*****
Total No. of significant predictors	8			7			9			6		

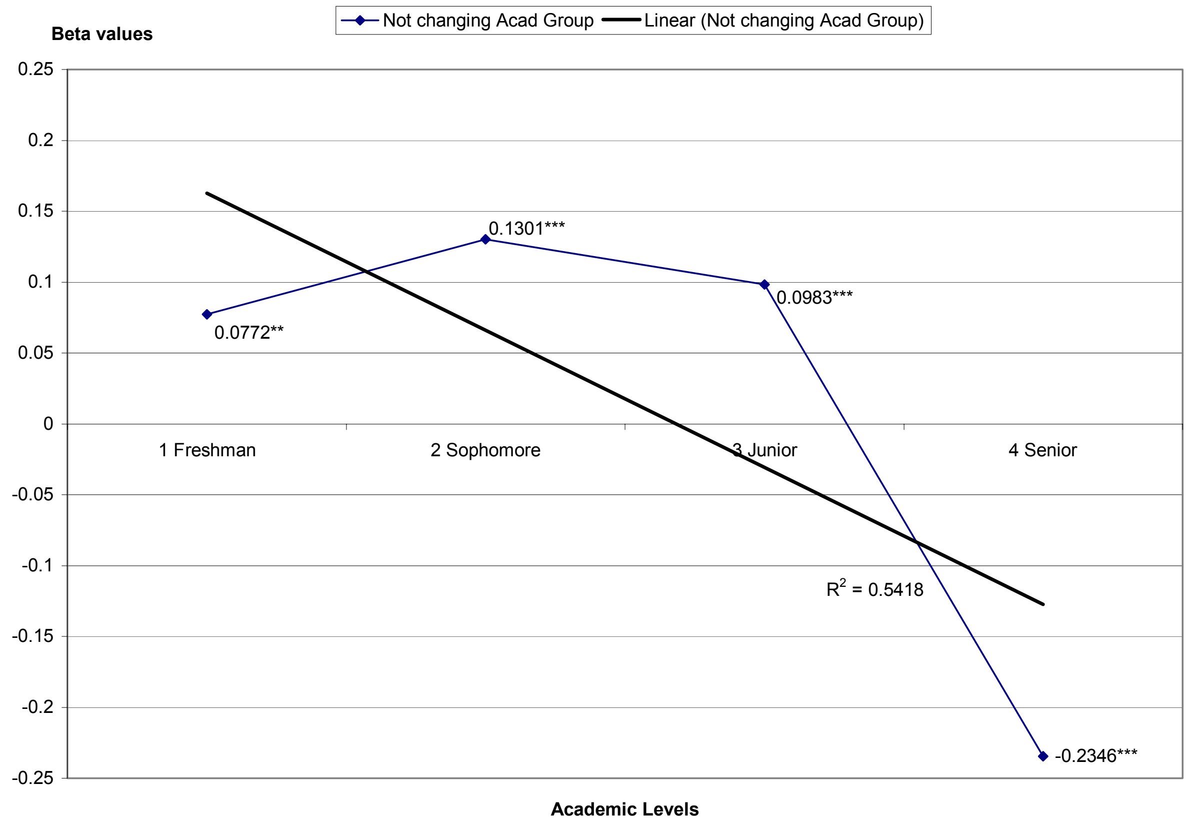
Note: significant p-values: * $P<0.05$, ** $P<0.01$, *** $P<0.001$ & $P<0.0001$

Note: Freshman (n): 21 observations were deleted in the logistic procedure due to missing values
Sophomore (n): 12 observations were deleted in the logistic procedure due to missing values
Junior (n): 4 observations were deleted in the logistic procedure due to missing values
Senior (n): 4 observations were deleted in the logistic procedure due to missing values

<u>Change Academic Group</u>: This variable was entered as a class variable in the model. Logistic regression modeled the effect on the DV of *not* changing academic groups vs. changing academic groups. In interpreting the results for this variable it is important to note that this is the only post-admission variable that it is not time-dependent. Finding out whether or not students changed academic groups in one specific academic level required extensive recoding and beyond the scope of this study.

The (+) sign of the beta coefficients indicates a negative contribution to probability prediction for each state of the dependent variable. In other words, '*not*' changing academic groups predicts lower academic level transitions when compared to changing academic groups. This implies that '*changing*' academic groups is favorable for advancing to higher academic levels. The beta values preserve the order of the (+) sign across the first three academic levels and the weight of the beta values is highest for the sophomore level. At the senior level the sign switch to (-). Thus '*changing*' academic groups no longer favors transitions. As seen in Graph 10, the R^2 of .54 indicates that a linear model explains about half of the variation in the estimated beta coefficients.

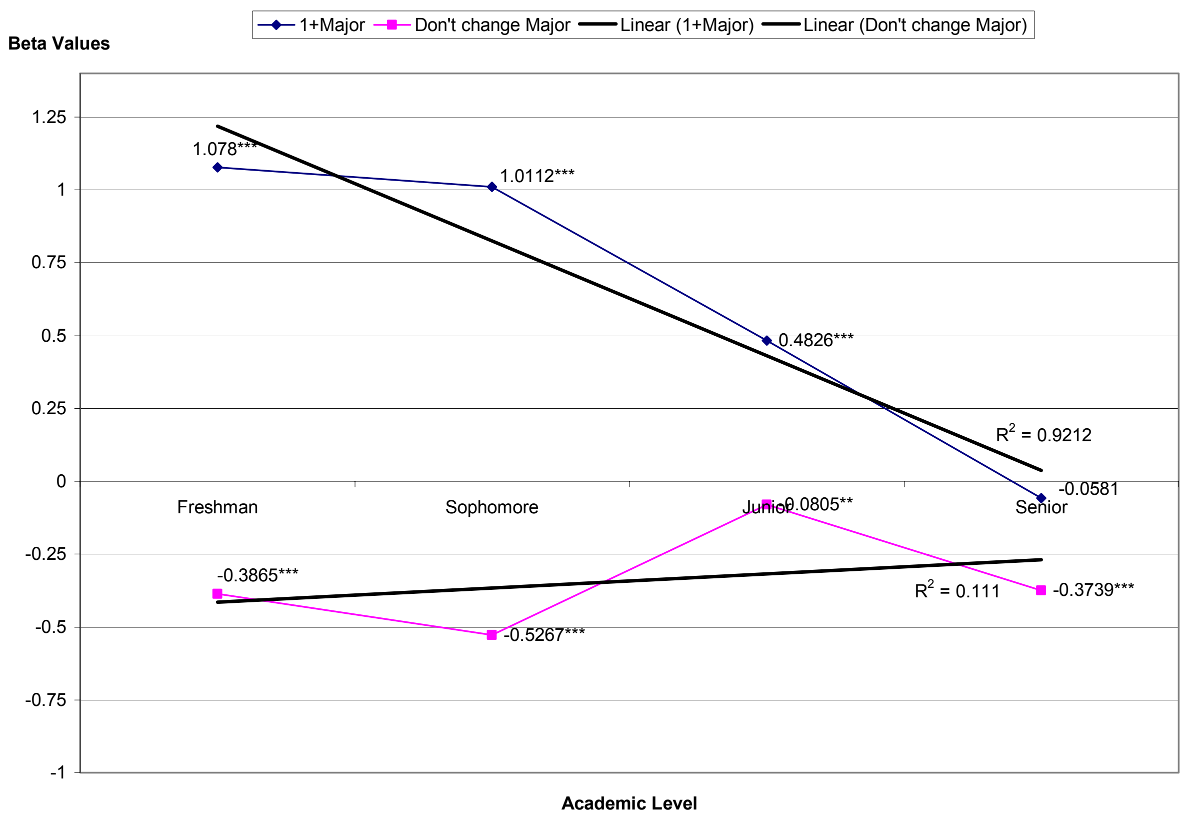
Graph 10 Beta Values for Change Academic Group across Academic Levels



Changing majors: This variable was entered in the model as a class variable. Logistic regression modeled the effect on the DV of having dual majors or more vs. changing majors, and the effect of *not* changing majors vs. changing majors. Having dual majors (i.e., more than one major) has (+) beta values at the freshman, sophomore, and junior levels and these values are statistically significant. This indicates a negative persistence effect on transition probabilities when compared to changing majors, while holding other predictors in the model constant. Thus, lower transition probabilities are predicted for having more than 1 major in the first three academic levels.

On the other hand, the (-) sign for Not Changing Majors predictor indicates a positive effect on transition probabilities. The effect is consistent and significant for all four levels, when compared to changing majors, and while holding all other variables in the model constant. This means that higher transition probabilities are expected when no change of major occurs at either academic level. Graph 11 shows the weight of the beta values for both predictors and how these vary across levels. The effect of having dual majors or more is almost linear over academic levels. Therefore the R-square is high (.921). Not changing majors has more weight at the sophomore level, dropping for students who achieve junior level and picking up some strength at the senior level. These fluctuations make the value of the R^2 smaller, explaining only 11% of the variance.

Graph 11 <u>Beta values for Changing Majors across Academic Levels</u>

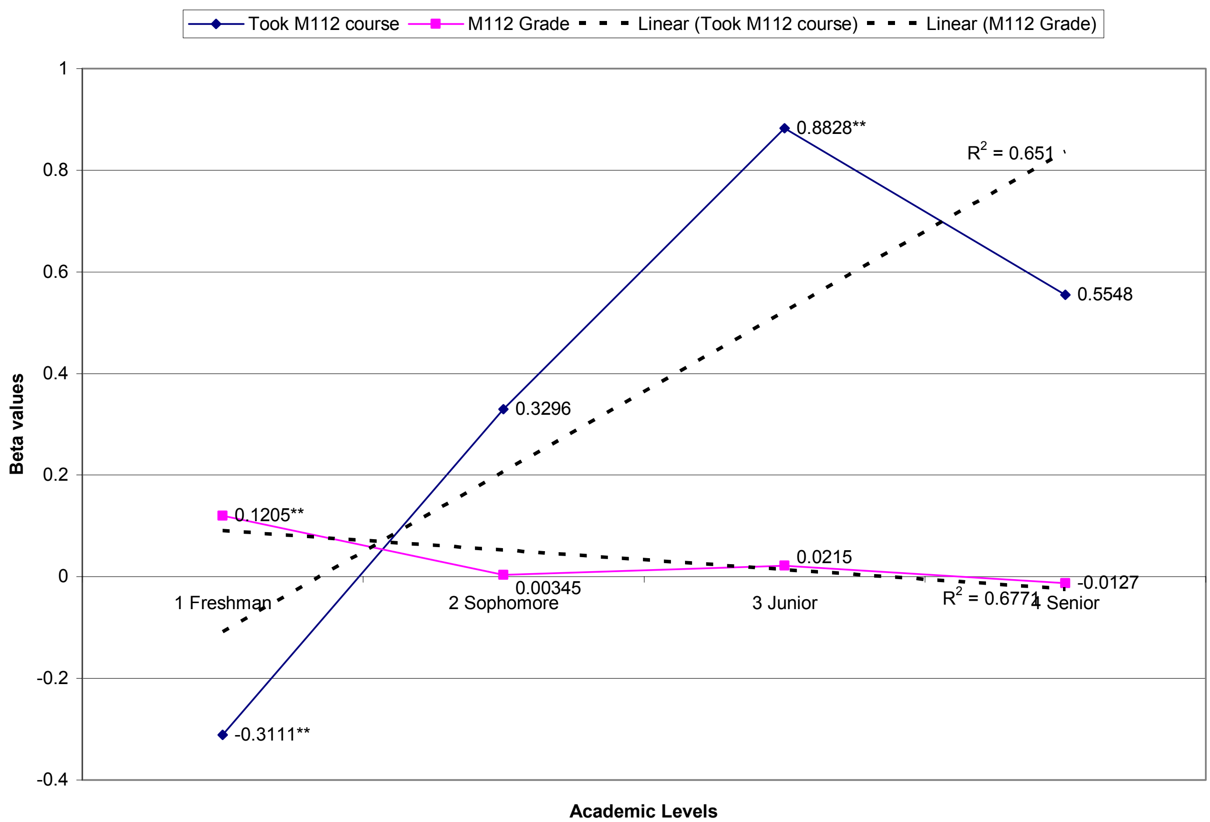


<u>All courses and course Grades</u>: The following paragraphs present the results for the courses and grades. There is one graph per course and grade. The results for the grades need to be read and interpreted with caution because the data is skewed towards showing a negative effect on transition probabilities. This occurs for two reasons: first, the majority of the students do not take these courses. As seen in Table 8b, between 20% and 35% of the students take these courses, and the majority takes them at freshman level, fewer at the sophomore level and even fewer at the junior and senior levels. Second, grades are only for those who take the course at any time, and a zero was assigned to those who did not take the course at all. Therefore, the negative effect of grades is likely the result of the frequent occurrence of zeros for those not taking the course, combined with the positive effect that not taking these courses has on persistence transitions.

<u>M112 Course</u>: Taking this course has a significant effect at the freshman and junior level. The (-) sign of the beta value at the freshman level indicates a positive contribution to persistence transition probabilities. Thus, higher transition probabilities are expected as a function of taking this course at the freshman level, holding other predictor variables constant. The order of the beta value sign changes to (+) and it becomes larger and significant at the junior level, which indicates that lower transition probabilities are expected as a function of taking this course at the junior level holding other variables in the model constant. At the senior level it is smaller and not significant. These deviations at the junior and senior levels affect the linear fit to the data, reducing the R^2 to .65 (see Graph 12). In summary, in order for M112 to positively affect transitions, this course needs to be taken at the freshman level as the positive persistence effect drifts at the junior and senior level.

<u>M112 Grade</u>: The grade received in M112 has a significant effect on transition probabilities at the freshman level, and the effect is close to zero and insignificant for the other levels. The (+) sign of the beta value at the freshman level indicates a negative contribution to transition probabilities. Lower grades, in this case a zero (i.e., not taking the course) favors transition probabilities at the freshman level.

Graph 12 <u>Beta Values for M112 Course and Grade across Academic Levels</u>

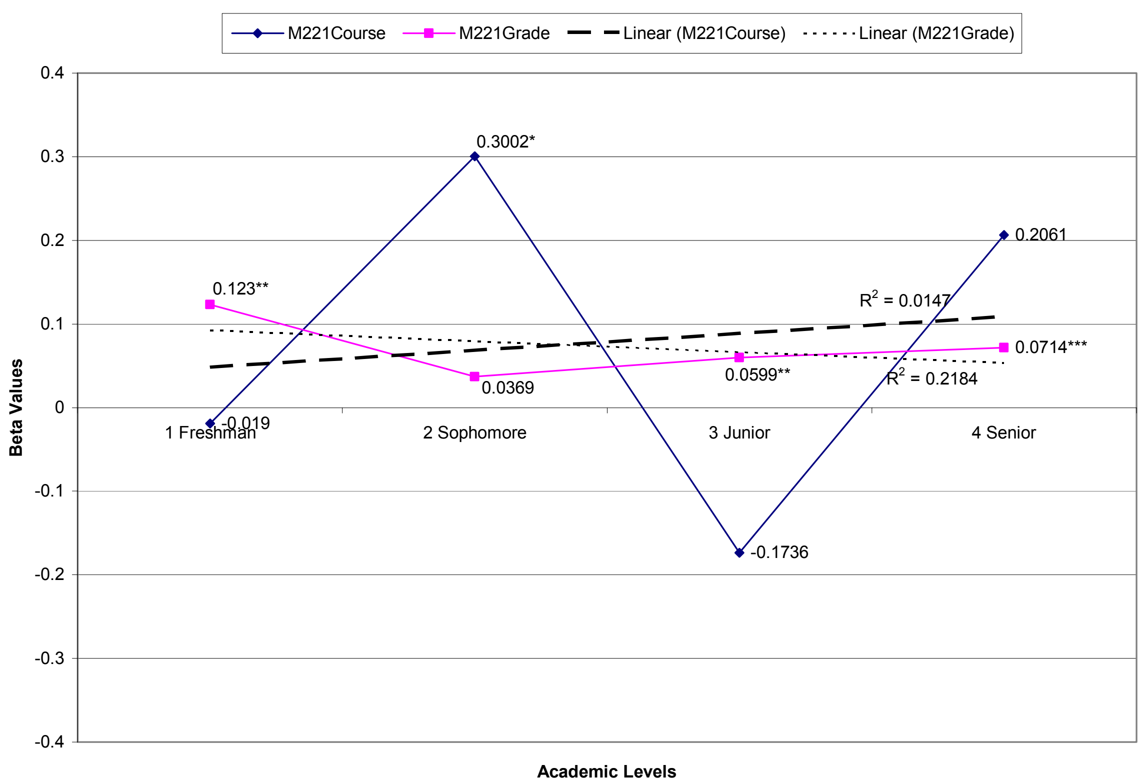


<u>M221 Course</u>**:** taking this course has a significant negative effect on persistence transition probabilities when the course is taken at the sophomore level. The (+) sign of the beta value at the sophomore level indicates a negative contribution to transition probabilities. Thus, lower academic level transition probabilities are expected as a function of taking this course

at the sophomore level, other predictor variables held constant. The order of the beta signs varies from level to level, which affect the fitted line reducing the R-square (R^2 = .015), not surprising since the beta is only significant at one academic level. The values of the parameter estimate suggest that taking this course at the freshman and junior level is more likely to favor persistence (see Graph 13). This is given as a tentative statement because the betas at these levels are not statistically significant.

M221Grade: : The grade received for M221 has a significant effect on transition probabilities at the freshman level, junior and senior levels, and while holding other variables in the model constant. However, the (+) sign for the beta values at these three levels, indicates a negative contribution to transition probabilities. The effect is small for all levels.

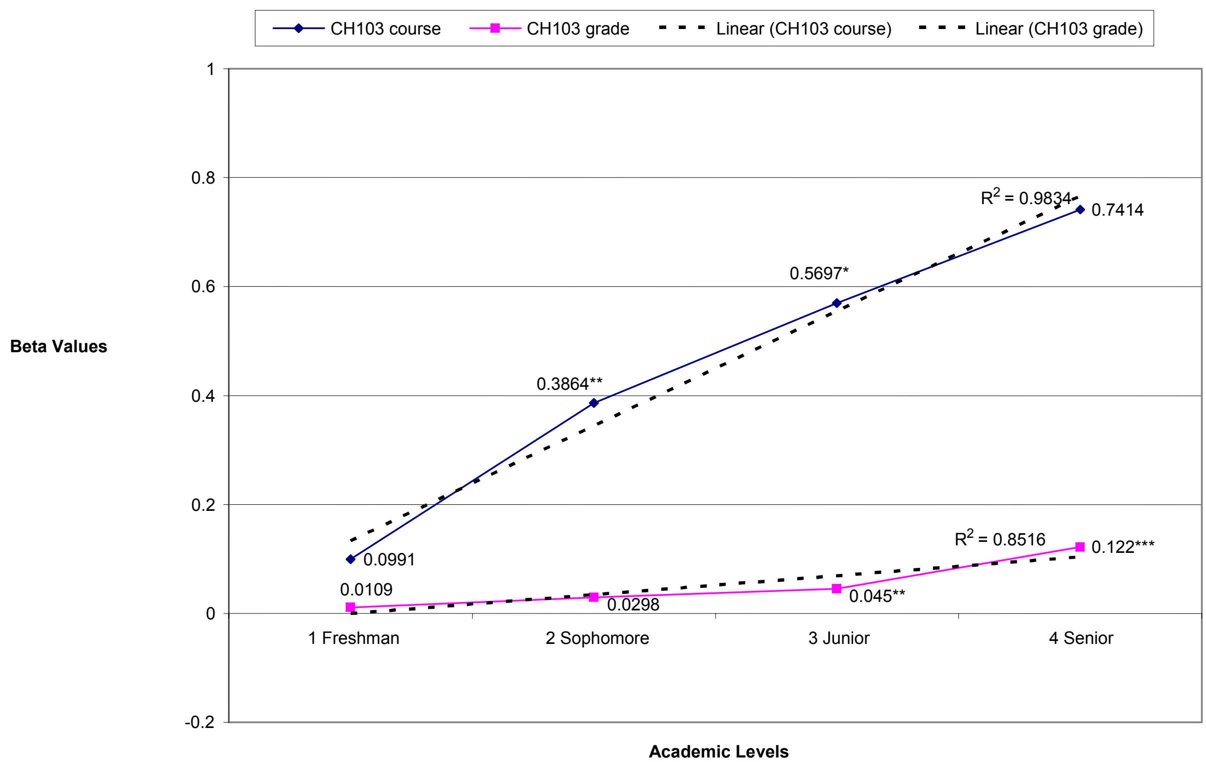
Graph 13 Beta Values for M221 Course and Grade across Academic Levels



CH103 Course: Taking this course has a significant effect on transition probabilities when taken at the sophomore and junior levels. The (+) sign of the beta value at these levels indicates a negative contribution to transition probabilities. Thus, lower academic level transition probabilities are expected as a function of taking this course at the sophomore and junior levels, holding other variables in the model constant. The beta values maintain the (+) sign across all four levels. The values increase steadily in almost a linear trend $R^2 = .983$ The values of the parameter estimate from this predictor suggest that taking this course at the sophomore and junior levels (see Graph 14) results in reduced probability of persisting.

