

ABSTRACT

CHINASA VICTOR UKPABI. Formulating a Prediction Model of Retention Rate in the University of North Carolina System (Under the direction of Drs. Robert Serow and Duane Akroyd)

Institutional effectiveness has become an issue in American Higher Education as governing bodies now require evidence of quality in accountability and funding issues. One measure widely used today to assess effectiveness is retention rate. In response to the above as well as to demographic changes, increasing costs, and the intense competition for new students, educational institutions are seeking new methods to increase the retention rate of their institutions.

Most research on retention focused on individual student-level variables, which only predict persistence. There is an obvious need to understand the impact of some uncontrollable external influences on retention rates. In light of the importance that retention rate has assumed, this study sought to develop a predictive model of retention rate in the 16-campus University of North Carolina System.

In an effort to develop a comprehensive model, this study employed selected institution-level variables. The study will fill a void, as the UNC system does not currently have a general statistical model for predicting retention rates in its multi-campus system.

The central research question for the study is whether an institution's retention rate is a function of the demographic characteristics, economic conditions, college management, and fiscal policy of the state in which the institution resides? Pooled cross-sectional time series technique was employed and the method of Ordinary Least Squares (OLS) was used in the estimation of the regression equation. Data was pooled for ten years to provide greater number of data points to overcome a potential degree of freedom problem that would arise studying only 15 institutions.

Four themes emerged from the analysis: headcount enrollment, amount of education and general expenditure on instruction and academic support, the county population where institution is located and the rate of unemployment in the county, are significant predictors of retention rate for an institution.

Future research including developing prediction models for minority institutions and employing only external variables will perhaps provide additional insights in our understanding of retention rate behavior.

**FORMULATING A PREDICTION MODEL OF RETENTION
RATE IN THE UNIVERSITY OF NORTH CAROLINA
SYSTEM**

By

CHINASA VICTOR UKPABI

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

EDUCATIONAL RESEARCH AND POLICY ANALYSIS

Raleigh

2005

APPROVED BY:

Dr. Troy Chen

Member of Advisory Committee

Dr. Dennis Daley

Member of Advisory Committee

Dr. Robert Serow

Co-Chair of Advisory Committee

Dr. Duane Akroyd

Co-Chair of Advisory Committee

DEDICATION

This Dissertation is dedicated to my late father, Engr. Emmanuel Ozuomba Ukpabi (1930-2002), who would never read this work that he very much wanted to read. It is also dedicated to my mom, Mrs. Rose Ulunma Ukpabi, who would read it in his behalf. May the good Lord bless them both.

BIOGRAPHY

Chinasa Victor Ukpabi was born on July 5, 1959 in Port Harcourt, Rivers State of Nigeria. He graduated from the College of Technology, Owerri, Imo State of Nigeria with a degree in Marketing in 1987. In search of greener pastures, he emigrated to the United States of America in 1989.

In the fall of 1990, he enrolled at the North Carolina Central University, Durham, where he completed his Masters in Business Administration (MBA) in 1994. And in the fall of 1996, he enrolled at the North Carolina State University, Raleigh and completed the course work and preliminary examinations in 1998 and 2004, respectively, for a Doctor of Philosophy in Educational Research and Policy Analysis.

Mr. Ukpabi worked in the Offices of Research, Evaluation and Planning at the North Carolina Central University, Durham and Institutional Effectiveness at Bennett College in Greensboro, North Carolina, between 2000 and 2004. He currently holds the position of Director of the Center for Survey Research at Shaw University, Raleigh and also runs an Institutional Effectiveness Consulting firm in Durham, North Carolina. He holds an active membership in the American Educational Research Association (AERA). He is happily married with three children, a beautiful daughter and two fine boys.

ACKNOWLEDGEMENTS

First of all, I would like to thank all four members of my dissertation committee— Dr. Robert Serow, Dr. Duane Akroyd, Dr. Dennis Daley, and Dr. Troy Chen, for their constructive comments. I would especially like to thank Dr. Serow, the Chair of my committee, and Dr. Akroyd, the co-chair, for their invaluable guidance throughout the process. They reviewed drafts in good time and were always available to discuss and resolve issues of concern. Dr. Serow has been my mentor since I entered the Educational Research and Policy Analysis program and I am eternally indebted to him.

Many other individuals also provided invaluable help. I wish to thank Drs. Fitzpatrick Anyanwu and Ita Ekanem, for their suggestions during the initial stages of this work. I want to thank Drs. Cheryl Thompkins and P.M. Mutisya for their advice. Dr. James Guseh was there for me. I cannot thank him enough.

I also wish to thank my five brothers and a sister, namely, Kachi, Nkem, Chinedu, Uzonna (Rev. Fr.), Ugwunna, and Akudo, for coming around with their emotional support after so many years of sitting on the fence. Above all, I want to thank my wife, Tochukwu, and my two sons, Chinasa II and Chukwuemeka, and my only daughter, Chidinma, for all their impatience during the course of this study.

TABLE OF CONTENTS

	PAGE
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii

CHAPTER ONE

Introduction.....	1
Conceptual Framework.....	7
Statement of the Problem.....	11
Purpose of the Study.....	14
Research Questions.....	14
Significance of the Study.....	18
Limitations of the Study.....	19
Definition of Terms.....	20
Summary.....	24

CHAPTER TWO

Review of the Literature	
Introduction.....	26
Relevant Research on Retention.....	26
Overview of Retention Models.....	30
Spady's Model.....	30
Tinto's Model.....	34
Pascarella's Model.....	37
Wyman's Model.....	40
Relationship Between Retention and Predictor Variables.....	42
Retention and Enrollment.....	42
Retention and Employment.....	44
Retention and Funding.....	46
Retention and Financial Aid.....	47
Effects of Low Retention.....	49
Effects of Low Retention on the Institution.....	50
Effects of Low Retention on Students.....	52
The Need for Increased Retention.....	53
Summary.....	55

CHAPTER THREE

Methodology	
Sample.....	57
Overview of Statistical Methods.....	63
Pooled Cross-section Analysis.....	66
Selection of Variables.....	73
The Dependent Variable.....	73
Independent Variables.....	75
Data and Methodological Framework.....	78
Assumptions of the Model.....	81
Data Source and Quality.....	85
Test of Stability of Model.....	88
The Two Big Schools (NCSU and UNC-CH).....	89
North Carolina State University.....	90
University of North Carolina.....	91

CHAPTER FOUR

Results	
Introduction.....	94
Data and Findings.....	95
The Full Regression Model.....	97
Impact of the Two Big Schools.....	104
Differences Between the Thirteen Small Schools and the Two Big Schools.....	107
Stability of the Model.....	112

CHAPTER FIVE

Summary, Discussion, Recommendations and Implications	
Rationale of the Study.....	115
Summary of Results.....	118
Discussion.....	121
Recommendations and Implications for Practice.....	126
Recommendations for Further Research.....	134
Overall Conclusions.....	140
References.....	143

APPENDICES

Appendix A.....	163
Appendix B.....	164
Appendix C.....	165
Appendix D.....	166

LIST OF TABLES

Table		Page
3.1	UNC Institutions and County Locations.....	61
3.2	Ten Retention Rates for all Schools in the Study.....	62
3.3	Pearson's Correlation Matrix For the Full Model.....	76
3.4	Independent Variables.....	77
3.5	Abbreviations of Variables.....	87
4.1	Descriptive Statistics of Variables.....	96
4.2	Regression Result of the Full Model.....	99
4.3	Regression Result of the Thirteen Schools...	106
4.4	Regression Result of the Two Big Schools....	107
4.5	Regression Result-First 5 Yrs (1991-1995)...	113
4.6	Regression Result-Last 5 Yrs (1996-2000)....	114

LIST OF FIGURES

Figure		Page
1.0	An Adaptation from Wyman's Predictive Model of Retention rate.....	13
2.1	A Replication of Spady's Model of College Dropout Process.....	34
2.2	A Replication of Tinto's Model of College Student Attrition.....	36
2.3	A Replication of Pascarella's Model of College Student Attrition.....	39
2.4	Work Experience and Average Annual Earnings of Workers 25 to 60 Years Old by Educational Attainment.....	54
3.1	NC Map of Public 4-Yr. Institutions and County Locations.....	60
3.2	Model of Internal and External Variables.....	74
3.3	Ukpabi's Model of Retention Rate.....	82

CHAPTER ONE

INTRODUCTION

The issue of retention of students has become the focus of increased attention in recent years, partly because of a general reduction in both the pool of available students and the level of allocable funding for institutions of higher education (McGrath & Braunstein, 1997). This situation, however, is changing, as there have been appreciable increases in the number of high school graduates who are college-bound. In a report published in 2003, the National Center for Education Statistics (NCES) stated that the total number of high school graduates increased by 3 percent between 1987-88 and 2000-2001, and is projected to increase 11 percent between 2000-2001 and 2012-13 (NCES, 2003). According to the report, the trend in North Carolina is similar to the national pattern. The number of high school graduates in North Carolina was projected to increase by about 20.6 percent between 2000-2001 and 2012-13 (NCES, 2003).

Another important factor that has had a significant impact on the business of higher education within the past two decades is the general public erosion of trust in all types of public institutions (Penn 1999). This public erosion of trust in higher education may not be unrelated to the general low rates of retention recorded in these institutions. For

instance, only about 53 percent of postsecondary students entering four-year institutions in pursuit of a bachelor's degree in 1995-1996 earned a degree within five years (National Center for Education Statistics, 2002). About 20 percent of the 1999 beginning students left postsecondary education before the beginning of their second year of study at the same institution in the year 2000 (CSRDE, 2001). Sixteen percent of students enrolled in the four-year sector left, while a whopping forty-two percent of students enrolled in the public two-year sector left their institutions (National Center for Education Statistics, 1998). According to NCES (2004) approximately 20 percent of the freshmen enrolled in colleges and universities drop out before completing their programs.

Generally, from 1983 to 1999, research showed very little change in the first-year attrition rate in the United States. During this period, for institutions offering a Ph.D., first-year attrition ranged from 23 percent to 27 percent. Over the same period, however, graduation rates decreased from 58 percent to 52 percent (ACT, 2000). For bachelor's degree granting institutions, however, the freshmen-to-sophomore attrition rate in 2000 was 25.8 percent. In the same period, the percentage of students who earned a bachelor's degree within five years was 51.2 percent (ACT, 2001). In the period

between 1977 and 1997, the number of students enrolled in college increased by about 26 percent to an estimated 14.6 million. This number is projected to rise to approximately 17.5 million students by the year 2010 (U.S Department of Education, 1999). Retention rates, however, have not recorded a similar increasing trend with college enrollment rates.

According to Peltier, Laden, and Matranga (1999), degree completion rates have been decreasing. Furthermore, the United States Department of Education (1999) noted that only 43 to 50 percent of freshman classes would persist to degree completion. Murtaugh et al. (1999) noted that the greatest attrition tends to occur between the freshman and sophomore years. Given the disproportionate number of students who leave college prior to their sophomore year (Levitz et al., 1999), this time period appears to be an appropriate focus for retention studies. Whatever is the cause of this low retention rate phenomenon, the problem must be addressed considering the increasing significance of retention rates in accountability and funding issues in higher education.

Higher retention rate is important as most states now use a funding formula to determine appropriations for public institutions of higher education, and most of these formulas give some consideration to the number of students enrolled in determining the budget for that institution (McKeown, 1996).

As a consequence, if the retention rate of an institution declines, public appropriations for that institution usually fall as well. This situation signaled the start of increased competition among institutions for eligible potential students. As a reaction to the dearth of funds, legislators began introducing legislation demanding greater accountability and mandating reports of certain statistical measures.

This overall situation has heightened the interest of public colleges and universities in institutional effectiveness. As noted by Stedman (2002), through the Reauthorization of the Higher Education Act of 1998, the United States Congress provided yet another opportunity for states to scrutinize their public institutions of higher education using a broad variety of measures. According to Wyman (1997), one measure widely used as an indicator of effectiveness is retention rate, the percent of entering students graduating or persisting in their studies at an institution (or attrition rate, the complement of retention rate).

In its policy on institutional effectiveness, the Commission on Institutions of Higher Education (CIHE) reaffirmed the importance of each institution measuring its effectiveness. An institution's efforts and ability to assess its effectiveness and use the obtained information for its

improvement are important indicators of institutional quality (Commission on Institutions of Higher Education, 1992). The Southern Association of Colleges and Schools (SACS) (1995), the major accreditation body in the southern United States, equally requires colleges and universities in its jurisdiction to measure the effectiveness of education they provide. The state of Virginia currently uses predicted rates as an element in performance funding for its university system (State Council of Higher Education of Virginia, 2004). The states of New York and North Carolina had also adopted similar uses in their university systems budgets (Burke and Minassians, 2003).

The calculation of retention rates of institutions of higher education has become one of the major components in evaluating institutional effectiveness. In light of the increasing importance that retention rate has assumed in recent times, especially as it relates to accountability in higher education, it has become necessary to have a fuller understanding of retention rate behavior and its use in evaluating institutional effectiveness.

A considerable amount of research on student retention and persistence has been conducted in the last several decades, but much of the research has focused on individual student characteristics. Few studies have focused on retention at the institution level. Although such individual level

research may provide some explanation about differences in the retention of individual students, it has not been able to explain in a conclusive way the differences in retention rates among colleges and universities. Even the few studies that have utilized institutional level predictor variables have sometimes combined them with student characteristics or used them to study community colleges and not four-year colleges. There is a dearth of research studies that employed institutional-level predictor variables in studying four-year colleges.

To this end, there is an opportunity to employ only institution level variables to predict retention rate at four-year colleges and universities. This study employed a set of predictor variables already identified in an earlier study (Wyman, 1997) as internal and external variables in modeling retention rate. The external variables are those that are beyond the control of the institutions, and internal variables are those that are somewhat controllable by the institutions. Both sets of variables are combined to formulate a predictive model of retention rate for the colleges and universities in the University of North Carolina system. Employing both external and internal variables has resulted in a model that increased the amount of retention rate variance explained.

Furthermore, such a model could be used to obtain a more accurate description of the degree to which institutions are locked into a certain retention range due to the influence of the uncontrollable forces. In other words, the study was able to formulate a predictive model of retention rate as well as identify what external institutional factors account for retention rate differentials among colleges and universities in the University of North Carolina system.

Additionally, the University of North Carolina system does not currently have a predictive statistical model of retention rate. This study is therefore timely, as it fills this void by producing a regression model for predicting retention rates in the system. It also provides an alternative check for the accuracy of retention rates derived by other methods.

Conceptual Framework

There are a number of theories on the issue of student retention. Notable among these theories are the works of several scholars, which are highlighted here.

Spady (1970) provided the first theoretical model of the drop out process in higher education, (see Figure 2.1). This model proposed that social integration increases institutional commitment, which in turn, reduces the likelihood that a

student will drop out. The model also included student's background characteristics as some of the factors influencing retention.

Based on Spady's research, Tinto (1975) introduced an explicit theoretical model that provided a description of the process of student integration into academic and social systems. The model suggests that personal characteristics and integration of students in their new environment combine to influence their success. This model is provided in Figure 2.2.

Pascarella (1980) adapted the earlier works of Spady (1970), Astin (1970), and Feldman (1970) and used them to support the inclusion of different sets of variables in his model. He zeroed in on the student-faculty interaction aspect in Tinto's model and suggests that student persistence is influenced by the interaction between a student's pre-enrollment dispositions and institutional variables, including the peer culture, the classroom, extracurricular activities, etc. He believes that these experiences during college influence the degree and quality of interactions among students, faculty, and various outcome measures.

This model proposed that to fully understand the influence of this student-faculty non-classroom interaction on educational outcomes and institutional persistence, it is important to give consideration to both the student's

background characteristics, college experiences and other institutional factors. Pascarella's model is represented in Figure 2.3.

Although the above models are some of the most widely cited, their predictive powers account for less than 50 percent of the variance in departure rates (Grayson and Grayson, 2003). Several studies have been done to better understand and increase the amount of variance that is explained by these theories. Building on the work of Tinto (1975), Bean and Metzger (1985), and Cabrera et al. (1992), Sandler (2000) adds the concepts of "career decision making self-efficacy", "perceived stress", and "financial difficulty" to the evolving model in an effort to increase the amount of variance that is explained by these theories. Allen (1999) employed a combination of motivation and background factors based on Bean's (1982) Student Attrition Model in a bold effort to have a fuller explanation of retention and persistence, while Tomlinson-Clarke and Clarke (1996) introduced gender as a variable in their research on alienation and persistence of students.

Most of these theories, unfortunately, are not applicable to the study of retention rate because samples are typically drawn at the individual student level. While such retention studies may help explain differences in the retention

likelihood of individual students, they are deficient in explaining the underlying determinants of inter-institutional differences in retention rates.

Wyman (1997), however, proposed the utilization of institutional characteristics to develop a predictive model of retention rate. Although there are several conceptualizations in the retention literature reviewed in Chapter Two which make contributions to the study of student retention, Wyman's model, illustrated in Figure 1.0 below, offered the most appeal. This model is adopted for this study because it employed institutional level measures, which were also divided into internal and external sets of variables. Employing both external and internal variables has resulted in a model that maximizes the amount of retention rate variance explained. Furthermore, such a model seems appropriate for obtaining a more accurate description of the degree to which institutions are locked into a certain retention range due to the influence of the uncontrollable external forces. In other words, the study produced a predictive model of retention rate as well as identified what external institutional factors accounted for retention rate differentials among institutions.

The Wyman model also explained about 66 percent of retention rate variance. Although the study was done at the community college level in South Carolina, it is replicated

here at the university level in the state of North Carolina. As in the Wyman study, this study also defined each observation as a single institution characterized by retention rate and by the variables that represent the demographic and economic indices of the institution's location. This kind of sampling framework that involves the aggregation of such indices had been used in the past for modeling homelessness trends over time (Kemp et al, 2001) and unemployment persistence in Italy (Amisano & Serati, 2003).

Statement of the Problem

The primary problem that prompted this study is the need to develop a predictive model of retention rate in the University of North Carolina System. The new model would explain retention rate variance more than past retention models using the 16-campus University of North Carolina system as a sample. It would also be beneficial to understand the impact of uncontrollable external influences on the retention rates of these institutions.

Institutions of higher education need accurate predictions of retention and enrollment rates to make sound academic and financial policy decisions. Educational administrators and planners have the challenge of designing and, of course, applying research on retention. There is a

problem of which retention model to use. Most of the existing models have been criticized for their poor predictive ability. Additionally, many of the models focused on student-level variables as opposed to institution-level variables. Those that employed institution-level variables have tended to generate models that only explain part of retention rate variance.

As reported by Grayson and Grayson (2003) most of the early models account for less than 50 percent of the variance in departure rates. According to them, the student attrition model (SAM) (Bean and Metzger, 1982) explained 44 percent of the variance in persistence, while the student integration model (SIM) (Tinto, 1975) explained only 38 percent. Furthermore, a model that resulted from a combination of the models of Tinto (1975), Bean and Metzger (1982), and Cabrera et al. (1992), accounted for 43 percent of the variance in persistence (Sandler, 2000). There is indeed a need for a retention rate model that would explain retention rate more than these earlier ones.

Educational institutions come in various sizes with differing budgets, funding capabilities, and enrollments. They are also located in areas with differing demographic, social, and economic characteristics. These differing circumstances may have differential impacts on their individual retention

rates. In fact, it is possible that using retention rates to compare institutions inappropriately penalizes institutions for “low” retention rates that, after correcting for external forces beyond institutional control, are actually acceptable or even meritorious (Wyman, 1997).

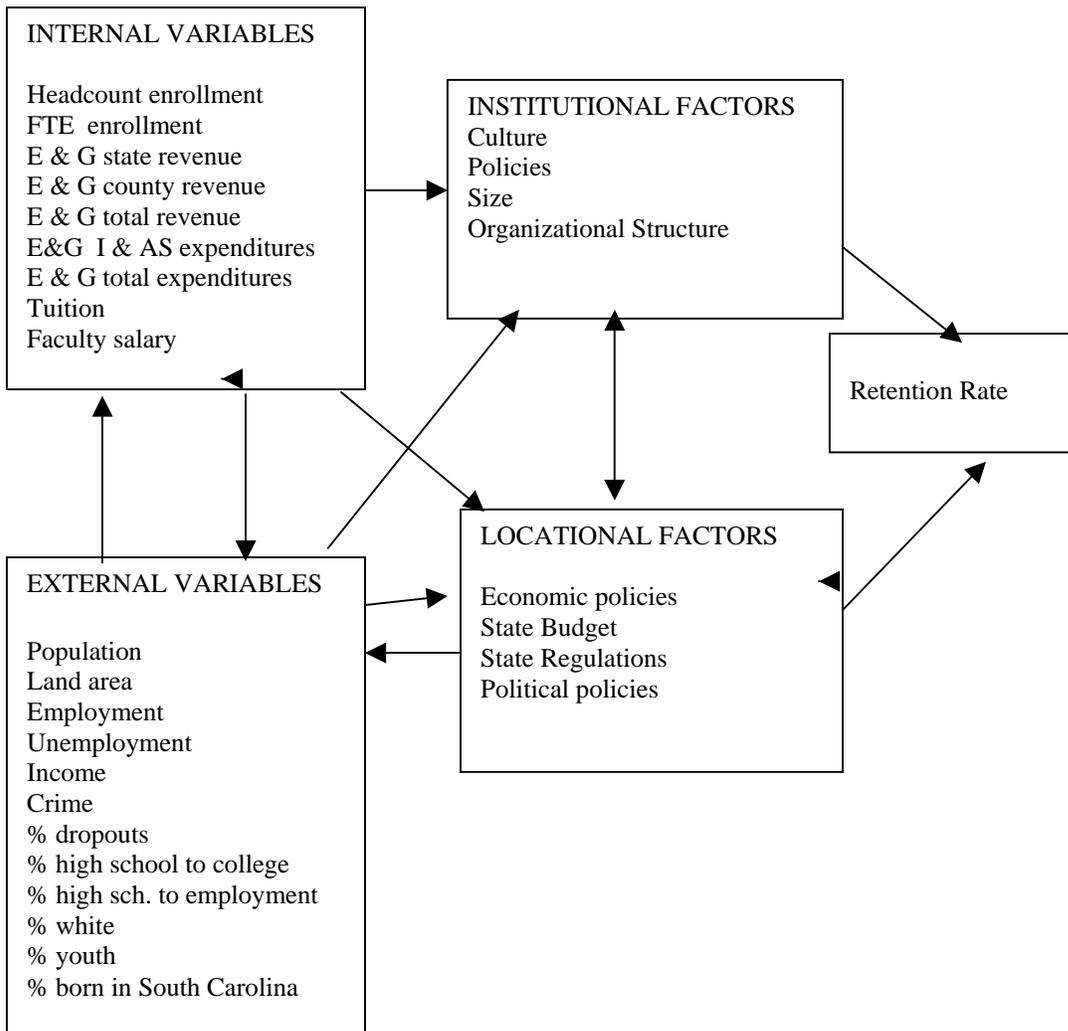


Figure 1.0: An Adaptation from Wyman’s Predictive Model of Retention Rate (1997)

Purpose of the Study

The purpose of this study is to provide a better predictive model of retention rates for colleges and universities, especially those in the University of North Carolina system. This venture was accomplished by developing a predictive model utilizing institutional-level variables that are within and those that are beyond the control of the various institutions in the system.

