

To the Graduate Council:

I am submitting herewith a dissertation written by Ken Rayburn Adcock entitled "Predicting Student Cumulative Grade Point Averages at the University of Tennessee for Graduates of Tennessee Public High Schools." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy with a major in Education.

C. Glennon Rowell
C. Glennon Rowell, Major Professor

We have read this dissertation
and recommend its acceptance:

Mary Jane Connelly
LeRoy J. Bellon
Robert K. Roney
Lyn Davis

Accepted for the Council:

C. W. Muntel
Associate Vice Chancellor
and Dean of The Graduate School

PREDICTING STUDENT CUMULATIVE GRADE POINT AVERAGES
AT THE UNIVERSITY OF TENNESSEE FOR GRADUATES OF
TENNESSEE PUBLIC HIGH SCHOOLS

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Ken Rayburn Adcock

May, 1992

Copyright 1991 by Ken R. Adcock

All rights reserved

* SAS, SAS/IML, SAS/PC, and SAS/STAT are registered
trademarks of SAS Institute Inc.

ACKNOWLEDGEMENTS

The author expresses gratitude and appreciation to those who assisted in the preparation of this dissertation:

To Dr. Glennon Rowell, chairman of my committee, who provided guidance, direction, and encouragement in completing this dissertation. His willingness to chair my committee will always be appreciated and remembered.

To Drs. Jerry Bellon, Mary Jane Connelly, Lynn Ourth, and Robert Roney, for serving on my committee and offering much needed support, suggestions, and constructive comments.

To Dr. Dan Quarles, for serving as my chairman during the beginning of my dissertation. His personal guidance and efforts in obtaining support for this research will always be remembered.

To Dr. Robert Levy, Associate Vice President for Academic Affairs and Research at the University of Tennessee, for providing the support and authority for obtaining student record information.

To Mr. Larry Stiles of the University of Tennessee Student Affairs Division, for providing computerized student record information.

To Mr. Don Cunningham, the University of Tennessee Office of Institutional Research, for his help and advice regarding the many pitfalls and nuances regarding the use and availability of student records.

To Dr. Fred Crosson of the Tennessee Education Association, for providing access and assistance in the use of TEA's research library.

To Dr. Don Yates, the Superintendent of Cleveland City Schools, for making available his very complete library.

To my special friends, Ken Griffith, Steve Reynolds, and Bill Forrester, who have encouraged me along the way.

To my friends at AGC Corporation, for their help and encouragement. Very special appreciation is owed to Bob George, Jim Davis, Chris Vernon, Jane Green, and Kathy Gillenwaters.

And especially to my family. My wife Charlotte and my sons Ray and Jonathan. They have often doubted my sanity, but they have always supported my efforts.

ABSTRACT

Multiple linear regression models were produced to predict the collegiate academic success of graduates of Tennessee public high schools from 1986 through 1989. Data from the University of Tennessee student records system were merged with data supplied by the Tennessee State Department of Education to develop the statistically based equations.

The research data set contained records of 10,327 graduates of Tennessee public high schools. These students subsequently entered the University of Tennessee located in Knoxville in the fall or summer immediately following graduation. Stepwise regression models were specified for the respective response variables; freshman cumulative GPA and fourth semester sophomore cumulative GPA. The independent variables considered were ACT composite scores, high school GPA, whether or not the student's high school was accredited by the Southern Association of Colleges and Schools (SACS), the percentage of students on free or reduced lunches in the school system, school system adjusted expenditure per pupil, school system enrollment, school system twelfth grade graduation rates, and whether or not the student received financial aid while in college.

The stepwise regression technique used in the model for predicting first semester cumulative GPA eliminated many of the candidate variables leaving high school GPA, ACT composite scores, the percentage of students on free or reduced lunches

for the school system, school system adjusted expenditure per pupil, and whether or not the high school was accredited by SACS. The five variables accounted for about 29% of the variance in freshman cumulative GPA.

The regression equation constructed for the last semester of the sophomore year included all of the independent variables stated for the earlier model plus an indicator for whether or not the student received collegiate financial aid and the enrollment size of the students' public school system. The six independent variables accounted for over 37% of the variation in the fourth semester sophomore year cumulative GPA.