CH103 Grade: The grade received inCH103 has a significant effect on transition probabilities at the junior and senior levels, while holding other variables in the model constant. The effect is the largest at the senior level. The (+) sign for the beta values at these two levels, indicates a negative contribution to transition probabilities.

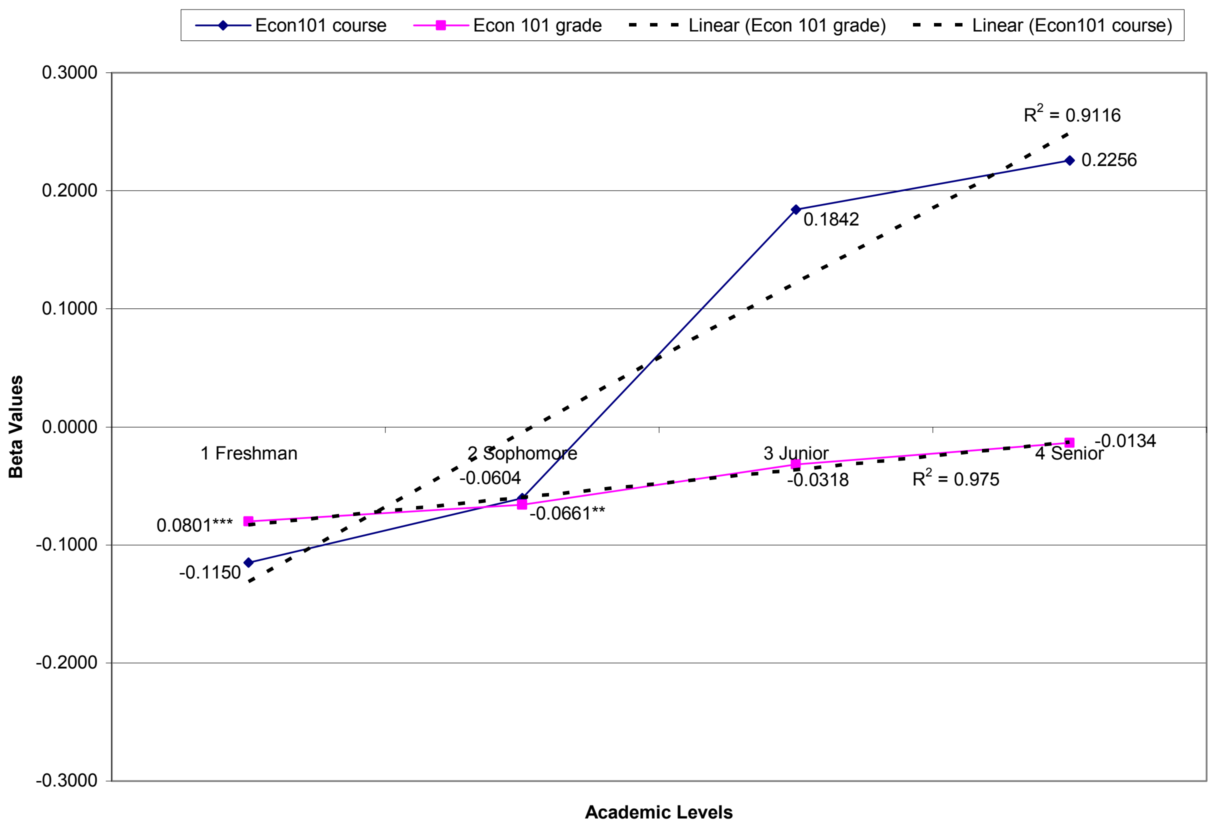
Graph 14 Beta Values for CH103 Course and Grade across Academic Levels



ECON 101 Course: Taking this course at a particular academic level does not have a significant effect on transition probabilities The values of the parameter estimate from this predictor suggest that this course is irrelevant for predicting the probabilities of persisting or not (Refer to Table 13)

ECON 101 Grade: The grade received in Econ101 has a significant effect on transition probabilities at freshman and sophomore levels, while holding other variables in the model constant. The effect is very small in all levels. The (-) sign for the beta values at these two levels indicates a positive contribution to transition probabilities.

Graph 15 Beta Values for Econ 101 Course and Grade across Academic Levels

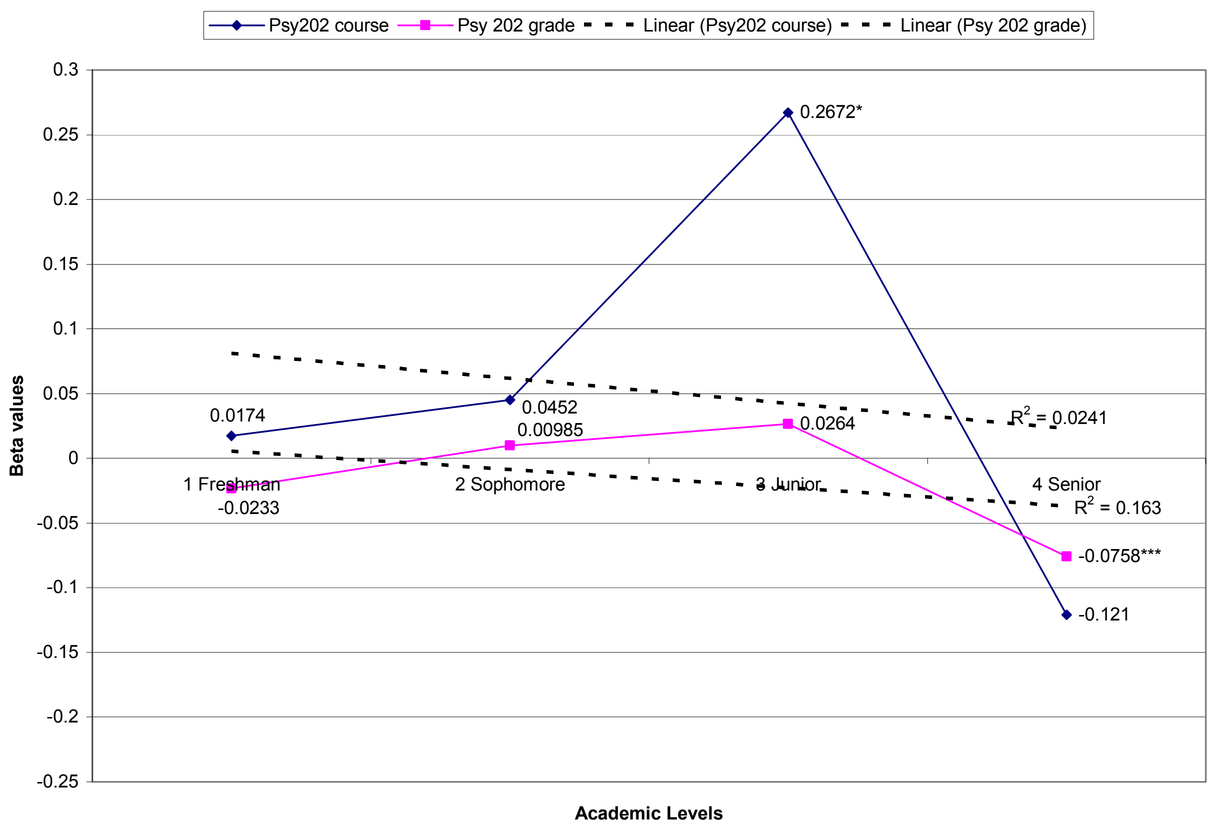


Psy202 Course taking this course has a significant effect at junior level. The (+) sign of the beta value at the junior level indicates a negative contribution to transition probabilities. Thus, lower transition probabilities are expected as a function of taking this course at this level, holding other predictor variables in the model constant. The order of the beta value sign changes from (+) in the first three levels to (-) but not significant in the senior level. In summary, this result indicates that for Psy 202 negatively, affects transition probabilities when taken at junior level. The effect change at the junior level is significant and of larger weight when compared to other levels. This variation contributes to the

reduction of the R^2=.024 when a linear model it fitted on the estimated beta coefficients (see Graph 16).

<u>Psy202 Grade</u>: The grade received in Psy202 has a significant effect on transition probabilities at the senior level, although the effect is small. The (-) sign of the beta value at the senior level indicates a positive contribution to graduating transition probabilities.

Graph 16 <u>Beta Values for Psy 202 Course and Grade across Academic Levels</u>



Confirmation of hypothesis 2 and hypothesis 2.1: test of the significance for the beta coefficients

The same binomial probability distribution used to compute the probability of k significant betas from a population of n binary variables used for the pre-admissions variables, was used to test the betas in the post-admissions variables. For the 14 post-admission variables, using the 0.05 level of significance, the probability of rejecting the null hypothesis is p = 0.05 and the probability of acceptance of the null hypothesis is q = 0.95 for each of the n = 14 post-admission variables. Thus, the probability of observing k significant post-admission variables by chance is found using equation 4.1. Writing out the binomial coefficient and using p=0.05 and q=0.95, this becomes:

$$\mathrm{P(k)} = \frac{14!}{k!(14-k)!}(.05^{k})(.95)^{14-k} \tag{4.5}$$

Following the same procedure as with equations 4.3 and 4.4, the probability of observing more that *j* significant betas by pure chance can be expressed symbolically as:

$$\sum_{i=j+1}^{14} P(k=i) = 1 - \sum_{i=0}^{j} P(k=i) \tag{4.6}$$

For the post-admissions variables, the academic level with the lowest number of significant betas is the senior level (i.e., 6 significant betas), therefore, it is reasonable to use the probability that more than 3 significant betas occur by chance to reject the null hypothesis. The four probabilities P(k=0), P(k = 1), P(k=2) P(k=3) computed are:

$$P(k=0) = \frac{14!}{0!(14-0)!}(.05^0)(.95^{14-0}) = .4877$$

Using the same equation:

P (k=1) = .3593
P (k=2) = .1229
P (k=3) = .0259

Which when summed equal 0.4877 + 0.3593 + 0.1229 + 0.0259 = 0.9958 From this it can be concluded that the probability of observing four or more significant betas by pure chance from a total of 14 variables is:

$$\sum_{i=4}^{14} P(k=i) = 1 - .9958 = .0042$$

Thus, it can be concluded that the probability of four or more significant betas happening by chance alone is 0.0042 or less. This is the same as saying that if we have four betas significant at the 0.05 level, we can be 99.58% confident that they did not occur by chance. This allows us to reject the null hypothesis and conclude that post-admission variables have an effect on transition probabilities (i.e., something other than chance is operating). As with the pre-admission variables, it is important to remember that because the binomial distribution assumes independence and many of the variables in this study may be somewhat correlated, the above probability estimates must be considered as approximations.

Simulation Application

The final set of analyses was conducted to examine the predicted transition probabilities for any given set of score profiles for a four academic level trajectory. These simulations serve two purposes: first, to show how a higher education institution might use the model to evaluate the effects of policy decisions on flow transition probabilities across academic levels; and second, to demonstrate that the interpretation of the direction of the effect (i.e., the beta sign (+) or (-) from the parameter estimates is being done correctly.

Only one simulation will be run for each Policy Option. To really investigate a policy, an institution would run multiple simulations with varying student characteristics in order to establish appropriate benchmarks for a given group of predictors. The first policy option will be followed by an example of how the probabilities presented would be calculated in the absence of the simulation tool.

<u>Policy option 1</u>: maintaining the current age at entrance of FYR students at 18 years old.
<u>Institutional question</u>: Given an exact set of characteristics for two students, what are the probabilities for a typical four academic level trajectory, of two individuals aged 18 and 25 years to stop-out, remain at the same academic level, advance to the next level or graduate? Given their age differences, how will their probabilities of graduation look at 4 and 6 years?

Students' hypothetical score profiles are presented in Table C-1. Note in these profiles that both students are white, female, not a targeted minority, not a Pell grant recipient, have a HS ranking of 75%, and a NEW_ACT score of 23. As freshmen, they both had a GPA of 3.0, did not change groups, did not change majors, took math 112, received a grade of 4, did not take m221 so received grade of 0, took ch103 and received a grade of 3, took econ101 and received a grade of 3, and did not take Psy202 so received a grade of 0. As sophomores, they earned a GPA of 3.1, did not change academic groups, did not change majors, took m221 and received a grade of 4, and took p202 and received a grade of 4. As juniors, they earned a GPA of 3.2, did not change academic group or major and did not take any of the courses so received grades of 0. As seniors, they earned a GPA of 3.3, did not change academic group or major and did not take any of the courses.

The effect of age is shown by comparison of the transition probability matrices in Table 14, for the student 18 years old and Table 15 for the student age 25. As was presented in the results, Age_at _Entrance had a positive and significant beta weight in all academic levels (see Table 9), so the predicted transition probabilities are expected to be lower for the older student.

Table 14 Transition probability matrix for 18 year old student

	F	S	J	Sr.	S_O	G	Probability
F	0.0369407	0.8995926	0.0246998	0.000	0.0387669	0.000	1.000
S	0.000	0.1087618	0.8127312	0.0042579	0.0742491	0.000	1.000
J	0.000	0.000	0.1060318	0.8709348	0.0210257	0.0020077	1.000
Sr.	0.000	0.000	0.000	0.2152252	0.0086526	0.7761223	1.000
S_O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0.000	1.000	1.000

Table 15 Transition probability matrix for 25 year old student

	F	S	J	S	S-O	G	Probability
Fr	0.0769262	0.8232754	0.0102817	0.000	0.0895167	0.000	1.000
S	0.000	0.2311132	0.5208614	0.0010446	0.2469808	0.000	1.000
J	0.000	0.000	0.2308081	0.7124425	0.0560219	0.0007275	1.000
Sr.	0.000	0.000	0.000	0.3759428	0.0193996	0.6046576	1.000
S_O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0.000	1.000	1.000

The numbers in the first row (Table 14) indicate that the student who is 18 years old at entry has a predicted probability of remaining a freshman of 0.0369407, whereas the 25 year old has a probability of 0.0769262 (Table 15). The 18 year old at entry had a predicted probability of advancing to the sophomore level of 0.8995926 (Table 14) compared to 0.8232754 (Table 15) for the 25 year old. The 18 year old had a predicted probability of advancing from freshman to junior level in one academic year of 0.0246998, compared to 0.0102817 for the 25 year old. The 18 year old had a predicted probability of stopping out during the first year of 0.0387669 compared to 0.0895167 for the 25 year old. Therefore, as age increases and all other variables in the model are held constant the transition probabilities for academic level decrease, and the probabilities of remaining in the same academic level or stopping out increase.

The probabilities in the 6x6 matrix shown in Tables 14 and 15 were obtained from a SAS/IML program (IML is a SAS based programming language), which calculates the expected transitions using all 25 beta values for the four logistic regression models (Tables B-1 to B-4). The following paragraphs present an example of how probabilities could be calculated by hand and with the use of formulas.

Example: how to calculate probabilities by hand

Say, for example, that one is calculating the probabilities for the first row in Table 14. Data for this row was obtained for the freshman dataset. In that dataset, the four states of the dependent variable were ordered as 1=Stop out, 2=FF, 3=FS and 4=FJ. Response level ordering is important when using SAS because, “PROC LOGISTIC always models the probability of response levels with *lower* Ordered Value. By default, response levels are assigned to Ordered Values in ascending, sorted order (that is, the lowest level is assigned

Ordered Value 1, the next lowest is assigned 2, and so on)" (Response Level OrderingSAS, 1999). The ascending default order was used based on the logical nature of the dependent variables, with stopping out exhibiting the least persistence, staying in the same grade the next, moving one grade as the next level, and moving two grades as the most persistence.

For this example, as seen in Table B-1 there are three intercepts: Intercept F-StopOut=2.8240, intercept F-F= 3.5325, and intercept F-So= 9.7106. The highest ordered value F-Jr is the reference variable by default, so there is no associated intercept. In order to compute transition probabilities for a given vector of scores on the 25 independent variables for a given subject, a value for $\sum_{i=1}^{25} \beta_i \cdot X_i$, needs to computed, where X_i is the subject score for the i^{th} independent variable and β_i is the corresponding beta value [e.g., $\beta_1 X_1$= 0.1273 (beta coefficient for age) × 18 (age at entrance)]

For a selected profile of independent variable scores (i.e., the first row of scores in Table C-1) $\sum_{i=1}^{25} \beta_i \cdot X_i$ = -6.03. Given this information, the transition probabilities can be computed.

First, a linear regression score needs to be computed for each of the k states of the dependent variable according to the familiar linear equation, for k=1,2,3:

$$\alpha_k + \sum_{i=1}^{25} \beta_i \cdot X_i$$

Using the estimates of the intercepts from the SAS® output and $\sum_{i=1}^{25} \beta_i \cdot X_i$ =-6.03, we have:

k = 1 2.8240 + -6.03 = -3.2060

k = 2 3.5325 + -6.03 = -2.4975

k = 3 9.7106 + -6.03 = 3.6806

The formula for computation of the cumulative probability for each of the ordered states of the dependent variable is

$$\Pr(State \leq k) = \frac{1}{1 + \exp[-(\alpha_k + \sum_{i=1}^{25} \beta_i \cdot X_i)]}$$

Thus, for state 1 (k=1), Pr(State ≤ 1) = 1/(1 + exp(3.2060) = .0389406;

for state 2 (k=2), Pr(State ≤ 2) = 1/(1 + exp(2.4975) = .0760336;

for state 3 (k=3), Pr(State ≤ 3) = 1/(1 + exp(-3.6806) = .975412.

By the rules of probability, we know that Pr(State ≤ 4) = 1.00000.

Knowing also that: Pr(State = 1) = .0389406 and Pr(State ≤ 2) = .0760336 the following can be deduced: Pr(State = 2) = .0760336 - .0389406 = .037093.