The whole study was strengthened by a detailed analysis of data obtained on 15 of the 16 institutions in the UNC system between fall of 1991 and fall of 2000. The hope is that this model would have the ability to provide better prediction of retention rate variance at these institutions. The study was also aimed at shedding more light on the direct impact of external forces that are uncontrollable by the institutions, which is not the case with most models previously employed in retention studies.

Research Questions

The central question of this study is the extent to which an institution's retention rate is a function of the demographic characteristics, economic conditions, management, and fiscal policy of the county in which the institution is

located. Drawing from this question, the specific research questions are as follows:

1. What is the predictive ability of the internal variables (headcount enrollment, education and general expenditures on instruction and academic support, average financial aid) and external variables (population, unemployment, crime) in determining an institution's retention rate?
2. Is the model of retention rate developed from the above variables stable over time?

Significance of Variables in Retention Studies

A number of studies have been done on retention in higher education using various variables related to student demographics, social as well as economic characteristics. This study has employed six independent variables that are significant in the study of retention rate, including the following: headcount enrollment, expenditures on instruction and academic support, financial aid, population, unemployment, and crime.

Headcount enrollment

Many research studies show evidence of the relationship between an institution's retention rate and student enrollment. According to Wetzal, et al. (1999), retaining current students into a second and a third year of study and

perhaps, even to graduation, may be desired to maintain enrollment levels in general. Wyman (1997) also stated that retention is a key issue in enrollment planning. It seems obvious that enrollments are stabilized by the retention of current students as well as the steady inflow of new students.

Unemployment rate

The vagaries of employment in recent times could impact the student's educational plans, in light of the increasing cost of going to college. This seems likely because an average student would consider the possibility of finding a good job on graduation from college. According to Wyman (1997), there is a very strong positive relationship between the institution's retention rate and regional employment per capita.

Funding

The issue of the retention of students and its relationship to the level of funding for institutions of higher education has received increased attention in recent years (McGrath and Braunstein, 1997). Retention experts agree that many colleges and universities lack the funds to develop truly effective programs. The perception of poor quality or ineffectiveness can have significant negative consequences on funding and on the institution's ability to attract and retain students.

Financial aid

A few studies have shown that meeting the financial need of minority students may be a key factor in addressing the problem of attrition in minority institutions. Georges (2000) found that there is a positive and statistically significant correlation between minority retention rate and the average scholarship and fellowship of students.

Population

Research on the relationship between retention and the population around an institution's location is scanty. Wyman (1997) found this relationship non-significant. The inclusion of this variable is based largely on its theorized value in retention literature.

Crime

As with population above, the relationship between crime and student retention has been largely unexplored in retention research. Wyman (1997) found this relationship non-significant. The inclusion of this variable is based on its theorized value in retention literature.

It is true that significant progress has been made in understanding individual student retention, but these key questions raised above about retention rate have not been fully addressed. If all the selected independent variables play a role in shaping retention rate, then what role does

each play and in what measure? What actions can policy makers and administrators take to increase retention rate? In fact, if we are to continue comparing institutions based on their retention rates, it becomes necessary to answer these important questions.

It is necessary to test for the stability of a regression model. This is done by testing for the equality of the coefficients between two periods. A change in parameters between two periods is an indication of structural change. In Ordinary Least Squares (OLS) estimation, the Chow test statistic for structural change is employed for this test of stability.

Significance of the Study

The retention of students is widely acknowledged as one of the knottiest problems in higher education. As stated earlier, retention rates of colleges and universities have become important in evaluating their effectiveness. In that case, any study that aims at improving the state of knowledge in this area is useful.

Institutions of higher learning pride themselves in achieving high retention and graduation rates. High retention and graduation rates are indicative of strength and portray a college or university as meeting the primary objectives for

which it was established. Such an institution will certainly acquire a lot of goodwill that will be helpful in attracting potential students.

Research findings emerging from retention research are very valuable to university administrators in their efforts to better direct their scarce retention resources.

The model produced in this study will identify some uncontrollable external factors that contribute to retention rate values, thereby revealing the degree to which colleges and universities are powerless in terms of their ability to determine their retention rates.

Limitations of the Study

The following are the limitations of this study:

1. The sample size for this study (15 institutions), is not large enough for a study employing regression analysis. This problem, however, is mitigated by pooling the data for ten years, yielding 150 data points.
2. This study assumed that retention rate is only affected by those demographic and economic characteristics occurring in the year of cohort formation. The possible effects that these variables

might have in the years prior to and following cohort formation are ignored simply because of the complexity of the exercise.

3. The retention rates that were employed in the study will likely understate the "true rates" of retention, because they are "within system" only and do not take account of students who may have transferred to another institution outside the system (and may have graduated from that institution). "True rates" are actual rates of retention that include even students that transferred to and graduated from any college.
4. The conclusions drawn in this study can only be generalizable to North Carolina. But the overall ideas emanating from the study can be replicated nationwide.

Definition of Terms

The following are the definitions of the significant terms used in this study:

Attrition: The term refers to any voluntary or non-voluntary departure of students prior to completion of their intended degree.

- Autocorrelation:** The correlation between two values of the same variable at times X_i and X_{i+k} .
- Cohort:** "Cohort" in this study refers to a group of students who matriculated at an institution in the same semester and year. Only students that have no previous college experience will be included in the freshmen cohort.
- Degrees of freedom:** A measure of the amount of information from the sample data that has been used up. Every time a statistic is calculated from a sample, one degree of freedom is used up. The term is used by statisticians to describe the number of values in the final calculation of a statistic that are free to vary.
- Departure rate:** The total number of students from a specific cohort who leave the institution before the regular period it takes to complete undergraduate study divided by

the total number of students admitted to that cohort.

Dummy variables: A way of adding the values of a nominal or ordinal variable to a regression equation. Each value of the categorical independent except one is entered as a dichotomy (e.g., East=1 if unit is in the East, otherwise 0; West=1 if unit is in the West, otherwise 0; etc.). One class must be left out to prevent perfect multicollinearity in the model.

Graduation rate: The total number of students from a specific cohort who earn a baccalaureate degree in four years divided by the total number of students admitted to that cohort.

Heteroscedasticity: A condition when the variance of residual error is not constant for all values of the independent variable(s).

Homoscedasticity: A condition when the variance of residual error is constant for all values of the independent variable(s). The residuals are supposed to be dispersed randomly throughout the range of the estimated dependent variable.

Multicollinearity: This refers to a condition when explanatory variables are so highly correlated that none of them has much unique explanatory power. The addition of more variables to a multiple regression model provides only a small boost in R^2 because of the presence of certain important predictors already in the model.

Persistence: Enrollment or graduation of a student at a specific interval of time after his or her first semester of enrollment. Students that are enrolled at census date in a given semester and those that graduated in previous semesters will be counted as persisters.

Retention: This refers to the continued enrollment of a student in the institution. It does not necessarily imply that degree completion will be achieved but that academic progress is being made.

Retention rate: The percent of entering students graduating or persisting in their studies at an institution.

Withdrawal rate: This refers to the percentage of entering students who leave the institution before attaining an undergraduate degree.

SUMMARY

This study aims to provide a better predictive model of retention rates for colleges and universities, especially those in the University of North Carolina System. This purpose was accomplished by developing a predictive model utilizing institutional-level variables that are within and those that are beyond the control of the various institutions in the system. The study also aimed at shedding more light on the

direct impact of those external forces that are uncontrollable to the institutions.

The next chapter reviews the literature that is relevant to retention studies in American higher education, by providing an overview of the works of some influential theorists. It also focuses briefly on the relationship between retention and some of the independent variables employed in the development of the model in this study. This is followed by the methodology employed in this study, which is a pooled cross-sectional time series technique.

CHAPTER TWO

REVIEW OF THE LITERATURE

Introduction

This chapter reviews the literature that is relevant to retention studies in American higher education. The review provides an overview of the works of some influential theorists that have paved the way for the model developed in this study. It also focuses briefly on the relationship between retention and some of the independent variables employed in the development of the model in this study. There is also a review of the effects of low retention on both the student and the institution in general, as well as the need for increased retention.

Relevant Research on Retention

Research related to retention in higher education goes back over seven decades (Braxton, 2000). A good amount of such research was done before 1970 by scholars such as Astin (1964), Heilburn (1965), Vaughan (1968), and Chickering and Hannah (1969). These studies focused on identifying and understanding the psychological variables that differentiated

between those who leave school and the ones that stay (Heilbrun, 1965).

Much of the existing body of research on student retention rates is based on the influential works of Spady, Tinto, and Pascarella. The works of these theorists have helped improve our understanding of the interaction between student and institutional culture. These models, however, have tended to be linear in nature and focused mainly on individual student variables. More recently, however, a number of models such as Wyman (1997), Berger (1998), and Tinto (1998), have given consideration to the role of institutional factors on retention rates.

Obviously, research on retention has a long history. But in spite of all these years of research and the series of interventions that had taken place to improve student persistence, many institutions are still struggling to improve on their low retention rates. The problem of student withdrawal from college became more intractable in the last few decades. As earlier stated, this problem was exacerbated initially by declines in student enrollment and, more recently, by the external demand for institutional accountability and effectiveness. Given the significance of retaining students in school, the reasons for and the process of student withdrawal from higher education have attracted the

special interest of administrators, institutional researchers, as well as scholars in higher education. The new culture of institutional accountability and effectiveness has pushed researchers to do more to improve our understanding of why some students persist and others do not.

A study conducted by the American Association of State Colleges and Universities (1997) showed an overall average six-year graduation rate that ranged from 40.6 percent in 1993 to 42.7 percent in 1996 for large institutions (enrollments over 12,000) and 41.5 percent in 1993 to 44.5 percent in 1996 for medium sized institutions (enrollments between 5,000 and 11,999). The study, however, indicated that though students may be taking longer to graduate, graduation rates in America's colleges and universities may be declining. This decline can be seen by comparing data from two distinct studies. A study of 344 colleges and universities by the Consortium for Student Retention Data Exchange (CSRDE) (2001) shows that the average six-year graduation rate for students who entered college in 1994 was 54.1 percent and the ACT (2001) report of five-year graduation rates shows the overall percentage of students earning a bachelor's degree within five years was 51.2 percent, while the rate for public institutions was 49.1 percent (ACT Newsroom 2001). The average six-year

graduation rate for the same period was basically similar, though not reported.

In terms of second year retention, the consortium provides some retention information from its member universities that could be compared. The 2001 national second year retention rate (students returning to college for a second year) for 1999 first-time freshmen cohorts was 79.8 percent for all member institutions, 86.8 percent for highly selective member institutions (SAT above 1,100), and 68.7 percent for less selective institutions (SAT below 990) (CSRDE, 2001). The ACT also provided second year retention statistics, which showed that in 2001, the second year retention rate for first-year students at four-year institutions was 74.2 percent, and 72.1 percent for public institutions (ACT Newsroom, 2001).

Students who left an educational institution did so for a number of reasons. There were those who left voluntarily for personal reasons ranging from insufficient financial aid, low aspirations, ineffective study skills, low institutional commitment, and poor school support services. According to Braxton and Lien (2000), when a student exhibits greater commitment and integration into the academic life of the institution, the greater is the likelihood that he/she will persist.

Braxton (2000, cited in Reason, 2003) stated that scholarly research on student persistence in college slowed down considerably in the mid-1990s with the total embrace and acceptance of Tinto's model. According to Keller (2001), the lull and the changing demographics of college students, may conspire to mitigate the effects of several student retention variables.

Overview of Early Retention Models

There are several theoretical models put forth by scholars to enhance the understanding of student retention. The positions and propositions of four extant models of retention in the literature are explored.

The seminal works of Spady (1970), Tinto (1975), and Pascarella (1980), which generally serve as foundational knowledge on issues related to retention in higher education, are discussed. There is also a brief discussion of a more recent study of retention by Wyman (1997). This is included to show how far scholars have come in studying this seemingly elusive subject of student retention.

Spady's Model

The first significant research on the issue of student retention was put forth by Spady (1970). It provided the first

theoretical model of the dropout process in higher education. Spady's (1970) model of the undergraduate dropout process, provided in Figure 2.1, was based on Durkheim's (1897/1966) theory of suicide. In this model, Spady compared Durkheim's concept of anomic suicide to the concept of involuntary student departure.

Durkheim concluded that "suicide varies inversely with the degree of integration of the social groups of which the individual forms a part" (Durkheim, 1966, p.209). Spady likened suicide to dropout from college. He stated that "although dropping out is clearly a less drastic form of rejecting social life than suicide, we assume that the social conditions that affect the former parallel those that produce the latter: a lack of consistent, intimate interaction with others, holding values and orientations that are dissimilar from those of the general social collectivity, and lacking a sense of compatibility with the immediate social system" (p. 78).

Spady's work emphasized the importance of these interactions. He conceived adjustment to college as a longitudinal process of interaction between the student and the academic and social systems of the institution. The degree to which students become integrated into these systems, together with background characteristics that they bring with

them, influences the students' decision to withdraw or persist.

Durkheim (1897/1966) hypothesized that individuals were more prone to commit suicide if they were not properly integrated into the fabric of society. This notion of suicide was brought to the college environment when Spady (1970), conceptually equated the dropout decision to suicide. He then argued that dropping out increased when the person's normative congruence was low. Normative congruence was defined as a situation when moral consciousness was lacking and friendship support was poor. Spady drew an analogy between poor grade performances and restricted intellectual development and Durkheim's lack of occupational success which Durkheim considered necessary to integration into society.

Spady (1970) proposed an explanatory sociological model for persistence and attrition at the tertiary institution (Figure 2.1), which comprises five independent variables, namely: normative congruence, friendship support, grade performance, intellectual development, and social integration. Two intervening variables (satisfaction and institutional commitment) were eventually added to the model. And then, two more variables (academic potential and family background), which linked all other variables together, were later included in a way that combined these variables sequentially and

causally. Institutional commitment was proposed to feedback into normative congruence.

Spady's model was not without weaknesses. The weaknesses came up in his 1971 empirical study using his model. His multiple regression analysis produced additional interactions among the variables in his original model; however, the measures used for his proposed dimensions were never factor analyzed. Spady (1971) accepted his questionnaire items as representative of the theorized retention constructs without testing them first. In other words, he conducted his study with the assumption, but not the verification, that his constructs were valid.

Spady's model has also been criticized for overemphasizing the role of the sociological process at the expense of the contribution of person-environment interactions (Tinto, 1975). Spady suggests that entrance into institutions of higher education, which of course, constitutes a new society for the student, necessitates to some extent a breaking of ties with his or her past society. This breaking of ties, he argues, catalyzes the creation of confusion and insecurity, which can result in anomic suicide in the form of attrition.

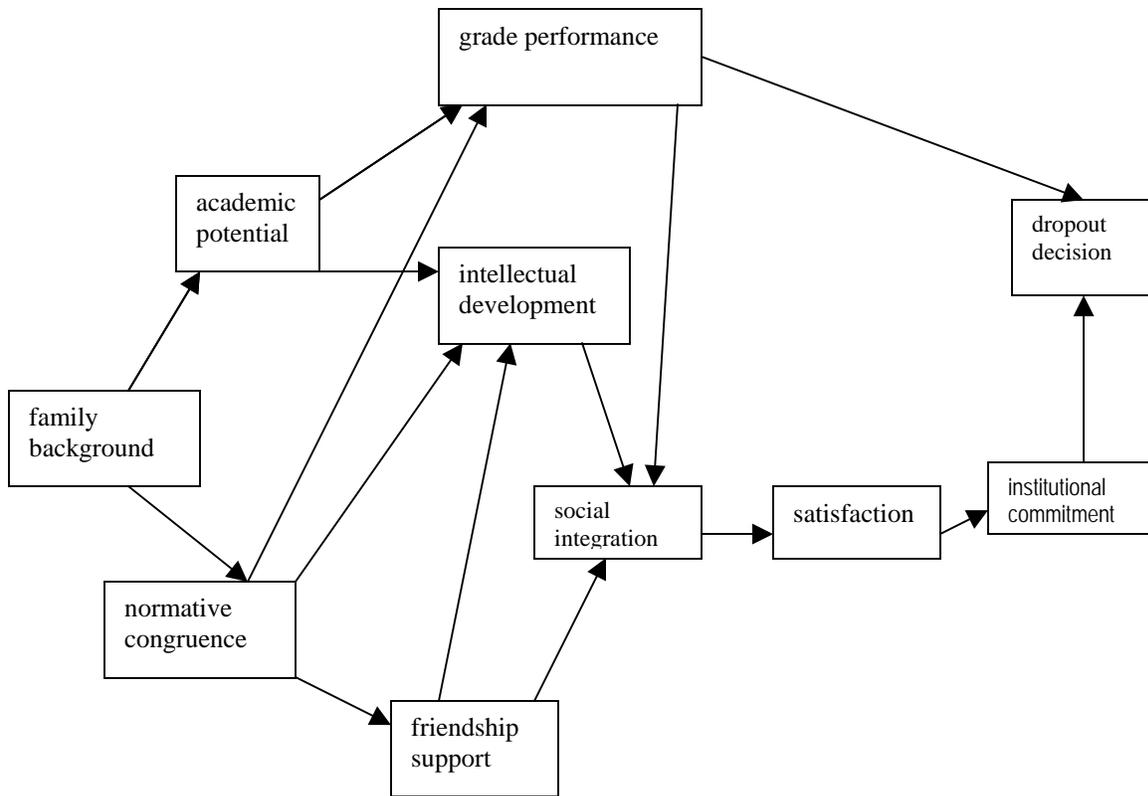


Figure 2.1:
An Replication of Spady's Model of the College Dropout Process (1970)

Tinto's Model

Tinto's (1975) model was an expansion of the earlier work of Spady on the attrition-retention process at the tertiary institutional levels. Tinto applied the social exchange theory to Durkheim's (1966) theory of suicide.

The exchange theory posits that human beings are likely to avoid costly behavior but tend to seek rewarding situations.

Tinto regarded persistence largely as an outcome of the

student's academic and social experiences after enrollment. According to the model, the personal characteristics and histories which students bring to the university influence how they interact with, and subsequently become integrated into, the academic and social systems of the institution.

According to Tinto, students apply the exchange theory in determining their academic and social integration, which are interpreted as goals and levels of institutional commitment. If the perceived benefits of college are higher than the costs, the student remains in school, but if other activities are perceived as having higher rewards and less cost, the student will decide to drop out (Andres and Carpenter, 1997).

Tinto accepted Spady's perception of society as analogous to the college environment. He likened integration into the society to being assimilated into college. He reasoned that for a student to be successful in college, some level of both academic and social integration is required. He defines academic integration essentially in terms of scholarly achievement but includes also the student's involvement with the institution's intellectual activities and services. The indicators of social integration reflect the student's participation with peers in the extra-curricular life of the campus. Tinto built upon the social integration variable employed in Spady's model by his inclusion of frequency and

quality of student-faculty contact. The rationale is that the level of commitment to both the institution and to graduation is determined by the extent to which students develop a sense of socialization and achievement.