It was found that for the given data set, high school GPA and ACT composite scores provided a more accurate prediction equation than could be accomplished with a single variable. However, the predictability contribution of the high school GPA was far superior to the ACT composite scores or to any other combination of variables.

TABLE OF CONTENTS

CHAPTER	PAGE
1. INTRODUCTION	1
Statement of the Problem	2
The Purpose	2
Significance	2
Definition of Terms	5
Research Questions	6
Limitations and Delimitations	8
Procedures	9
Organization of the Study	11
2. A REVIEW OF RELATED LITERATURE	13
Historical Overview	13
Topics Relevant to Current Research	13
Critique of Current Status of Research	23
Contribution of Research to the Literature	23
3. METHODOLOGY	25
Introduction	25
Statistical analysis	28
Assumptions	33
4. RESULTS AND ANALYSIS OF THE DATA	35
Data Preparation and Validation	35
Findings	38
Unanticipated Results	86
5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS.....	88
Summary	88
Conclusions	99
Recommendations	100
LIST OF REFERENCES	103
VITA	109

LIST OF TABLES

	PAGE
3.1 GNP Annual Price Deflators and Adjusted Academic Year Price Deflators	34
4.1 High School GPA by Year for First-Time Freshmen as Reported by the University of Tennessee as Compared to High School GPA for First-Time Freshmen Who are Graduates of Tennessee Public High Schools	36
4.2 Composite ACT scores by Year for First-Time Freshmen as Reported by the University of Tennessee Compared to High School Composite ACT Scores for First-Time Freshmen Who are Graduates of Tennessee Public High Schools	36
4.3 Number of First-Time Freshmen and Their Respective High Schools and School Systems by Year	39
4.4 Descriptive Statistics for Freshman First Semester GPA by High School Graduation Year	39
4.5 Descriptive Statistics for SACS High School Accreditation by High School Graduation Year	41
4.6 Descriptive Statistics for ACT Composite Scores by High School Graduation Year	41
4.7 Descriptive Statistics for High School GPA by High School Graduation Year	43
4.8 Descriptive Statistics for Percentage of Students on Free or Reduced Lunches by High School Graduation Year	43
4.9 Descriptive Statistics for Adjusted Expenditures Per Pupil by High School Graduation Year	44
4.10 Descriptive Statistics for School System Enrollment by High School Graduation Year	46
4.11 Descriptive Statistics for School System Twelfth Grade Percentage Enrollment by High School Graduation Year	46
4.12 Descriptive Statistics for Students Receiving Financial Aid by High School Graduation Year	48

4.13	Simple Correlation Coefficients: Freshman Cumulative GPA Versus Other Independent Variables	49
4.14	Regression Analysis for First Semester Freshman Year Cumulative GPA Model	53
4.15	Regression Parameter Estimates for First Semester Freshman Model by Year	56
4.16	Descriptive Statistics at End of Sophomore Year: Cumulative GPA by High School Graduation Year	59
4.17	Descriptive Statistics at End of Sophomore Year: SACS Accredited High Schools by High School Graduation Year	59
4.18	Descriptive Statistics at End of Sophomore Year: ACT Composite Scores by High School Graduation Year ..	60
4.19	Descriptive Statistics at End of Sophomore Year: High School GPA by High School Graduation Year ...	60
4.20	Descriptive Statistics at End of Sophomore Year: Percentage of Students on Free or Reduced Lunches by High School Graduation Year	62
4.21	Descriptive Statistics at End of Sophomore Year: Adjusted Expenditures Per Pupil by High School Graduation Year	62
4.22	Descriptive Statistics at End of Sophomore Year: Students Receiving Financial Aid by High School Graduation Year	63
4.23	Descriptive Statistics at End of Sophomore Year: School System Enrollment by High School Graduation Year	65
4.24	Descriptive Statistics at End of Sophomore Year: Twelfth Grade Enrollment as a Percentage of School System Enrollment by High School Graduation Year	65
4.25	Correlation Coefficients: Sophomore Cumulative GPA Versus Other Independent Variables	66
4.26	Regression Analysis for End of Sophomore Year Cumulative GPA Model	70
4.27	Regression Parameter Estimates for End of Sophomore-Year-Model by High School Graduation Year.....	73