By similar reasoning, Pr(State = 3) = .975412 - .0760336 = .899378; and Pr(State = 4) = 1.000000 - .975412 = .024588.

As a check, we note that .038941 + .037093 + .899378 + .024588 = 1.000000 as it should be since all four transition probabilities have been accounted for. Finally, it can then be stated that: Pr (Stopout) = .038941; Pr(FF) = .037093; Pr(FS) = .899378; and Pr(FJ) = .024588.

Note that these four estimated probabilities allow to fill in the first row of the transition probability matrix associated with the given score profile **X** that produced

$\sum_{i=1}^{25} \beta_i \cdot X_i$ =-6.03.

Proceeding in the same manner for each of the remaining three datasets allows us to estimate the remaining rows of state transition probabilities in the matrix of transition probabilities. The particular state transition is a function of the score vector **X** and will change as **X** is altered.

Four year and six year transition probabilities

Up to this point a state transition probability matrix (**TP)** has been completed for all four dataset, and these are yearly steps transitions. Thus, to find out the two-step (read two year) transition probabilities, the **TP** matrix would be raised to the second power, $(\mathbf{TP})^2$, and so on for each additional year. So to find out the six year transition probabilities, (**TP)** would be raised to the sixth power, $(\mathbf{TP})^6$. Each entry in this matrix is the probability that a subject with score profile **X** who started in a given row state will be in the corresponding column state after six years. As written in the IML simulation program, the $(\mathbf{TP})^4$, $(\mathbf{TP})^5$ and $(\mathbf{TP})^6$ are part of the output. Tables E-1 and E-2 in Appendix E present the 4 year and 6 year transition probabilities respectively for policy option 1. Raising the annual transition matrix by powers relies on the inherent Markov assumption that the annual transition matrix remains constant over time.

For example, following up on policy 1, to find the estimated probability of starting as a freshman and graduating after being in school for up to six years, we can look in the $(\mathbf{TP})^6$ matrix in the first row and last column, which corresponds to the graduating state. The estimated probability is (P=0.8273162) which compared with the $(\mathbf{TP})^6$ matrix (Table E-2) for the student 25 years old is (P=0.4657288). This means that the probability that a 25 year old student will graduate within six years is only a little over half the probability for an 18 year old student. In this sense the simulation tool puts a risk value to being 7 years older than the average age at entrance.

To emphasize the results from four, five and six year transitions, a comparison can be made with the descriptive data in this study based on the data shown in (Graph 1). From the three cohorts (N=16,345), it is known that 9,206 of the students (56.3%) graduated in $\leq$ 4 years. These numbers include students of all ages. In the 18 year old four step transition matrix (Table E-1 first row, last column) the probability of starting as a freshman and graduating after four years is 0.5228237, so the predicted result is comparable to the actual graduation rate. The results in these matrices are also comparable with time to degree percentages reported by (Choy, 2002) for 1992-93 graduates.[3] According to her results, 78% of the students who were ≤20 years old at entrance graduated in 4 to 5 years, compared to 35% of those who were 21-24 at entrance. The five step transition matrix for those entering at 18 years of age shows a probability of 0.7555078 and for those entering at 25 years of age is 0.3642498, again comparable numbers

Policy option 2: Identify at the time of admissions students who are at risk of taking longer to graduate (here, at risk means a certain population who demonstrates a greater tendency to take longer to graduate).

Institutional question: Is there an empirical way of finding out whether a university with strict admission policies has an at-risk population?

---

[3] The source of Choy (2002) data comes from the B&B/94 data.

Students' hypothetical score profiles for policy 2 are given in Table C-2. Note in these profiles that both students are 18 years old. However, student A is African American and student B is White. Student A and student B are females. Student A is a Target Minority and student B is not. Student A receives a Pell Grant and student B does not. Students A and B have the same HS% of 80%, a New_ ACT score of 24, and their score profile for the post-admission variables is the same. The post admission variables are the same as the ones from the students in Policy option 1. As freshmen they both had a GPA of 3.0, did not change groups, did not change majors, took math 112, received a grade of 4, did not take m221 so received grade of 0, took ch103 and received a grade of 3, took econ101 and received a grade of 3, and did not take Psy202 so received a grade of 0. As sophomores, they earned a GPA of 3.1, did not change academic groups, did not change majors, took m221 and received a grade of 4, and took p202 and received a grade of 4. As juniors, they earned a GPA of 3.2, did not change academic group or major and did not take any of the courses so received grades of 0. As seniors, they earned a GPA of 3.3, did not change academic group or major and did not take any of the courses.

The effect of the differences in pre-admission variables between students A and B is shown by comparison of the transition probability matrices in Table 16 for student A and Table 17 for student B. The differences in probabilities, starting at the sophomore level and becoming large by the senior level, are an empirical proof that student A is at a higher risk of not graduating than student B. So, just as the literature suggested student A has pre-admission characteristics which decrease her probabilities of persisting.

Table 16 Transition probability matrix for student A

	F	S	J	Sr	S-O	G	Probability
F	0.022557	0.9128013	0.0417324	0.000	0.0229157	0.000	1.000
S	0.000	0.1244764	0.7838135	0.003534	0.0881762	0.000	1.000
J	0.000	0.000	0.1434655	0.825094	0.0300477	0.0013928	1.000
Sr.	0.000	0.000	0.000	0.4665776	0.0288525	0.5045699	1.000
S-O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0000	1.000	1.000

Table 17 Transition probability matrix for student B

	F	S	J	Sr	S-O	G	Probability
F	0.0321285	0.9057683	0.0287555	0.000	0.0333477	0.000	1.000
S	0.000	0.1045769	0.8202124	0.004486	0.0707247	0.000	1.000
J	0.000	0.000	0.1069928	0.8697756	0.021245	0.0019866	1.000
Sr.	0.000	0.000	0.000	0.2378722	0.0098663	0.7522615	1.000
S_O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0.000	1.000	1.000

The numbers in the first row of Tables 16 and 17 indicate that student A has a predicted probability of remaining a freshman of 0.022557, about a third lower than the probability of 0.0321285 for student B. Student A had a predicted probability of advancing to the sophomore level of 0.9128013, slightly higher than the probability of 0.9057683 for student B. Student A had a predicted probability of advancing from freshman to junior level in one academic year of 0.0417324, again slightly higher than the 0.0287555 probability for student B. Student A had a predicted probability of stopping out after acquiring freshman status of 0.0229157, which again is about a third lower than the 0.0333477 probability for student B. Up to the first year, then, being a female, African-American, Target minority and Pell Grant receiver predicts slightly <u>higher</u> retention and advancement probabilities than being a White female, not Targeted Minority and not a Pell Grant receiver, holding other variables in the model constant.

However, after student A completes her freshman credits, the transition probabilities start reducing in comparison to student B, while holding other variables constant. Note for example that the probability of staying at junior level for student A is 0.1434655, higher than that of student B at the same level (0.1069928). This trend continues until the probability of student A graduating after having achieved senior credits (0.5045699) is much lower than the probability for student B (0.7522615).

Table 18 shows the calculated probabilities that a student entering as freshman with A and B characteristics will stop out within 4, 5 and 6 years, and the probabilities of graduating. Also, there is a column for the differences. Data was pooled from the first row and last two columns of the matrices $\mathbf{(TP)^4}$, $\mathbf{(TP)^5}$, $\mathbf{(TP)^6}$ (Appendix E, Tables E-3 and E-4).

Table 18 Probabilities of stopping out and graduating from $(TP)^4$, $(TP)^5$, $(TP)^6$

Entering as Freshman Probability of S_O	Student A ↑risk	Student B ↓risk	Difference
Within 4 years	0.1653199	0.1353598	0.03
Within 5 years	0.1800545	0.1392197	0.04
Within 6 years	0.1875275	0.1403285	0.05
Entering as Freshman Probability of Graduating			
Within 4 years	0.3302071	0.518104	0.19
Within 5 years	0.5618165	0.7545591	0.19
Within 6 years	0.687944	0.8309627	0.14

As seen in the second column of Table 18, at 4 years student A has a little over 16% chances of stopping out, at 5 years her chances increase to 18% and to nearly 19% at 6 years. For student B the chance of stopping out starts lower (13.5% after four years), and the yearly increment is smaller as well. In general, the probabilities for student A or B of stopping out do not increase much more after the fourth year, and the increment is not linear from 4 to 5 to

6 years. However, it is evident that for student A the probability of stopping out within 6 years is about a third higher than that of student B. Also, interesting is the difference in graduation probabilities for students A vs. B. Table 18 provides two important facts about graduation for these students. First, the numbers in the table show that allowing two more years than the traditional 4 years increases the probabilities of graduating for both students. For student A, the probability more than doubles, from 0.330 in four years to 0.688 in six years. The increase from years four to six for student B is not quite as high, but the probability of graduating after 6 years relative to 4 years also increases. Second, the graduation probabilities are always higher for student B when compared to student A, although the difference gets smaller after 6 years. These results provide empirical evidence that in this university with strict admission policies students with the characteristics of student A could be considered an at risk population, since about 17% fewer of these students are expected to graduate.

<u>Policy option 3</u>: For retention purposes, identify protective variables for students who are at risk of showing a greater tendency to stop out or take longer to graduate.

<u>Institutional question</u>: What can the school do to foster academic resilience?

Students' hypothetical score profiles for policy 3 are given in Table C-3. Note in these profiles that both students have identical pre-admission variables: they are 19 years old, belong to the same ethnic group, American Indian, are both males and are both Target Minorities. Neither student received a Pell Grant Pell, their HS% is 71%, and their ACT is 26. Their score profile for the post-admission variables is different. As freshmen, student X got a GPA of 2.7 while student Y got a GPA of 3.0. They did not change academic groups, had only one major, and did not change majors. Student X took math 112, received a grade of 2.8, and did not take any other critical courses while being a freshman. Student Y took M112 and received a grade of 3, and did not take any other critical courses. As sophomores student X earned a GPA of 2.8 and student Y earned a GPA of 3.0. Student X still had one major, and did not change majors. Student Y also kept one major, and did not change majors, but changed to another college (changed academic groups) (Note: realistically speaking this student would eventually have to select another major). Student X took Math 221 and received a grade of 2.9, and did not take any other critical courses. Student Y took Econ101 and received a grade of 3 and did not take any other critical courses. As juniors, student X earned GPA of 3.0, did not change academic group, continued with one major, but changed majors. Student Y earned a GPA of 3.1, did not change academic groups, did not pick or changed majors. Student X took CH103 and received a grade of 3, student Y did not take any critical courses while being a junior. As seniors, student X earned a GPA of 3.1, and student Y a GPA of 3.2. They did not change academic groups, did not pick another major or change majors. Student X did not take any critical courses and Student Y needed a quick number of credits so he took PSY 202 and received a grade of 3. The effect of the differences in post-admission variables between student X and Y is shown by comparison of the transition probability matrices in Table 19 for student X and Table 20 for student Y.

Table 19 Transition probability matrix for student X

	F	S	J	S	S-O	G	Probability
F	.0692692	0.8400683	0.0117874	0.000	0.0788751	0.000	1.000
S	0.000	0.1642569	0.7036814	0.0022946	0.1297671	0.000	1.000
J	0.000	0.000	0.2615869	0.6703258	0.0674907	0.000596	1.000
Sr.	0.000	0.000	0.000	0.5084705	0.0347315	0.456798	1.000
S-O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0.000	1.000	1.000

Table 20 Transition probability matrix for student Y

	F	S	J	S	S-O	G	Probability
F	0.0484325	0.8811358	0.018194	0.000	0.0522377	0.000	1.000
S	0.000	0.0651102	0.8862447	0.0080376	0.0406074	0.000	1.000
J	0.000	0.000	0.1258343	0.8468571	0.0256714	0.0016372	1.000
Sr.	0.000	0.000	0.000	0.4156914	0.023112	0.5611966	1.000
S_O	0.000	0.000	0.000	0.000	1.000	0.000	1.000
G	0.000	0.000	0.000	0.000	0.000	1.000	1.000

The numbers in the first row of Table 19 indicate that student X has a predicted probability of remaining a freshman of 0.0692692, nearly 50% higher than student Y with a probability of 0.0484325 (Table 20). Student X had a predicted probability of advancing from the freshman to the sophomore level of 0.8400683, lower than the probability of 0.8811358 for student Y. Student X had a predicted probability of stopping out during the first year of 0.0788751, higher than 0.0522377 for student Y. In the first year, student Y has higher chances of advancing and lower chances of stopping out, holding other variables in the model constant. The differences are not very high, because we know from the beta coefficients that GPA is one of the stronger predictors increasing transition probabilities towards graduation and these two students have similar GPA's and took the same course receiving a similar grade.

Doing a similar comparison for the sophomore level, we see that post-admission differences affected student X by tripling his probability of stopping out at the sophomore level (0.1297671) when compared to the probability for student Y (0.0406074). The probability of advancing to the junior level for student X is 70%, compared to 88% for student Y. In general terms, lower probabilities towards persisting are predicted for student X when compared to student Y. After the students enter the third academic level the differences in transition probabilities are about as high as in the previous level, and are still lower for student X. Note for example, how student X has a 67% chance of advancing to the senior level and student Y has an 85% chance. After achieving senior status, 11% more students with the characteristics of student Y would be expected to graduate.

In doing this example, the intention was to show how small changes (e.g., changing academic groups, not taking certain courses) increase the persistence probabilities for student Y. Of course these changes could have been exaggerated if the differences in CUM_GPA would have been larger all along, since we know that this is one of the strongest post-admission predictors.

The differences in graduation rates between student X and Y are more evident if the analysis is done by years, using the $\mathbf{(TP)^4}$, $\mathbf{(TP)^5}$,$\mathbf{(TP)^6}$ matrices. Referring to Tables E-3 and E-6, one can see that after four years a student with X characteristics has more than double the probability of stopping out, for student X (P=0.2991) and for student Y (P=0.1366). Similarly, the probability of student Y starting as a freshman and graduating after four years is double (P=0.3925) that of student X (P=0.1897). In the 4, 5, and 6-year matrices we can also see that, similar to the previous examples, the probabilities of graduating increase for both student X and Y as more years are allowed for graduation. However, after 6 years it is estimated that nearly 33% of the students with the characteristics (pre-and post admission variables) of student X have stopped out, while only 15% of the student with the characteristics of student Y have stopped out. In terms of graduation, students with X characteristics have only about two-thirds the probability of graduating (P=0.498553), compared to students with Y characteristics (p=0.7562195). Table 21 presents a summary of the differences in graduation probabilities for entering freshman with X and Y characteristics after 4, 5, and 6 years.

Table 21 Probabilities of graduating for student X and Y, from $(TP)^4$, $(TP)^5$ & $(TP)^6$

Entering as Freshman Probability of Graduating	Student X ↑risk	Student Y ↓risk	Difference
After 4years	0.1897192	0.3925725	0.20
After 5 years	0.3736785	0.6389802	0.26
After 6 years	0.498553	0.7562195	0.26

In this table we can see how allowing two more years beyond the traditional four greatly increases the probabilities of graduating for both types of students. Also, since the different post-admissions paths for students X and Y result in greatly different 4 and 6 year graduation probabilities, this example shows how the simulation tool can provide insight into what post-admission variables might serve as protective factors.

In summary, the findings demonstrate how, depending on the type of effect the parameter estimates have on the dependent variable, a combination of pre- and post-admission variables predict lower or higher transition probabilities and lower or higher graduation rates. As exemplified in Policy Option 1, Age_at_Entrance had positive and significant beta weights at all academic levels, so the predicted transition probabilities were lower for the older student. In Policy Option 2, where both students differ only in their pre-admission variables, the effect of being an African American, Targeted Minority, female, receiving Pell Grants (student A) vs. being a white, female, not a Target Minority, not receiving Pell Grants (student B), was slightly higher on transition probabilities for student A than for student B in the first year. However, after student A completes her freshman credits, the transition probabilities start lowering in comparison to student B and while holding other variables constant. These differences were expected based on Pell_Grant's positive and significant beta weight for all academic levels. Also, at the senior level the large, significant and positive beta value for Target Minority predicted lower transition probabilities. In Policy Option 3, where the students differ only in their post-admission variables, having a higher GPA in the first semester, changing academic groups, and taking critical courses when the effect on transitions was expected to be positive on transitions, did in fact increase transition probabilities for student Y when compared to student X. In the same sense, changing majors and taking some critical courses are detrimental to the transition probabilities of student X when compared to student Y. The difference is more evident at the sophomore level, when student X signs up for one of the critical courses M221. As expected, this increases the student X's probability of stopping out.

By comparing graduation rates after 4, 5 and 6 years we can see how the difference in graduation rates between a pair of students can be interpreted as a risk value. As seen in the results, the student who was more 'at risk' (e.g., have at risk characteristics, in policy option 1 this is the older student, in policy option 2 it is student A, and in policy option 3 it is student X) had a higher 'risk value' of lower graduation probability than his/her counterpart.

CHAPTER 5: SUMMARY & DISCUSSION OF RESULTS

This study employed a Markov student flow model to estimate the transition probabilities of stopping out, staying at the same academic level, or advancing to the next higher academic level up to graduation, for first year undergraduate students at a large Midwestern university. A core research question guiding the initiation of this study was to understand what type of students are at risk of not persisting, and why some at-risk students do persist or show more academic resilience. In this context, risk implies that a designated group of students demonstrates a greater tendency to stop out or not advance to the next academic level, in comparison with a base line group.

Theoretical models of flow processes of students in higher education have been proposed to explain what constitutes being “at risk” of not persisting, and have been tested in qualitative and quantitative ways. While the outcome of these theories has had a positive impact in the design and execution of programs and policies aimed at helping at-risk students defy the odds, criteria under which certain populations can be classified as “at risk” of not persisting in their particular university setting are still unclear to university administrators. Not as much is known about how institutions can use their own historical data for predictive purposes. It is tempting for administrators to base their decisions for policy changes and program allocation on common indicators that apply across a range of institutions or on focused descriptive studies within the institution. While these decision options are valid and can be useful for interpreting predictive approaches, they are prone to overlooking important predictors available within the institution. Consequently, this study offers a process for assessment of at risk students using pre- and post-admission predictive variables.

Logistic regression was used to predict expected probabilities of specific flow-model event given a profile of independent variable scores. The logistic regression model assigns a risk value in the form of a probability to each profile with respect to the risky event, be it stopping out, staying in the same academic level, advancing to the next academic level or graduating. Therefore, risk can be quantified by the transition probabilities associated with each profile.

For example, consider two profiles differing only in age, as presented in the Results in policy option 1. The profile where age at entry is 18 years results in a predictive probability = 0.0387669 of stopping out at the freshman level, compared to the profile where age at entrance is 25 years, for which the probability is 0.0895167 all other predictors being equal. From the results one can conclude that the risk of stopping out at the freshman level is 2.3 times greater for the 25 year old student than for the 18 year old, holding all other variables constant. The beta coefficients discussed in this chapter reflect the relative contribution that each of the independent predictor variables makes to the predicted probability in each of the four undergraduate academic levels (i.e., freshman, sophomore, junior and senior).

The summary and discussions of this study are organized into 3 parts. The first describes anticipated relationships resulting from the logistic regression analysis. The second analyzes and explains unexpected relationships and findings. The third and final section discusses possible uses of the simulation tool for the predicted persistence probability rates.