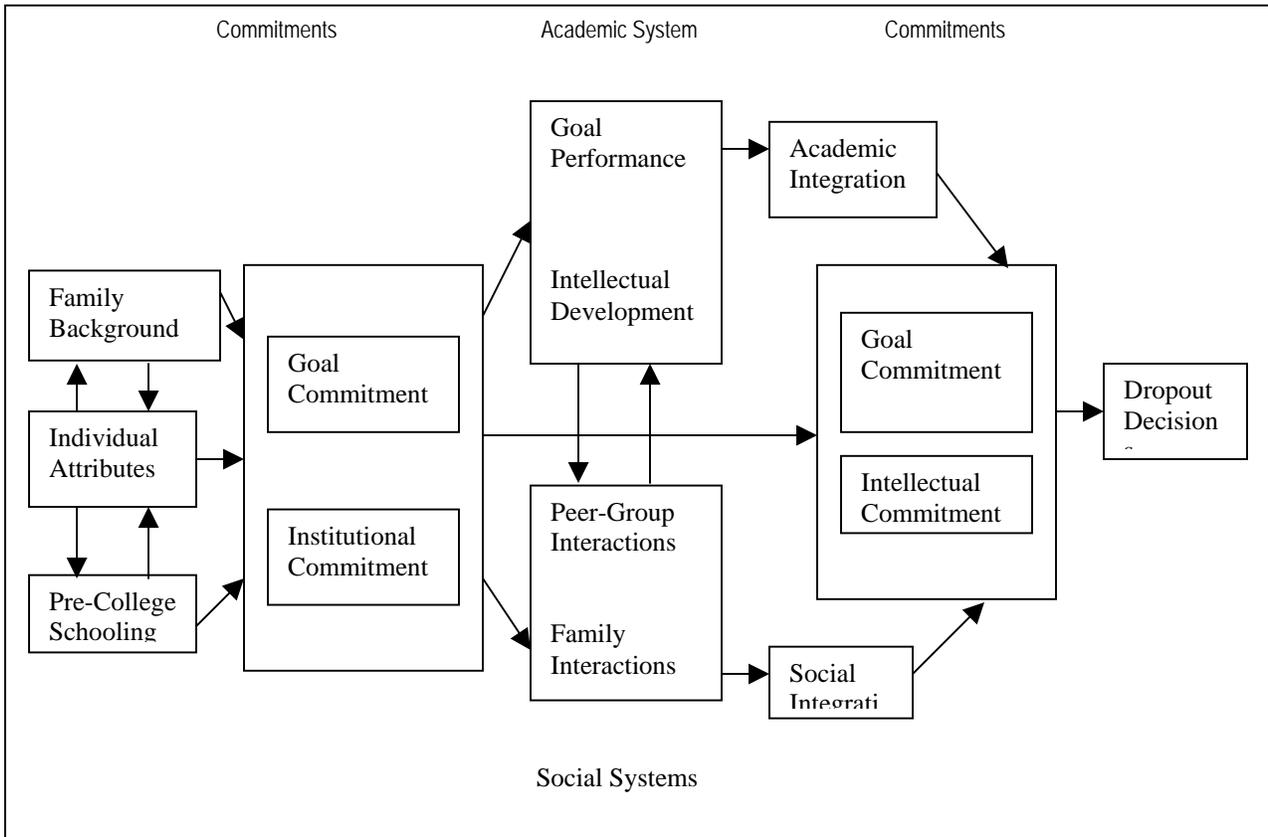


Figure 2.2: A Replication of Tinto's Model of College Student Attrition (1975)

Tinto's model presented in Figure 2.2 comprises five sets of variables, namely: individual and background characteristics, initial goal and institutional commitments,

academic and social integration, subsequent commitments, and a dropping-out decision. The model posits that people come to college possessing a variety of personal and family characteristics as well as some pre-college academic experiences. It further states that those pre-college experiences are the basis for the initial commitment to earning a degree and the continued commitment to earning it in the same institution. The model further assumes that these commitments in turn interact with the academic and social environment of the institution to produce differing levels of academic and social integration for an individual student.

Pascarella's Model

Pascarella's (1980) model was an adaptation of the earlier works of Spady (1970), Astin (1970), and Feldman (1970) which he used to back up the inclusion of different sets of variables in his model. His model emphasized the interaction between student and faculty as being important in students' educational outcomes as well as persistence in college. Pascarella posits that students who are better integrated into the social and academic life of the institution are likely to experience more academic development during their college studies. He expands this by including personal growth and social functioning. Pascarella's model

argued that cognitive outcomes are largely influenced, directly and indirectly, by five sets of variables, namely: organizational characteristics, institutional environment, personal characteristics, campus socialization agents, and the quality of student effort as presented in Figure 2.3.

According to Pascarella (1980), the background characteristics that a student brings to college interacted with institutional factors. He proposed that the interactions that take place between students' pre-college variables and the institution itself were influential in the students' persistence or withdrawal decision.

The dependent variable, persistence or withdrawal decision, is seen to be influenced by student pre-enrollment characteristics, interactions with faculty and peers, and quality of effort. The model assumes interactions within and without the classroom as well as the quality of the instruction received. It is noteworthy that both institutional environment and organizational characteristics show an indirect effect on cognitive outcomes. Their effect is believed to be indirect and mediated by student-faculty interactions and the quality of student effort. The model is seen to have placed heavy emphasis on student-faculty interactions. According to Pascarella (1980), "in order to understand the unique influence of student-faculty non-

classroom contact on educational outcomes and institutional persistence, it is necessary to take into account, not only background characteristics which students bring to college, but also actual experiences of college in other areas, as well as salient institutional factors” (p.568).

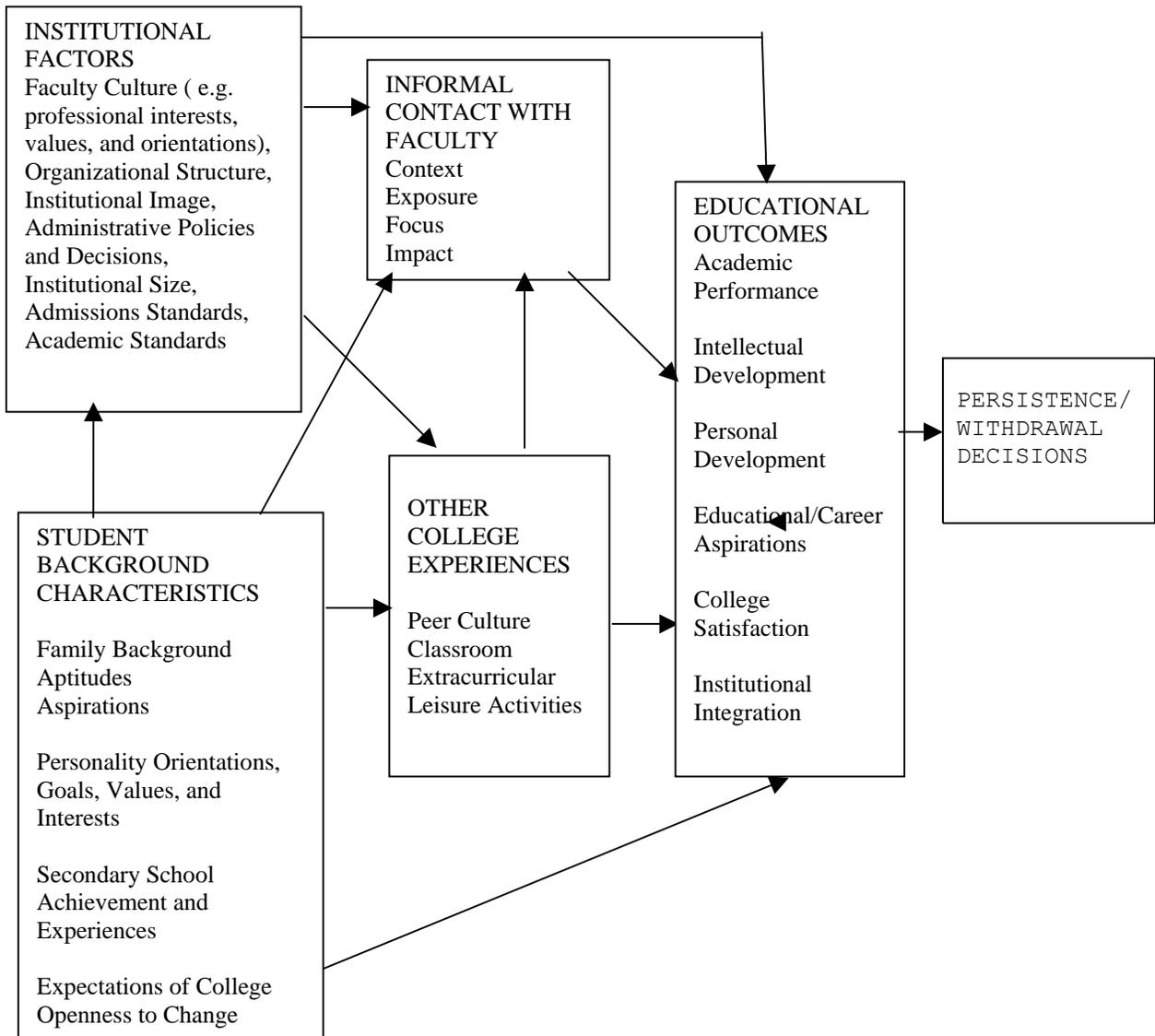


Figure 2.3:
An Replication Pascarella’s Model of College Student Attrition
(1980)

This model has several significant contributions to retention theory. It is an expansion of the student-faculty interaction concept adapted from both Spady (1970) and Tinto (1975). It is also a longitudinal model that employed many independent variables. The use of many variables makes it a comprehensive model that is robust.

Pascarella's model, however, has been criticized as having a narrow focus by looking at retention at a specific institution. This shortcoming was, however, overcome in this study because of the approach of looking at a university system and not an individual institution.

Wyman's Model

One of the most comprehensive studies on retention rate in community colleges was undertaken by Wyman (1997). It is among the most significant research efforts to examine retention rate using institutional level variables as opposed to student level variables. A representation of Wyman's model is provided in Figure 1.0.

In this model, however, only two variables were significant. The first significant variable, regional employment per capita, was critical in Wyman (1997) study. This finding suggests that student retention is strongly related to that portion of the county population that is

employed. The other variable that was significant in that study is the ratio of academic support spending per student to regional income per job. This variable is also vital in that study, being one of only two variables that were significant in a model that explained 66 percent of retention rate variance in a model. Conducted in the South Carolina Technical College System, the study found that retention rate could be partially controlled through college and government fiscal policy, even though forces external to the institution largely predetermine retention rate variance.

According to Wyman (1997), in spite of the considerable advance made in understanding individual retention rate, some important questions about retention rate have not been answered conclusively. Some of these questions include: Does retention rate measure institutional effectiveness? If so, does a high retention rate portray an institution particularly effective at retaining students by shielding them from the grasp of attrition, or does it simply reflect the selective admission policy of that institution? He maintained that until some of these questions are fully addressed, it would be inappropriate to continue to compare the effectiveness of institutions on the basis of their retention rates.

Relationship Between Retention and Predictor Variables

This section looks at some of the predictor variables employed in this study and explores their individual theoretical relationships with retention.

Retention and Enrollment

Numerous surveys and studies over the course of 15 years show the impact of enrollment management systems on colleges and universities. One important factor that needs close examination is the relationship of student retention to enrollment. Changes in the age distribution of the population impact the enrollment projections of colleges and universities, but low retention also contributes to reduced enrollment levels.

The traditional view of undergraduate college students as 18-22-year-old White, full-time students attending residential colleges conforms only to a small part of the contemporary college population (Keller, 2001). Most retention research in higher education was based on the traditional view of students, rather than the reality of the student diversity on our campuses. Furthermore, the demographic characteristics of undergraduate students continue to change (National Center for Education Statistics (2001)).

According to Wetzel, et al. (1999), retaining current students into a second and a third year of study and perhaps even to graduation, may be desired to maintain enrollment levels in general.

The increasing pressure for increased accountability by colleges and universities, coupled with the changing demographics of the student body, has provided the catalyst for the development of more detailed and sophisticated retention models. Regardless of the number of students an institution enrolls, its inability to retain students will have serious negative consequences for institutional stability.

In a study conducted by the American Council on Education, about 62 percent of college administrators believed that enrollment challenges were one of the most important factors facing their schools. According to the administrators, this concern was attributed to changes in demography, technology, and federal and state financing policies for aid and higher education funding (Dennis, 1998). This appears to be a correct observation because institutional enrollments are stabilized by the retention of currently enrolled students as well as the steady inflow of new students. The relationship of retention to enrollment was also cited by Wyman (1997), who stated that retention is a key issue in enrollment planning.

Retention and Employment

Labor market outcomes continue to be strongly affected by educational attainment. In November 2004, the greatest monthly increase in employment went to 4-year college graduates. The number of employed persons age 25 and older with a four-year or more advanced college degree increased by 74 percent. The 40.0 million four-year college graduates age 25 and older employed in November 2004 represent a 15.4 percent growth in employment compared to five years earlier. Over the same period, employment of persons in that age-group who had completed some college but less than a four-year degree increased by about 5.7 percent. However, the 35.8 million employed persons in the same age-group whose highest educational attainment was a high school diploma declined by 1.8 percent. This implies that education is key to lower unemployment (Employment Policy Foundation, 2004).

With the availability of transportation and ease of relocation within the United States, workers with higher educational attainment can move without difficulty to areas with jobs that typically require a high level of training and pay higher wages. Therefore the educational attainment of an area's population is indicative of the quality, and at times quantity, of jobs available in an area.

Low high school graduation rates may suggest the absence of jobs requiring a high school diploma that pay substantially higher than those that do not require a diploma, thus lowering the incentive to graduate. The overall low rate of graduation may be attributed to the lack of overwhelmingly better employment opportunities resulting from a college education (Indiana Economic Development Council, 2004).

Considering the above information, it would seem plausible that the availability of employment opportunities in and around the location of an educational institution may be a factor in attracting and retaining prospective students. The vagaries of employment in recent times could impact the student's educational plans, in light of the increasing cost of going to college. This seems likely because an average student would be expected to consider the possibility of getting a good job on graduation from college.

According to Wyman (1997), there is a very strong positive relationship between the institution's retention rate and regional employment per capita. Therefore, regional employment per capita at the time of cohort formation is an important determinant of first-year retention rate.

Retention and Funding

The issue of the retention of students and its relationship to the level of funding for institutions of higher education has received increased attention in recent years (McGrath and Braunstein, 1997; Bogue, 1998; Burke and Seban, 1998; and Lau, 2003). According to Roach (1999), higher education experts agree that interest in student retention has reached an all-time high. Yet there is a growing frustration among retention experts that many colleges and universities lack either the will or the funds to develop truly effective programs.

Some state legislatures have begun to measure and translate success in school performance using retention and graduation rates. These states have employed these rates in their institutional funding formulas. Perception of poor quality or ineffectiveness can have significant negative consequences on funding and on an institution's ability to attract and retain students. To maintain their effectiveness as well as funding levels, these institutions have embraced several quality control movements, including total quality management (TQM), peer evaluations, and the performance indicator movement (Bogue, 1998).

In the past twenty years, state lawmakers have become increasingly interested in linking funding for higher

education with performance on key quality indicators. They have oftentimes used performance reports to determine funding levels of educational institutions, as is done in the state of South Carolina (Burke and Seban, 1998).

The implication of all this is that higher educational institutions that fail to improve on their student retention rates will suffer financially. Each state funds public education by establishing a dollar amount per student for school operating expenses. State support for higher education is directly related to the general condition of a state's economy, state tax capacity, and availability of revenues, among many other factors.

Retention and Financial Aid

Over the past decade, institutional researchers have disagreed about the impact of student financial aid on attendance and persistence in institutions of higher education. A recent national study (St. John, 2000) illustrated the evolving nature of research on financial aid and its impact on persistence. St. John, Paulsen, and Starkey (1996) indicated that aid and various financial variables contributed more to the variance in student persistence than did those of social and academic integration processes.

At the institutional level, however, there is still some controversy over whether student aid has an effect on student persistence. Studies that have investigated the role of financial aid on undergraduate student persistence have been conflicting, at best. Research by Jones and Moss (1994) indicates that recipients and non-recipients of federal financial aid persist at comparable rates. Perna (1998) found that the effect of financial aid on persistence depended on the type and package of financial aid received. He also indicated that both receiving work-study and receiving an aid package that contains only grants had positive direct effects on persistence. He however, concluded that the direct impact of receiving aid toward bachelor's completion is marginal.

Students have repeatedly reported that finances are a primary reason for dropping out of college, but there is a consistent pattern in the literature stating that financial aid is not a significant correlate of attrition. This may not be the case in all types of institutions. We may find different results when we examine the impact of financial aid on the retention rates of institutions that engage large numbers of low- income students.

In a study that examined the impact of financial aid resources on minority retention rate variability across institutions, Georges (2000) indicated that meeting the

financial need of minority students may be a key factor in addressing the problem of attrition observed in Historically Black institutions and other minority institutions. According to the study, the correlation between minority retention rate and the average scholarship and fellowship is positive and statistically significant; that is, the minority retention rate tends to be higher at those institutions with high average financial aid awards.

The conflicting findings about the impact of financial aid on student retention may be due to the fact that family income or other correlates of persistence are not controlled for in these studies (St. John, 2000). According to Alkerheim, Berger, Hooker, and Wise (1998) some of these conflicting results may be explained by establishing standards that determine what constitutes adequate or inadequate aid for an average student. It may be necessary to further explore the significance of inadequate levels of financial aid in studies of persistence. Mortenson's (1999) research drew attention to the impact of levels of unmet need on persistence.

Effects of Low Retention

Included in majority of the early retention models such as (Spady, 1970; Tinto, 1975; and Pascarella, 1980), it seems obvious that institutional commitment is an important

condition for student retention. In fact, institutions that are committed to the goal of increasing student retention seem to find a way to achieve it. Attrition has serious consequences for both students and their institutions. It exacts financial costs on both student and institution.

According to Tinto (1993), institutions depend on tuition revenues to support academic programs, manage physical plants, and deliver other student services for continued improvement and survival. Institutional commitment must go beyond mere rhetoric; it must involve the willingness to invest resources as well as provide the incentives necessary to improve the school's ability to enhance student retention.

Effects of Low Retention on the Institution

Student performance has become an increasingly important topic in higher education. This is largely due to its relationship to retention and graduation rates, and also the increasing emphasis placed upon performance as a basis for funding. These performance-based funding formulas usually tie appropriations to measures of student success such as retention and graduation rates (McKeown, 1996). Pressure has been mounted on public colleges and universities to demonstrate institutional effectiveness. Major regional accrediting organizations, including the Southern Association

of Colleges and Schools (SACS) (1995) require colleges and universities in their jurisdictions to measure the effectiveness of the education they provide. This became even more important in light of dwindling state funding. Sanders and Burton (1996) reported that institutions of higher education are under increasing pressure from their constituents to demonstrate student success. According to Astin (1997), retention and graduation rates are the most common measures of student success.

When students fail to persist in school, enrollment falls and consequently, tuition revenues dwindle. Educational institutions expend a lot of resources to stem the tide of student attrition because of its obvious financial implications to them. Public educational institutions get much of their total revenues from tuition revenues, and this is more so with private institutions.

As mentioned earlier, improved retention rates are important as most states now use some funding formula in the determination of appropriations for public institutions of higher education, and most of these formulas factor in student enrollment in determining budgets (McKeown, 1996). This implies that state appropriations decline with declining student enrollments.

In a study based on Mountain Empire Community College, Sydow and Sandel (1998) reported that the total net revenue gained by retaining one first-year student that persists to graduation is about \$4,025. The report further indicated that the institution would save about \$94,588 if it succeeds in reducing its first-to-second year dropout rate by about 10 percent.

The Federal Student Right-to-Know and Campus Security Act of 1991 requires educational institutions to disclose data on the quality of their educational programs. This law requires institutions to release information about the graduation rates of students that they admit. This disclosure enables potential students to make informed decisions about colleges they might attend (McLaughlin, Brosovsky, and McLaughlin, 1998). It is obvious that such disclosure can affect public perception of an institution, which can in turn, affect the institution's enrollment.

Effects of Low Retention on the Students

Just as for institutions, attrition can have serious implications for students. A student that drops out of school may suffer tremendous emotional damage, not to mention the future financial cost due to low educational attainment. According to Tinto (1993), males age twenty-five and above who

have college degrees earned 23 percent more than their counterparts who have between one and three years of college education. The US Census Bureau (2002) reported that adults from ages 25 to 64 who worked at any time during 1997 to 1999 had average earnings ranging from \$18,900 for high school dropouts to \$25,900 for high school graduates, \$45,400 for college graduates, and \$99,300 for workers with professional degrees. As shown in Figure 2.4, with the exception of workers with professional degrees who have the highest average earnings, each successively higher education level is associated with an increase in earnings.

The Need for Increased Retention

As already stated, student retention has become a priority concern for educational institutions across the United States. Although college administrators have always done something to enhance student success, student retention is now a matter of economic survival. In spite of the recent increase in the proportion of youth going to college, rates of completion have not changed (NCES, 2004). This means that colleges still scramble to increase their enrollments in order to replace dropouts. Dropouts only mean lost students and lost revenues.

According to the US Census Bureau (1998), about 82.8 percent of all adults ages 25 and over had completed high school and 24.4 percent had completed a Bachelor's degree or more. Although nearly 75 percent of high school graduates go on to some form of postsecondary education, over 50 percent who enter college fail to complete a degree, and 33 percent never even make it to the sophomore year (Early College High School Initiative, 2003). Other very recent studies indicate similar trends of low graduation and retention rates in America's colleges and universities.



Figure 2.4: Work Experience and Average Annual Earnings of Workers 25 to 64 Years Old by Educational Attainment: 1997-1999

Source: U.S. Census Bureau, 1998-2000

According to the National Center for Education Statistics (NCES) (2004), among students who started at a four-year college or university in 1995-96, 53 percent had completed a bachelor's degree at the end of 5 years, 24 percent were no longer enrolled and had not earned a degree and 13 percent were still enrolled, but had not yet earned a degree.

In light of the practical significance of low retention rate to institutions of higher learning, the need to find ways to increase it cannot be overemphasized. Modeling retention rate in an effort to identify the variables that influence retention rate seems to be a step in the right direction.

SUMMARY

Student retention is a very complex subject. There are many variables that can have direct and indirect influence on the persistence or otherwise of a student. Thus, colleges and universities employ different models in an effort to have a better understanding of the retention phenomenon and possibly reduce it. An effective retention model would involve a system-wide approach using institution-level variables. In addition, it has been seen from the descriptions of earlier models that student-level variables cannot adequately provide information for a comprehensive study of student retention. Additionally, the earlier models were also largely

atheoretical and descriptive in nature. More recent studies, however, have begun to employ statistical approaches in studying student retention.