4.28 Analysis of Statistical Difference in First Semester Cumulative GPA and End of Sophomore-Year Cumulative GPA	76
4.29 Analysis of Persistence by SACS High School Accreditation	79
4.30 Analysis of Persistence by ACT Composite Scores	79
4.31 Analysis of Persistence by High School GPA	80
4.32 Analysis of Persistence by Percentage of Students In School System on Free or Reduced Lunches	80
4.33 Analysis of Persistence by Adjusted Expenditures Per Pupil	81
4.34 Analysis of Persistence by Students Receiving Financial Aid During Their Freshman Year	81
4.35 Analysis of Persistence by School System Enrollment	82
4.36 Analysis of Persistence by Students in Twelfth Grade as a Percentage of Total School System Enrollment .	82
4.37 First-Time Freshmen Supplied by the Top Feeder High Schools and Schools Systems by Year	87

CHAPTER 1

INTRODUCTION

Colleges and universities are very interested in recruiting and retaining students who will excel academically and who are motivated to remain in school until they graduate. Likewise, high schools strive to provide students with the requisite skills necessary to be successful at the collegiate level. Predicting academic performance has been a difficult, if not a seemingly impossible, task for many years. Numerous variables and complicating factors enter into each student's life that affect his/her ability not only to remain in school but also to achieve academic levels at his/her individual potential.

Admission into college is based upon the hope and anticipation of academic success. Are there a set of variables that will contain the components of collegiate readiness and will perhaps provide insight into the selection process? More specifically, are there variables that are readily available through The Tennessee State Department of Education and from the University of Tennessee that can be used for constructing and utilizing academic prediction methodologies without having to resort to lengthy and costly surveys and data collection schemes?

STATEMENT OF THE PROBLEM

The University of Tennessee system has been providing data for the past three years to Tennessee high schools regarding the collegiate first semester performance of their graduates. These results are of interest to both the higher education institutions and to the secondary school systems. High drop out rates and poor performances of many students often are perplexing to school administrators. The cost to universities to recruit and retain students continues to increase. The inability to predict academic achievement of college bound students is an ongoing problem that can be made less difficult with systematic research.

THE PURPOSE

The purpose of this study was to construct a mathematically based model that could be used to predict student cumulative grade point average (GPA) at the University of Tennessee for graduates of Tennessee public high schools.

SIGNIFICANCE

Significance can be addressed from the students, parents, high school, and university perspectives. First, public K-12 schools may be forced into proving that they can more aptly prepare students for college than their sister public institutions in the surrounding geographic area. The "freedom to choose" movement advocates that parents and students should

not be forced to attend a particular school but rather should have the right to choose among a set of potential candidate institutions. Even more serious problems will be encountered if parents are provided vouchers that will be honored by not only other public schools but also by the private school sector. Public high schools could conceivably lose their monopoly position and rightly or wrongly could be forced to compete for students in a free market economy.

Public school systems have been criticized heavily due to the public's perceived failure for them to produce students who are prepared for not only college but the work place. Coleman and Hoffer (1987) in Public and Private High Schools: The Impact of Communities state that graduates of Catholic high schools out perform their contemporaries from public schools. They attribute this primarily to the self discipline that is instilled into Catholic school graduates which subsequently results in better study habits and higher grades. However, Ravitch (1985) in The Schools We Deserve: Reflection on the Educational Crises of our Time stated that public schools should view this report as good news since it appears that schools can help determine educational achievement despite the family backgrounds of students. Freedom to choose and voucher systems are each emotional and complex issues, but either can place pressure on public high schools to demonstrate that their graduates can be successful in higher education. Representatives of higher education on the other

hand would like to spend their time and resources attracting and retaining students who can successfully complete their chosen undergraduate program. The majority of studies in recent years have emphasized instructional and curriculum processes and individual student characteristics rather than attempting to identify school system classifications and broad indicators of student performance. Earlier studies have generally been restricted to identifying relationships with college freshmen and other secondary school variables. This study provided an opportunity for a longitudinal look over a three-year time period and eliminated short term influences.