Validation of the Persistence Probability Model for the Pre-Admission Variables

Expected results

Age: The older the student the lower the probabilities of persisting towards graduation for all four academic levels. In the literature of persistence, it is common to see that age is a significant predictor (Astin, 1971; Doolittle, 1996; Leppel, 2002; Tinto, 1993). An explanation for this may be that the older the student the more responsibilities he/she may have, which serve as a force pulling the student away from school responsibilities. Also, it was found that the effect of age is largest in the sophomore year. This finding is new in the literature, mainly because most studies do not analyze the student flow by academic level transitions. In general terms the findings for this predictor, although interesting for policy decisions, do not have major implications for the university being studied because the majority of the student are 18 years old.

Ethnicity: Being American Indian was found to lower the transition probabilities for this group when compared to whites, at the freshman year. Data from this sample shows that not only the proportion of students in this ethnic group is small, but nearly 50% of those who enter never receive a degree from this university. Among the total number of white students, 20% never receive a degree from this university. Of course this could mean that student who stop out go somewhere else and finish their degree, but clearly they do not get a degree from their first choice university. Padilla (1997), citing others, reports that the estimated college degree completion rate over four years for Native Americans in the US is 25%. Non-persistence of Native Americans has also been addressed in research done by Gloria (2001). The author found that self-beliefs, social support and comfort in the university were each significant predictors of academic non-persistence decisions of American Indian undergraduates. Gloria's research also highlights the importance of addressing the school environment related issues in the first year of college. The results for the other ethnic groups (African American, Hispanic, Pacific Islander and all Targeted Minority) will be presented under unexpected findings.

Gender: At the junior and senior levels, being female is more favorable than being male when it comes to increasing the probabilities of advancing to the next academic level and graduating. Gender was not found to be a significant predictor at the freshman and sophomore levels. The finding is consistent with lower attrition rates for females reported by Astin (1993) and by Tinto (1993); and gender differences in persistence behavior (Pascarella & Terenzini, 1980). Recent research on persistence is clear about how the gap in graduation attainment between males and females is closing. However, the research is not always consistent. In a recent study by Horn (2004), which compared students who first began their postsecondary education in 1989-90 with those who first enrolled 6 years later in 1995-96, it was found that women in the second cohort may be taking longer than 5 years to attain a credential, but have similar rates of degree attainment as their counterparts in the first cohort. On the other hand, a similar study also comparing students in the same cohorts (Horn, L. & Peter, 2005) found that among the 1989-90 postsecondary students, women were more likely than men to complete some type of degree. Among the 1995-96 cohort women were still more likely than men to complete a degree within 5 years, but no significant differences were detected in the percentages from both genders that completed a degree or were still enrolled 5 years after they started their postsecondary education.

An interesting finding of this study is that gender is a significant predictor in the highest academic level. This could be interpreted as meaning that males attain the correct number of credits to achieve the senior academic level, but do not have the required credits for degree completion. Thus, graduation favors females. Adelman (1999) also found, using longitudinal data from the National Center for Education Statistics, that males take longer to complete degrees. However, he suggests that part of the reason may be that males are over-represented in fields where time-to-degree is traditionally longer (e.g. engineering, and architecture).

Target Minority: Being part of this group has a positive and significant effect on transition probabilities in the first year. In other words, at the freshman level targeted minorities are being retained and moving to the sophomore level, when compared to non-minorities and holding other variables constant. Being a targeted group means being part of a minority group that qualifies for school diversity initiatives and special support programs. Often schools consider this group at risk of not persisting. As expected, the results in this study suggest that the system is effectively implementing support programs offered to them, as seen in the increase of the transition probabilities, in the first year of college. However, the linearity of the beta coefficients (Graph 5) suggests that the probability of advancing to the next academic level diminishes linearly over academic levels. The expected aspect of this finding is that it is known that targeted minorities are part of an at-risk group for lower persistence, for many institutional programs this is one reason why they are “targeted”. The novel finding is the unanticipated increase of risk of grade retention and/or stop out over time. This point will be further analyzed under unexpected findings.

Pell grant indicator: This indicator exerted a negative effect on persistence probabilities, with beta coefficients being statistically significant at three out of the four academic levels (i.e., Sophomore, Junior, Senior). This result was not surprising in that the Pell grant indicator was chosen to serve as a proxy for socioeconomic status (SES). Being from a low-income background is a risk eligible characteristic for Pell grant, a need-based student aid program. Even though this study did not employ a SES predictor, the results may be compared to findings of previous studies which examine the effect of family income and persistence (Pascarella & Terenzini, 1980).

Based on the results for the Pell Grant indicator, one might initially argue that the financial aid system is not effectively compensating for the disadvantage of low-income background by boosting recipients' likelihood to persist when compared to those not receiving the grant. While this argument is intuitively correct, it is not accurate considering the complexity of financial aid programs. In order to make any conclusions with respect to financial aid, other forms of aid awards including other grants (e.g., state grants, institutional grants, and Federal Supplemental Educational Opportunity Grants), loans (e.g., Subsidized and Non-subsidized Federal Stafford Loans, State loans, and institutional loans) and College work study awards would have to be taken in to account. Only then could it be determined whether this finding indicates that money is being distributed to students who are stopping out or not advancing to the next academic level, or that a grant for students of low socioeconomic status is not enough to compensate for their disadvantaged background. For now, it is safe to interpret that in this study students receiving the Pell grant are the least academically resilient students when compared to those who do not receive this grant.

Finally, with respect to the almost perfect linearity of the beta coefficients across academic levels for the Pell Grant predictor, this could be interpreted as receiving a grant over time increases the likelihood of the student staying longer in school. In this study a positive beta at the senior level is more likely to mean a "jam" in the senior status and thus taking longer to graduate than an increase probability of stopping out. With this in mind one could also argue that the negative effect of the Pell Grant indicator increases over time because as the student advances, the expenses increase and therefore the financial need becomes more evident. In light of this situation the student may start working, going part time which are factors that negatively affect persistence (Bean, 1980). To make any further conclusions in regard to this point more information such as number of terms the student receives a Pell Grant and amount in dollars received, would be needed.

<u>High School Percentile</u>: The negative sign and small value for the beta indicates that at the freshman level this predictor has a slight contribution on persistence (i.e., the higher the HS% scores the higher the probabilities of advancing to the next academic level and graduation). This finding was expected because HS% scores are correlated to high school grades (not available for this study). As reported by (Astin, 1993; Astin, Korn, & Green, 1987) the most important freshman predictors of retention are the students' high school grades and admissions test scores. Others (Bean, 1980; Hossler et al., 1999) have also found a direct relation between high school performance and college performance. On the other hand, the reverse of the sign at the senior level is not an expected result, and it will be addressed under unexpected results.

<u>ACT scores</u>: Similarly to high school percentile, ACT test scores were found to have a positive and significant contribution to transition probabilities at the freshman level and an

unexpected reverse of sign at the senior level. The beta for ACT scores is smaller, but still significant at the sophomore level. The initial positive contribution is expected, since students with high scores are usually those with higher mathematical and verbal aptitude, and these characteristics correlate with greater chances of meeting the formal academic demands of the university (Astin et al., 1996). The sign change at the senior level will be discussed in the following section.

Unexpected results

A major unexpected finding in this study is the increase in predictability of pre-admission variables at the senior level, along with the reversal of the signs for several of the beta values. This finding corroborates that the structure of the pipeline persistence model organized by academic level was necessary, since data show that predictors' effect on the dependent variable varies across academic levels. In other words, it is not sufficient to conclude that variables that have an effect on persistence do so at all academic levels.

- Significant effects and counter-intuitive sign reversal for pre-admission predictors:

<u>Hispanic</u>: The sign for this predictor goes from a low negative in the freshman year to a higher positive at the sophomore level, reducing in size but with the same positive sign at the junior level (i.e., positive sign (+) = negative predicted transition probabilities), back to a negative, higher value and significant. After all senior credits have been achieved, transition probabilities are favorable for Hispanic groups when compared to Whites. As will be explained later, this could indicate that after this group travels a "bumpy" road towards graduation, those who make it to senior status graduate promptly.

Target minority: Interestingly, for this group the reverse of the sign from negative to positive (i.e., + = negative predicted transition probabilities) occurs at the junior level and the effect becomes significant at the senior level. As mentioned before, the expected aspect of this finding is the fact that minorities struggle persisting or acquiring a degree in a shorter amount of time than non-minorities. This finding is consistent with other studies (Astin, 1993). The unexpected finding is the timing of the negative effect on transition probabilities, at the senior level. So, the positive effect of being a Target Minority when compared to the non-Targeted minority does not hold at the senior level and in fact the sign reverses. The beta value is one of the strongest in this study and is statistically significant showing increase risk for persistence. Three points that will be further developed at the end of this section are suspected to be in play. First, the strength of the beta for this predictor and its significance may help justify the argument that few beta values are statistically significant when the independent effect of ethnic groups is assessed on persistence, and this happens because the sample for these groups is too small for a statistical test to capture any differences. Second, if assumed that the risk of persistence in fact increases as a function of academic levels for Target Minorities when compared to non Targeted Minorities, then the data is indicating that what is effective for the first year of college to retain minorities is not equally effective as the student advances academic levels. Third, transition to the next academic level at the senior level means graduation, so a negative effect on persistence at this level is perhaps indicating that these are the students who take longer to graduate when compared to their counterparts and holding all other variables in the model constant.

ACT and High school Percentile: An important finding for these two predictors is the reverse in signs of the beta value, from negative to positive (i.e., + = negative predicted transition probabilities) and significant in the last academic level (See Graphs 7 & 8). The sign reversal for these two predictors at the senior level suggests that those who come in more prepared, students with higher ACT and higher High School % are stopping out at the very end or more likely staying longer as seniors. In a recent research study at a western university, Arredondo and Knight, 2005, found that students with higher SAT scores and higher high school grades, for whom the expected six-year graduation rate was 81.4%, had an actual graduation rate of 61.3%. In their study these students did not meet the expected graduation rates, and the authors concluded either these well-prepared students are not performing to the level expected or they are transferring out and getting their degree in other more prestigious universities. For this institution we know that among those who achieve senior credits, 2.6% never graduate from this university[4] (Refer to Table 9, DV4 stop-outs).

In studying why math and science students leave the engineering field, Seymour & Hewitt (1997) reported that the students leaving engineering were academically no different than those that remained, noting that students left for reasons relating to perceptions of the institutional culture and career aspects. Perhaps these authors' conclusions can provide a similar explanation as to why well prepared students may not finish their degree where they started.

---

[4] There is a possibility that a few students who entered in the last cohort 1998-99 might have come back a few years later to attain a degree. The cube from which this data was obtained was last updated in may of 2004

- Non-significant effects, but interesting sign reversal among the pre-admission predictors occurs for the following:

Ethnic Groups Graph 3 shows interesting interactions across academic levels between the coefficients for the all the ethnic groups. The following paragraphs present some 'speculations' that have been made with respect to the sign change for each ethnic group. However, it is important to keep in mind when analyzing these fluctuations that only a few coefficients have statistical significance. One could argue that it is possible that the signs do not change. An insignificant test at a conventional alpha level such as 0.05 suggests that the effect of the predictor on the dependent variable is not statistically different from zero. Knowing that the confidence interval around the coefficient includes the value of zero, the parameter estimate could be in any range from (-) to zero or even (+). Therefore commenting on these findings may not be meaningful (Liao, 1994). Finaly, with such small sample per ethnic group, it is difficult to accept the finding that the effect of ethnicity on persistence is some cases is null or small.

Ethnic group- American Indian: for this group the beta values changes from positive at the freshman and sophomore level to negative at the junior level and back to positive at the senior level. The pattern of the sign suggests that this group has a 'break' at the junior level and then again struggles at the senior level. A good speculation in this case is that those who stop out at the freshman level do so for good and never come back. Those left are the truly resilient students

Ethnic group-Asian/Pacific Islander: the reverse of the sign occurs for this group in the last two academic levels, from negative to positive at the senior level, although not statistically significant. This suggests that in the last stretch being in the Asian/Pacific islander ethnic group it is no longer as favorable for persistence as it was at the freshman and sophomore levels when compared Whites and holding other variables constant. A speculation on this respect is that Asians are taking longer to graduate because they are graduating from fields where degree completion traditionally takes longer (e.g., science, and engineering)

African American: for this group the interpretation of the signs suggests a good start (e.g., - = positive predicted transition probabilities), to a greater struggle in the sophomore year (see also in graph 3 the deviation from zero for the beta value), to a more even path in the last two levels. Suggesting that the effect of being African American when compared to and Whites is almost null (i.e., the parameters are so close to zero). This close in the racial gap could simple be possible. The few African-Americans who come to this university and keep enrolled have good chances of graduating when compared to Whites. What we do know is that the school where the data comes from is a close admission university with very high standards of admissions independent of ethnicity. Nevertheless, the ethnicity results are difficult to interpret in light of the strong negative effect that being a Target Minority has on persistence. It is known from the descriptive data that the Majority of the Asian _Americans are not Target Minorities and with the exception of Hispanics, the beta values for the other groups are not significant or do not have a (+) beta value. The question arises, where does the significant positive beta for Target Minorities comes from?

To conclude, in an effort to make sense of why several of the predictors in this study re-gain predictability power at the senior level, and in most cases with a change in the sign (e.g., from negative betas to positive and vice versa) it is concluded that this is happening because once a student reaches the necessary amount of credits to hold senior status the flow movement is reduced, there is no other level to move to. There is a fundamental difference between the senior level and the earlier levels: to advance between earlier levels a student only needs a specific <u>number</u> of credits, but to advance from the senior level (i.e., graduate) a student needs both the specific number of credits and the right <u>type</u> of credits -- that is, credits from the classes that satisfy graduation requirements for their chosen degree. All students at the senior level have achieved the number of credits needed to hold senior status, but many may not have taken the specific classes they need to graduate. Table 9 showed that few students stop out at this point, so although stopping out is a path it is a very small one; more frequently followed paths are: staying longer vs. graduating. Thus, the results at the senior level need to be interpreted slightly differently from those of the preceding levels. In this last academic level, the results more likely illustrate what happens with the students who stay longer vs. those who graduate in $\leq$ 4 years. In this sample, about 40% of students who graduate from the institution being studied take longer than 4 years to graduate.

It is also possible that the reversal of the signs in the beta values occurs because of the structure of the databases. An example of this is that students who come in with advanced credit may take longer than a year to graduate after they have attained senior status, simply because they haven't yet taken the classes needed to graduate. As seen in Table 6, each level is unique in that it encompasses only students from one academic level (e.g., database III has only juniors), but as the academic level advances so does the number of students who achieve the right number of credits to be in the corresponding level but are in different academic terms. For example, in data base IV c. the students have enough credits to be seniors, but they are starting their fifth term/ third year. However, the plausibility of this theory loses strength in light of the knowledge that model fitting procedures (e.g., logistic regression) generally tailor themselves to where most of the data points are.

Validation of the Persistence Probability Model for the Post-Admission Variables

Expected results

CUM_GPA: Cumulative grade point average as a post-admission predictor makes a positive contribution to the prediction of persistence, and its positive effect is maintained throughout all four academic levels although at a decreasing rate. GPA has long been mentioned as a predictor of college retention (Arredondo & Knight, 2005; Astin, 1993; Bean & Metzner, 1985; Tinto, 1993; Zhu, 2002). While it is true for this study that Cum_GPA is highly correlated for the four academic levels, the way data was separated by academic level allows interpretation of its unique contribution at each academic level. As expected, the effect of Cum _GPA decreases as the student advances. This indicates that by the time students achieved senior status GPA is not as important. In this study, Cum _GPA was the significant predictor with the highest beta weight among all pre-admission and post-admission variables, and the peak of its weight happens at the sophomore level, by the end of the 4th term. This information is more detailed than what other researchers have found and more reassuring when considering that in most other studies the GPA used is a student-reported GPA or simply a first semester GPA.

The reason why CUM_GPA is a stronger predictor of persistence at the sophomore level than at the freshman level may be because prior to the junior year students are more likely to be making decisions about selecting majors, changing academic groups or selecting more credits. An interesting follow up of the results from this data would be to analyze what connection the strengthening of CUM_GPA at this level has with the fact that changing academic groups and not changing majors also pick up prediction strength at the sophomore level.

<u>Changing academic groups</u>: Changing academic groups has a positive effect on persistence, and it has a higher effect on transition probabilities when this change happens at the sophomore level. As expected the effect on persistence is negative once students have achieved senior status. This finding seems at first counterintuitive. However, moving across programs does not always involve picking up more credits. The findings may suggest that some programs might be bottlenecks for the transition flow of student towards graduation, while other programs allow the flow at a faster rate. Thus, giving flexibility to the students by changing academic groups can speed the process of advancing through the levels until graduation. In light of the resilience theory, this is considered a protective factor. Academically successful students with more sense of future and skills to adapt (Wang et al., 1997) may be finding a better fit by changing academic programs. This pattern of flexibility can be used in advising students.

Further investigation in this area could help point out programs that are (and are not) letting students through more smoothly while still retaining the students. Two points need to be kept in mind. First, independent of whether or not academic programs retain their students, years to degree has traditionally been longer for programs like Engineering, Business and Heath Sciences (Adelman, C, 1999). Second, as found in this study changing academic groups has a positive effect on transition probabilities, but supporting a finding like this could directly or indirectly affect departures from fields that already have problems attracting students, such as the Math and Science and Technology (MST) fields (Seymour & Hewitt, 1997).

Not changing majors: has a facilitative effect on persistence. At first it seems as if this predictor should yield similar results than that of changing academic groups. However, when a student changes programs many credits are passed on, but when a student changes majors more credits are usually required. Having more credits is not only more academically challenging, but also challenging in terms of time, since for example for a science course this also means extra lab hours. In this study the effect of not changing major is strongest at the sophomore level, perhaps because this is when students are more likely to change or select a major. The results of having dual majors are in agreement with not changing majors. Having dual or more majors has a negative effect on persistence with the exception of the senior level, where the effect is non-significant. Sanford and Rivera (1994) found in surveying parents of young adults who had been enrolled in a university for at least 8 consecutive terms that the predominant reason to justify why their children took longer to graduate was “changing majors”, while the second highest reason was "felt no pressure to finish in 4 years”. On the other hand, Fredda (2000) found no significant difference in drop out rates at the freshman levels between those who changed majors and those who did not. The effect of changing majors could be one of those effects that varies from institution to institution, in other words it may be institution specific.

Courses and Grades

The expected aspect of the findings concerning the five critical courses chosen in this study was that these are in fact critical for predicting persistence. Taking or not taking these courses was expected to have some effect on persistence. Informal interviews with administrators and students from the university being studied, along with the available literature, led to the classification of these courses as critical. It was suspected that some Math courses could negatively affect persistence if the student failed (Levin & Wyckoff, 1990), and that the timing for when these courses are taken might differ among persisters and nonpersisters (Zhu, 2002). In general terms, according to the results obtained in this study taking the selected five critical courses has for the most part a negative impact on persistence The exceptions are:

M112 course: Taking this course makes a positive contribution to the probabilities of advancing to the next graduation level and graduating, when it is taken at the freshman level.

ECON 101 grade: receiving a higher grade in this course has a significant positive persistence effect on transition probabilities at the freshman and sophomore levels.

PSY202 grade: receiving a higher grade in this course has a significant positive graduating effect on transition probabilities at the senior level.

Unexpected Results

In this section, findings for Courses and Grades are discussed. Some of what is discussed here may be repetitive from the previous statements made under expected findings. Yet they are repeated in order to show the variables' effects from one academic level to the next, or the variation in effect between grade and course. Thus, the unexpected findings have to do with the mixed results concerning these predictors.