The present study is the first attempt at statistically predicting retention rate in the UNC System using institution-level variables. Wyman (1997) study of student retention offers a particularly appropriate model for this study in that it distinguishes between two sets of predictors: internal (controllable) and external (uncontrollable) variables. By so doing, this study may demonstrate the degree to which colleges and universities are powerless in terms of their ability to really determine their retention rates.

CHAPTER THREE

METHODOLOGY

Sample

This study consists of the fifteen universities in the University of North Carolina system presented in Table 3.1. It is based on institution-level variables which indicate that institutions constitute the sample and not students. The study excluded the North Carolina School of the Arts, because it is not a four-year institution, bringing the number to fifteen. This sample is ideal for this study for a number of reasons. First, the institutions are all the members that make up the entire University of North Carolina system and each maintains the necessary data to compute retention rate. Second, each institution's location is defined along county boundaries, and this makes it possible to easily marry county-based demographic and economic data to the institution. Location refers to the immediate county boundary of the institution. Lastly, though there is some overlapping among some of the 15 county locations, the majority of them are non-overlapping and they constitute the entire state of North Carolina as displayed in Figure 3.1.

By their share size, the biggest schools in the system (NCSU and UNC-CH) do have some influence beyond their immediate environment. Their impact is explored later in this

chapter. In spite of this, however, they are also influenced by the county-based demographic and economic factors, by virtue of their physical location in those counties.

Virtually all the observations in this study are non-overlapping, except North Carolina A & T University and University of North Carolina-Greensboro, which are both in Guilford County. The fact that these observations are non-overlapping is important in this study. It means that they are mutually exclusive. The independence among them strengthens the internal validity of the study. Similarly, the fact that the observations constitute the entire membership of the statewide campus system ensures substantial external validity. This also means that the results have wider generalizability within the state.

The observations in this study are identified both by institution and by year. The limited number of institutions (15) available to conduct this study and do the subsequent analysis would be a serious constraint on the regression modeling. To overcome this problem, the measures for each of the 15 institutions were pooled for ten (10) separate years (1991-2000), to provide greater number of data points of 150 observations (15 institutions times 10 years) given in Table 3.2. In practical terms, this means that each institution was

counted as a separate observation for each of the years which data were pooled.

This provided three simultaneous benefits: First, a ten-year pooled analysis provides a very robust analysis of the relationship between each of the measures and retention rate, as it combines several years of data. Second, the increased number of observations undoubtedly becomes adequate for the application of more complex statistical modeling methods. Third, this pooled cross-sectional analysis incorporates the effects of time into the estimated coefficients of the independent variables.

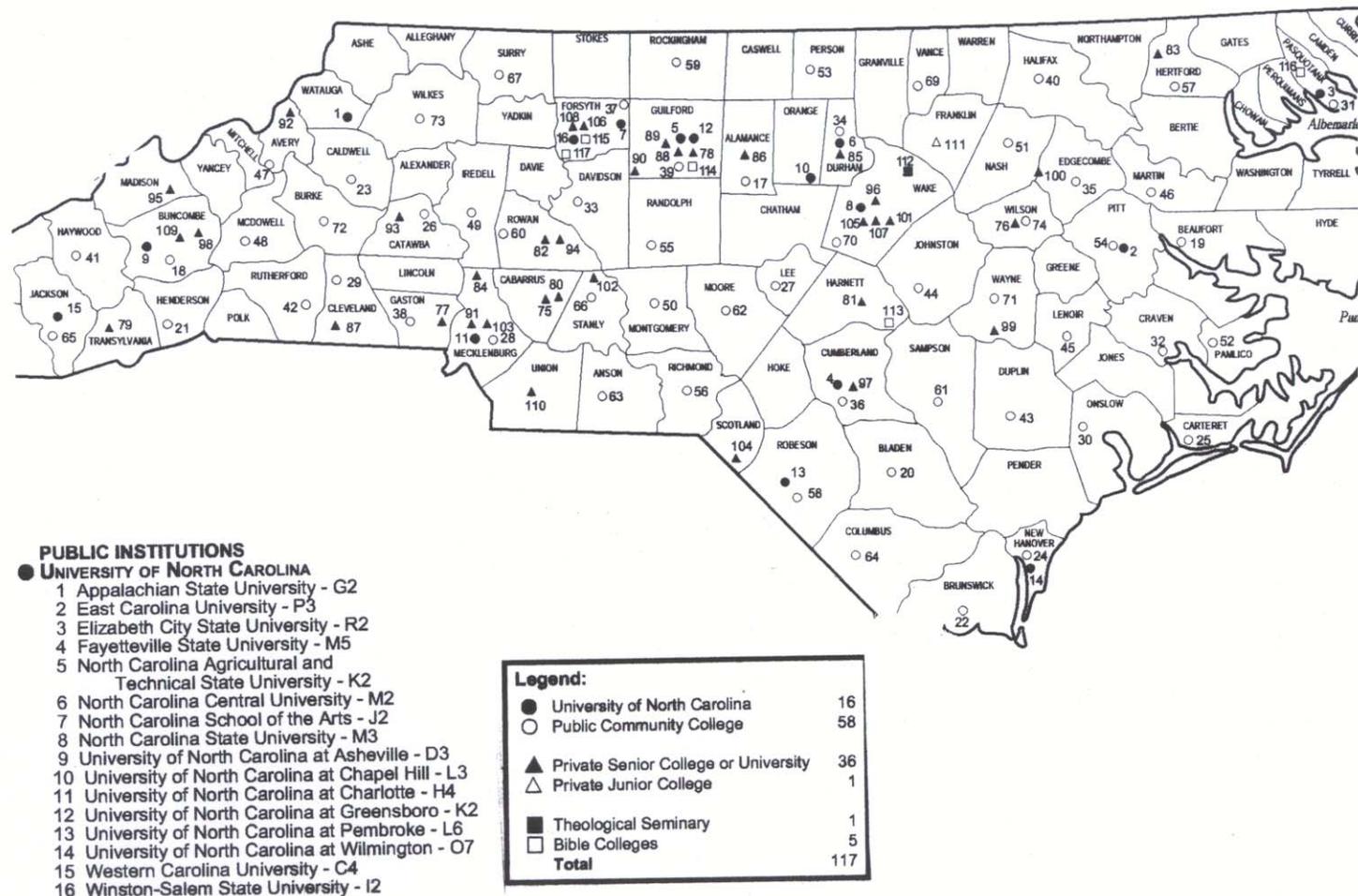


Figure 3.1: Map of 4-Year Institutions and their County Locations
 Source: Statistical Abstracts of Higher Education in North Carolina (2001-2002)

TABLE 3.1: UNC Institutions and their County Locations

1.	Appalachian State University	Watauga
2.	East Carolina University	Pitt
3.	Elizabeth City State University	Pasquotank
4.	Fayetteville State University	Cumberland
5.	North Carolina A & T University	Guilford
6.	North Carolina Central University	Durham
7.	North Carolina State University	Wake
8.	University of North Carolina-Asheville	Buncombe
9.	University of North Carolina-Chapel Hill	Orange
10.	University of North Carolina-Charlotte	Mecklenburg
11.	University of North Carolina-Greensboro	Guilford
12.	University of North Carolina-Pembroke	Robeson
13.	University of North Carolina-Wilmington	New Hanover
14.	Western Carolina University	Jackson
15.	Winston-Salem State University	Forsyth

TABLE 3.2: Ten Retention Rates for all Schools in the study (1991-2000)

SCHOOLS	YEARS									
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
ASU	86.4	85.9	85.1	82.1	83.7	82.3	82.9	81.1	84.7	83.2
ECU	77.6	77.4	76.8	77.4	77.1	77.7	78.2	79.0	76.4	78.0
ECSU	77.9	77.2	73.9	66.8	77.4	75.2	76.4	72.9	77.3	81.5
FSU	77.5	74.8	72.5	64.6	69.7	71.2	72.8	74.2	72.6	71.0
NCAT	76.1	75.7	74.9	74.1	74.1	75.0	76.7	75.5	72.3	76.6
NCCU	83.3	79.3	77.4	78.2	72.5	84.1	77.2	77.0	72.3	78.5
NCSU	89.1	87.5	89.0	86.7	86.8	88.1	87.8	88.0	88.9	88.7
UNCA	78.3	78.0	80.2	78.3	77.0	80.4	77.0	77.8	76.8	79.6
UNCCH	94.6	93.2	93.3	92.8	93.7	93.6	94.8	93.9	93.9	95.0
UNCC	80.8	78.7	78.1	80.1	79.3	76.1	78.7	73.3	73.1	77.7
UNCG	78.9	75.9	75.4	76.0	71.4	72.2	73.9	74.0	73.9	74.9
UNCP	74.1	69.5	69.7	63.8	71.7	68.4	71.0	66.6	67.4	68.6
UNCW	77.7	77.0	78.4	79.5	77.8	79.6	78.3	80.0	79.8	81.7
WCU	71.4	72.1	73.8	70.4	69.0	67.0	67.5	69.6	71.5	69.4
WSSU	67.1	75.7	72.6	71.6	74.4	74.8	68.0	72.9	71.8	73.1

Ten retention rates for all schools in the study (University of North Carolina General Administration, 2002).

Overview of Statistical Methods

Educational researchers have employed several statistical methods in the prediction of retention rates. Logistic, probit, and linear regressions are the three most popular of these methods. In most of these studies, the dependent variable is determined by tracking student enrollment from one academic year to another based on registration records. Logistic regression is usually applied when the independent variable is categorical and the dependent variables are dichotomous (Huesman et. al., 1994). Liu and Liu (1999) employed the probit method in the determination of the maximum likelihood estimates for their retention model. Ordinary least squares (OLS) regression has also been employed in formulating a model of persistence (Eaton & Bean, 1995).

Various other statistical procedures, based on assumptions of linearity, such as path analysis and step-wise regression have also been used (Cabrera, Nora and Castaneda, 1993; Berger and Braxton, 1998; and Elkins, Braxton, and James, 2000.) French et al. used path analysis to model success and persistence of engineering students. The model, based on theoretical and empirical evidence, included both cognitive and noncognitive factors. Cognitive factors included High School Rank, Scholastic Aptitude Scores, and University Grade Point Average.

Noncognitive factors included motivation, as well as faculty and student integration. Outcome variables in the model were grade point average, enrollment at the university, as well as within engineering. The result indicated significant relationships among the factors were found (French, et al., 2003). Some scholars have used discriminant analysis, which also assumes linear relationships, to classify variables into the values of a categorical dependent (Huesman, et. al., 1994; and Ryland, et al., 1994).

Murtaugh, Burns, and Schuster (1999) employed the survival analysis technique, which is prevalent in biostatistical research, for modeling the retention of undergraduate students at Oregon State University for the time period 1991-1996. This model included only demographic and academic variables, and the analysis showed that the estimated probability of attrition was approximately 40%. The test results predicted a statistically significant relationship between retention and the explanatory variables.

Matrix modeling has also been recently employed in estimating retention rates. Luna (1999) investigated retention and withdrawal rates at the University of Louisville from fall 1984 to fall 1995 using a student flow matrix model. A principal justification for using this model is that it is capable of graphically showing the relationships between the

multidimensional variables. Luna identified three variables for this study: remaining in the department; external retention, that is, changing departments but remaining at the same institution; and withdrawal from the institution. The results show that attrition is greater among first-time freshmen compared to total undergraduates. They also indicated that, with respect to retention, the variables had differential impacts on the ten schools observed.

Political economists, and more recently, educational researchers, have tended to investigate the relationship between institutions and socio-economic variables by comparing observations across space and time. Until recently, the space and time domains have rarely been studied simultaneously in comparative research. However, new approaches in quantitative methodology have emphasized the significance of being sensitive to space and time effects. Pooled time series cross-section analysis (TSCS) is probably the most important procedure to examine time and space dimensions simultaneously (Podesta, 2000).

In light of the significance of the pooled time series cross sectional procedure in this study, it becomes necessary to provide an extended discussion of this procedure. The objective is to present a comprehensive explanation of TSCS for a fuller understanding of its import in this study.

Pooled Cross-section Analysis

Pooled cross-section regression analysis, also called 'panel analysis', is an area of statistical/econometric analysis that has recently experienced rapid expansion, resulting in major advances in theory and estimation. The reason for this is the difficulty in obtaining data for most applied research. Often, we can only get few years worth of data, which is not adequate for doing a time series analysis. In addition, simple cross-section regression cannot incorporate the effects of time into the estimated coefficients. Pooling, therefore, enables us to stack 5 years or less of data into one data set, thus allowing the application of some quite sophisticated statistical techniques to exploit its information content whilst incorporating the effects of time and space (Kemp, et al., 2001).

According to Pindyck and Rubinfeld (1998), there are three basic approaches to analysis of pooled data.

1. Simple regression analysis
2. Fixed-Effects model
3. Random Effects model

The first approach involves doing a simple regression analysis on the pooled data using a dependent variable and various independent variables. The problem with this approach is that

there is no consideration for the effect of time on retention rate or for any variation in retention rate across institutions.

In light of the above shortcoming therefore, the fixed-effects model, a more efficient method of analysis that takes into consideration the possibility that retention rate will vary across time and institutions, is usually adopted. This technique, however, presents a problem of its own, with the inclusion of additional terms known as dummy variables to account for time and institutional variations in retention rates. Large numbers of time dummy variables and institution dummy variables will substantially reduce the number of degrees of freedom.

Again, Pindyck and Rubinfeld (1998) noted that potential degrees of freedom problems could be overcome by omitting dummy variables for time even while including those of cross-section units. The fixed-effects model assumes that it is possible to disentangle the effects of time and units. In other words, the error term in the regression specification is uncorrelated across time and institutions (Johnston and Di Nardo, 1998).

In the event that there is some correlation over time and across institutions, then the random effects model is adopted instead. But because this model allows for correlation over

time and units, one of the assumptions of classical regression analysis would be violated and OLS will no longer provide correct estimates of the standard errors and therefore t-statistics.

It seems to be a good idea at this juncture to explore some of the advantages and possible disadvantages of the pooled cross section analysis technique.

The first advantage concerns the "small N" problem suffered by both time series and cross-sectional analysis. The limited number of spatial units and the limited number of available data over time led data sets of these two techniques to violate basic assumption of standard statistical analysis. Most specifically, the small sample of conventional comparisons shows an imbalance between too many explanatory variables and too few cases. Consequently, within the context of the small sample the total number of the potential explanatory variables exceeds the degree of freedom required to model the relationship between the dependent and independent variables. Fortunately, in using pooled TSCS designs, we can greatly relax this restriction. This is possible because, within the pooled TSCS analysis, the cases are "institution-year" (NT observations) starting from the institution i in year t , then institution i in year $t+1$ through institution z in the last year of the period under

investigation. This allows us to test the impact of a large number of predictors on the level and change in the dependent variable within the framework of a multivariate analysis (Schmidt 1997, p. 156).

Two, pooled models have become popular because they allow inquiry into variables that are difficult to study in simple cross-sectional or time-series analysis. This is because there is a virtual absence of variability across either time or space. In practice, many institutional characteristics tend not to change over time. Therefore, regression analysis of pooled data combining space and time may rely upon higher variability of data with respect to a simple time series or cross-section design research (Hicks 1994, p. 170-71).

Three, pooled TSCS analysis has the ability to capture not only the variation of what emerges through time or space, but also the variation of these two dimensions simultaneously. This is because, instead of testing a cross-section model for all institutions at one point in time or testing a time series model for one institution using time series data, a pooled model is tested for all institutions through time (Pennings, Keman & Kleinnijenhuis 1999, p. 172). Given these advantages, pooled analysis has become central in quantitative studies of comparative political economy, as well as educational research in the past few decades. A number of researchers have employed

pooled models to address classical questions of these and other disciplines.

In spite of these reasons in support of the TSCS, it is pertinent to mention the following shortcomings of this statistical procedure. The pooled TSCS design often violates the standard OLS assumptions about the error process. In fact, OLS regression estimates, commonly used by social scientists to link potential causes and effects, are likely to be biased, inefficient and/or inconsistent when they are applied to pooled data. This is because the errors for regression equations estimated from pooled data using OLS procedure tend to generate five complications (Hicks 1994, p. 171-72).

One, errors tend to be not independent from a period to the next. In other words, they might be serially correlated, such that errors in institution i at time t are correlated with errors in institution i at time $t+1$. This is because observations and traits that characterize them tend to be interdependent across time. For example, temporally successive values of many institutional traits (e.g., population size) tend not to be independent over time.

Two, the errors tend to be correlated across institutions. They might be contemporaneously correlated, such that errors in institution i at time t are correlated with

errors in institution j at time t . Commonsensically, one could not expect errors in the statistical model for, say, UNC-CH to lack some resemblance to that for NCSU or errors for NCCU and NCA&T to be altogether independent. Instead, we would expect disturbances for such institutions to be cross-sectionally correlated.

Three, errors tend to be heteroschedastic, such that they may have differing variances across institutions. In other words, institutions with higher values on variables tend to have less restricted and, hence, higher variances on them. This means that the variance in employment rates will tend to be greater for larger service areas with large heterogeneous labor forces than for small, homogeneous counties. Moreover, errors of a TSCS analysis may show heteroschedasticity because the scale of the dependent variable, such as the rate of retention, may differ between institutions.

Four, errors may contain both temporal and cross-sectional components reflecting cross-sectional effects and temporal effects. Errors tend to conceal unit and period effects. This is because the heteroschedasticity and autocorrelation we observe are a function also of model misspecification. The misspecification that is peculiar with pooled data is the assumption of homogeneity of level of dependent variable across units and time periods. In

particular, if we assume that units and time periods are homogeneous in the level (as OLS estimation requires) and they are not, then least squares estimators will be a compromise, unlikely to be a good predictor of the time periods and the cross-sectional units, and the apparent level of heteroschedasticity and auto-correlation will be substantially inflated (Stimson 1985, 919).

Five, errors might be nonrandom across spatial and/or temporal units because parameters are heterogeneous across subsets of units. In other words, since processes linking dependent and independent variables tend to vary across institutions or/and period, errors tend to reflect some causal heterogeneity across space, time, or both. Therefore, this complication, like the previous one, could be interpreted as a function of misspecification. If we estimate constant-coefficients models, we cannot capture the causal heterogeneity across time and space.

In summary, these errors can and do pose serious problems in model specification, if they are not carefully watched or resolved when encountered. The errors represent issues of precaution in dealing with pooled time series of cross-section analysis. Fortunately, this particular study did not encounter much of these problems, so we are able to proceed without fear of serious model misspecification.

Selection of Variables

This section focuses on the dependent and independent variables in terms of how and why they were selected for this particular study. Figure 3.2 represents the model showing the dependent and independent variables.

The Dependent Variable

The dependent variable for this study is retention rate. Ten retention rates (1991-2000) were computed for each of the 15 institutions in the study. Each retention rate was derived for a cohort of first-time freshmen enrolled full-time in the 1991 fall semester, the next for a 1992 cohort, and all the way to the 2000 cohort defined in a similar way. Retention rate here is defined as the percent of the cohort enrolling in the same institution the following fall semester. This retention rate data was obtained from the database of the University of North Carolina General Administration's Student Retention Tracking System and the Statistical Abstract of Higher Education in North Carolina. Retention rates for all schools are provided in Table 3.2.

First-year retention rate was chosen as the dependent variable for two reasons. One, the largest decrease in retention rate generally occurs during the first year.

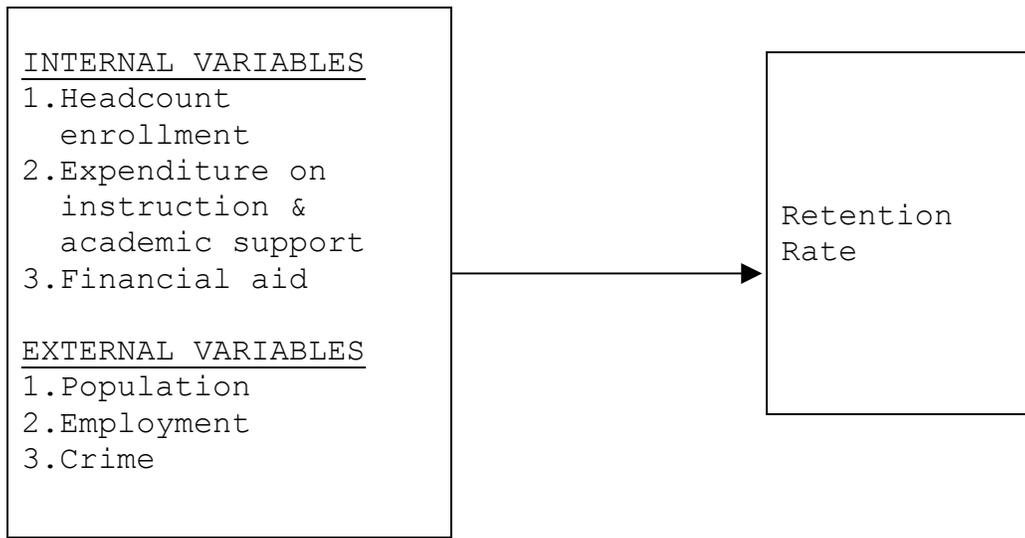


Figure 3.2: Model of Internal and External Variables

This is supported by Snyder & Hoffman (1995) who estimated that over 800,000 entering freshmen in 1993 withdrew from post-secondary educational institutions in the United States prior to their second year. About 20 percent of the 1999 beginning students left postsecondary education before the beginning of their second year of study at the same institution in the year 2000 (CSRDE, 2001). Furthermore, considering the disproportionate number of students who leave college prior to their second year (Levitz et al., 1999), this time period appears to be an appropriate focus for retention studies. Two, it may be more straightforward to isolate and interpret the effects of one-year changes in independent

variables on retention rate. This might not be the case for multi-year retention rates as a one-year change in an independent variable may unduly affect a two-year retention rate, thereby muddling the interpretation of some relationships. In light of the above, therefore, the greatest opportunity for improving retention rate lies, perhaps, in understanding first-year retention.