In addition, the entire population of Tennessee high school seniors who elected to attend the University of Tennessee located in Knoxville was examined over these years. This study should be of interest to other southern state universities which depend upon public high schools for a majority of their student populations.

Hopefully, the research provides some answers and points the way for additional study to make the higher education experience more meaningful and successful for not only the student, but also for the institutions that are charged with providing education to the student population.

DEFINITION OF TERMS

ACT (The American College Testing Program) of Iowa City, Iowa administers the ACT Assessment test to high school juniors and seniors throughout the nation. The test "is a comprehensive evaluative, guidance, and placement program used by over a million college-bound students each year" (Noble, 1990, p. 3).

Expenditures per pupil is the total dollar expenditures in a school district divided by the average daily attendance. The total dollar expenditure is the sum of the current operating expenditures, administration cost for the State Department of Education, monies for the Retirement and Social Security of teachers in grades K-12, and monies from the U. S. Department of Agriculture (Annual Statistical Report of the Department of Education, 1987, July, p. 32).

GNP (Gross National Product) is the total market value of goods and services produced in the U.S. during a calendar year (Amos, 1987, p. 766).

GPA (Cumulative Grade Point Average) is the measurement used in this research to express academic achievement in high school and college. It is based upon a 4.0 scale.

SACS (Southern Association of Colleges and Schools) is the organization that accredits elementary schools, secondary schools, and colleges in the eleven states of Alabama, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Texas, and Virginia.

SAS (Statistical Analysis System) is the statistical software package that was used to process the data and to calculate the statistical results used in this research. SAS is a product of SAS Institute, Inc. of Cary, NC.

SAT (Scholastic Aptitude Test) is the test "sponsored by the College Entrance Board, administered by the Educational Testing Service, taken by nearly 1.5 million persons a year, and used in admissions by more than 1,500 colleges and universities" (Crouse & Trusheim, 1988, p. xi).

SES (socioeconomic status) is a socioeconomic index used to categorize individuals and groups into social classes (Hess, Markson, & Stein, 1988, p. 189). There is not a universal standard SES index that is used by researchers. This investigation utilized the percent of students on free or reduced lunches within a school district as the SES variable.

TSDOE (Tennessee State Department of Education) located in Nashville is the state agency empowered to implement and monitor the state's education policies.

UTK Refers to the University of Tennessee located in Knoxville.

RESEARCH QUESTIONS

The following independent variables were identified: school system annual expenditure per pupil, percentage of students in school system on free or reduced lunch, school system twelfth grade enrollment size, total school system

enrollment, whether or not the students' high school is accredited by SACS, students' high school GPA, students' composite ACT score, and whether or not the student is receiving financial aid in college.

1. Is there a statistically significant relationship among any of the selected independent variables and students' first semester cumulative GPA?

2. Is there a regression model that utilizes the stated independent variables that will predict first semester GPA for a student?

3. Is there a statistically significant relationship among any combination of the independent variables and students' fourth semester cumulative GPA?

4. Is there a regression model that utilizes the stated independent variables that will predict fourth semester GPA for a student?

5. Is there any statistically significant difference in students' fourth semester cumulative GPA and their first semester cumulative GPA?

6. Is there any statistically significant difference in any of the independent variables used to predict first semester GPA and the independent variables used to predict fourth semester cumulative GPA?

LIMITATIONS AND DELIMITATIONS

This research must be viewed as descriptive in nature and should be used as prescriptive only after very careful consideration. Regression equations result from fitting equations to data where the data values are within discrete ranges. The model is then useful in predicting a response variable when values for the independent variables are within the range restrictions (bounds) of the data that was used to produce the response function. For example, if a regression equation had adjusted expenditures per pupil as an independent variable and the values used to calculate the parameter estimate for adjusted expenditure per pupil had a minimum value of \$1,500 and a maximum value of \$4,000, then the regression model is valid only when a value is substituted into the equation for adjusted expenditure per pupil that is within these bounds. Thus, it would not be mathematically correct if an analyst attempted to use the model with a value of \$7,000 for adjusted expenditure per pupil.