M112 Course: The data indicates that this course has a positive persistence effect on transition probabilities if the course is taken at the freshman level. The effect is significant but negative at the junior level, which suggests that the most vulnerable students should probably defer taking the course until the junior level.

M112 Grades: The data indicates that it is better to not have a grade in this course, which because of scoring peculiarities translates in to higher transition probabilities when this course is not taken. Knowing the limitations that the coding for the Grade predictors may be imposing a more realistic conclusion is that taking this course is not recommended to improve transition probabilities, but if this course has to be taken it should be taken at the freshman level.

M221 Course: This course does more to improve persistence when taken at the junior level. However, most students take the course at the sophomore level, an event that predicts lower transition probabilities associated with persistence.

M221 Grades: The beta values indicate that grades negatively affect persistence at the freshman, junior and senior levels. The smallest effect, though non-significant is for the sophomore level. This suggests that the most vulnerable students probably take the course at the sophomore level. Those who take the course early (at the freshman level) or late (at the junior level) are the least vulnerable and therefore show a positive contribution to persistence. Although it is beyond the scope of this study, a follow-up to investigate how this course is affecting transition probabilities is recommended, perhaps by simply running a similar analysis with only students who take at least one of the courses chosen for this study.

<u>CH103 Course</u>: The significant and (+) beta values at the sophomore and junior levels predict lower transition probabilities. The almost perfect linear trend of the beta coefficients across academic levels (See Graph 14) suggests that this course is time-dependent. In order for this course to be a "protective factor" against risk persistence probabilities it has more favorable results when taken earlier, at the freshman level.

<u>CH103 Grades</u>: The statistically significant and (+) beta values at the junior and senior levels predict lower transition probabilities in moving to higher academic levels. This suggests that vulnerable students take this course at the junior and senior level. This course does more to improve persistence when taken at the freshman level.

<u>ECON 101 Course</u>: The values of the parameter estimate from this predictor suggest that this course is irrelevant for predicting the probabilities of persisting or not.

<u>ECON 101 Grade</u>: receiving a grade in this course has a positive effect on persistence at the freshman and sophomore levels. Thus, taking this course at these levels improves transition probabilities. The significance of this predictor at earlier levels perhaps corroborates that this in fact a 101 course. For a 101 course to play a protective role and positively affect persistence the

<u>Psy 202 Course</u>: The statistically significant (+) beta value at the junior level predicts lower persistence transition probabilities. This suggests that risk persistence increases when this course is delayed.

<u>PSY 202 Grade</u>: Receiving a grade in this course has a positive effect on persistence when this occurs at the senior level. This is the only course that does not seem to be time

dependent like the other course, meaning that students who are graduating sooner somehow are taking this course at the senior level. Perhaps this course provides a fast avenue to students looking for fast credits necessary to complete the graduation requirements.

In interpreting the results obtained for Course and Grades, it is important to keep in mind that although many students (between 20% to 40%) take at least one of these courses, the majority does not take them. Because of the dichotomous coding imposed on the courses (1=took the course in X academic level vs. 0=did not take the course), there is a preponderance of 0 scores, which could be pooling the data to a negative effect on transitions. This is more likely to be happening for the junior and senior levels. However this can be further investigated, because logistic regression is robust for unequal cell size (Long, 1997).

A similar logic applies to the coding for the course grades. As was explained at the beginning of Chapter 5, the grades were re-coded to a numeric scale, and in order to include everyone who did not have a grade, a 0 was assigned to those who did not have a grade. This was necessary, otherwise students who did not take all of these courses would have been excluded from the entire study -- it was not possible to exclude them from the analysis for only the portion of the study involving the courses they didn't take. This is why a positive (+) beta value for a course grade would be better interpreted as an indicator for not taking the course, rather than an indication that poor grades lead to higher transition probability. Also, it should be noted that to avoid having very small n for each grade-course these are not academic level-dependent. The grade refers to the score the student received when he/she took the course, independent of the level the student was when taking the course. With these limitations in mind, knowledge of the study's design and common sense can help to interpret what is important to improve transition probabilities. An empirical tool can then validate and complement common sense.

Limitations

The findings of this study are expected to be most valid within the context of the institution that generated the data. Nevertheless, some findings may be applicable to other institutions, and both the approach and the model itself could easily be adapted for use by other institutions. There are two challenges in using a student flow model: One is to understand the structure and content of the student record database and find appropriate independent variables to use as predictors. Because the data is likely to be pre-collected, the use this type of data imposes methodological limitations, common in studies that rely on secondary data (Kiecolt & Nathan, 1985). As was the case in this study, extensive data manipulation and reorganization are often necessary. Manipulation of a database can lead to errors, so consistent checks against the source are crucial.

The use of pre-collected data also means that some of the variables available were never intended to be used to analyze casual relationships, thus extensive recoding and caution interpreting the results are necessary. Restricted availability of measures is another characteristic of pre-collected data. Therefore desirable ways of measuring certain constructs is sometimes not possible. An example in this study is socio-economic status (SES). In most of the datasets developed by the National Center of Education Statistics (NCES), SES is a composite of measures like parental education, parental occupation, household items and parental income. Limited information concerning wealth or SES was available to the researcher, and what was available came from the financial aid office. No type of financial aid information alone is enough to determine need, because not everyone who applies for financial aid is in economic need, nor does every student in economic need apply for financial aid/grants

Also, specific to this study is the limitation to academic records only. Data available could not be connected to any other databases at the university that may have information on attitudinal and behavioral variables, such as those collected in satisfaction surveys, faculty evaluation and class assessments. However, if available such data could be used in a student flow model. Despite these data-related limitations, the proposed prediction model is innovative and practical. It uses available data and addresses relevant institutional persistence issues.

The second challenge regards recognizing that the student flow in higher education is a very dynamic process, and a study like this only capture one aspect. A recent article in the Chronicle of Higher Education (February, 24, 2006, p A37) quoted Michael Kirst; a professor of education at Stanford University, as suggesting that in describing the student flow, “the metaphor we ought to use is that of a path with twists and turns and not the traditional pipeline, where you put oil in and it flows out.” Cross-sectional regression modeling like the Markov student flow model used in this study provides only a snapshot of the dynamics of student flow. While better than smaller dissections of descriptive data, it is only a picture frozen in time. To be useful it requires constant monitoring. Following up on the metaphor, several photos make the filmstrip.

Among some of the shortcomings, this study fails to account for students entering in the spring, for what happens when a student stops out and comes back, for transfers to another institution, especially those who attain a degree, and for those who stagger in an academic level and regain momentum, even graduating on time. Nevertheless, this study's results have strong face validity in light of what it is known from the literature of persistence, and recent findings based on national data from the The American Freshman: National Norms for Fall 1994. (Astin & Oseguera, 2002) and grade-cohort data collected by the NCSE (NELS: 88/2000) (Adelman, Clifford, 2006). In fact, a finding of Adelman's study and relevant to this study is that continuous enrollment (e.g., not stopping out, however in his study stopping out for one term it is still considered continuous enrollment) increases the probability of degree completion by 43%, holding 16 other variables constant. The present study, however, views institutional persistence beyond retention rates and provides a constructive point of view and data to help all students become more academically resilient.

CHAPTER 6: IMPLICATIONS OF THE STUDY

The following paragraphs present the most important implications for theory, practice and research.

Implications for Theory

From a theoretical perspective, the results of this study indicate that a more comprehensive understanding of institutional persistence can be achieved when data is divided into academic levels. One of the most interesting findings of this study is that some predictors that are statistically significant at the freshman level lose their effect on persistence by the time the student achieves the senior level, and vice versa. Additionally, the sign of some significant betas changed as the students progressed from one academic level to another, meaning that some factors can have either a positive or a negative effect on persistence, depending upon when they occur. Looking at the results in detail, most of the beta sign switches occur after the end of the sophomore level, suggesting that what happens between the sophomore and junior levels may help to explain the real-life movement of students, where 5 and 6 year paths to graduation are becoming the norm.

The model also provides an empirical way of assigning a risk value to persistence. The simulation tool is a consistent method that can help to evaluate what characteristics constitute risk and which do not, and what enrollment or academic paths increase or decrease risk (i.e., becoming a protective factor). Using the theory of resilience as a lens for viewing the problem of persistence helped to keep the focus on success. Focusing on educational resilience and being able to assign a value to risk contributes to giving a real meaning to the otherwise vague term “at-risk”. Without having empirical knowledge of what constitutes risk, the term “at-risk” can be misleading and has the tendency to lump people who otherwise would not be grouped together. In the absence of a value for risk, counselors, professors and administrators can fall into the error of making false, a priori ideas about those labeled at-risk. In this sense, the term is only useful if it is specific (e.g., at risk for...) and if it has a value. Finally, the findings of this study provide a map for further investigation, and a tool to complement the use of descriptive analysis.

Implications for Practice

A main benefit of this study is the utility of the Markov student flow model as a simulation tool. As previously discussed, the model allows prediction of the effect of a set profile of pre-and post- admissions variables on transition probabilities. Decision makers can propose different policies, alter variable profiles and explore the predicted effects on the institutional persistence of at-risk students.

The fourteen logistic regressions, one for each transition probability, constitute the core of the simulation tool. Given a profile of independent variable values, transition probabilities can be estimated for any subject or group of subjects with an identical profile. These fourteen estimated probabilities can then be used to populate the transition matrix. As presented in the three policy options, by comparison of students 18 years old vs. 25 years old, students A vs. B, and students X vs. Y, this model was used to analyze differences in retention. The information obtained can be used to improve the retention of students with characteristics that leave them with a low probability of achieving graduation. In an evaluative process, their predicted career outcome can be improved by altering post-admission variables that yield a more favorable outcome.

The policy implications of the simulation are far-ranging. Other persistence scenarios can be proposed and compared, and the effects of a change in the sub-set of independent variables' scores on graduation probabilities determined. This can help accomplish two of the goals of this study: being able to identify students who are at-risk, and studying the effects of certain protective factors on the persistence of at-risk students. An example of this was presented by the comparison of students X and Y. Because of their pre-admission variables, their persistence probabilities were expected to be the same, but lower when compared to other ethnic groups. By making changes in the post-admission variables it was possible to see the effect of, for example, first semester GPA, which resulted in higher transition probabilities for student Y.

To use an analogy, the simulation tool has a similar utility to the prediction of tropical storms, where a storm path is proposed but there is always a margin of error. As with the storm models, the predictions of the simulations are not expected to be exact. It is recommended for the proposed tool that the data be complemented with other ways of identifying risk characteristics in students or protective factors in the school environment. For example, if the simulation tool is used at the time of admissions to identify incoming at-risk students, numbers can be plugged into the model and results can be used to support decisions made by experienced admission counselors. In addition, based on the number of predicted probabilities the tool can be used to rank students from most to least at-risk. This first form of flagging students can help administrators determine what groups of students should be offered entrance to special programs or university services. The information obtained from the protective factors can be shared with programs serving at-risk students, academic counselors and advisors. Of course, the literal use of the simulation results may be irrelevant in some cases. For example, advising a student to change academic groups if the student is happy with his/her career would probably be inappropriate. However, if a student is in the process of making this decision, but is not sure how this change might affect his/her future performance, academic counselors can analyze the circumstances and encourage a different path based on the probabilities for change of academic group (i.e., contrary to the results on persistence based on national data (Adelman, Clifford, 2006) in this study changing colleges was shown to increase transition probabilities).

Implications for Research

A common question after an extensive study like this one is how can this research be taken further? The answer to this is almost endless when considering the extensions of the possible research using the same data, other cohorts, and the same student flow model applied to other universities with similar and different levels of admission selectivity (e.g., open admissions, liberal, selective and highly selective). The following paragraphs discuss research possibilities that could take the present study a step further within the institution and to broader generalizations.

The Markov student flow model created in this study has great potential for use in institutions of higher education. It is innovative in that it is one of few models [i.e., only one other study known by the author uses a Markov-chain analysis of student flow in graduate education (Borden & Dalphin, 1998)] that follows student flow from enrollment through graduation. It also departs from the traditional dichotomous dependent variable (persisters vs. non-persisters) by broadening the analysis with an ordinal dependent variable that permits assessing the effect of pre- and post-admission variables on the different levels of undergraduate student progression.

Needless to say, the internal validity and external validity of a model like this can be improved in endless ways. Two forms of validation that are simple to perform based on the current study, and perhaps necessary and recommended for applicability are:

1- Do a cross-validation of this model using another set of cohorts from the same institution. After considering historical changes in the demographics from the two different cohorts, if both models yield the same significant predictors and similar fluctuation in the ordering of the beta signs, the predictive relationships obtained would be more stable for use within the institution.

2- Improve the statistical power of the probability model by further testing the results for the proportional or parallel odds assumption test that were obtained for the four logistic regression models. The results of these tests are presented for each of the models in Tables B-1 to B-4. This test is based on the assumption that the explanatory variables have the same effect on the odds or parallel slopes regardless of the levels in the dependent variables (e.g., it assumes that for each ordinal level SO, FF, FS, FJ the effect of the predictor variables is the same). According to the results obtained in this study, a significant test statistic provides evidence that the parallel

regression assumption has been violated. This implies that at least one of the variables in the model may be having a differential effect across the academic levels. How much the slope for each of the ordinal levels deviates, and which predictor contributes to this deviation is unknown. O'Connell (2006) recommends not to make final conclusions from simply interpreting this test, and suggests different approaches to testing parallelism. When the assumption of parallelism is rejected, Long (1997) suggests considering alternative models that do not impose the constraint of parallel slopes, such as models for nominal data. Both of these options were beyond the scope of this study, but they provide an interesting follow-up that would strengthen the internal validity of the flow model.

Other ways to expand the current research are as follows:

- To extend the data analysis to the 10^{th} term. This extension would allow the completion of the flow model in one more academic year. It is clear from the results obtained that students are staying enrolled for longer than four years. This trend is not surprising when compared to other American universities, where recent findings indicate that the majority of students who enter college attain a degree, but not during the 'traditional' four years(Choy, 2002; Horn, L. & Berger, 2004, HERI, 2006). The US Department of Education calculates retention rates up to 6 years.
- To try different types of recoding for the dependent variables. For example, adding a level for the "true" stop out vs. those who stop but come back. It would be interesting to see whether/how this changes the results.
- To augment the work with a qualitative component to improve the understanding of the meaning of at-risk at the particular institution. This could be done by further exploring the concept of resilience. Going back to the original idea, now that there is a quantitative way of determining who is at-risk, it would be interesting to contact students for interview or surveys. The following could be investigated: intentions to leave or stop, attitudes with respect to college life, feeling emotionally or

academically prepared for school, experiences with courses, advising, college life, etc. This information can also be cross-validated with what advisors or outreach administrators perceive of these students.

- Further investigate the results for ethnic group and courses.

In summary, this study explored the feasibility of developing a user-friendly routine for prediction of graduation probability as a function of a profile of independent variables. Also, this study utilizes a resource (i.e., student record data) that has historically been used for descriptive purposes only. The recompilation of individuals' academic progress has little use for policy formulation other than providing material for reports that broadly describe general trends. From a policy perspective, historical data is of interest to the extent that is predictive of the future. In this dissertation design, the use of logistic regression permits prediction of the expected probability of specific flow-model events given a profile of independent variables scores. Based on this model, each individual is assigned a transition probability using the logistic regression developed for each of the four datasets. The probabilities obtained assign a risk status to each observation with respect to the risk event (i.e., the different levels of the dependent variable). These predicted probabilities help to define the at-risk status for each individual. Thus, individual subjects with high predicted stop-out probabilities have a higher chance for the risk event (i.e., not progressing in a timely manner, dropping out, or not advancing). Those with low stop-out probabilities have a lower chance for the risk event. Because of the historical context of the data, knowing whether or not a degree was conferred allows some validation of the logistic prediction equations.

REFERENCES

Adelman, C. (1999). *Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment* (ED publications). Washington, DC: U.S. Department of education.

Adelman, C. (2006). *Tool Box Revisited: Paths to degree completion from high school to through college*. Washington, DC: U.S. Department of Education.

Afolayan, P. O. (1996). *Defying the rule of conventional wisdom. A study of the resiliency factors in the school experiences of African American female high school students with a prior record of academic failure and grade retention.* Unpublished Dissertation, University of Wisconsin-Madison, Madison.

Alexander, B., et al. (1998). *Minority undergraduate retention at UW-Madison: A report on the factors influencing the persistence of today's minority undergraduates* (No. 1. Prepared for Associate Vice-Chancellor Paul Barrows and the Committee on Minority Retention at the UW-Madison). Madison, WI: University of Wisconsin-Madison, the LEAD Center.

Arredondo, M., & Knight, S. (2005). Estimating degree attainment rates of freshman: a campus perspective. *Journal of college student retention, 7*(1-2), 91-115.

Astin, A. (1971). *Predicting academic performance in college*. New York: The free press.

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

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

Astin, A., et al. (1987). Retaining and Satisfying Students. *Educational Record, 68*(1), 36-42.

Astin, A., & Oseguera, L. (2002). *Degree attainment rates at American colleges and universities*. Los Angeles, CA: Higher Education Research Institute (HERI), UCLA.

Astin, A., et al. (1996). *Degree attainment rates at American colleges and universities: Effects or race, gender, and institutional type*. Washington, DC: American Council of Education.

Balistreri-Clarke, M. (1996). *Patterns of student success: a life task analysis.* Unpublished Dissertation, University of Wisconsin, Madison.

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

Bean, J., & Metzner, B. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research, 55*, 485-505.

Benard, B. (1991). *Fostering Resilience in kids: protective factors in the family, school and community.* Portland OR: Northwest Regional Educational Laboratory (ED 335 781).

Benard, B. (1995). *Fostering Resilience in Children*. Retrieved July, 2001, from http://www.parenthoodweb.com/library/ERIC_ResilienceChildren.htm

Benard, B., & Marshall, K. (1997). A framework for practice: tapping innate resilience. *Research Practice, 5*(1).

Bernal, E., et al. (2000). *The Relationship between race and socioeconomic status (SES): Implications for institutional research and admissions policies* (Report-Research): Paper presented at the Annual Forum of the Association for Institutional Research (AIR, 2000).

Borden, V. M., & Dalphin, J. F. (1998). *Simulating the Effect of Student Profile Changes on Retention and Graduation Rates: A Markov Chain Analysis.* Paper presented at the Annual Forum of the Association for Institutional Research (38th, Minneapolis, MN May 17-20, 1998).

Braxton, J., & Lien, L. (2000). The viability of academic integration as a central construct in Tinto's interactionalist theory of college student departure. In J. M. Braxton (Ed.), *Reworking the student puzzle* (First ed.). Nashville, TN: Vanderbilt University Press.

Braxton, J., & Mundy, M. (2001). Powerful institutional levers to reduce college student departure. *Journal of College Student Retention, 3*(1), 91-118.

Braxton, J., et al. (1997). Appraising Tinto's theory of College student departure. In J. Smith (Ed.), *Higher education: Handbook of theory and research* (Vol. 12). New York: Agathon.

Brower, A. (1990). Student perceptions of life-task demands as a mediator in the freshman year experience. *Journal of The Freshman Year Experience, 2*(2), 7-30.

Brower, A. (1992). The "second half" of student integration. *Journal of Higher Education, 63*(4), 441-462.

Cabrera, A., et al. (1992). The convergence between two theories of college persistence. *Journal of Higher Education, 63*(2), 143-164.