Independent Variables

The independent variables that were employed include six measures, characterized either as internal or external to the institutions, which were adapted from the Wyman (1997) study of retention rate at regional two-year colleges in South Carolina. The selection process was based on the bivariate relationship between each of the measures and retention rate, provided in Table 3.3. In other words, a correlation was computed for each pair and the ones that had significant coefficients at $\alpha=0.05$ level of confidence were selected. Some others were selected based on their hypothesized value and/or correlation with retention rate in previous research. For example, in Wyman (1997), regional employment per capita at the time of cohort formation and the ratio of institutional instruction and academic support spending per headcount student to regional income per job at the time of cohort

formation, were highly correlated with retention rate. Population and amount of financial aid available were also included because of their hypothesized value in previous research, bringing the total number of independent variables to six. All the independent variables are listed in Table 3.4. The first three are the internal variables. They are called internal because they are somewhat controllable by the management of the institution. The last three are the external variables. They are referred to as external because they are somewhat beyond the control of the institutions.

TABLE 3.3: PEARSON CORRELATION MATRIX FOR THE FULL MODEL

	RETRATE	HECENROL	EGIASEXP	AVGFIN	POPULATN	UNEMPLMT	CRIME
RETRATE	1.0000						
HECENROL	0.7397 <.0001	1.0000					
EGIASEXP	0.7135 <.0001	0.7547 <.0001	1.0000				
AVGFIN	-0.2547 <.0017	-0.3469 <.0001	-0.1507 0.0657	1.0000			
POPULATN	0.0599 <.4662	0.3676 <.0001	0.1068 0.1931	-0.0471 0.5671	1.0000		
UNEMPLMT	-0.5616 <.0001	-0.4506 <.0001	-0.4435 <.0001	-0.0749 0.3621	-0.3206 <.0001	1.0000	
CRIME	0.2165 0.0078	0.3434 <.0001	0.2492 0.0021	-0.1472 0.0723	0.4822 0.5579	-0.0955 0.2448	1.0000

Table 3.4: INDEPENDENT VARIABLES

1. **Headcount enrollment (HCE)** (total number of students enrolled in fall semester). This is the actual student population of the institution. Data was obtained from the Statistical Abstracts of Higher Education in North Carolina (2000).
2. **E&G I&AS expenditures (EGIASEXP)** (education and general expenditures on instruction and academic support in dollars). This is the amount of dollars spent directly on instruction and academic support activities. Data was obtained from the Integrated Postsecondary Education Data System (IPEDS) reports (UNC General Administration, 2000).
3. **Average Financial aid (AVGFIN)** (total amount of financial aid available in dollars divided by the total number of students enrolled). Data on this was obtained from the Statistical Abstracts of Higher Education in North Carolina (2000).
4. **Population (POPULATN)** (number of persons living in the area). This variable represents the total population of people residing in the county where the institution is located. Data on the variable was obtained from the U.S. Bureau of Economic Analysis (2000).

5. **Unemployment Rate (UNEMPLMT)** (the percentage of unemployed persons living in the area). This is the actual level of unemployment in the county where the institution resides. The source of data is the U.S. Bureau of Economic Analysis (2000).

6. **Crime (CRIME)** (the number of crimes committed on the campus). This represents the level of crime on each campus. Data was obtained from the U.S. Bureau of Economic Analysis (2000).

Data and Methodological Framework

In this study, a pooled cross-sectional time series technique was employed in the analysis of data. As stated earlier, the limited number of observations would be a problem but this was overcome by pooling the time series and cross section together to give a greater number of data points over time.

Pooled data are characterized by having repeated observations (years) on fixed units (institutions). In other words, pooled arrays of data combine cross-sectional data on N spatial units and T time periods to produce a data set of $N \times T$ observations. As mentioned earlier, pooled data are equally

important in providing the necessary information to deal with the inter-temporal dynamics as well as the individuality of the entities under investigation (Dielman, 1983).

The method of Ordinary Least Squares (OLS) was used in the estimation of the pooled data set, subject to the assumption that regression coefficients are equal across all institutions. To use the OLS procedure, the regression coefficients are assumed to be equal across all units. This means that the time-series estimates of a coefficient do not differ from the cross-section estimates of the same coefficient; i.e., the structure of the underlying equation does not change over time and across institutions. Considering the diversity of institutions to be pooled, this assumption of homoscedasticity does not always hold and could result in serious model misspecification. The problem of heteroscedasticity, if present, can be resolved by introducing dummy variables into the regression. The fixed effects approach has the capability of maintaining the stability of the coefficients through the introduction of dummy variables. Generally, however, time series data usually produce stable coefficients due to the fact that changes in both the dependent and independent variables are likely to be of the same magnitude (Pindyck and Rubinfeld, 1998).

Drawing from the matrix modeling form, a generic pooled linear regression model can be written, thus:

$$Y_{it} = \beta_0 + \sum \beta_k x_{kit} + \epsilon_{it}$$

where $i=1,2,\dots,N$; refers to a cross-sectional unit; $t=1,2,\dots, T$; refers to a time period and $k=1,2,\dots, K$ refers to a specific explanatory variable. Thus, Y_{it} and X_{it} refer respectively to dependent and independent variables for unit i and time t ; and ϵ_{it} is a random error and β_0 and β_k , refer respectively to the intercept and the slope parameters. Ordinary least squares linear regression model was fitted to the data using various combinations of the predictor variables listed in Table 3.4. It is worth noting that pooled cross-sectional time series represents a set of partitioned matrices that can be presented thus:

$$\begin{array}{ccccccc}
 & \mathbf{Y}_{11} & & \mathbf{X}_{111} & \mathbf{X}_{121} & \dots & \mathbf{X}_{1K1} & & & \epsilon_{11} \\
 & \mathbf{Y}_{12} & & \mathbf{X}_{112} & \mathbf{X}_{122} & \dots & \mathbf{X}_{1K2} & \mathbf{B}_0 & & \epsilon_{12} \\
 & \dots & & \dots & \dots & \dots & \dots & \mathbf{B}_1 & & \dots \\
 \mathbf{Y} = & \mathbf{Y}_{1t} & \mathbf{X} = & \mathbf{X}_{11t} & \mathbf{X}_{12t} & \dots & \mathbf{X}_{1Kt} & \mathbf{B} = \mathbf{B}_2 & \epsilon = & \epsilon_{1t} \\
 & \mathbf{Y}_{21} & & \mathbf{X}_{211} & \mathbf{X}_{221} & \dots & \mathbf{X}_{2K1} & \dots & & \epsilon_{21} \\
 & \dots & & \dots & \dots & \dots & \dots & \mathbf{B}_K & & \dots \\
 & \mathbf{Y}_{Nt} & & \mathbf{X}_{N1t} & \mathbf{X}_{N2t} & \dots & \mathbf{X}_{NKt} & & & \epsilon_{Nt}
 \end{array}$$

Considering the variables used in the study, the model of retention rate to be estimated was specified as follows:

$$\text{RETRATE}_{it} = \beta_0 + \beta_1\text{HECENROL}_{it} + \beta_2\text{EGIASEXP}_{it} + \beta_3\text{AVGFIN}_{it} + \\ \beta_4\text{POPULATN}_{it} + \beta_5\text{UNEMPLOYMT}_{it} + \beta_6\text{CRIME}_{it},$$

where, $i=1, \dots, N$ and $t=1, \dots, T$.

Definitions of these abbreviations are in Table 3.5.

Assumptions of the Model

The model used pooled time series and cross section data and allows the intercept to vary with each institution and year. The regression coefficients are assumed to be equal across all cross-sectional units, which implies that the structure of the underlying equation does not change over time, and across institutions. This assumption, however, does not always hold and may not be well-suited to empirical research. Therefore, the model was tested for the stability of the coefficients between two subsets of the sample. The sample was divided into two equal groups, with each group pooled for 5 years. This grouping made it possible to compare coefficients.

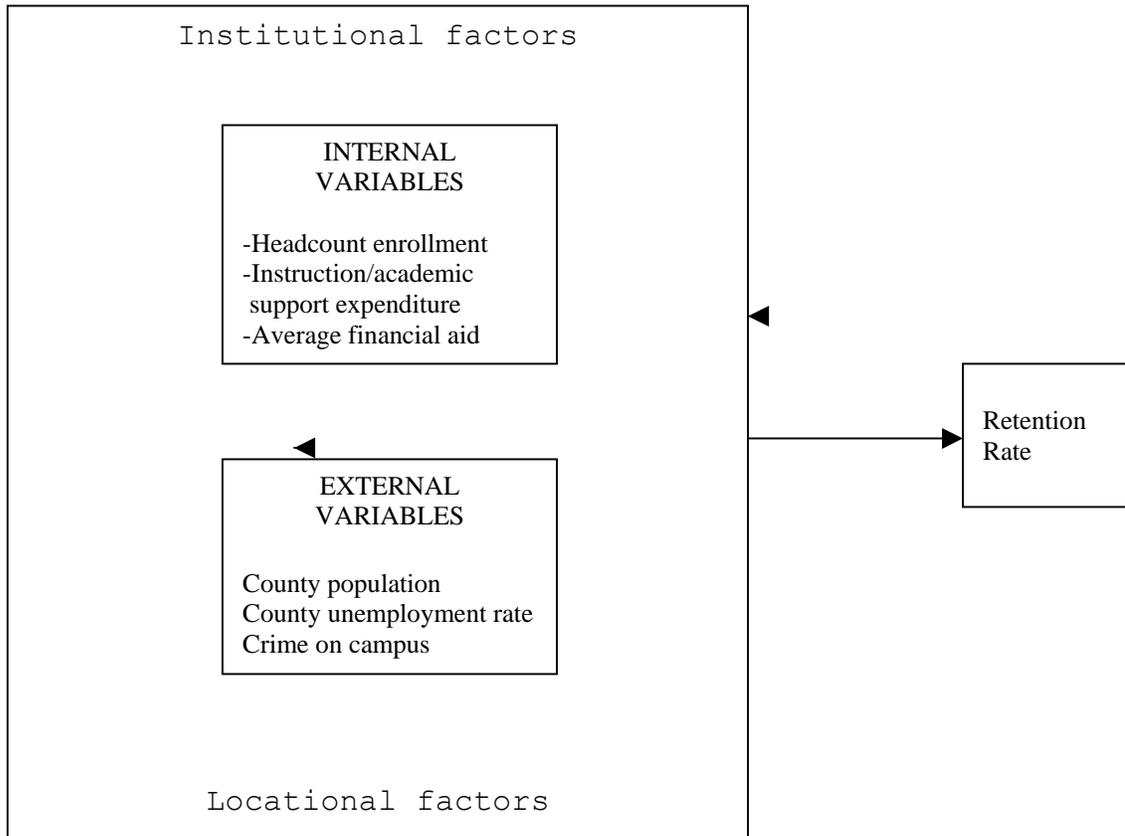


Figure 3.3: Ukpabi's Model of Retention Rate

The disturbance term in the model was assumed to satisfy the assumptions of the classical linear regression model. That is, the disturbance term is independent and identically

distributed with mean zero and common variance. Violations of these assumptions could lead to problems, such as heteroscedasticity and autoregression. The former will lead to inefficient estimates while the latter will lead to biased estimates. Appropriate diagnosis would be applied to detect these problems, if it became necessary.

The fixed effects specification has the advantage of addressing some of the problems that result when OLS assumptions are violated. Heteroscedasticity, for example, does not usually occur in time series data. According to Pindyck and Rubinfeld (1998), this is because changes in both the dependent variable and one or more of the independent variables are likely to be of the same order of magnitude. Considering the foregoing statement, therefore, the inference can be drawn that since this study involves time series data and the institutions are pooled for a period of ten years, it may be safe to assume a homoscedastic error term.

According to Berry (1993, p. 81), "when the homoscedastic assumption is violated, conventionally computed confidence intervals and conventional t-tests for OLS estimators can no longer be justified." In the event that this assumption is violated in a pooled data set, as in this particular study, one good way to address the problem in ordinary least squares

regression is by introducing dummy variables as a way of conditioning the variance to the sample (Sayrs, 1989).

A considerable number of explanatory variables were employed in this study, so we make an additional assumption that no exact linear relationship exists between the independent variables. The violation of this assumption would result in perfect multicollinearity and it would be difficult to calculate the OLS estimates of the parameters. Although in practice perfect multicollinearity is very unlikely to occur, some degree of multicollinearity is usually present in most regressions. Very high cases of multicollinearity may make it difficult, if not impossible; to isolate the effect each of the collinear variables has on the dependent variable. The result is unbiased parameters and larger standard errors, hence the t ratios will be smaller and the regression parameters will not be statistically significant. Although a high degree of multicollinearity may have adverse effects on regression results, this is by no means inevitable. The implication is that, if regression equations have low estimated standard errors and high t ratios, one should not be too concerned about the presence of multicollinearity (Thomas, 1997).

Fortunately though, multicollinearity can be corrected when detected. One approach is to eliminate one of the highly

collinear variables. This approach, of course, might result in specification bias, if the literature speaks strongly to the inclusion of that variable in the model. This problem will, however, be resolved if the variable with the larger coefficient relative to the dependent variable is chosen since both correlate to this dependent variable. Another approach in dealing with multicollinearity is to enlarge the sample size. This would not be a good option in this case since the data are already pooled. The third approach is to transform the multicollinear variables by forming a linear combination of them.

Efforts will be made to resolve the problem of multicollinearity through any of the above approaches, only when it presents serious consequences. These consequences consist of low and insignificant t scores, or a big enough change in the parameters to make them differ significantly from expectations. It must be borne in mind, however, that the mere existence of multicollinearity does not necessarily indicate that the estimates are invalid.

Data Source and Quality

The data used in this study include regional demographic and economic data that are available for the fifteen institutions over the period 1991 to 2000 from the Bureau of

Economic Analysis (BEA) of the U.S. Department of Commerce (2000). Retention rate data were obtained from the Program Assessment Abstract of University North Carolina General Administration Reports (2000). Data on institution funding were obtained from the University of North Carolina General Administration (Integrated Postsecondary Education Data System (IPEDS) reports (1991-2000). With regards to the quality of the information, the documentation is reproduced from original materials, which are also published by these institutions.

In developing a predictive model of retention rate, the modeling process is governed by the following five objectives:

1. To maximize the adjusted R squared statistic.
2. To include only predictor variables with regression coefficients having at most a .05 probability of equaling zero. A t-test will be used to make this determination.
3. To keep the number of predictor variables that exhibit very high multicollinearity to the barest minimum.
4. To ensure the reliability of findings by requiring the final model to be supported at the 0.05 level of significance via an F-test on subsets of the sample.
5. Finally, the optimal model will need to meet the classical assumptions of linear regression.

TABLE 3.5: ABBREVIATIONS OF VARIABLES

RETRATE	Retention rate
HECENROL	Headcount enrollment
EGIASEXP	Education & General expenditure on Instruction and Academic Support
AVGFIN	Average Financial aid
POPULATN	Population
UNEMPLMT	Unemployment Rate
CRIME	Crime

As part of the modeling process, computing Pearson's correlation coefficients for all combinations of the predictor variables bolsters the internal validity of this study. Taking steps to protect against very high multicollinearity, also strengthens the internal validity. As already stated, an F-test of the equality of regression coefficients across sample subsets establishes model stability.

All combinations were tested to find the model that explains the most variance. No limitations were placed on the number of variables that ended up in the final regression model. Theory was allowed to dictate the number of variables in the model. On finding the model that meets the objectives

of this study, the adjusted R-squared statistic associated with that model is assumed to quantify the proportion of retention rate variance that is accounted for by the model.

Test of Stability

The sizes of the educational institutions that make up the sample for this study show numerous variations. The variables employed also exhibit similar variations. As a result of these differences, it is hypothesized that the regression coefficients of the independent variables might differ across the model.

To test the stability or equality of the coefficients, we must perform a Chow test on the regression. The test involves computing two separate regressions for two subsets of the sample and one regression for the combined observations, and applying the appropriate F test. One subset involves the first five years and the second subset is the last five years of the dataset. The test is expressed as follows:

$$F_{k, N+M-2K} = \frac{[(ESS_F - ESS_P)/K]}{[ESS_P / (N+M-2k)]}$$

Where;

ESS_F is the residual sum of squares in the full model;

ESS_p is the residual sum of squares in the partial model;

K is the number of parameters estimated including the intercepts;

N and M are the number of observations for the sub samples, respectively.

The following hypothesis is tested:

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_k = 0.$$

H_A : At least one of the coefficients is nonzero.

The test involves adding the residual sum of squares of the two equations of the sub samples (partial models). The error sum of squares of the full model would also be obtained. Since there are $k = 7$ restrictions, and $N + M - 2K = 136$ degrees of freedom, the appropriate F statistics with 7 restrictions and 136 degrees of freedom are calculated thus:

$$F_{7, 130+20-14} = \frac{[(2011.23-1774.77)/7]}{[1774.77/(130+20-14)]}$$

$$F = 33.78/13.05 = 2.59$$

If $f > F(1-\alpha; K, N+M-2K)$, then reject the hypothesis that all regression coefficients are equal to 0.

The Two Big Schools (NCSU and UNC-CH)

North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC-CH) are the two biggest

schools in the study. By virtue of their large sizes and budgets, it was considered necessary to look at them separately to better understand their collective impact on the study.

North Carolina State University, Raleigh

North Carolina State University (NCSU) founded in 1887 and located in Raleigh, North Carolina, is a member of the 16-campus University of North Carolina system. As enunciated in its mission statement, NC State educates students for the 21st century and inspires future leaders.

The mission of North Carolina State University is to serve its students and the people of North Carolina as a doctoral/research-extensive, land-grant University. Through the active integration of teaching, research, extension, and engagement, North Carolina State University creates an innovative learning environment that stresses mastery of fundamentals, intellectual discipline, creativity, problem solving, and responsibility. Enhancing its historic strengths in agriculture, science, and engineering with a commitment to excellence in a comprehensive range of academic disciplines, North Carolina State University provides leadership for intellectual, cultural, social, economic, and technological

development within the state, the nation, and the world (North Carolina State University, 2004).

The University has a current enrollment of 29,637 students which makes it the biggest educational institution in the state of North Carolina. With an annual budget of approximately \$820 million and an endowment valued at more than \$312 million, NC State has total research expenditures of \$444 million. The university has a faculty population of 1,685 and a diverse student body that comes from all 50 states of the federation and 110 countries around the world (North Carolina State University, 2004).

University of North Carolina, Chapel Hill

The University of North Carolina at Chapel Hill (UNC-CH), the nation's first state university, was founded in 1789. It is a member of the 16-campus University of North Carolina system.

The mission of the University is to serve all the people of the State, and indeed the nation, as a center for scholarship and creative endeavor. The University exists to teach students at all levels in an environment of research, free inquiry, and personal responsibility; to expand the body of knowledge; to improve the condition of human life through service and publication; and to enrich our culture. To fulfill

this mission, the University must, among other things, address, as appropriate, regional, national, and international needs.

The University has a current enrollment of 26,000 students, which makes it the second biggest educational institution in the state of North Carolina, behind NC State. With an annual budget of approximately \$1.5 billion and an endowment valued at more than \$1.1 billion, UNC Chapel Hill has the largest total research expenditure (\$479 million) in the 16-campus UNC system. The university has a faculty population of 3000, and a diverse student body that comes from all 50 states of the federation and 100 countries around the world (University of North Carolina at Chapel Hill, 2004).

In conclusion, North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC-CH) are also the only two Research I institutions in the University System of North Carolina. They also have stated missions that include an international focus. In addition to the above characteristics and their sizes in terms of their very large student enrollments, it was considered important to run a separate regression model for them to check if they would unduly influence the results.

To achieve this goal, two regressions were run: one involving just the 20 observations of the two schools,

referred to as the Two Big Schools and the other represents the remaining thirteen institutions (excluding the big two), whose headcount enrollments range from 2,035 to about 18,750.

CHAPTER FOUR

RESULTS

Introduction

This chapter presents the results of the regression equation modeling conducted to address the research questions identified in Chapter One. The two specific research questions proposed in Chapter One are as follows:

Question 1: What is the predictive ability of the internal variables (Headcount enrollment, Education and general expenditures on instruction and academic support, Average Financial aid) and external variables (Population, Unemployment, Crime) in determining an institution's retention rate?

Question 2: Is the regression model of retention rate developed from the above variables stable over time?

In formulating the model, a pooled cross-sectional time series technique was employed and the Ordinary Least Squares (OLS) method was used in the estimation of the data. The modeling was conducted to develop a statistical model as explained in Chapter Three for the purpose of predicting retention rates in a universal sample of colleges and

universities in the UNC system. The results from the analysis are provided below.

To anticipate, the findings indicate that headcount enrollment, amount of education and general expenditure on instruction and academic support, county population and unemployment rate, are significant predictors of retention rate for an institution. The b coefficients in the regression model were found not to be structurally similar but the model was found to be reliable ($F=58.66$, $p < 0.0001$) in estimating retention rate for the range of values examined for the institutions studied.

Data and Findings

The sample means and standard deviations for the dependent and the six independent variables are provided in Table 4.1. The regression analysis assessed the individual importance of each of the independent variables as well as the combined predictive capability of all the variables.

Five regression models were run in this study. One, which has all the 150 observations resulting from pooling the variables for a ten-year period (1991-2000), referred to here as The Full Model, is presented in Table 4.2. The second is the 130 observations that result from running the model without the two biggest institutions (NCSU and UNC-CH)

provided in Table 4.3, and the third model represents the 20 observations of the two biggest institutions by themselves given in Table 4.4. The fourth represents the 75 observations of the first five years of data (1991 - 1995) provided in table 4.5, and the fifth model represents the other 75 observations of the last five years of data (1996 - 2000) given in table 4.6.