The investigation included only students at the University of Tennessee at Knoxville who graduated from locally funded and supported Tennessee public high schools. This restriction eliminated students from private secondary schools, vocational schools, university-sponsored secondary schools, correctional institutions and the various schools for the blind. Students who elected to attend private and public institutions of higher education other than the University of

Tennessee located at Knoxville were not considered in any of the analyses. Data were used for high school seniors who graduated in the spring of 1986 through 1989 and who entered the University of Tennessee the subsequent summer or fall. This provided a minimum of two years of university history for students entering immediately upon graduating from high school and persisting for four contiguous semesters. No attempt was made to identify or rank individual high schools or school districts in the investigation.

Attempting to examine academic achievement at the end of the sophomore year with the same basic set of explanatory variables that were used in predicting first semester academic achievement carries some risk. With the exception of whether or not a student receives financial aid, the variables all occur prior to entering college. A myriad of events and influences impact the student after beginning college. The contribution of these factors on academic achievement are unknown. Thus, the maturation threat to the internal validity of the research can not be ignored (Gay, 1987, p. 313).

PROCEDURES

Data were collected from several sources. Information regarding school systems and secondary schools were obtained primarily from publications of the Tennessee State Department of Education. School system data included an identification number of each local school system, the percentage of students

in each system on free or reduced lunches, the annual expenditures per pupil, twelfth grade enrollment size, and the school system enrollment. Individual high school data included an identification number for each high school and whether or not the school was accredited by the Southern Association of Colleges and Schools (SACS). Warren, Innis, and Quarles (1977) identified these and several other categories as useful for examining the performances of Tennessee public schools.

Two types of data were obtained from The University of Tennessee at Knoxville. The first type was data used to classify and categorize the student. Necessary data were graduating high school, ACT Composite scores, high school GPA, and any pertinent financial aid data. The second type was data used to track persistence and academic advancement of each student. Data included term identification code, academic year, term semester hours taken/passed, semester GPA, total semester hours taken/passed, and cumulative GPA for the term.

Computer files containing public school system and high school data were manually constructed from State Department of Education source documents. The University of Tennessee data were available on computer readable media and were manipulated as necessary to match the collegiate data with the corresponding high school data. Data were gathered from the total population of incoming UTK freshmen who graduated from

Tennessee public high schools the prior spring for the years 1986 through 1989. Multivariable regression analysis was performed using first semester GPA and fourth semester GPA as dependent (response) variables. Independent variables were high school GPA, ACT composite scores, percentage of students on free or reduced lunches by system, adjusted expenditures per pupil by system, school system enrollment, percentage of students in system that were in the 12th grade, whether or not the high school was accredited by the Southern Association of Colleges and Schools, and whether or not the student was receiving financial aid.

ORGANIZATION OF THE STUDY

Chapter 1 contains an introduction of the subject, a statement of the problem, and the purpose and significance of the study. The chapter also contains definitions of pertinent terms, a statement of the research questions, limitations and delimitations of the research, and an overview of the procedures used in the study.

Chapter 2 is a general review of the literature. The first section contains a brief historical overview. The second and largest section contains the status of current research as it applies to predicting student achievement. A critique is provided, and this is followed by a summary of what is known and unknown regarding prediction equations for student achievement. The final section of the chapter is

devoted to what contribution the research results will hopefully make to the literature.

Chapter 3 contains a detailed definition of the mathematical models and the statistical analysis techniques and procedures used to address each of the research questions. Each of the model variables are discussed along with basic assumptions regarding the problem domain.

Chapter 4 contains the results and analysis of the data. Each research question is stated and evidence presented that addresses the pertinent areas of interest.

Chapter 5 is devoted to the summary of the findings, conclusions, and recommendations of the research.

CHAPTER 2

A REVIEW OF RELATED LITERATURE

HISTORICAL OVERVIEW

The research regarding achievement has come primarily from two directions. Efforts to identify and measure the factors that affect academic success were the first area (Coleman, Campbell, Hobson, McPartland, Mood, Frederic, & York, 1966; Hanushek, 1986). Construction of statistically based models whose variables could be used to predict a response variable such as achievement test scores or GPA was the second approach (Noble, 1990; Sawyer & Maxey, 1979; Willingham, Lewis, Morgan, & Ramist, 1990). This dissertation research was aimed at prediction models, but naturally the variables used for predictive purposes must be causal and well grounded in theory and empirical research.