Cabrera, A., & La Nasa, S. (2000a). *On the path to college: Three critical tasks facing America's disadvantaged*. University Park, PA: Center for the Study of Higher Education.

Cabrera, A., & La Nasa, S. (2000b). Overcoming the tasks on the path to college for America's disadvantages. In A. Cabrera & S. L. Nasa (Eds.), *Understanding the college choice of disadvantage students* (Vol. 107). San Francisco.

Cabrera, A., & La Nasa, S. (2000c). Understanding the College-Choice Process. In A. Cabrera & S. L. Nasa (Eds.), *Understanding the College Choice of Disadvantaged Students* (Vol. 107). San Francisco: Jossey-Bass.

Cabrera, A., & La Nasa, S. (2000.). Three critical tasks America's disadvantaged face on their path to college. In A. Cabrera & S. L. Nasa (Eds.), *Understanding the College Choice of Disadvantaged Students*. San Francisco: Jossey-Bass.

Cabrera, A., et al. (2001). *Pathways to a Four-Year Degree: The Higher Education Story of One Generation* (Report prepared with the support of the Association for Institutional Research (AIR)).

Cabrera, A., et al. (1993). Structural Equations Modeling Test of an Integrated Model of Student Retention. *Journal of Higher Education, 64*(2), 123-139.

Cabrera, A., et al. (1990). Exploring the Effects of Ability to Pay on Persistence in College. *The Review of Higher Education, 13*(3), 303-336.

Callan, P., & Finney, J. (2003). *Multiple Pathways and State Policy: Toward Education and Training Beyond High School.* Retrieved February, 2003, from http://www.jff.org/jff/kc/library/0185

Choy, S. (2002). *Access & Persistence: Findings from 10 years of longitudinal research on students.* (No. 309375). Washington DC: American Council on Education (ACE),.

Choy, S., et al. (2000). Transition to college: What helps at-risk students and students whose parents did not attend college. In A. F. C. a. S. M. L. Nasa (Ed.), *New Directions for Institutional Research* (Vol. 107): Jossey-Bass.

Colbeck, C., et al. (2001). Learning Professional Confidence: Linking Teaching Practices, Students' Self-Perceptions, and Gender. *The Review of Higher Education, 24*(2), 173-191.

College Board, S. (2004). *SAT Program Handbook: A comprehensive guide to the SAT Program for school counselors and admissions officers.* Retrieved 08-30-04, 2005, from http://www.collegeboard.com/prof/counselors/tests/sat/sat-prog-handbook.pdf

Cyrulnik, B. (2000). *Etología tres entrevistas con Boris Cyrulnik* (W. González & L. H. Hernández, Trans.). Santiago de Cali: CEIC -Departamento de Filiosofia, Universidad del Valle.

Delwiche, L. D., & Slaughter, S. J. (2003). *The little SAS book: A primer* (Third ed.). Cary, NC: SAS Institute Inc.

DeMaris, A. (1993). Odds versus probabilities in logit equations: A reply to Roncek. *Social Forces, 71*(4), 1057-1065.

Doll, B., & Lyon, M. (1998). Risk and resilience: Implications for the delivery of educational and mental health services in schools. *School psychology review, 27*(3), 348-363.

Doolittle, M. (1996). *Predicting the decision to go to graduate school among college seniors in Engineering.* Unpublished Dissertation, North Carolina State University, Raleigh.

Finley, M. (1999). Cultivating Resilience: An overview for rural educators and parents. In B. Cesarone (Ed.), *Resilience Guide:* . Champaign, IL: ERIC Clearing house on Elementary and Early Childhood Education. ED 436 307.

Fletcher, J. T., et al. (1999). *High school and college grades: Is past performance a predictor of future performance?* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Point Clear, AL.

Ford-Edwards, J. F. (2002). *Towards a theoretical model of minority student success: Successful black students and their perspectives on the variables that promote and retard their retention.* Unpublished Dissertation, Drexel University.

Fredda, J. (2000). *An examination of first-time in college freshmen attrition within the first year of attendance.* (No. Report 00-25). Fort Lauderdale, FL: Nova Southeastern University.

Gandara, P. (1995). *Over the ivy walls: The educational mobility of low-income Chicanos*. Albany: State University of New York Press.

Gloria, A. M., et al. (1999). African American Students' persistence at a predominantly white University: influences of social support, university comfort, and self-beliefs. *Journal of College Student Development, 40*(3), 257-268.

Gloria, A. M., & Robinson Kurpius, S. (2001). Influences of Self-Beliefs, Social Support, and Comfort in the University Environment on the Academic Non-persistence Decisions of American Indian Undergraduates. *Cultural Diversity and Ethnic Minority Psychology; APA, 7*(1), 88-102.

Green, R. (1998). *Retention and achievement in one first year program: a multivariate analysis.* Unpublished Dissertation, North Carolina State University, Raleigh.

Hagedorn, L., et al. (2001). Correlates of Retention for African-American Males in Community Colleges. *Journal of College Student Retention, 3*(3), 243-263.

Harper, V. R. (2003). *Most likely to succeed: the self-perceived impact of involvement on the experiences of high-achieving African American undergraduate men at predominantly white universities.* Unpublished Doctoral Dissertation, Indiana University, Bloomington.

Horn, L., & Berger, R. (2004). *College Persistence on the Rise? Changes in 5-Year Degree Completion and Postsecondary Persistence Rates Between 1994 and 2000* (No. NCES 2005-156; U.S. Department of Education). Washington D.C.: U.S. Government printing office.

Horn, L., & Peter, K. (2005). *Gender differences in participation and completion of undergraduate education and how they have changed over time* (No. NCES 2005-169; U.S. Department of Education). Washington D.C.: U.S government printing office.

Horn, L. J. (1997). *Confronting the odds: Students at risk and the pipeline to higher education* (Statistical Analysis Report): National Center for Education Statistics.

Horn, L. J., & Chen, X. (1998). *Toward resiliency: At-risk students who make it to college*: U.S. Department of Education.

Hossler, D., et al. (1999). *Going to college: how social, economic, and educational factors influence decision students make*. Baltimore: Johns Hopkins University Press.

Jones, L. (2001). *Retaining African Americans in higher education : challenging paradigms for retaining students, faculty, and administrators* (1st. ed.). Sterling, VA: Stylus Publications.

Kiecolt, K. J., & Nathan, L. E. (1985). *Secondary analysis of survey data* (Vol. 53). Newbury Park: Sage Publications.

LePage-Lees, P. (1997). *From disadvantage girls to successful women. Education and women's resiliency*. Westport, Connecticut: Praeger.

Leppel, K. (2002). Similarities and differences in the college persistence of men and woman. *Review of Higher Education, 25*(4), 433-450.

Levin, J., & Wyckoff, J. (1990). *Student Characteristics that predict persistence and suscess in baccalaureate engineering.* Paper presented at the American Educational Research Association, Boston, MA.

Liao, T. (1994). *Interpreting Probability Models logit, probit, and other generalized linear models* (Vol. No.07-101). Thousand Oaks, CA: Sage Publications.

Ling Yeh, T. (2002). Asian American College Students Who are Educationally at Risk. *New directions for students services, 97*(Spring), 61-71.

Long, J. S. (1997). *Regression models for categorical and limited dependent variables* (Vol. 7). Thousand Oaks, CA: SAGE publications.

Masten, A. (1994). Resilience in individual development: successful adaptation despite risk and adversity. In M. C.Wang & E. W. Gordon (Eds.), *Educational Resilience in Inner City America. Challenges and Prospects* (pp. 194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Masten, A. (2001). Ordinary magic. *American Psychologist, 56*(3), 227-238.

Mc Grath, M., & Braunstein, A. (1997). The prediction of freshmen attrition: An examination of the importance of certain demographic, academic, financial and social factors. *College Student Journal, 31*(3).

Menard, S. (1995). *Applied Logistic Regression* (Vol. 106). Thousand Oaks: Sage.

Miller, D., & MacIntosh, R. (1999). Promoting resilience in urban African-American adolescents: Racial socialization and identity as protective factors. *Social Work Research, 23*(3), 159-169.

Newman, L. C. (1997). *Quality Improvement/assessment of educational system for students of color in the University of Wisconsin college of engineering.* Unpublished Doctoral Dissertation, University of Wisconsin, Madison.

Nonis, S., & Wright, D. (2003). Moderating effects of achievement striving and situational optimism on the relationship between ability and performance outcomes of college students. *Research in Higher Education, 44*(3), 327-346.

Nora, A. (2001). The depiction of significant others in Tinto's "rites of passage": A reconceptualization of the influence of family and community in the persistence process. *Journal of College Student Retention, 3*(1), 41-56.

Nora, A., et al. (1996). Differential impacts of academic and social experiences on college-related behavioral outcomes across different ethnic and gender groups at four-year institutions. *Research in Higher Education, 37*(4), 427-451.

O'Connell, A. A. (2006). *Logistic Regression models for ordinal response variables* (Vol. No.07-146). Thousand Oaks, CA: Sage publications.

O'Connor, C. (2002). Black women beating the odds from one generation to the next: how the changing dynamics of constraint and opportunity affect the process of educational resilience. *American Educational Research Journal, 39*(Winter No. 4), 855-903.

Office of the Registrar. (2002). *Enrollment Report* (No. Fall Semester 2002-2003, 10/10/2002). Madison: University of Wisconsin.

Padilla, R. V. (1999). College student retention focus on success. *Journal of College Student Retention, 1*(2), 131-145.

Padilla, R. V., et al. (1997). Developing local methods of minority student success in college. *Journal of College Student Development, 38*(2), 125-135.

Pascarella, E., & Terenzini, P. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *Journal of Higher Education, 51*(1), 60-75.

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

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43*(2), 187-205.

Resilience NET. (2003). Retrieved 05-2005, from http://www.resilnet.uiuc.edu/abtrnet.html

Rhodes, W. A., & Hoey, K. (1994). *Overcoming childhood misfortune. Children who beat the odds*. Westport, Connecticut: Praeger.

Riehl, R. (1994). The academic preparation, aspirations and first-year performance of first-generation students. *College and University, 70*(1), 14-19.

Rodriguez, S. (2003). What helps some first generation students succeed? *About Campus, 18*(4), 17-22.

Rutter, M., M.D. (1993). Resilience: some conceptual considerations. *Journal of Adolescent Health, 14*, 626-631.

Sanford, T. R., & Rivera, N. (1994). *Parents' Perceptions of Students' Time to Degree.* U.S.
North Carolina: AIR 1994 Annual Forum Paper.

SAS. (1999). *SAS OnlineDoc, V8*, 2006, from http://v8doc.sas.com/sashtml/

Seymour, E. (2001). Tracking the process of change in US undergraduate education in science, mathematics, engineering, and technology. *Science Education, 85*(6), 79-105.

Seymour, E., & Hewitt, N. (1997). *Talking about leaving: why undergraduates leave the sciences*. Boulder, CO: Westview.

Stancill, J. (2006, March 24th). UNC Targets graduation. *The News & Observer,* p. B1.

The College Board. (2003). *Trends in college pricing*. www.collegeboard.com: The College Entrance Examination Board.

Thielemann, J. S. (1999). *The role of resilience as a determinant of persistence for minority college students: A logistic regression analysis of a theoretical model of student persistence.* Unpublished Dissertation, University of Houston, Houston.

Tierney, W. G. (1992). *Official encouragement, institutional discouragement : minorities in academe-- the Native American experience*. Norwood, NJ.: Ablex Pub. Corp.

Tinto, V. (1975). Dropout from Higher Education: a theoretical syntheses of recent research. *Review of Educational Research, 45*, 89-125.

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

Tinto, V. (1998). Colleges as communities: taking research on student persistence seriously. *The Review of Higher Education, 21*(2), 167-177.

Tinto, V. (2002). *Enhancing student persistence: Connecting the dots*. Retrieved 12/2002, 2002, from http://http://www.wiscape.wisc.edu/calendar/details.asp?idEvents=71

U.S. Department of Education. (2000a). *Condition of Education, 2000* (No. NCES 2000-062). Washington, DC: NCES.

U.S. Department of Education. (2001a). *Students whose parents did not go to college: Post secondary access, persistence and attainment,* (No. 126). Washington, D.C.: NCES.

U.S. Department of Education. (2002). *Projections of education statistics to 2012* (No. 2002-030). Washington DC: NCES.

U.S. Department of Education, National Center for Education Statistics (NCES). (2000b). *Descriptive Summary of 1995-96 Beginning Postsecondary Students: Three years later* (No. 2000-154). Washington, DC: NCES.

U.S. Department of Education, National Center for Education Statistics (NCES). (2001b). *Credits and Attainment: Returns to Postsecondary Education Ten Years After High School* (No. 2001-168). Washington: NCES.

U.S. Department of Education, National Center for Education Statistics (NCES). (2001c). *Educational Achievement and Black-White Inequality* (No. 2001-061). Washington: NCES.

U.S. Department of Education, National Center for Education Statistics (NCES). (2001d). *High School Academic Curriculum and the Persistence Path Through College* (No. 2001-163). Washington: NCES.

U.S. Department of Education, National Center for Education Statistics (NCES). (2003). *A Descriptive Summary of 1999-2000 Bachelor's Degree Recipients 1 Year Later, With an Analysis of Time to Degree.* (No. 2003-165). Washington: NCES.

U.S. Department of Education National Center for Education Statistics, N. (2001). *Educational Achievement and Black-White Inequality* (No. 2001-061): NCES.

Wang, M. C., & Gordon, E. W. (Eds.). (1994). *Educational resilience in inner city America challenges and prospects*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wang, M. C., et al. (1994). Educational resilience in inner city schools. In M. C. Wang & E. W. Gordon (Eds.), *Educational resilience in inner-city America* (pp. 45-72). Hillsdalle, NJ.: Lawrence Erlbaum Associates.

Wang, M. C., et al. (1997). *Fostering educational resilience in inner-city schools*. Retrieved 7/17, 2003, from http://www.temple.edu/lss/htmlpublications/publications/pubs97-4htm

Wang, M. C., et al. (1998a). *Educational Resilience*. Retrieved August, 2003, from http://www.temple.edu/lss/htmlpublications/spotlights/100/spot11.htm

Wang, M. C., et al. (1998b). *Educational resilience: an emergent construct*. Retrieved 7/10, 2003, from http://www.temple.edu/lss/htmlpublications/spotlights/100/spot105.htm

Winfield, L. F. (1994). *Developing resilience in urban youth*.Unpublished manuscript, Oak Brook, IL.

Worrel, F. (1996). *The risk-resiliency paradigm in research on dropping out.* Paper presented at the Annual meeting of the American psychological association, Toronto, Canada.

Ybarra, M. A. (2000). *Ecological factors that predict persistence among college students.* Unpublished Dissertation, University of Wisconsin-Madison, Madison, WI.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy- beliefs of women in mathematical, scientific and technological careers. *American Educational Research Journal, 37*(1), 215-246.

Zhu, L. (2002). *How the first-year college experience contributes to persistence.* Paper presented at the Annual Forum for the Association for Institutional Research, Toronto- Ontario, Canada.

APPENDICES

Appendix A: Glossary of Terms

Academic year: two consecutive terms beginning with the fall and ending with the spring .

Academic group: the College or School in which the student is enrolled in the given term

Academic level: the student's academic status, based on their achieved credits (e.g., freshman)

Educational Resilience is defined as success in school and in other life accomplishments, despite environmental adversities brought about by early traits, conditions and experiences. In this study, retention and resilience are different in the sense that resilient students are those who persist despite having at risk conditions. Educational resilience then refers to students who completed a bachelor's in a timely manner.

First year students (FYR): students who have registered at the present university without having previously earned college credits at another university. These students are almost always between 17 and 18 years old and come to the university directly from high school. It is necessary to differentiate between freshmen and first year students because increasingly many students come to the University with advanced placement and/or credits that may put them as sophomores or higher standing at the time of admissions.

Pell Grant: Pell Grant was named after Rhode Island Senator Clairborne Pell and formerly called the Basic Educational Opportunity Grant program (BEOG). Pell provides grants to assist qualified undergraduate students who attend public, independent or proprietary postsecondary educational institutions based on financial need. Need is determined by applying a formula to income, assets, and other information provided by the student on the Federal Financial Aid Application. A Federal Pell Grant, unlike a loan, does not have to be repaid. Generally, Pell Grants are awarded only to undergraduate students who have not earned a bachelor's or professional degree.

Appendix B: Supplementary Tables

Table B-1 Logistic regression results for freshmen, pre-admission and post-admission variables.

Predictor Variables
Pre-admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1stop out	1	2.8240	0.8713	10.5049	0.0012
Intercept	2FF	1	3.5325	0.8712	16.4397	<.0001
Intercept	3FS	1	9.7106	0.8786	122.1617	<.0001
AGE_AT_ENTRANCE		1	0.1273	0.0450	8.0064	0.0047
ETHNIC_GROUP	American Indian/	1	0.9159	0.2704	11.4704	0.0007
ETHNIC_GROUP	Asian/Pacific Is	1	-0.5455	0.1460	13.9691	0.0002
ETHNIC_GROUP	Black	1	-0.00025	0.1870	0.0000	0.9989
ETHNIC_GROUP	Hispanic	1	-0.0230	0.1819	0.0160	0.8994
ETHNIC_GROUP	Unknown	1	-0.1686	0.2339	0.5198	0.4709
GENDER	F	1	0.0297	0.0246	1.4598	0.2270
TARGETED_MINORITY_IN		1	-0.6858	0.2711	6.4008	0.0114
PELL_Ind		1	0.1216	0.0645	3.5536	0.0594
new_hsp		1	-0.0106	0.00236	20.2787	<.0001
NEW_ACT		1	-0.1032	0.00797	167.6642	<.0001

Predictor variables
Post-admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
CUM_GPA_2ND_TERM		1	-1.5489	0.0441	1231.4410	<.0001
ChgAcdGrp_Ind	No changed	1	0.0772	0.0276	7.8044	0.0052
ChangeMajor_Acadyear	+1major	1	1.0784	0.0653	272.8819	<.0001
ChangeMajor_Acadyear	No changed	1	-0.3865	0.0586	43.4661	<.0001
m112course_Acadyear_		1	-0.3111	0.1192	6.8124	0.0091
m112Grade		1	0.1205	0.0454	7.0544	0.0079
m221course_Acadyear_		1	-0.0190	0.1161	0.0269	0.8698
m221Grade		1	0.1230	0.0396	9.6498	0.0019
ch103course_Acadyear		1	0.0991	0.1058	0.8782	0.3487
ch103Grade		1	0.0109	0.0348	0.0981	0.7542
Econ101course_Acadye		1	-0.1150	0.0670	2.9482	0.0860
Econ101Grade		1	-0.0801	0.0206	15.1334	0.0001
Psy202course_Acadyea		1	0.0174	0.0625	0.0776	0.7806
Psy202Grade		1	-0.0233	0.0211	1.2193	0.2695

Overall Model Evaluation

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4178.8721	25	<.0001
Score	4068.7262	25	<.0001
Wald	3021.8438	25	<.0001

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
2283.6214	50	<.0001

Table B-2 Logistic regression results for sophomores, pre-admission and post-admission variables.