TABLE 4.1. DESCRIPTIVE STATISTICS OF VARIABLES

VARIABLE	n	Mean	SD	Min	Max
1. RETRATE	150	77.73	6.83	63.80	95.00
2. HECENROL	150	10.23	7.76	1.77	28.61
3. EGIASEXP	150	85.39	109.87	8.74	565.67
4. AVGFIN	150	832.87	591.19	55.00	3567.00
5. POPULATN	150	233.85	177.45	27.46	699.47
6. UNEMPLMT	150	3.63	1.88	1.00	1.00
7. CRIME	150	32.01	77.82	0.00	938.00

As was stated earlier, the presence of heteroscedasticity produces biased standard errors, which in turn leads to bias in test statistics and confidence intervals. Also, OLS is not optimal when severe heteroscedasticity is present. Fortunately, however, time-series data are usually safe from heteroscedasticity, because changes in both the dependent and

independent variables are likely to be of the same order of magnitude (Pindyck and Rubinfeld, 1998). In addition, unless heteroscedasticity is severe, significance tests are virtually unaffected, and thus OLS estimation can be used without concern of serious distortion (Williams, 2003).

In light of the above explanations, it is safe to assume a homoscedastic error term, since this study involves time-series data and institutional data are pooled for a period of ten years.

The Full Regression Model

The predictive model of first-year retention rate and the model's associated statistics as shown in Table 4.2 was generated to address the research question: What is the predictive ability of the internal and external variables in determining an institution's retention rate? In more general terms, the model will help to determine whether each of these variables that relate to the demographic characteristics, economic conditions, college management, and fiscal policy of the institutions plays a role in predicting retention rate and by how much.

The adjusted R-square for all variables in the full regression model was .6990 which means that almost 70 percent of the variance in retention rate is explained by the six

independent variables. This is a high level of explained variance for this kind of study, and it provides the predictive ability of the variables used in this study in determining an institution's retention rate. It answers in the affirmative the general research question of whether an educational institution's retention rate is a function of the demographic characteristics, economic conditions, college management and fiscal policy of the state in which the institution resides.

The full model shows that four variables had significant independent effect on retention rate. Headcount enrollment ($t=6.03$), education and general expenditure on instruction and academic support ($t=2.66$), county unemployment rate ($t=-6.15$), and county population ($t=-4.99$), were significant at the $p<.05$ level. The independent effect of headcount enrollment underscores the importance of number of students admitted in determining an institution's retention rate.

The relationship between each of these four predictors and retention rate is strong as evidenced by the larger t -values ($t=6.03$, $t=2.66$, $t=-6.16$, and $t=-4.99$, respectively) associated with their estimated regression coefficients.

TABLE 4.2. REGRESSION RESULTS OF THE FULL MODEL

VARIABLE	B	SE	t-value
INTERCEPT	79.66	1.40	56.81 ^a
HECENROL	0.46	0.08	6.03 ^a
EGIASEXP	0.01	0.01	2.66 ^a
AVFINAID	-0.00	0.00	-1.74
POPULATN	-0.01	0.00	-4.99 ^a
UNEMPLMT	-1.21	0.20	-6.15 ^a
CRIME	-0.00	0.00	-0.91

Adjusted $R^2 = 0.70$ (n = 150, p < 0.05). ^ap < 0.05.

Notes: Hecenrol in thousands
 Populatn in thousands
 Avgfin in thousands of dollars
 Egiasexp in millions of dollars

Hecenrol and egiasexp each has a direct relationship with retention rate as evidenced in their positive b coefficients, while the relationship between each of unemployment rate and population is indirect, given their negative b coefficient. The other two predictor variables namely, average financial aid (t=-1.74) and crime (t=-0.91) were not significant in the model.

This finding on financial aid corroborates much of the findings in the retention literature that financial aid is a

negative correlate of attrition (Paulsen and St. John, 2002). DesJardins, Ahlburg, and McCall (2002) found that loans have a negative effect on persistence, although this effect diminishes over the years in college. As stated previously, studies that have investigated the role of financial aid on undergraduate student persistence have been conflicting, at best. Research by Jones and Moss (1994) indicates that recipients and non-recipients of federal financial aid persist at comparable rates. This result, however, is not surprising as it only confirms what this researcher has always argued in less formal settings prior to undertaking this study. There is ample empirical evidence resulting from simulations of real studies that indicate that amount of financial aid received does not really prevent an academically unprepared student from dropping out. This finding may be due to the contemporary attitude of students to financial aid money. Most students consider it free money which can be used any kind of way and this might result in less serious commitment to their education which in turn causes them to dropout. This attitude is not surprising in light of the lack of direction of financial aid policy in general. The strongest evidence for the lack of direction is the absence of any effort to define and measure progress in meeting the goals of the program (Kane, 1999).

Crime ($t = -0.91$, $p = 0.3642$), though not significant in this study has an inverse relationship with retention rate, which means that as the level of crime increases, retention rate decreases, other things being equal. It is noteworthy that the number of crimes committed on an institution's campus was also expected to have an inverse relationship with retention rate. The implication is that institutions that have high incidence of crime are more likely to experience lower retention rates than those that experience lower incidence of crime.

Population ($t = -4.99$, $p < .0001$) is negative and significant in the model. This suggests that as the population in the institution's location increases, retention rate decreases. This finding is contrary to what the literature says about the relationship between population and retention rate. The possible explanation for this unexpected result is that a high population around an educational institution might attract a mix of elements not conducive to learning. The lure of the glamour of a big city coupled with the pressure of negative influences that it brings, might impact on the commitment of students, especially marginal ones.

Unemployment ($t = -6.15$, $p < .0001$) was negative and significant in the model. This shows that as the rate of unemployment in the location of the institution increases, the

retention rate for the institution decreases. This is in line with the literature on retention, Wyman (1997) which suggests that as the number of employed people grows, retention rate increases as well. This means that retention rate is directly related to the percent of the county population that is employed and inversely related to the unemployment rate (the percent of the labor force that is unemployed). This finding might be counter-intuitive but it is supported by the regression result.

It is important to note that the impact of unemployment is strong in this study. There is no evidence of collinearity between unemployment and population as evidenced by the significant correlation between them ($r=-0.3206$, $p<0.0001$). The presence of multicollinearity could mitigate the individual impact of these variables on retention rate. Despite this possibility, however, these variables are not eliminated because the literature speaks strongly to their inclusion in similar retention studies. Furthermore, the presence of multicollinearity, as was noted earlier, does not necessarily make our estimates invalid in this study.

The significant independent impacts of headcount enrollment, unemployment, and population represents the defining characteristics of this study, with betas ranging from -0.02 to 6.03, for headcount enrollment; -6.15 to 0.31,

for unemployment; and -1.42 to -4.99, for population in the five regressions. Beta represents the average amount by which retention rate increases or decreases when any independent variable increases or decreases one standard deviation and other independents are held constant. Therefore, this variability in beta weights for these variables implies that there is room to improve retention rate in this model. This also enhances the importance of these variables relative to this model.

In his impressive research effort at formulating a predictive model of retention rate at regional two-year colleges, Wyman (1997) found that regional employment per capita was a critical predictor variable. In the Wyman study, education and general expenditure on instruction and academic support and employment were the only significant variables. The model explained 66% of the variance in retention rate in the 16-campus system of South Carolina Community Colleges. But headcount enrollment and population were not significant in that study. In this particular study, however, three additional variables, headcount enrollment, unemployment rate, and population were found to be strong predictors of retention rate. It is noteworthy that headcount also was not significant in Wyman (1997).

The variable, headcount enrollment, as was expected contributed very significantly to explained variance in retention rate. Retaining current students into a second and a third year of study and beyond, perhaps even to graduation, may be desired to maintain enrollment levels in general. It would do it in a way that may be cheaper than recruiting larger and larger numbers of new and perhaps even more academically marginal students who will also leave after a short period of time (Wetzels, et al, 1999).

In relative magnitude, headcount enrollment, education and general expenditure on instruction and academic support, unemployment rate, and population, contributed significantly in explaining the variance in retention rate. Higher educational institutions that have higher headcount enrollments and spend more of their funds directly on instruction and academic support activities are more likely to experience higher retention rates. Additionally, the situation of unemployment and size of the population in an institution's location would likely have significant impacts on its retention rate.

Impact of the Two Big Schools

The results of the models resulting from the exclusion of the two big schools indicated that they have some limited

influence on the study. Two regressions were run: one involving the thirteen institutions after excluding the big two, referred to as the Thirteen schools, is given in table 4.3, and the other representing just the 20 observations of the two schools, referred to as the Two Big Schools provided in table 4.4.

The structure of the two regressions for the 13 schools and the two big schools are fundamentally different in terms of the differing b coefficients for each of them. In the 13 schools, the coefficient for hecenrol is positive ($b=0.2921$) whereas it is negative ($b=-0.0183$) in the two big schools. Also, in the two big schools, the b coefficients for avgfin ($b=0.0012$) and unemployment ($b=0.1916$) are positive, whereas these same coefficients avgfin ($b=-0.0015$) and unemployment ($b=-0.0084$) are negative in the 13 schools. Furthermore, the F values, though significant in each of the regressions, show a wide variation ($F=37$), for the two big schools and ($F=13$) for the 13 schools. The R-squares for the regressions, which indicate the percent of explained variance in retention rate, are significantly different. $R^2=0.92$ for the two big schools and $R^2=0.35$ for the 13 schools.

Comparing the 13 schools (i.e. excluding the two big schools) with the full model, there is not much difference in the structure of the two regression equations. The b

coefficients are similar. The betas for the intercept are $b=80$ for the full model and $b=81$ for the 13 schools. The coefficients for the independent variables are similar in magnitude and significance, except for *egiasexp*, which is negative ($b=-0.0029$) for the 13 schools but positive ($b=0.0124$) in the full model.

TABLE 4.3. REGRESSION RESULTS OF THE THIRTEEN SCHOOLS

VARIABLE	B	SE	t-value
INTERCEPT	81.37	1.50	54.33 ^a
HECENROL	0.29	0.09	3.27 ^a
EGIASEXP	-0.00	0.01	-0.43
AVGFIN	-0.00	0.00	-2.49 ^a
POPULATN	-0.01	0.00	-3.89 ^a
UNEMPLMT	-1.20	0.20	-6.11 ^a
CRIME	-0.00	0.00	-0.25

Adjusted $R^2 = 0.3528$ ($n = 75$, $p < 0.0001$). ^a $p < 0.05$.

Notes: Hecenrol in thousands
 Populatn in thousands
 Avgfin in thousands
 Egiasexp in millions of dollars

TABLE 4.4. REGRESSION RESULTS OF THE TWO BIG SCHOOLS

VARIABLE	B	SE	t-value
INTERCEPT	98.12	14.62	6.71 ^a
HECENROL	-0.02	0.78	-0.02
EGIASEXP	-0.00	0.01	-0.15
AVGFIN	0.00	0.00	0.56
POPULATN	-0.01	0.01	-1.42
UNEMPLMT	0.20	0.63	0.31
CRIME	-0.04	0.04	-1.05

Adjusted R² = 0.9189 (n = 75, p < 0.0001). ^ap < 0.05.

Notes: Hecenrol in thousands
 Populatn in thousands
 Avgfin in thousands
 Egiasexp in millions of dollars

Differences in the Results of the Thirteen Small Schools and the Two Big Schools

It is interesting to note that none of the variables are significant in the regression involving only the two big institutions. Headcount enrollment, an important variable in retention research, was not significant in the regression for the two big schools, in spite of its considerable significance

in the regression of the thirteen smaller institutions. This may be due to the fact that the big schools have large student populations, which might have mitigated impact on headcount. The same thing can be said about the influence of population as well. The budgets and per student spending of these bigger schools are so huge in comparison to the smaller ones that when the bigger schools are isolated, impacts of average financial aid and spending on instruction and academic support are moderated.

The two big institutions, which are the largest in the state, have considerable amounts of dollars appropriated for research and student work-study as well as internship programs. They are also located in or around cities and local governments where significant economic activities take place. These two reasons may be important in accounting for high level of job availability in their locations, thereby weakening the impact of unemployment in the model.

Furthermore, the regression result of the thirteen institutions differs greatly from that of the two bigger ones because of degrees of freedom issues. The small sample size ($n=20$) for the bigger institutions is too small to generate statistically meaningful results. This is because when $n < 30$, the t distribution takes on a standard deviation that is increasingly greater than one. This means that the

distribution is more spread out than the standard normal distribution.

Multiple regression has been described as a seductive technique. If one plugs in as many predictor variables as one can think of, usually at least a few of them will come out significant. This is because one is capitalizing on chance when simply including as many variables as one can think of as predictors of some other variable of interest. This problem is compounded when, in addition, the number of observations is relatively low ($n=20$) as in the two bigger institutions. Intuitively, it is clear that one can hardly draw conclusions from an analysis of six questionnaire items based on 20 respondents. Most authors recommend that one should have at least 10 to 20 times as many observations as one has variables, otherwise the estimates of the regression line are probably very unstable and unlikely to replicate if one were to do the study over (Berry, 1993).

According to Tabachnick and Fidell (2001), a rule of thumb for testing b coefficients is to have $N \geq 104 + m$, where m = number of independent variables. A rule of thumb for testing R -square is $N \geq 50 + 8m$. Where $m \geq N$, regression gives a meaningless solution with R -square = 1.0. In general, you need a larger N , especially when the dependent variable is skewed.

When the population standard deviation (μ) is unknown and the sample size is less than 30 ($n < 30$), the distribution of the test statistic cannot be guaranteed to be normal. In fact, the test statistic can be said to conform to a t distribution. The t distribution is similar to the standard normal distribution in that it is symmetrically distributed around a mean value. But where the t distribution differs from the standard normal is that the standard deviation of the t distribution is determined by the number of degrees of freedom.

Degrees of freedom are calculated from the size of the sample. They are a measure of the amount of information from the sample data that has been used up. Every time a statistic is calculated from a sample, one degree of freedom is used up. They are, in deed, a measure how much precision an estimate of variation has. A general rule is that the degrees of freedom decrease when we have to estimate more parameters. Before you can compute the standard deviation, you have to first estimate a mean. This causes you to lose a degree of freedom and you should divide by $n-1$ rather than n (Green, 1991). In more complex situations, like Analysis of Variance and Multiple Linear Regression, we usually have to estimate more than one parameter. Measures of variation from these procedures have

even smaller degrees of freedom than our standard deviation formula.

When you have a very large sample drawn from your data set, the difference between t values and Z values is miniscule. But as the size of your sample falls, the t distribution takes on a standard deviation increasingly greater than 1. In other words, this means that the t distribution when $n < 30$, is more spread out than the standard normal distribution. To show this, take a look at the explanation of the normal distribution curve to confirm that 95% of all Z values lie between 1.96 and -1.96. It is the case that in a t distribution with 20 degrees of freedom, 95 % of all t values will lie between 2.09 and -2.09. But with only 5 degrees of freedom, 95 % of all t values will lie between 2.57 and -2.57.

Furthermore, a closer look at the retention rates of all institutions for the ten-year period shows that the retention rates for the two bigger schools do not have much variation. There is a difference of only an 8-percentage point between the highest and the lowest rates. For the thirteen smaller schools, however, there is 22-percentage point variation in retention rates. The implication of this is that the bigger schools have less room to vary and as such produced meaningless results. Considering the fact that regression

analysis is really about measuring the variance in the dependent variable, the inadequate variability in the retention rates for the bigger schools produced an unreliable result.

Stability of the Model

The result of the Chow test performed in Chapter Three to test the stability of the regression model was not significant. Two regression models were run: one utilized the 75 observations that were pooled for the first five years (1991 - 1995), referred to as the First five years, provided in Table 4.5; and another that used the other 75 observations pooled for the last five years (1996 - 2000), called the Last five years, which is given in Table 4.6. The two sub-samples were found to be structurally dissimilar and this shows that the model might not be stable. But it is important to recognize that the two equations are very close structurally as evidenced by the close numbers in terms of their various F values, t values, R-squares, and coefficients of each variable. This is given in Tables 4.5 and 4.6 respectively.

TABLE 4.5. REGRESSION RESULTS-FIRST 5 YEARS (1991-1995)

VARIABLE	B	SE	t-value
INTERCEPT	81.61	2.22	36.83 ^a
HECENROL	0.32	0.20	1.62 ^a
EGIASEXP	0.00	0.01	0.55
AVGFIN	-0.00	0.00	-1.72
POPULATN	-0.01	0.00	-3.08 ^a
UNEMPLMT	-1.24	0.31	-3.95 ^a
CRIME	0.05	0.06	0.92

Adjusted $R^2 = 0.6871$ ($n = 75$, $p < 0.0001$). ^a $p < 0.05$.

Notes: Hecenrol in thousands
 Populatn in thousands
 Avgfin in thousands
 Egiasexp in millions of dollars

Furthermore, going by the high R-Square ($R^2=0.6990$) and highly significant F value ($F=58.66$, $p<.0001$), it can be concluded that the regression equation developed is useful for predicting retention rate for the range of values examined for the institutions in the study.

From an F distribution table, provided in Appendix D, the critical F-value at the $\alpha= 0.05$ level is $F (0.95; 7, 136)= 2.01$.

TABLE 4.6. REGRESSION RESULTS-LAST 5 YEARS (1996-2000)

VARIABLE	B	SE	t-value
INTERCEPT	79.06	1.76	44.89 ^a
HECENROL	0.34	0.12	2.92 ^a
EGIASEXP	0.02	0.01	3.04 ^a
AVGFIN	0.00	0.00	0.18
POPULATN	-0.01	0.00	-3.66 ^a
UNEMPLMT	-1.58	0.27	-5.93 ^a
CRIME	-0.00	0.00	-0.46

Adjusted $R^2 = 0.76$ ($n = 75$, $p < 0.0001$). ^a $p < 0.05$.

Notes: Hecenrol in thousands
 Populatn in thousands
 Avgfin in thousands
 Egiasexp in millions of dollars

The value of the calculated F, is 2.59. Since $f=2.59 > 2.01$, we reject the null hypothesis that both β_1 and β_2 are equal to 0. This implies that there is a probability that the sub-samples are structurally dissimilar and so the model might not be good enough for estimating retention rates for both sub-samples. The full model, however, is very reliable considering the considerably high F-value ($F=58.66$) and the corresponding P-value ($p < .0001$), which is highly significant.

CHAPTER FIVE

SUMMARY, DISCUSSIONS, RECOMMENDATIONS, AND IMPLICATIONS

This chapter presents a summary of the study's rationale and conclusions. It also summarizes the implications of these conclusions and makes some recommendations for further research on student retention in higher education in the United States.

Rationale of the Study

Student retention is widely acknowledged as one of the knottiest problems in higher education. The changing age demographics and its impact on college student enrollment have been noted in a series of retention studies. Further more, research on student retention in higher education in the United States indicate that a large number of students fail to persist to graduation from their initial institutions of enrollment. In fact, about 20 percent of the 1999 beginning students left postsecondary education before the beginning of their second year of study at the same institution in the year 2000 (CSRDE, 2001). According to NCES (2004), approximately 20 percent of the freshmen enrolled in colleges and universities drop out before completing their programs.

The twin problems of low enrollment and high attrition have adverse impacts on both the image and fiscal well-being of an institution. Retention rates of colleges and universities have become important in evaluating their effectiveness as most states now use formulas to determine budgets and appropriations for their public institutions. According to McKeown (1996), most of these formulas give some consideration to the number of students enrolled in determining the budget for that institution. A public institution that increases its enrollment and retention will obviously attract increased funding from the state. Furthermore, high graduation rate is indicative of strength and portrays a college or university as meeting the primary objective for which it was established. Such an institution will certainly acquire considerable goodwill that will be helpful in attracting potential students.

In light of the increasing importance that retention rate has assumed in recent times, especially as it relates to institutional effectiveness and accountability in higher education, it has become necessary to have a fuller understanding of retention rate behavior. Considerable research has been conducted on student retention and persistence in the last several decades, but much of the research have largely focused on individual student

characteristics. Although such individual level research provides some explanation about differences in the retention of individual students, it has not conclusively explained the differences in retention rates among colleges and universities.

This particular study is expected to produce a model that would have the power to identify some uncontrollable external factors that contribute to retention rate values. Such a model will probably produce measurements that will be more representative of actual effectiveness of the institution. Furthermore, such a model could be used to obtain a more accurate expression of the degree to which colleges and universities are powerless in terms of their ability to really determine their retention rates.

To this end, there is an opportunity to employ only institution level variables to predict retention rate at 4-year colleges and universities. This study employed a set of predictor variables already identified in an earlier study (Wyman, 1997) as internal and external variables in modeling retention rate. The external variables are those that are beyond the control of the institutions, and internal variables are those that are somewhat controllable by the institutions. Both sets of variables are combined to formulate a predictive model of retention rate for the colleges and universities in

the University of North Carolina system. Employing both external and internal variables resulted in a model that maximized the amount of retention rate variance explained.

Furthermore, such a model would be used to obtain a more accurate description of the degree to which institutions are locked into a certain retention range due to the influence of the uncontrollable forces. In other words, the study was able to formulate a predictive model of retention rate as well as identify what external institutional factors account for retention rate differentials among colleges and universities in the University of North Carolina system.

Additionally, the University of North Carolina system does not currently have a predictive statistical model of retention rate. This study is therefore timely, as it fills this void by producing a regression model for predicting retention rates in the system. It also provides an alternative check for the accuracy of retention rates

Summary of Results

This study addressed a central research question as follows: what is the predictive ability of the selected internal and external variables in determining an institution's retention rate? The second research question addressed the need to know whether the model developed from

this study was stable over time. The results given in Chapter Four indicate that the retention rate of an institution can be shaped by some of these variables.

It was found that headcount enrollment, amount of education and general expenditure on instruction and academic support, unemployment, and size of population, are significant predictors of retention rate for an institution. The Full-model regression analysis provided in Table 4.2, shows an adjusted R-squared of .70, indicating that 70 percent of the variance in retention rate is explained by the model.

The other variables that are significant in this study are education and general expenditure on instruction and academic support, unemployment and population. To increase retention rate, therefore, colleges and universities must secure appropriations sufficient to increase or formulate budgets so as to increase per student spending on instruction and academic support (Wyman, 1997). Also, the situation of unemployment and size of population in an institution's location are expected to have an impact on retention rate. Locations that have high unemployment rates are likely to experience lower retention rates. The same is also true that institutions located in counties that have large populations will likely experience reduced retention rates. According to Mohammadi (1996), external forces, particularly those related

to community forces in the immediate geographical environment of the college's location, are also important in understanding and interpreting the retention patterns at two-year public community colleges. This will also probably be applicable to four-year public institutions.

As can be seen in Table 4.1, the mean first-year retention rate for the 150 observations is 77.67 percent. This implies that more than 20 percent of students enrolled would not be retained in the university system. The standard deviation of retention rate, 6.84 percent, shows a considerable variation in retention rate across observations. In fact, as evident in Table 4.1, first-year retention rate for all the 1991-2000 cohorts indicates a variation range of 31 percentage points, from a low 64 percent to as high as 95 percent.

The regression coefficient for predictor 1 (hecenrol), 0.4576, indicates that retention rate improves 0.45 percent for every 1000 increase in headcount enrollment. The large t-value associated with the coefficient for hecenrol, $t=6.03$ ($p < .0001$), implies that this relationship is very strong. Headcount enrollment of an institution, then, is an important determinant of first-year retention rate. Headcount enrollment is partially a function of both external forces and

institutional policy, and so is somewhat controllable by university policy makers.

The full regression model also indicates that retention rate increases about 0.01 percent for every \$1 million increase in expenditure on instruction and academic support. Also, the large t-value associated with the coefficient of this predictor, ($t=2.66$, $p= .0003$), implies that its relationship with retention rate is very strong. For the variable, unemployment, retention rate decreases 1.21 percent for every 1 percent increase in unemployment. The large t value for this variable ($t=-6.15$) indicates a very strong inverse relationship with retention rate. Lastly, for population, retention rate decreases 0.01 percent with every 1000 increase in population. Also, the large t value ($t=-4.99$) shows a strong but inverse relationship with retention rate.

Discussion

An interpretation of the results of this study yielded the following conclusions:

Analysis using multiple regression indicated that four of the six independent variables were significant at the five percent level. The significant variables include headcount enrollment, expenditure on instruction and academic support, unemployment rate, and size of population. Generally,

researchers have studied inter-institutional differences in retention rate (Wyman, 1997). With the exception of size of population and headcount enrollment, the findings of this study corroborate the findings of Wyman (1997) who concluded that regional employment per capita and spending on instruction and academic support per student as a ratio of regional mean income influence retention rate. Cambrera, Castaneda, Nora, and Hengstler (1992) affirmed the impact that forces external to the institution (Bean and Metzner, 1985) also have on student attrition. They concluded that the interplay between institutional, personal, and external factors needed to be considered in developing programs to increase persistence.

Headcount enrollment is an important variable in this study. An institution that has high student enrollment would enjoy higher retention rate than one that has low enrollment. Headcount enrollment, however, was not significant in Wyman (1997) study that was based on community colleges in South Carolina. It appears that enrollment does not impact retention rates in two-year colleges and this may be due to the fact that by their nature, they are set up to feed four-year colleges. As originally conceived, two-year colleges were to provide education equivalent to the first two years at a four-year institution (Kane and Rouse, 1999). From a two-year

college, students would transfer to a four-year college, where work towards a baccalaureate degree would conclude. Thus, the intended purpose of two-year colleges was to create more Bachelor's-earning students by extending opportunities for entrance into higher education.

Another very important variable in the study is expenditure on instruction and academic support. Institutions should make serious efforts to secure appropriations sufficient to positively impact student spending on instruction and academic support. This is consistent with the findings of Wyman (1997), who concluded that spending on instruction and academic support per student as a ratio of regional mean income has a positive influence on institutional retention rate. The secured appropriations should be targeted to the specific aspects of the program that have the potential for program improvement and impact on retention. To achieve the desired end, this effort should begin with developing an assessment system for instruction and academic services as suggested by Astin's I-E-O model of evaluation. According to Astin (1993), the model looks at Input, Environment and Output. Input refers to what the students bring with them as they enter the program or institution, such as, demographics, learning styles, motivation, etc. The environment includes the programs and services that an institution provides for

students including instruction, tutoring, computer-assisted learning, etc, and output includes the outcomes of what has happened to the students who were served by the program. The purpose of this kind of program analysis is to appropriately identify and target instructional spending so as to improve institutional retention rate. In response to research findings (Picklesimer & Miller, 1998; Ryland, Riordan, & Brack, 1994), colleges and universities have developed several intervention programs and have often used inventories to assess student needs when they enter the institution in order to increase their retention rates.

The situation of unemployment and the size of population in an institution's location were significant in this study. Institutions in counties that have high unemployment rates are likely to experience lower retention rates. Students increasingly straddle the boundary between wage labor and college (American Council on Education (ACE), 2000) and education's opportunity cost may influence retention. Purdue University (2001) found that 17-35% of Indiana students cite employment as the reason for dropping out. St. John et al. (2001) found that changes in labor markets partially explain changes in retention. Mohanty and Finney (1997) studied students' wage-based opportunity costs and found that higher

wage rates initially enhance retention up to a point, after which the probability of retention would begin to decrease.

Institutions located in counties that have large populations would likely experience lower retention rates. This finding is consistent with Mohammadi (1996), who observed that external forces, particularly those related to community forces in the immediate geographical environment of the college's location, are also important in understanding and interpreting the retention patterns at two-year public community colleges. The same is applicable to four-year public institutions. Additionally, Bean and Metzner (1985) also found that environmental variables, such as, employment have an impact on institutional retention rates.

Going by the mean first year retention rate of 77.67 percent, over 20 percent of first year students would not be retained in the University of North Carolina system in any given year. If we should assume that this is the final year for the first-time, full-time freshmen in the institutions of the UNC system, this retention rate is not very impressive and really reveals that the university system has some room to improve on student retention. This retention rate is unimpressive because it is way below the national average. The freshman to sophomore year CSRDE benchmark retention rate for highly selective public schools is about 87 percent in the

most recent available national data. Even when compared with the national average freshman to sophomore year retention rate for all Consortium for Student Retention Data Exchange schools, which is 79 percent, it still shows that there is some room for improvement (Consortium for Student Retention Data Exchange) (CSRDE, 2002).

Higher education institutions can substantially increase their retention rates. The standard deviation of retention rate, 6.84 percent, shows a considerable variation in retention rate across observations. In fact, as evident in Table 4.1, first-year retention rate for all the 1991-2000 cohorts indicates a variation range of 31 percentage points, from a low 64 percent to as high as 95 percent. This wide range of variation suggests that considerable amount of changes in retention rate can be achieved. This observation is consistent with the findings of Wyman (1997) who concluded that a variation range of about 6.41 percent in retention rate is possible in colleges and universities.

Recommendations and implications for practice

The results of this study have several implications for higher education practitioners. The following action strategies linked to the findings and conclusions of this

study are hereby put forward for university management and policy makers, and indeed other retention practitioners.

When institutions of higher learning increase their enrollments, the opportunity exists for improving retention. Thus, there should be increased student recruitment efforts to beef up enrollment figures. A variety of enrollment models have dealt with the factors that affect college enrollment which have included larger retention. Some of these models, such as, Hearn and Longanecker (1985), Leslie and Brinkman (1987), and Becker (1990) have a general consensus that enrollment has two major components, initial enrollment and continual enrollment or retention. The initial enrollment decision is essentially a discrete process. The student has to decide to attend college, to which colleges to apply, and then, if accepted at more than one college, which college to attend. Once the student enrolls, that initial decision to attend a particular college is completed. However, once enrolled at a particular college, there is first the ongoing decision to continue in college in general, and second, the decision to continue at a particular college. The student retention decision is continually updated with the arrival of new information such as academic status, grades, and satisfaction with the social life or student peer group, i.e., information not present in the initial enrollment decision. In

addition, the longer a student remains at a particular institution, the higher is the transaction cost of switching to a competing institution, since there is a higher probability of losing more credit hours the longer one stays at the original institution. Thus, colleges and universities face a two-fold enrollment problem: first, getting students to enroll initially, and second, keeping those students enrolled or retained.

For major universities, with a surplus of applicants and high rates of retention, attrition may not be a major problem, especially if there is a waiting pool of applicants or transfer students to fill any positions that may become open. However, retention may be an issue at those major schools to the degree that they want to expand the diversity of the student body. Colleges and universities with a more open admissions policy, and without a substantial waiting list of first-time applicants or transfers, may find that attrition is a more serious issue, since overall enrollment numbers drive financial support from the public sector. For this second group of institutions, retention is an important, perhaps even critical, issue to ensure their continual enrollment numbers and financial viability. In this case, retaining current students into a second and a third year of study and beyond, perhaps even to graduation, may be desired to maintain

enrollment levels in general. Retention may serve to maintain enrollment levels in a manner that may be cheaper than recruiting larger and larger numbers of new and perhaps even more academically marginal students who will also leave after a short period of time (Wetzel, O'Toole, & Peterson, 1999). College administrators need to engage in better marketing of their program offerings so as to differentiate themselves in the eyes of potential students, thereby attracting more students.

Another key variable in this study is expenditure on instruction and academic support. This variable should be carefully evaluated. The implication of this for university policy makers is that to increase retention by say, one percent, expenditures on instruction and academic support must increase by about ten million dollars. Considering this substantial amount of money required to influence such a key variable, it becomes more obvious why most poor institutions are unable to considerably improve their retention rates. This finding is consistent with Wyman (1997) who concluded that a huge investment in instruction and other methods of academic support only makes a very small difference in retention rate. With the general reduction in the level of allocable funding for institutions of higher education (McGrath & Braunstein, 1997; Callan, 2002), most institutions, especially the poorer

ones, would be hard-pressed to accommodate the current and future growth in college enrollments. It is important to note that under our federal system, responsibility for education is with the fifty states. The diversity of state higher education structures is mirrored by the diversity of state revenue structures, which are characterized by their low elasticity. This means that state and local revenues from existing taxes do not grow as rapidly as personal income because they are based on taxes on goods sold. From the foregoing, therefore, university administrators have to devise ingenious ways to raise additional funding, if they are to cope with the new demands of enrollment growth.

The influences of unemployment in and around the location of the institution need to be critically examined. From a policy standpoint, the influences of this variable are virtually outside the control of university administrators. As already stated, students increasingly straddle the boundary between wage labor and college (American Council on Higher Education (ACE) 2000) and education's opportunity cost may influence retention. Purdue University (2001) found that 17-35% of Indiana students cite employment as the reason for dropping out. St. John et al. (2001) found that changes in labor markets partially explain changes in retention. In light of the above situations, therefore, institutions need to

explore ways to collaborate with local businesses and governments to improve the employment situation in their locations.

The size of population in the institution's location is equally important and should be taken into consideration in matters of retention rate. We recognize that this variable is virtually outside the control of college administrators. But, by way of a mutually beneficial relationship, college management could work with local governments, businesses, and organizations to try to influence the impact of this variable on institutional retention rates. With the indirect relationship between size of population in the college location and the institution's retention rate, retention rate decreases as population increases around campus location. This finding is consistent with the Little Hoover Commission Report (2000) which concluded that as the diversity and population size of California and other states increase, there is a need to provide relevant, effective educational experiences for students who may otherwise not attend and successfully complete college. Population, however, was not significant in Wyman (1997) study.

Educational institutions should consider going through a process of benchmarking to establish a baseline or standard by which to judge their retention rates. They must determine

where they are and also where they want to be regarding their retention rate. This kind of goal setting means that the institution must first define its acceptable level of student retention, a standard that reflects institutional mission and the unique organizational context or market niche in which the institution is positioned.

College administrators should realize that persistence in higher education is the result of a complex set of interactions that occur over time. And so, it is important for them to adopt a wider perspective in thinking about retention. In fact, using a model to evaluate one retention program is not enough to do the job. To have a meaningful impact on institutional retention and graduation rates, college administrators must focus on developing system-wide strategies. This probably will require a re-evaluation of how they allocate their time and resources, and also how they relate to their external governing bodies.

This study offers an integrative framework in understanding the interplay between institutional and environmental variables in developing institutional retention rates. The results also underscore the need for college administrators to focus on variables, which are highly predictive of retention rate as the target variables to address intervention strategies. It is clear that intervention

strategies would be better off addressing those variables that can be manipulated and which have been found to be the strongest predictors of institutional retention rate.

Many of the studies that form the foundation of our knowledge about retention in higher education assumed the traditional view of students, which is 18-to-22-year-old White, full-time students attending residential colleges, rather than the reality of today's diverse student population (Keller, 2001; Pascarella & Terenzini, 1998). As the demographic characteristics of undergraduate students continue to change (National Center of Education Statistics, 2001), it is likely that older studies that relied on those traditional samples no longer represent the contemporary higher education landscape. Thus, in consideration of the ever-increasing diversity of today's student population, there is a need, therefore, to review the effects of variables on retention and adjust our current mode of practice of student affairs.

In conclusion, more than two-thirds of the variance in retention rate in the University of North Carolina System is explained largely by four variables. Although forces external to the domain of university management influence determine a considerable amount of the variance in retention rate among the institutions, retention rate evidently is somewhat

controllable through government and university fiscal policies.

Recommendations for further research

In the course of this study, certain other related issues that seem worthy of investigation came to the fore. I recommend additional research in the following areas:

Studies should focus on developing a model that will predict retention rate in minority institutions in the University of North Carolina system, if only to serve as comparisons to this particular study. This study is in no way suggesting that the model will have the same power of explaining retention rate variance in both predominantly black and predominantly white institutions. Some of the minority institutions in this study might have other unique factors that differentiate them from the majority institutions. According to the American College Testing Program (1993) freshman-to-sophomore attrition is highest at those four-year institutions with open admissions policies and at historically black colleges and universities. Consequently, these colleges must look for better models to predict their retention rates. Eimers and Pike (1997) studied the differential impacts of variables for both black and white students and found that the academic performance of minority students did not help to

predict their intentions to stay or leave the institution. It appears that many retention variables may have different influences on the retention rates of minority students and institutions. A model that would identify preconditions for attrition could be used in concert with traditional models and might prove helpful in understanding the complex decisions to persist in school and offer possible solutions to increase retention rates in minority institutions.

Efforts should be made to develop a model of retention rate employing only external variables, with a view to learning how much of the variance in retention rate is explained solely by those variables. This knowledge would probably indicate how much institutions are limited in terms of deciding what their retention rate should be. By restricting the predictors to be composed exclusively of external measures and invoking the iterative modeling process, Wyman (1997) believes that the resulting model would maximize the amount of variance explained solely by external measures. Furthermore, the utility of such a model would lie not in its power to explain retention rate variance in general but in its use to remove unwanted external sources of variation from observed retention rate values. A models so derived (Wyman, 1997) suggests would produce measurements that are more representative of actual performance.

It is also suggested that future studies formulate predictive models of retention rate by pooling more than ten years of data where possible and employing more predictor variables. Traditional approaches to inquiry whereby small number of variables are isolated to examine their impacts will no longer suffice. Studies must include as many variables and interactions as possible in order to have a fuller understanding of retention rate in light of the increasingly diverse student population. This might produce a more robust and informative result, as this particular study has proved. While the use of more variables is suggested, a word of caution is important. Care should be taken to keep the number of variables at the optimal level, because as Woodard, Mallory, and De Luca (2001) stated, the researchers may have run full course on trying to explicate high unexplained variance by regressing large sets of numerical variables.

There is a need to develop a model specifically for the North Carolina Community Colleges system that would be sensitive to the unique characteristics of these two-year colleges. Across several decades of research, community college attrition rates have been consistently reported as very high. Early research by Clark (1960) and Thornton (1966) found that more than 40 percent of community college freshmen either did not complete their educational objectives or did

not return for their second year. Later research yielded similar results, finding that only 45 percent of community college first-time, full-time freshmen who intended to earn a degree or certificate graduated in the period from 1998 to 2001 and that 32 percent of students failed to return for their second year at a community college or enroll at another institution of higher education (SREB, 2003). Overall, researchers and practitioners continue to find that community college student dropout rates are significantly higher than those of senior institutions (Mohammadi, 1994).

Defining student success at the community college is as complex and diverse as the students who choose to attend these institutions. Community colleges serve every type of student from the well prepared high school graduate to the under-prepared high school graduate, from the academically gifted to the academically at risk, from the high school student taking a few courses, to the senior citizen interested in personal enrichment. This makes the student body at any given community college remarkably complex (McCabe, 1999), and also creates unique challenges for those interested in trying to understand student persistence and retention in the community college setting. To fully appreciate the challenges community colleges face in terms of defining and influencing student retention, it is first important to understand the unique identity of the

community college in the higher education market. According to (Neutzling, 2003) student retention is in itself a complicated subject, but the dynamic and diverse nature of the community college makes understanding retention rate even more complex. From the foregoing, understanding why students persist and graduate from community colleges is a question that merits exploration.

It is strongly recommended that focus be directed at developing a model for predicting retention rate in private institutions within the state of North Carolina, if only to serve as a comparison to this particular study. Private institutions have differing characteristics from public ones, especially in terms of many state rules and regulations, as well as funding sources. Public institutions receive most of their funding from state appropriations and as such may be less impacted by attrition than private institutions. These differences would likely justify the use of a different model to predict retention rates in private colleges and universities. According to (Pascarella and Terenzini, 1991; Astin, et al., 1996) public and private institutions, when matched with equivalent cohorts, would produce different but higher retention results in public institutions.

With the increasing need to enhance student enrollment in colleges and universities, it becomes necessary to consider

ways to improve our understanding of the impact of out-of-state students in the enrollment pattern of the University of North Carolina system. This effort is important in light of the significance of headcount enrollment in this study as well as the increasing percentage of the nonresident segment of our student body. Out-of-state students generally pay higher tuition fees (Rizzo & Ehrenberg, 2002). Mixon and Hsing (1994) studied the determinants of nonresident enrollments and found that higher nonresident tuition levels were associated with higher nonresident enrollments. Siow (1997) found, after controlling for student body ability, that universities with more successful researchers were more likely to have larger shares of nonresident and foreign students. Geene (1994) studied out-of-state tuition levels at public universities using cross-section data and found that states with many private colleges, lower tax rates, poor labor markets, and with strong in-migration of both population and students, charged higher nonresident tuition. He also found that higher regional tuition was associated with higher nonresident tuition levels but that was not statistically significant. Murtaugh et al. (1999) found that in-state students have lower attrition rates than nonresidents. From the foregoing, further research is needed to explore the impact of nonresident tuition on nonresident student persistence. Murtaugh et al.

(1999) suggests that lower persistence experienced by out-of-state students may be due to the fact that scholarship opportunities for them are more limited and summer orientation programs may be inconvenient for them.

Overall conclusions

For several decades, scholars have been trying to understand and interpret college student retention behavior. Efforts have been made to improve theories that explain college student persistence because attrition exerts some costs to both the individual student and the institution. Furthermore, retention rates of colleges and universities have become important in evaluating their effectiveness as most states now use formulas to determine budgets and appropriations for their public institutions. In light of the increasing importance that retention rate has assumed in recent times, especially as it relates to institutional effectiveness and accountability in higher education, it has become necessary to attempt to gain a fuller understanding of retention rate behavior.

Although considerable progress has been made in understanding individual student retention, significant questions about retention rate have not been conclusively answered. To address such questions, this particular

investigation produced a multiple regression model to predict retention rate in colleges and universities in the University of North Carolina system utilizing institutional-level variables. The study found headcount enrollment, spending on instruction and academic support, size of population and unemployment rate in its location, to be significant predictors of retention rate for an institution.

The model explained about 70 percent of the variance in retention rate in the system. It also answers in the affirmative the general question of whether an educational institution's retention rate is a function of the demographic characteristics, economic conditions, college management and fiscal policy of the state in which the institution is located. This level of explained variance indicates the combined predictive ability of the variables used in this study in determining an institution's retention rate.

This study reveals that the amount an institution can change its retention rate is rather limited. While retention rate varies considerably across institutions, it varies only slightly across time for each institution. An institution that serves a region of low unemployment should be expected to enjoy a higher retention rate than one that serves a region of high unemployment. To increase retention rate, universities must secure adequate appropriations to increase per student

spending on instruction and academic support at a faster rate than average local income. As the population in the institution's location increases, its retention rate usually plummets. Conversely, increasing enrollment levels for an institution boosts its retention rate. In fact, retaining current students longer, perhaps, even to graduation, helps to maintain enrollment levels in general.

Finally, two thirds of the variance in retention rate across the universities in the University of North Carolina System is largely explained by the four variables. Although much of the variance in retention rate among universities is predetermined by forces beyond the sphere of university influence, retention rate is found to be at least partially controllable through government and university fiscal policies.