Predictor Variables
Pre-admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1stop out	1	0.0795	0.9985	0.0063	0.9365
Intercept	2SS	1	1.1066	0.9984	1.2285	0.2677
Intercept	3SJ	1	8.0574	1.0047	64.3174	<.0001
AGE_AT_ENTRANCE		1	0.2012	0.0520	14.9617	0.0001
ETHNIC_GROUP	American Indian/	1	0.3632	0.3325	1.1928	0.2748
ETHNIC_GROUP	Asian/Pacific Is	1	-0.3925	0.1700	5.3287	0.0210
ETHNIC_GROUP	Black	1	0.2345	0.2114	1.2310	0.2672
ETHNIC_GROUP	Hispanic	1	0.1745	0.2070	0.7107	0.3992
ETHNIC_GROUP	Unknown	1	-0.2527	0.2604	0.9420	0.3318
GENDER	F	1	0.0146	0.0265	0.3050	0.5808
TARGETED_MINORITY_IN		1	-0.3190	0.3147	1.0271	0.3108
PELL_Ind		1	0.1970	0.0677	8.4690	0.0036
new_hsp		1	-0.00315	0.00255	1.5220	0.2173
NEW_ACT		1	-0.0368	0.00852	18.6296	<.0001

Predictor variables
Post-admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
GPA_So		1	-1.6661	0.0548	925.0135	<.0001
ChgAcdGrp_Ind	No changed	1	0.1301	0.0286	20.6668	<.0001
ChangeMajor_Acadyear	+1major	1	1.0112	0.0486	432.3140	<.0001
ChangeMajor_Acadyear	No changed	1	-0.5267	0.0408	166.3522	<.0001
m112course_Acadyear_		1	0.3296	0.1765	3.4870	0.0619
m112Grade		1	0.00345	0.0232	0.0221	0.8817
m221course_Acadyear_		1	0.3002	0.1350	4.9441	0.0262
m221Grade		1	0.0369	0.0231	2.5611	0.1095
ch103course_Acadyear		1	0.3864	0.1240	9.7112	0.0018
ch103Grade		1	0.0298	0.0185	2.5865	0.1078
Econ101course_Acadye		1	-0.0604	0.0753	0.6432	0.4225
Econ101Grade		1	-0.0661	0.0208	10.1145	0.0015
Psy202course_Acadyea		1	0.0452	0.0747	0.3664	0.5450
Psy202Grade		1	0.00985	0.0197	0.2508	0.6165

Overall Model Evaluation

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	2398.5492	25	<.0001
Score	2439.6263	25	<.0001
Wald	2007.1832	25	<.0001

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
1043.7166	50	<.0001

Table B-3 Logistic regression results for juniors, pre-admission and post-admission variables

Predictor Variables
Pre-admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1Stop Out	1	-1.9045	1.0115	3.5453	0.0597
Intercept	2JJ	1	0.00903	1.0106	0.0001	0.9929
Intercept	3JSr	1	8.1450	1.0337	62.0896	<.0001
AGE_AT_ENTRANCE		1	0.1452	0.0526	7.6242	0.0058
ETHNIC_GROUP	American Indian/	1	-0.5137	0.4229	1.4756	0.2245
ETHNIC_GROUP	Asian/Pacific Is	1	0.0140	0.1803	0.0060	0.9381
ETHNIC_GROUP	Black	1	-0.0157	0.2377	0.0043	0.9475
ETHNIC_GROUP	Hispanic	1	0.1359	0.2223	0.3738	0.5409
ETHNIC_GROUP	Unknown	1	0.2014	0.2594	0.6027	0.4376
GENDER	F	1	-0.1962	0.0261	56.6936	<.0001
TARGETED_MINORITY_IN		1	0.2582	0.3199	0.6511	0.4197
PELL_Ind		1	0.2913	0.0687	17.9917	<.0001
new_hsp		1	0.00401	0.00272	2.1686	0.1409
NEW_ACT		1	-0.00945	0.00866	1.1893	0.2755

Predictor Variables
POST-Admission

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
GPA_Jun		1	-1.4478	0.0619	546.3481	<.0001
ChgAcdGrp_Ind	No changed	1	0.0983	0.0273	12.9531	0.0003
ChangeMajor_Acadyear	+1major	1	0.4826	0.0521	85.8722	<.0001
ChangeMajor_Acadyear	No changed	1	-0.0805	0.0380	4.4812	0.0343
m112course_Acadyear_		1	0.8828	0.3779	5.4566	0.0195
m112Grade		1	0.0215	0.0228	0.8875	0.3462
m221course_Acadyear_		1	-0.1736	0.3068	0.3203	0.5714
m221Grade		1	0.0599	0.0203	8.7151	0.0032
ch103course_Acadyear		1	0.5697	0.2799	4.1437	0.0418
ch103Grade		1	0.0450	0.0169	7.1111	0.0077
Econ101course_Acadye		1	0.1842	0.1110	2.7528	0.0971
Econ101Grade		1	-0.0318	0.0181	3.1090	0.0779
Psy202course_Acadyea		1	0.2672	0.1211	4.8670	0.0274
Psy202Grade		1	0.0264	0.0177	2.2183	0.1364

Overall Model Evaluation

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	1002.5369	25	<.0001
Score	1001.7805	25	<.0001
Wald	931.7682	25	<.0001

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
430.2377	50	<.0001

Table B-4 Logistic regression results for seniors, pre-admission and post-admission variables.

Predictor Variables
PRE-Admissions

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1Stop Out	1	-6.2973	0.9478	44.1473	<.0001
Intercept	2SrSr	1	-2.7993	0.9447	8.7801	0.0030
AGE_AT_ENTRANCE		1	0.1169	0.0496	5.5588	0.0184
ETHNIC_GROUP	American Indian/	1	0.0181	0.3754	0.0023	0.9615
ETHNIC_GROUP	Asian/Pacific Is	1	0.2722	0.1946	1.9563	0.1619
ETHNIC_GROUP	Black	1	-0.0492	0.2554	0.0372	0.8471
ETHNIC_GROUP	Hispanic	1	-0.5278	0.2362	4.9950	0.0254
ETHNIC_GROUP	Unknown	1	0.00103	0.2557	0.0000	0.9968
GENDER	F	1	-0.1664	0.0225	54.7228	<.0001
TARGETED_MINORITY_IN		1	1.0599	0.3597	8.6838	0.0032
PELL_Ind		1	0.3674	0.0648	32.0989	<.0001
new_hsp		1	0.0192	0.00257	55.9235	<.0001
NEW_ACT		1	0.0365	0.00774	22.2208	<.0001

Predictor Variables
POST-Admissions

Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
GPA_Sr		1	-0.7086	0.0611	134.6368	<.0001
ChgAcdGrp_Ind	No changed	1	-0.2346	0.0233	101.7391	<.0001
ChangeMajor_Acadyear	+1major	1	-0.0581	0.0491	1.4024	0.2363
ChangeMajor_Acadyear	No changed	1	-0.3739	0.0344	117.9349	<.0001
m112course_Acadyear_		1	0.5548	0.5437	1.0414	0.3075
m112Grade		1	-0.0127	0.0201	0.3987	0.5277
m221course_Acadyear_		1	0.2061	0.4755	0.1878	0.6648
m221Grade		1	0.0714	0.0168	18.1251	<.0001
ch103course_Acadyear		1	0.7414	0.5115	2.1011	0.1472
ch103Grade		1	0.1220	0.0139	77.1545	<.0001
Econ101course_Acadye		1	0.2256	0.1860	1.4719	0.2250
Econ101Grade		1	-0.0134	0.0145	0.8426	0.3586
Psy202course_Acadyea		1	-0.1210	0.1888	0.4111	0.5214
Psy202Grade		1	-0.0758	0.0142	28.3991	<.0001

Overall Model Evaluation

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	815.8082	25	<.0001
Score	785.9919	25	<.0001
Wald	752.8783	25	<.0001

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Score Test for the Proportional Odds Assumption

Chi-Square	DF	Pr > ChiSq
337.3759	25	<.0001

Appendix C: Hypothetical Profile Scores for Policy Options 1,2, and 3

Table C-1 Hypothetical profile scores for policy option 1

Hypothetical profile scores for a student 18 years old

Obs	V1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	18	-1	-1	-1	-1	-1	1	0	0	75	23	3.0	1	0	1	1	4	0	0	1	3	1	3	0	0
2	18	-1	-1	-1	-1	-1	1	0	0	75	23	3.1	1	0	1	0	0	1	4	0	0	0	0	1	4
3	18	-1	-1	-1	-1	-1	1	0	0	75	23	3.2	1	0	1	0	0	0	0	0	0	0	0	0	0
4	18	-1	-1	-1	-1	-1	1	0	0	75	23	3.3	1	0	1	0	0	0	0	0	0	0	0	0	0

Hypothetical profile scores for a student 25 years old

Obs	V1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	25	-1	-1	-1	-1	-1	1	0	0	75	23	3.0	1	0	1	1	4	0	0	1	3	1	3	0	0
2	25	-1	-1	-1	-1	-1	1	0	0	75	23	3.1	1	0	1	0	0	1	4	0	0	0	0	1	4
3	25	-1	-1	-1	-1	-1	1	0	0	75	23	3.2	1	0	1	0	0	0	0	0	0	0	0	0	0
4	25	-1	-1	-1	-1	-1	1	0	0	75	23	3.3	1	0	1	0	0	0	0	0	0	0	0	0	0

Table C-2 Hypothetical profile scores for student with differences in preadmission variables

Hypothetical profile scores for student A

Obs	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	18	0	0	1	0	0	1	1	1	80	24	3.0	1	0	1	1	4	0	0	1	3	1	3	0	0
2	18	0	0	1	0	0	1	1	1	80	24	3.1	1	0	1	0	0	1	4	0	0	0	0	1	4
3	18	0	0	1	0	0	1	1	1	80	24	3.2	1	0	1	0	0	0	0	0	0	0	0	0	0
4	18	0	0	1	0	0	1	1	1	80	24	3.3	1	0	1	0	0	0	0	0	0	0	0	0	0

Hypothetical profile scores for student B

Obs	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	18	-1	-1	-1	-1	-1	1	0	0	80	24	3.0	1	0	1	1	4	0	0	1	3	1	3	0	0
2	18	-1	-1	-1	-1	-1	1	0	0	80	24	3.1	1	0	1	0	0	1	4	0	0	0	0	1	4
3	18	-1	-1	-1	-1	-1	1	0	0	80	24	3.2	1	0	1	0	0	0	0	0	0	0	0	0	0
4	18	-1	-1	-1	-1	-1	1	0	0	80	24	3.3	1	0	1	0	0	0	0	0	0	0	0	0	0

Table C-3 Hypothetical profile scores for student with differences in post-admission variables

Hypothetical profile scores for student X

Obs	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	19	1	0	0	0	0	-1	1	0	71	26	2.7	1	0	1	1	2.8	0	0	0	0	0	0	0	0
2	19	1	0	0	0	0	-1	1	0	71	26	2.8	1	0	1	0	0	1	2.9	0	0	0	0	0	0
3	19	1	0	0	0	0	-1	1	0	71	26	3.0	1	0	-1	0	0	0	0	1	3	0	0	0	0
4	19	1	0	0	0	0	-1	1	0	71	26	3.1	1	0	1	0	0	0	0	0	0	0	0	0	0

Hypothetical profile scores for student Y

Obs	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25
1	19	1	0	0	0	0	-1	1	0	71	26	3.0	1	0	1	1	3	0	0	0	0	0	0	0	0
2	19	1	0	0	0	0	-1	1	0	71	26	3.0	-1	0	1	0	0	0	0	0	0	1	3	0	0
3	19	1	0	0	0	0	-1	1	0	71	26	3.1	1	0	1	0	0	0	0	0	0	0	0	0	0
4	19	1	0	0	0	0	-1	1	0	71	26	3.2	1	0	1	0	0	0	0	0	0	0	0	1	3

Appendix D: Routine to Estimate Transition Probabilities

This is the (simulation tool), written in IML/SAS

```
/*Routine to estimate higher education academic level transition probabilities*/
/*Transition probabilities estimated by logistic regression*/
/*Regression parameters are from Ms. Olga Herrera's unpublished dissertation results*/
/*Routine outputs one-step, four-step, five step, and six step transition probability
matrices*/
/*User is prompted to enter score profile for each of four academic levels*/
/*Upon completion of each year profile data is saved by entering <submit> at the command line
prompt*/
/*Upon entering the desired number of profiles, user must enter <exit> */
/*Data is best entered by using overwrite feature (insert key shown as fat cursor)*/
/*User can enter from one to four profiles, with each profile corresponding to an academic
level*/
/*No more than four profiles can be entered*/
/*User will have to enter <forward 2> at command line to gain access to last two variables*/
/*User must enter <top> at command prompt if desired to scroll to beginning of data window*/
/*Pre-admission data can not be altered across yearly profiles*/
/*Routine performs range checks for all variables*/
/*Pre-admission variables must remain constant over all academic year profiles*/
/*Post-admission courses can be entered as taken in only one academic year*/
proc iml;
/* module to initialize the independent variables*/
/*                                                */
start initial;
    v1 = 00.0000;
    v2 = 00.0000;
    v3 = 00.0000;
    v4 = 00.0000;
    v5 = 00.0000;
    v6 = 00.0000;
    v7 = 00.0000;
    v8 = 00.0000;
    v9 = 00.0000;
    v10 = 00.0000;
    v11 = 00.0000;
    v12 = 00.0000;
    v13 = 00.0000;
    v14 = 00.0000;
    v15 = 00.0000;
    v16 = 00.0000;
    v17 = 00.0000;
    v18 = 00.0000;
    v19 = 00.0000;
    v20 = 00.0000;
    v21 = 00.0000;
    v22 = 00.0000;
    v23 = 00.0000;
    v24 = 00.0000;
    v25 = 00.0000;
finish initial;
start initial1;
    v12 = 00.0000;
    v13 = 00.0000;
    v14 = 00.0000;
    v15 = 00.0000;
    v16 = 00.0000;
    v17 = 00.0000;
    v18 = 00.0000;
    v19 = 00.0000;
    v20 = 00.0000;
    v21 = 00.0000;
    v22 = 00.0000;
    v23 = 00.0000;
    v24 = 00.0000;
    v25 = 00.0000;
finish initial1;
```

```
/* module to collect values of independent variables*/
/*                                                   */
start dataget;
    /* define the data window*/
    window score_profile cmndline=cmnd msgline=msg
       group=ind_vars
       #2  "AGE_AT_ENTRANCE                    :" @40 v1 6.4
       #3  "ETHNIC_GROUP - Am. Ind             :" @40 v2 6.4
       #4  "ETHNIC_GROUP - As/Pi               :" @40 v3 6.4
       #5  "ETHNIC_GROUP - Black               :" @40 v4 6.4
       #6  "ETHNIC_GROUP - Hispanic            :" @40 v5 6.4
       #7  "ETHNIC_GROUP - Unknown             :" @40 v6 6.4
       #8  "GENDER                             :" @40 v7 6.4
       #9   "TARGETED_MINORITY_IN              :" @40 v8 6.4
       #10  "PELL_Ind                          :" @40 v9 6.4
       #11  "HIGH_SCHOOL_PERCENTI              :" @40 v10 6.4
       #12  "NEW_ACT                           :" @40 v11 6.4
       #13  "CUM_GPA_2ND_TERM                  :" @40 v12 6.4
       #14  "ChgAcdGrp_Ind - No changed        :" @40 v13 6.4
       #15  "ChangeMajor_Acadyear - +1 major   :" @40 v14 6.4
       #16  "ChangeMajor_Acadyear - No changed :" @40 v15 6.4
       #17  "m112course_Acadyear_              :" @40 v16 6.4
       #18  "m112Grade                         :" @40 v17 6.4
       #19  "m221course_Acadyear_              :" @40 v18 6.4
       #20  "m221Grade                         :" @40 v19 6.4
       #21  "ch103course_Acadyear              :" @40 v20 6.4
       #22  "ch103Grade                        :" @40 v21 6.4
       #23  "Econ101course_acadye              :" @40 v22 6.4
       #24  "Econ101Grade                      :" @40 v23 6.4
       #25  "Psy202course_Acadyea              :" @40 v24 6.4
       #26  "Psy202Grade                       :" @40 V25 6.4;
/*Collect user entries*/
do until (cmnd="EXIT");
    msg = "ENTER SUBMIT TO ADD ANOTHER YEAR, EXIT TO QUIT";
    /*                                             */
    /* loop until user types submit or exit        */
    /*                                             */
    do until (cmnd="SUBMIT"|cmnd="EXIT");
       display score_profile.ind_vars;
    end;
    if cmnd = "SUBMIT" then
       do;
          append;
          run initial1;
       end;
end;
window close=score_profile;
finish dataget;
run initial;
create score_matrix var('v1':'v25');
run dataget;
/*Create internal data matrix*/
read all var _NUM_ into X;
close score_matrix;
n = nrow(X);
/*Check that no more than 4 academic year profiles have been entered*/
if n > 4 then
    do;
       print "Number of user score profiles exceeds four!";
       abort;
    end;
p = ncol(X);
if p < 25 then
    do;
       print "Number of numeric variables less than 25!"
       abort;
    end;
/* Check for consistency of pre-admission variables across academic years*/
pre_adm = X[1:n,1:11];
row_sum = pre_adm[,+];
sum = 0;
```

```
do i = 1 to n;
    sum = sum + row_sum[i];
end;
mean = sum/n;
if mean ^= row_sum[1] then
    do;
       print "Preadmission variables are NOT constant across academic years!";
       abort;
    end;
/* Perform range checks for all independent variables*/
do i = 1 to n;
    if X[i,1] < 14 | X[i,1] > 40 then
        do;
           print "AGE_AT_ENTRANCE independent variable out-of-range!";
           abort;
        end;
    sum = 0;
    do j = 2 to 6;
        if X[i,j] ^= -1 & X[i,j] ^= 0 & X[i,j] ^= 1 then
            do;
               print "ETHNIC_GROUP must be scored either -1, 0, or 1!";
               abort;
            end;
         else
            sum = sum + X[i,j];
    end;
    if sum ^= -5 & sum ^= 1 then
            do;
               print "ETHNIC_GROUP is mis-scored!";
               abort;
            end;
    if X[i,7] ^= -1 & X[i,7] ^= 1 then
        do;
           print "GENDER must be scored either -1 or 1!";
           abort;
        end;
    if X[i,8] ^= 0 & X[i,8] ^= 1 then
        do;
           print "TARGETED_MINORITY_IN must be scored either 0 or 1!";
           abort;
        end;
    if X[i,9] ^= 0 & X[i,9] ^= 1 then
        do;
           print "PELL_Ind must be scored either 0 or 1!";
           abort;
        end;
    if X[i,10] < 5 | X[i,10] > 100 then
        do;
           print "HIGH_SCHOOL_PERCENTI independent variable out-of-range!";
           abort;
        end;
    if X[i,11] < 5 | X[i,11] > 40 then
        do;
           print "NEW_ACT independent variable out-of-range!";
           abort;
        end;
    if X[i,12] < 0 | X[i,12] > 4 then
        do;
           print "CUM_GPA independent variable out-of-range";
           abort;
        end;
    if X[i,13] ^= -1 & X[i,13] ^= 1 then
        do;
           print "ChgAcdGrp_Ind independent variable must be scored either -1 or 1!";
           abort;
        end;
    do j = 14 to 15;
        if X[i,j] ^= -1 & X[i,j] ^= 0 & X[i,j] ^= 1 then
            do;
                print "ChangeMajor_Acadyear_ independent variable must be scored either -1,
0, or 1!";
```

```
                abort;
             end;
     end;
     if X[i,16] ^= 0 & X[i,16] ^= 1 then
         do;
            print "m112course_Acadyear_ independent variable must be scored either 0 or 1!";
            abort;
         end;
     if X[i,17]  < 0 | X[i,17] > 4 then
         do;
            print "m112Grade independent variable must be scored between 0 and 4!";
            abort;
         end;
     if X[i,18] ^= 0 & X[i,18] ^= 1 then
         do;
            print "m221course_Acadyear_ independent variable must be scored either 0 or 1!";
            abort;
         end;
     if X[i,19] < 0 | X[i,19] > 4 then
         do;
            print "m221Grade independent variable must be scored between 0 and 4!";
            abort;
         end;
     if X[i,20] ^= 0 & X[i,20] ^= 1 then
         do;
            print "ch103course_Acadyear_ independent variable must be scored either 0 or 1!";
            abort;
         end;
     if X[i,21] < 0 | X[i,21] > 4 then
         do;
            print "ch103Grade independent variable must be scored between 0 and 4!";
            abort;
         end;
     if X[i,22] ^= 0 & X[i,22] ^= 1 then
         do;
            print "Econ101course_Acadye independent variable must be scored either 0 or 1!";
            abort;
         end;
     if X[i,23] < 0 | X[i,23] > 4 then
         do;
            print "Econ101Grade independent variable must be scored between 0 and 4!";
            abort;
         end;
     if X[i,24] ^= 0 & X[i,24] ^= 1 then
         do;
            print "Psy202course_Acadye independent variable must be scored either 0 or 1!";
            abort;
         end;
     if X[i,25] <  0 | X[i,25] > 4 then
         do;
            print "Psy202Grade independent variable must be scored etween 0 and 4!";
            abort;
         end;
end;
/*Check for consistency of post-admission variables across academic years*/
post_adm=X[1:n,12:25];
col_sum = post_adm[+,];
if col_sum[5] ^= 0 & col_sum[5] ^= 1 then
     do;
        print "m112course_Acadyear_ variable incorrectly entered!";
        abort;
     end;
if col_sum[7] ^= 0 & col_sum[7] ^= 1 then
     do;
        print "m221course_Acadyear_ variable incorrectly entered!";
        abort;
     end;
if col_sum[9] ^= 0 & col_sum[9] ^= 1 then
     do;
        print "ch103course_Acadyear variable incorrectly entered!";
        abort;
```

```
    end;
if col_sum[11] ^= 0 & col_sum[11] ^= 1 then
    do;
       print "Econ101course_Acadye variable incorrectly entered!";
       abort;
    end;
if col_sum[13] ^= 0 & col_sum[13] ^= 1 then
    do;
       print "Psy202course_Acadyea variable incorrectly entered!";
       abort;
    end;
/*Check for multiple grades for same course across academic years*/
/*Note: No differentiate is made between retaking course and inadvertant entry error.*/
do i = 17 to 25 by 2;
    count = 0;
    do j = 1 to n;
       if X[j,i] > 0 & X[j,i] <= 4 then count = count + 1;
    end;
    if count > 1 then
        do;
           print "Duplicate course grade!";
           abort;
        end;
end;
/*Set intercept values*/
Fr = {2.8240 3.5325 9.7106};
So = {0.0795 1.1066 8.0574};
Jr = {-1.9045 0.00903 8.1450};
Sr = {-6.2973 -2.7993};
/*Set beta values*/
Fr_beta = {0.1273, 0.9159, -.5455, -.00025, -.0230, -.1686, 0.0297, -.6858, 0.1216, -.01060,
-.1032, -1.5489,
0.0772, 1.0784, -.3865, -.3111, 0.1205, -.0190, 0.1230, 0.0991, 0.0109, -.1150, -.0801,
.01740, -.0233};
So_beta = {0.2012, 0.3632, -.3925, 0.2345, 0.1745, -.2527, 0.0146, -.3190, 0.1970, -.00315, -
.0368, -1.6661,
0.1301, 1.0112, -.5267, 0.3296, 0.00345, 0.3002, 0.0369, 0.3864, 0.0298, -.0604, -.0661,
0.0452, 0.00985};
Jr_beta = {0.1452, -.5137, 0.0140, -.0157, 0.1359, 0.2014, -.1962, 0.2582, 0.2913, 0.00401, -
.00945, -1.4478,
0.0983, 0.4826, -.0805, 0.8828, 0.0215, -.1736, 0.0599, 0.5697, 0.0450, 0.1842, -.0318,
0.2672, 0.0264};
Sr_beta = {0.1169, 0.0181, 0.2722, -.0492, -.5278, 0.00103, -.1664, 1.0599, 0.3674, 0.0192,
0.0365, -.7086,
-.2346, -.0581, -.3739, 0.5548, -.0127, 0.2061, 0.0714, 0.7414, 0.1220, 0.2256, -.0134, -
.1210, -.0758};
/*Multiply score vector times beta vector*/
if n = 1 then
    do;
       Fr_score = X[1,]*Fr_beta;
    end;
if n = 2 then
    do;
       Fr_score = X[1,]*Fr_beta;
       So_score = X[2,]*So_beta;
    end;
if n = 3 then
    do;
       Fr_score = X[1,]*Fr_beta;
       So_score = X[2,]*So_beta;
       Jr_score = X[3,]*Jr_beta;
    end;
if n = 4 then
    do;
       Fr_score = X[1,]*Fr_beta;
       So_score = X[2,]*So_beta;
       Jr_score = X[3,]*Jr_beta;
       Sr_score = X[4,]*Sr_beta;
    end;
/*Create shell 6 x 6 transition probability matrix*/
P_one = J(6,6,0);
```

```
P_one[5,5] = 1;
P_one[6,6] = 1;
/*Compute transition probabilities*/
if n = 1 then
    do;
       Fr_SO = Fr[1] + Fr_score;
       Fr_FF = Fr[2] + Fr_score;
       Fr_FS = Fr[3] + Fr_score;
       P_one[1,5] = 1/(1 + exp(-Fr_SO));
       P_one[1,1] = (1/(1 + exp(-Fr_FF))) - P_one[1,5];
       P_one[1,2] = (1/(1 + exp(-Fr_FS))) - (1/(1 + exp(-Fr_FF)));
       P_one[1,3] = 1 - (1/(1 + exp(-Fr_FS)));
    end;
if n = 2 then
    do;
       Fr_SO = Fr[1] + Fr_score;
       Fr_FF = Fr[2] + Fr_score;
       Fr_FS = Fr[3] + Fr_score;
       P_one[1,5] = 1/(1 + exp(-Fr_SO));
       P_one[1,1] = (1/(1 + exp(-Fr_FF))) - P_one[1,5];
       P_one[1,2] = (1/(1 + exp(-Fr_FS))) - (1/(1 + exp(-Fr_FF)));
       P_one[1,3] = 1 - (1/(1 + exp(-Fr_FS)));
       So_SO = So[1] + So_score;
       So_SS = So[2] + So_score;
       So_SJ = So[3] + So_score;
       P_one[2,5] = 1/(1 + exp(-So_SO));
       P_one[2,2] = (1/(1 + exp(-So_SS))) - P_one[2,5];
       P_one[2,3] = (1/(1 + exp(-So_SJ))) - (1/(1 + exp(-So_SS)));
       P_one[2,4] = 1 - (1/(1 + exp(-So_SJ)));
    end;
if n = 3 then
    do;
       Fr_SO = Fr[1] + Fr_score;
       Fr_FF = Fr[2] + Fr_score;
       Fr_FS = Fr[3] + Fr_score;
       P_one[1,5] = 1/(1 + exp(-Fr_SO));
       P_one[1,1] = (1/(1 + exp(-Fr_FF))) - P_one[1,5];
       P_one[1,2] = (1/(1 + exp(-Fr_FS))) - (1/(1 + exp(-Fr_FF)));
       P_one[1,3] = 1 - (1/(1 + exp(-Fr_FS)));
       So_SO = So[1] + So_score;
       So_SS = So[2] + So_score;
       So_SJ = So[3] + So_score;
       P_one[2,5] = 1/(1 + exp(-So_SO));
       P_one[2,2] = (1/(1 + exp(-So_SS))) - P_one[2,5];
       P_one[2,3] = (1/(1 + exp(-So_SJ))) - (1/(1 + exp(-So_SS)));
       P_one[2,4] = 1 - (1/(1 + exp(-So_SJ)));
       Jr_SO = Jr[1] + Jr_score;
       Jr_JJ = Jr[2] + Jr_score;
       Jr_Sr = Jr[3] + Jr_score;
       P_one[3,5] = 1/(1 + exp(-Jr_SO));
       P_one[3,3] = (1/(1 + exp(-Jr_JJ))) - P_one[3,5];
       P_one[3,4] = (1/(1 + exp(-Jr_Sr))) - (1/(1 + exp(-Jr_JJ)));
       P_one[3,6] = 1 - (1/(1 + exp(-Jr_Sr)));
    end;
if n = 4 then
    do;
       Fr_SO = Fr[1] + Fr_score;
       Fr_FF = Fr[2] + Fr_score;
       Fr_FS = Fr[3] + Fr_score;
       P_one[1,5] = 1/(1 + exp(-Fr_SO));
       P_one[1,1] = (1/(1 + exp(-Fr_FF))) - P_one[1,5];
       P_one[1,2] = (1/(1 + exp(-Fr_FS))) - (1/(1 + exp(-Fr_FF)));
       P_one[1,3] = 1 - (1/(1 + exp(-Fr_FS)));
       So_SO = So[1] + So_score;
       So_SS = So[2] + So_score;
       So_SJ = So[3] + So_score;
       P_one[2,5] = 1/(1 + exp(-So_SO));
       P_one[2,2] = (1/(1 + exp(-So_SS))) - P_one[2,5];
       P_one[2,3] = (1/(1 + exp(-So_SJ))) - (1/(1 + exp(-So_SS)));
       P_one[2,4] = 1 - (1/(1 + exp(-So_SJ)));
       Jr_SO = Jr[1] + Jr_score;
```

```
        Jr_JJ = Jr[2] + Jr_score;
        Jr_Sr = Jr[3] + Jr_score;
        P_one[3,5] = 1/(1 + exp(-Jr_SO));
        P_one[3,3] = (1/(1 + exp(-Jr_JJ))) - P_one[3,5];
        P_one[3,4] = (1/(1 + exp(-Jr_Sr))) - (1/(1 + exp(-Jr_JJ)));
        P_one[3,6] = 1 - (1/(1 + exp(-Jr_Sr)));
        Sr_SO = Sr[1] + Sr_score;
        Sr_SrSr = Sr[2] + Sr_score;
        P_one[4,5] = 1/(1 + exp(-Sr_SO));
        P_one[4,4] = 1/(1 + exp(-Sr_SrSr)) - P_one[4,5];
        P_one[4,6] = 1 - (1/(1 + exp(-Sr_SrSr)));
    end;
/*Output one-step and multiple step transition matrices*/
if n < 4 then print P_one;
else
    do;
       print "One-step transition";
       print P_one;
       print "Four-step transition";
       four = P_one**4;
       print four;
       print "Five-step transition";
       five = P_one**5;
       print five;
       print "Six_step transition";
       six = P_one**6;
       print six;
    end;
quit;
run;
```

Appendix E: Simulations Output TP, $(TP)^4$, $(TP)^5$, $(TP)^6$

Table E-1 Policy 1: Student 18 years old

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0369407	0.8995926	0.0246998	0	0.0387669	0
0	0.1087618	0.8127312	0.0042579	0.0742491	0
0	0	0.1060318	0.8709348	0.0210257	0.0020077
0	0	0	0.2152252	0.0086526	0.7761223
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
1.8622E-6	0.0017293	0.0321434	0.2997203	0.1435814	0.5228237
0	0.0001399	0.0040276	0.0900851	0.1134584	0.7922889
0	0	0.0001264	0.0161062	0.0340791	0.9496883
0	0	0	0.0021457	0.0110019	0.9868524
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

F	S	J	Sr	S_O	Gr
6.879E-8	0.0001898	0.0048138	0.0925096	0.146979	0.7555078
0	0.0000152	0.0005408	0.022897	0.114333	0.862214
0	0	0.0000134	0.0035766	0.0342211	0.962189
0	0	0	0.0004618	0.0110204	0.9885177
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
2.5411E-9	0.0000207	0.0006646	0.0241037	0.1478948	0.8273162
0	1.6552E-6	0.0000697	0.0053991	0.1145436	0.879986
0	0	1.4211E-6	0.0007814	0.0342523	0.9649648
0	0	0	0.0000994	0.0110244	0.9888762
0	0	0	0	1	0
0	0	0	0	0	1

Table E-2 Policy 1: Student 25 years old

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0769262	0.8232754	0.0102817	0	0.0895167	0
0	0.2311132	0.5208614	0.0010446	0.2469808	0
0	0	0.2308081	0.7124425	0.0560219	0.0007275
0	0	0	0.3759428	0.0193996	0.6046576
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
0.000035	0.0150464	0.0865844	0.282206	0.4225794	0.1935488
0	0.002853	0.0256683	0.1763948	0.3808991	0.4141849
0	0	0.0028379	0.084123	0.0987215	0.8143175
0	0	0	0.019975	0.0304654	0.9495596
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

	S	J	Sr	S_O	Gr
2.6938E-6	0.0035063	0.0278218	0.1677954	0.436624	0.3642498
0	0.0006594	0.0074105	0.0846045	0.3864637	0.520862
0	0	0.000655	0.0336473	0.1005125	0.8651852
0	0	0	0.0075095	0.0308529	0.9616377
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
2.0723E-7	0.0008126	0.0082478	0.0829066	0.442304	0.4657288
0	0.0001524	0.0020538	0.0370866	0.388683	0.5720242
0	0	0.0001512	0.0131161	0.1012019	0.8855308
0	0	0	0.0028231	0.0309986	0.9661783
0	0	0	0	1	0
0	0	0	0	0	1

Table E-3 Policy 2: Student A: African-American, female, Targeted minority, receiving a Pell Grant with a HS%80 and ACT score of 24

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0225507	0.9128013	0.0417324	0	0.0229157	0
0	0.1244764	0.7838135	0.003534	0.0881762	0
0	0	0.1434655	0.825094	0.0300477	0.0013928
0	0	0	0.4665776	0.0288525	0.5045699
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
2.586E-7	0.0021477	0.0434215	0.4589036	0.1653199	0.3302071
0	0.0002401	0.0075767	0.2570053	0.1643781	0.5707998
0	0	0.0004236	0.119935	0.0806605	0.7989809
0	0	0	0.047391	0.0515261	0.9010829
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

F	S	J	Sr	S_O	Gr
5.8317E-9	0.0002676	0.0079129	0.2499485	0.1800545	0.5618165
0	0.0000299	0.0012752	0.1261653	0.1720422	0.7004875
0	0	0.0000608	0.0563085	0.0841336	0.8594971
0	0	0	0.0221116	0.0528935	0.924995
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
1.315E-10	0.0000333	0.001345	0.1231502	0.1875275	0.687944
0	3.7198E-6	0.0002064	0.0599181	0.1757234	0.7641484
0	0	8.7194E-6	0.0263224	0.0857601	0.8879087
0	0	0	0.0103168	0.0535315	0.9361518
0	0	0	0	1	0
0	0	0	0	0	1

Table E-4 Policy 2: Student B: White, female, Not targeted minority, not receiving a Pell Grant with a HS%80 and ACT score of 24

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0321285	0.9057683	0.0287555	0	0.0333477	0
0	0.1045769	0.8202124	0.004486	0.0707247	0
0	0	0.1069928	0.8697756	0.021245	0.0019866
0	0	0	0.2378722	0.0098663	0.7522615
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
1.0655E-6	0.001482	0.0308087	0.3142444	0.1353598	0.518104
0	0.0001196	0.0038843	0.100324	0.110938	0.7847341
0	0	0.000131	0.0204061	0.0361305	0.9433323
0	0	0	0.0032017	0.0129043	0.983894
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

F	S	J	Sr	S_O	Gr
3.4234E-8	0.0001559	0.0045119	0.1015533	0.1392197	0.7545591
0	0.0000125	0.0005137	0.0272433	0.1120188	0.8602117
0	0	0.000014	0.004968	0.0363346	0.9586834
0	0	0	0.0007616	0.0129359	0.9863025
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
1.0999E-9	0.0000163	0.0006107	0.0280817	0.1403285	0.8309627
0	1.308E-6	0.0000652	0.0069273	0.1122994	0.8807068
0	0	1.5001E-6	0.001194	0.0363839	0.9624207
0	0	0	0.0001812	0.0129434	0.9868754
0	0	0	0	1	0
0	0	0	0	0	1

Table E-5 Policy 3: Student X

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0692692	0.8400683	0.0117874	0	0.0788751	0
0	0.1642569	0.7036814	0.0022946	0.1297671	0
0	0	0.2615869	0.6703258	0.0674907	0.0005966
0	0	0	0.5084705	0.0347315	0.456798
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
0.000023	0.0062343	0.1023587	0.402581	0.2990837	0.1897192
0	0.0007279	0.0285898	0.289801	0.2612982	0.419583
0	0	0.0046824	0.1687784	0.14289	0.6836492
0	0	0	0.0668441	0.0659368	0.8672191
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

F	S	J	Sr	S_O	Gr
1.5948E-6	0.0010434	0.0311629	0.2733286	0.320785	0.3736785
0	0.0001196	0.007991	0.1665215	0.2733874	0.5519806
0	0	0.0012248	0.0889576	0.149068	0.7607496
0	0	0	0.0339883	0.0682584	0.8977533
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
1.1047E-7	0.0001727	0.008886	0.1598712	0.3325169	0.498553
0	0.0000196	0.0021745	0.0900281	0.2797258	0.628052
0	0	0.0003204	0.0460533	0.1522402	0.801386
0	0	0	0.017282	0.0694389	0.9132791
0	0	0	0	1	0
0	0	0	0	0	1

Table E-6 Policy 3: Student Y

One-step transition

P_ONE

F	S	J	Sr	S_O	Gr
0.0484325	0.8811358	0.018194	0	0.0522377	0
0	0.0651102	0.8862447	0.0080376	0.0406074	0
0	0	0.1258343	0.8468571	0.0256714	0.0016372
0	0	0	0.4156914	0.023112	0.5611966
0	0	0	0	1	0
0	0	0	0	0	1

Four-step transition

FOUR

F	S	J	Sr	S_O	Gr
5.5024E-6	0.0006588	0.0311847	0.4389845	0.136594	0.3925725
0	0.000018	0.0033969	0.2111607	0.099355	0.6860693
0	0	0.0002507	0.0865064	0.0642469	0.848996
0	0	0	0.0298596	0.0383734	0.931767
0	0	0	0	1	0
0	0	0	0	0	1

Five-step transition

FIVE

F	S	J	Sr	S_O	Gr
2.6649E-7	0.0000477	0.0045081	0.2088963	0.1475674	0.6389802
0	1.1702E-6	0.0004434	0.0906545	0.1043233	0.8045776
0	0	0.0000315	0.0361723	0.0662526	0.8975435
0	0	0	0.0124124	0.0390635	0.9485241
0	0	0	0	1	0
0	0	0	0	0	1

Six_step transition

SIX

F	S	J	Sr	S_O	Gr
1.2907E-8	3.3434E-6	0.0006096	0.0906545	0.1525131	0.7562195
0	7.6189E-8	0.0000568	0.0380598	0.1064299	0.8554534
0	0	3.97E-6	0.0150632	0.0670895	0.9178433
0	0	0	0.0051597	0.0393503	0.9554899
0	0	0	0	1	0
0	0	0	0	0	1