REFERENCES

- ACT. (2000). National college dropout and graduation rates, 1999. Retrieved from the website <http://www.act.org/new/releases/2000/02-16b00.html>.
- ACT Newsroom. (2001). More first-year college students Return for the second year; fewer students graduate in five years. Retrieved from the website <http://www.act.org/new/releases/2001/04-25-01.html>
- ACT. (2002). College graduation rates steady despite increase in enrollment. Retrieved from the website <http://www.act.org/news/releases/2002/11-15-02.html>
- Akerhielm, K., Berger, J., Hooker, M., & Wise, D. (1998). Factors Related to College Enrollment: Final Report. Princeton, N.J.: Mathtech.
- Allen, D. (1999). Desire to finish college: An empirical link between motivation and persistence. Research in Higher Education, 40(4), 461-486.
- American Association of State Colleges and Universities (1997). Policies and practice: a focus on higher education retention. Washington, DC: AASCU.
- American Council on Education (ACE) (2000). More college students work. Higher Education and National Affairs, 49(12), 4.

- Amisano, G. & Serati, M. (2003). Unemployment persistence in Italy: An econometric analysis with multivariate time varying parameter models. *Luic Papers, Serie Economia e Impresa*, 33 (121).
- Andres, L., & Carpenter, S. (1997). Today's higher education students: issues of admission, retention, transfer, and attrition in relation to changing student demographics. Research report for the British Columbia Council on Admission and Transfer.
- Astin, A. W. (1964). Personal and environmental factors associated with college dropouts among high aptitude students. *Journal of Educational Psychology*, 55, 219-227.
- Astin, A. (1970). The methodology of research on college impact. *Sociology of Education*, 43, 223-254.
- Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25(4), 297-308.
- Astin, A. W. (1993). What matters in college: Four critical years revisited. San Francisco: Jossey-Bass.
- Astin, A. W. (1996). Involvement in learning revisited: lessons we have learned. *Journal of college student development*. 37 (2), 123-134.
- Astin, A. W. (1997). How "good" is your institution's Retention rate? *Research in Higher Education*, 38 (6), 647-658.

- Bean, J. P. (1982). Conceptual models of student attrition. New Directions for Institutional Research, 36, 17-33.
- Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. Review of Educational Research. 55(4), 485-540.
- Becker, W.E. (1990). The demand for higher education. In Hoenack, S.A. & Collins, E.L. (Ed.), The Economics of American Universities: Management, Operations, and Fiscal Environment, State University of New York Press.
- Berger, J. P. (1998). The impact of organizational behavior at colleges on student satisfaction. Paper presented at the annual meeting of the Association for the Study of Higher Education, Miami, FL.
- Berger, J. B., & Braxton, J. M. (1998). Revising Tinto's interactionalist theory of student departure through theory elaboration: Examining the role of organizational attributes in the persistence process. Research in Higher Education, 36(5), 595-612.
- Berry, W. D. (1993). Understanding regression assumptions. Beverly Hills, CA: Sage.
- Bogue, E. G. (1998). Quality Assurance in Higher Education: The Evolution of Systems and Design Ideals. New Directions for Institutional Research 25(3), 7-18.

Blose, G. (2000). Modeled retention and graduation rates:
Calculating expected retention and graduation rates for
multi-campus universities. In Gerald Gaither (ed.), The
New Three R's: Recruitment, Remediation and Retention,
San Francisco: Jossey-Bass.

Bonwell, C. C. & Eison. J. A. (1991). Active Learning:
Creating excitement in the classroom. (ASHE-ERIC Higher
Education Report No. 1) Washington, D.C. The George
Washington University.

Brawer, F.B. (1996). Retention-attrition in the nineties.
(ERIC Document Reproduction Service No. ED393510)

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

Braxton, J. M. & Lien, L. A. (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 departure
puzzle. Nashville, TN: Vanderbilt University Press.

Burke, J .C. & Minassians, H. (2003). Performance
reporting: "real" accountability or accountability
"lite": The Seventh Survey. Higher Education Program:
The Nelson A. Rockefeller Institution of Government.

Burke, J. C. & Serban, A. M. (1998). Current status and future prospects of performance funding and performance budgeting for higher education: The second survey. Albany, NY: Nelson A. Rockefeller Inst. of Government. (Eric Document Reproduction Service No. ED 423727).

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

Cabrera, A. F., Castaneda, M.B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. Journal of Higher Education, 63(2), 143-164.

Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. Journal of Higher Education, 64(20), 123-138.

Chickering, A. W. & Hannah, W. (1969). The process of Withdrawal. Liberal Education, 55, 551-558.

Commission on Institutions of Higher Education. (1992). Policy on Institutional Effectiveness, Bedford, MA.

Consortium for Student Retention Data Exchange (CSRDE) (2001). Executive summary. 2000-2001 CSRDE report. The retention and graduation rates in 344 colleges and universities. Retrieved from the website <http://www.occe.ou.edu/csrde/reports.htm>

- Consortium for Student Retention Data Exchange (CSRDE)
(2002). Executive summary. 2002 CSRDE report. The
freshman to sophomore year CSRDE benchmark retention
rate. Retrieved from the website
<http://www.udel.edu/IR/reports/persisters>
- Denis, M. J. (1998). A practical guide to enrollment and
retention management in higher education. Westport, CT:
Bergin and Garvey.
- DesJardins, S. L., Ahlburg, D.A., & McCall, B. P. (1999).
An event history model of student departure. Economics
of Education Review, 18(3), 375-390.
- Dielman, T. E. (1983). Pooled cross-sectional and time
series data: A survey of current statistical methodology.
The American Statistician, 37(2) 111-122.
- Durkheim, E. (1897/1966). Suicide, a study in sociology.
Translated by Spaulding, J. A. & Simpson, G. (Ed.).
New York: Free Press.
- Early College High School Initiative (2003). Why early
college high schools are necessary. Retrieved from the
website: <http://www.earlycolleges.org>
- Eaton, S. B., & Bean, J.P. (1995). An approach/avoidance
behavioral model of college student attrition. Research
in Higher Education, 36, 617-647.

Eimers, M.T. & Pike, G.R. (1997). Minority and nonminority adjustment to college: Differences or similarities? Research in Higher Education, 38 (1), 77-97.

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

Employment Policy Foundation (2004). Jobs report confirms continued labor market recovery. Retrieved from the website: <http://www.epf.org/pubs/snapshots/2004>

Feldman, K. A. (1970). Some methods for assessing college impacts. Sociology of Education, 44, 133-150.

French, B. F., Immekus, J. C., & Oakes, W. (2003). A structural model of engineering students success and persistence. Paper presented at the annual Frontiers of Education Conference.

Georges, A. (2000). Keeping what we've got: the impact of financial aid on minority retention in engineering. Hispanic Times Magazine. 23, (5).

Grayson, J.P, & Grayson, K. (2003). Research on retention and attrition. Canadian Millennium Scholarship Foundation: Montreal, Canada

Green, S. B. (1991). How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26, 499-510.

- Hearn, J.C. & Longanecker, D. (1985). Enrollment effects of alternative postsecondary pricing policies. Journal of Higher Education, 56, 485-508.
- Heilbrun, A. B., Jr.(1965). Personality factors in college dropouts. Journal of Applied Psychology, 49, 1-7.
- Hicks, A. (1994). Introduction to pooling. In Janoski, T. & Hicks, A. (Ed.), The comparative political economy of the welfare state, Cambridge University Press.
- Huesman, R., Moore, J., Druva-Roush, C., Wang, X., & Huang, C. (1994). A Comparison of decision theoretic approaches to studying student retention. AIR 1994 Annual Forum Paper, New Orleans, Louisiana.
- Indiana Economic Development Council (2004). Strategic planning for the 21st century. Retrieved from the website: <http://www.iedc.org>
- Integrated Postsecondary Education Data System (IPEDS) Reports (2000). University of North Carolina General Administration (UNC-GA).
- Johnson, J. & Di Nardo, J. (1998). Econometric methods, Singapore: McGraw-Hill.
- Jones, B. & Moss, P. (1994). The influence of financial aid on academic performance in medical school. Journal of Student Financial Aid, 24 (3), 5-11.

- Kane, T.J. & Rouse, C.E. (1999). The community college: Educating students at the margin between college and work, Journal of Economic Perspectives, 13, (1), 63-84.
- Keller, G. (2001). The new demographics of higher education. The Review of Higher Education, 24, 219-236.
- Kemp, P. A., Lynch, E., & Mackay, D. (2001). Structural trends in homelessness: A quantitative analysis. Scottish Executive Central Research Unit. Homelessness Task Force.
- Lau, L. K., (2003). Institutional factors affecting student retention. Education (Fall).
- Leslie, L.L. & Brinkman, P.T. (1987). Student price response in higher education: the student demand studies. Journal of Higher Education, 58, 181-204.
- Levitz, R. S., Noel, L., & Richter, B. J. (1999). Strategic moves for retention success. In G. H. Gaither (ed.), Promising practices in recruitment, remediation, and retention (pp. 31-50). (New Directions for Higher Education, n. 108) San Francisco: Jossey-Bass.
- Liu, E. & R. Liu (1999). An application of Tinto's model at a commuter campus. Education, 119, 537-41.
- Luna, A. (1999). Using a student flow matrix model to identify retention and attrition patterns. SAIR Conference Paper, Chattanooga, Tennessee.

Mayes, L., Williford, L., & Ussery, R. (2000) Policy Implications of using prediction models of student retention and graduation rates. Paper presented at the Association of Institutional Research meeting, Cincinnati.

McGrath, 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, 396-408.

McKeown, M. P. (1996). State funding formulas for public four-year institutions. Denver, CO: State Higher Education Executive Officers.

McLaughlin, G., Brosovsky, P., & McLaughlin, J. (1998). Changing perspectives on student retention: a role for Institutional research. Research in Higher Education, 39 (1), 1-17.

Mohanty, M. & Finney, M. (1997). Evidence on the effect of young adults' wages on their college attendance decisions. Applied Economics Letters, 4, 733-735.

Mohammadi, J. (1996). Exploring retention and attrition in a two-year public community college. VCCA Journal, 10 (1), 39-50.

Mortenson, T. (1999). Unmet and overmet financial need of undergraduate students. Postsecondary Education Opportunity, 88, 1-10

- Murtaugh, P.A., Burns, L.D., & Schuster, J. (1999).
Predicting the retention of university students. Research
in Higher Education, 40(3), 355-371.
- National Center for Education Statistics (1996). The
Condition of Education. Washington, D.C.: U.S. Government
Printing Office.
- National Center for Education Statistics (2001). The
Condition of Education. Washington, D.C.: U.S. Government
Printing Office.
- National Center for Education Statistics. (October, 2002).
High school graduates (U.S. Department of Education
Office of Educational Research and Improvement/NCES 03-
013). Washington DC: U.S. Government Printing Office
- National Center for Education Statistics. (October, 2003).
High school graduates (U.S. Department of Education
Office of Educational Research and Improvement/NCES 03-
013). Washington DC: U.S. Government Printing Office
- National Center for Education Statistics. (October, 2004).
High school graduates (U.S. Department of Education
Office of Educational Research and Improvement/NCES 03-
013). Washington DC: U.S. Government Printing Office
- National Center for Education Statistics. (June, 2004).
Postsecondary persistence and progress (U.S. Department
of Education Office of Educational Research and
Improvement/NCES 03-013). Washington DC: U.S. Government
Printing Office

North Carolina State University (NCSU) (2004). North Carolina State University: Mission statement.
<http://www.ncsu.edu/planning.html>

Pascarella, E.T. & Terenzini, P.T. (1980). Predicting Freshman persistence and voluntary dropout decisions from a theoretical model. Journal of High Education, 51, 60-75.

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

Pascarella, E. T., & Terenzini, P. T. (1998). Studying college students in the 21st century: Meeting new challenges. The Review of Higher Education, 21, 151-165

Peltier, G.L., Laden, R., & Matranga, M. (1999). Student Persistence in college: A review of research. Journal of College Student Retention, 1, 357-376.

Penn, G. (1999). Enrollment management for the 21st century: Delivering institutional goals, accountability and fiscal responsibility (ASHE-ERIC/Higher Education Report Series No. 26-7). Washington, D C: Association for the Study of Higher Education. (ERIC Document Reproduction Service ED 432939).

Pennings, P., Keman, H., & Kleinnijenhuis, J. (1999). Doing research in political science: An introduction to comparative methods and statistics, Sage.

- Perna, L. W. (1998). The contribution of financial aid to undergraduate persistence. Journal of Student Financial Aid, 28 (3), 25-40.
- Picklesimer, B.K. & Miller, T.K. (1998). Life-skills development inventory-college form: An assessment measure. Journal of College Student Development, 39(1), 100-110.
- Pindyck, R.S. & Rubinfeld, D.L. (1998). Econometric Models and Economic Forecasts, Singapore: McGraw-Hill.
- Podesta, F. (2000). Recent developments in quantitative Comparative methodology: The case of pooled time series cross-section analysis. DSS Papers SOC 3-02.
- Pogue, J. H. (2000). A comparative analysis of graduation rates of African American students at historically black colleges and universities and predominantly White institutions. Doctoral dissertation, The University of Arizona.
- Porter, K. (2002). The value of a college degree ERIC/Higher Education Report Series, Washington, D C: Association for the Study of Higher Education. (ERIC Document Reproduction Service ED 470038).
- Purdue University. (2001). Purdue University lilly retention proposal. Available:
<http://www.purdue.edu/Retention/pages/reports/proposal/proposal 1.html>

- Reason, R. D. (2003). Student variables that predict retention: Recent research and new developments. NASPA Journal, 40(4), 172-191.
- Roach, R. (1999). Succeeding on white campuses. Black Issues in Higher Education, 15,42-43.
- Ryland, E., Riordan, R., & Brack, J. (1994). Selected Characteristics of high-risk students and their enrollment persistence. Journal of College Student Development, 35, 54-58.
- Sanders, L., & Burton, J. D. (1996). From retention to Satisfaction: new outcomes for assessing the freshman Experience. Research in Higher Education, 37 (3), 555-567.
- Sandler, M. (2000). Career decision-making self-efficacy, perceived stress, and an integrated model of student persistence. Research in Higher Education, 41(5), 537-580.
- Sayrs, L. W. (1989). Pooled time series analysis, London: Sage.
- Schmidt, M. G. (1997). Determinants of social expenditure in liberal democracies. Acta Politica, 32(2), 153-173.
- Shelley, K. J. (1992). The future of jobs for college Graduates. Monthly Labor Review. 115(7), 13.

Snyder, T. D., & Hoffman, C. M. (1995). State comparisons of education statistics: 1969-70 to 1993-94. (NCES Report No. 95-122). Washington, DC: U.S. Department of Education.

Southern Association of Colleges and Schools. (1995). Criteria for Accreditation: Commission on Colleges. Decatur, GA: Southern Association of Colleges and Schools.

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

State Council of Higher Education of Virginia. (2004). Virginia Higher Education Performance Funding Model.

State University of New York. (1998). Task Force Report on Performance Indicators and Merit-Based Performance Funding. July 6.

Statistical Abstracts of Higher Education in North Carolina (2000). UNC-GA Progress/ABS AR002

Stedman, J. B. (2002). Higher Education Act: Reauthorization Status and Issues. Congressional Research Service.

Stimson, J. A. (1985). Regression in space and time: A statistical essay. American Journal of political sciences, 29 (4), 914-947.

- St. John, E. P., Paulsen, M. B., & Starkey, J. B. (1996). The nexus between college choice and persistence. Research in Higher Education, 37(2), 175-220.
- St. John, E.P., Hu, S. & Weber, J. (2001). State policy and affordability of public higher education: The influence of state grants on persistence in Indiana. Research in Higher Education 42(4), 401-428
- St. John, E. P. (2000). The impact of student aid on recruitment and retention: What the research indicates. New Directions for Student Services, 89, 61-76.
- Sydow, D. L., & Sandel, R. H. (1998). Making student Retention an institutional priority. Community College Journal of Research and Practice, 22(7), 637.
- Tabachnick, B. G., & Fidell, L. S. (2001). Using Multivariate Statistics (4th ed.). Needham Heights, MA: Allyn and Bacon.
- Thomas, R. L., (1997). Modern econometrics: An introduction. Harlow: Addison-Wesley.
- Thomlinson-Clarke, S., Clarke, D. (1996). Institutional effects on alienation and student effort at three campus environments. Journal of College Student Development, 37(1), 60-67.

- Ting, S.R. (1998). Predicting first-year grades and Academic progress of college students of first-generation and low-income families. Journal of College Admission, 158, 14-23.
- Tinto, V. (1975). Dropout from higher education: A Theoretical synthesis of recent research. Review of Educational Research, 45(1), 89-125.
- Tinto, V. (1993). Leaving college: rethinking the causes and cures of student attrition. Chicago: The University of Chicago Press.
- Tinto, V. (1998). Colleges as communities: Taking research on student persistence seriously. The Review of Higher Education, 21, 167-177.
- University of North Carolina, Chapel Hill (UNC-CH) (2004). University of North Carolina: Mission statement. <http://unc.edu/about/mission.html>
- University of North Carolina General Administration (UNC-GA) (2000). Ten retention rates for schools in the study.[http://www.northcarolina.edu/content.php/assessment/reports/student info/retention.htm](http://www.northcarolina.edu/content.php/assessment/reports/student%20info/retention.htm)
- United States Department of Education. National Center for Education Statistics. (1998). Stopouts or stayouts? Undergraduates who leave college in their first year. Washington, DC. (NCES Publication No. 1999-087)

United States Department of Education. (1999). Digest of Education statistics. Washington, DC: National Center For Education Statistics.

United States Department of Education. (2000). Digest of Education statistics. Washington, DC: National Center For Education Statistics.

U.S. Department of Commerce, Bureau of the Census. (1998). Economics and Statistics Administration.

U.S. Department of Commerce, Bureau of Economic Analysis. (2000). Economics and Statistics Administration.

U.S. Department of Commerce, Bureau of the Census. (2002). Economics and Statistics Administration.

Vaughan, R. P. (1968). College dropouts: Dismissed vs. withdrew. Personnel and Guidance Journal, 46, 985-989.

Wetzel, J. N., Dennis, O., & Peterson, S. (1999). Factors affecting student retention probabilities: a case study. Journal of Economics and Finance. 23, 45-55.

White, W. F., & Mosely, D. (1995). Twelve-year pattern of retention and attrition in a commuter type university. Education, 115 (3), 400-402.

Woodard, D. B., Mallory, S. L., DeLuca, A. M. (2001).

Retention and institutional effort: A self-study
framework, 39(1), 53-83.

[http://publications.naspa.org/naspajournal/vol39/iss1/art
4](http://publications.naspa.org/naspajournal/vol39/iss1/art4)

Wyman, F.J.(1997,Summer). A predictive model of retention
rate at regional two-year colleges. Community College
Review, 25, (1), 29-59.

APPENDICES

APPENDIX A
SAS Regression Output of the Full Model

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: retrate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	4950.57628	825.09605	58.66	<.0001
Error	143	2011.23045	14.06455		
Corrected Total	149	6961.80673			

Root MSE	3.75027	R-Square	0.7111
Dependent Mean	77.67133	Adj R-Sq	0.6990
Coeff Var	4.82839		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	79.66424	1.40221	56.81	<.0001
hecenrol	1	0.45763	0.07588	6.03	<.0001
egiasexp	1	0.01236	0.00464	2.66	0.0086
avgfin	1	-0.00102	0.00058476	-1.74	0.0832
unemplat	1	-1.21140	0.19686	-6.15	<.0001
populatn	1	-0.01007	0.00202	-4.99	<.0001
crime	1	-0.00385	0.00423	-0.91	0.3642

APPENDIX B
SAS Regression Output of the Thirteen Schools

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: retrate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	1033.60887	172.26814	12.72	<.0001
Error	123	1665.97736	13.54453		
Corrected Total	129	2699.58623			

Root MSE	3.68029	R-Square	0.3829
Dependent Mean	75.62538	Adj R-Sq	0.3528
Coeff Var	4.86647		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	81.36611	1.49775	54.33	<.0001
hecenrol	1	0.29213	0.08944	3.27	0.0014
egiasexp	1	-0.00289	0.00677	-0.43	0.6703
avgfin	1	-0.00150	0.00060273	-2.49	0.0139
unemplmt	1	-1.19558	0.19561	-6.11	<.0001
populatn	1	-0.00838	0.00216	-3.89	0.0002
crime	1	-0.00106	0.00420	-0.25	0.8016

APPENDIX C
SAS Regression Output: First 5 Years (1991-1995)

The SAS System

The REG Procedure

Model: MODEL1

Dependent Variable: retrate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2417.05792	402.84299	28.08	<.0001
Error	68	975.55754	14.34643		
Corrected Total	74	3392.61547			

Root MSE	3.78767	R-Square	0.7124
Dependent Mean	77.83733	Adj R-Sq	0.6871
Coeff Var	4.86613		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	81.61250	2.21602	36.83	<.0001
hecenrol	1	0.32388	0.20014	1.62	0.1102
egiasexp	1	0.00397	0.00718	0.55	0.5822
avgfin	1	-0.00256	0.00148	-1.72	0.0896
populatn	1	-0.00932	0.00303	-3.08	0.0030
unemplmt	1	-1.24005	0.31369	-3.95	0.0002
crime	1	0.05359	0.05845	0.92	0.3625

APPENDIX D
SAS Regression Output: Last 5 Years (1996-2000)

The SAS System

The REG Procedure
 Model: MODEL1
 Dependent Variable: retrate

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	2765.85177	460.97529	39.22	<.0001
Error	68	799.20610	11.75303		
Corrected Total	74	3565.05787			

Root MSE	3.42827	R-Square	0.7758
Dependent Mean	77.50533	Adj R-Sq	0.7560
Coeff Var	4.42327		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	79.06258	1.76143	44.89	<.0001
hecenrol	1	0.33731	0.11562	2.92	0.0048
egiasexp	1	0.01959	0.00645	3.04	0.0034
avgfin	1	0.00012618	0.00069422	0.18	0.8563
populatn	1	-0.00954	0.00261	-3.66	0.0005
unemplmt	1	-1.57892	0.26604	-5.93	<.0001
crime	1	-0.00180	0.00394	-0.46	0.6491