TOPICS RELEVANT TO CURRENT RESEARCH

Mathematically based prediction models have taken two distinct approaches. The first was to construct correlational associations as reported by Nobel (1990) and Willingham, Lewis, Morgan, and Ramist (1990). This methodology primarily involved either Pearson's product moment correlational coefficient r or in some cases, the multiple linear correlational coefficient R . The variables of interest were

in the majority either students' high school GPA and university freshman cumulative GPA (UFCGPA), high school rank and UFCGPA, ACT scores and UFCGPA, or SAT scores and UFCGPA. In some cases the models were developed as simple linear regression equations and as multiple linear regression equations with two independent variables.

The second approach was reported by Hanushek (1986) in an exhaustive study of econometric models in education. His research was based upon those efforts in which there was an attempt to measure academic achievement in terms of economic production functions. Hanushek reported that 147 different educational production functions have been published. The economic production functions, or input/output models, are based on certain economic factors producing educational outcomes. Hanushek's research updated the activities after the controversial Coleman (1966) study and involved a wide range of explanatory variables.

The econometric type models as reviewed by Hanushek (1986) were intended to be used in the formation and study of state/national education and economic policies. The research as reported by Nobel (1990) and Willingham, Lewis, Morgan, and Ramist (1990) was intended to be used by parents, students, and college officials for the process of college selection and admission decision making.

A review of the recent literature regarding the prediction of a student's readiness for college reveals a

myriad of directions that could be pursued. However, it is possible to identify a taxonomy of major categories that can support systematic research. There are intra-relations and even inter-relations between each area that add complexities and problems. An obvious example is that good families help make for good schools, and good schools may often be the reason that good families locate and live in a given neighborhood. However, in spite of the many complexities, the hierarchical categories such as the student, the family, the high school, and the school district do have contributions that researchers have attempted to encapsulate, identify, study, and measure their impact on student academic achievement (Chubb & Moe, 1990, p. 111).

The Family

The decline of the traditional family formations is a well-known concern and phenomenon of today's society that many feel is one of the greatest problems facing our nation. Goodlad in A Place Called School: Prospects for the Future (1984, p. 7-10) cited seven trends and factors. Some of these are outside the sphere of the family, but they are so intertwined that the inclusions in this section can be justified: (1) the weakening of the household and church influences, (2) the decline of traditions supporting relationships between the home and school, (3) changes in community structures due to economic conditions. (For

example, the replacing of the social aspects of local shops and stores with large impersonal supermarkets and malls.) (4) a lack of cooperation among business, education, and government regarding educational issues, (5) often there are divisions between educators. (For example, principals versus teachers, the English department versus the mathematics department, and specialists versus classroom teachers), (6) the problems resulting from handling a growing diverse ethnic student population, and (7) lack of research and planning in how the school fits in with the other sources of a child's learning experience. (Television was the example cited as a major unknown component of education.)

The influence of the family on academic achievement can not be ignored. Researchers such as Coleman, Campbell, Hobson, McPartland, Mood, Frederic, and York (1966), Hanushek (1986), and Chubb and Moe (1990) identified factors such as parents' education, family income, family work habits, family educational values, and even inherited intelligence passed on to the children that have significant contributions to the level of scholarship and readiness for college. Good and Brophy (1986) reported that non-school factors such as parent involvement and SES are much more important than effects produced by the schools themselves.

The Student

Prediction of student achievement must certainly be centered on how the student has performed up to the point of beginning the college experience. The two variables used most frequently in prediction equations are high school GPA and achievement test scores.

High school GPA is an overall measurement of students' academic performance over their high school career. The significance of its interpretation is that it is not a single point measurement but rather it is an indicator of long run performance. Comparisons of individual high school GPA is always a topic of controversy. However, one can argue that the long term nature of the number should dampen out the effect of different grading practices and teacher expectations. The primary drawback to the use of high school GPA is that the value often fails to reflect curriculum and individual course difficulty. For example, an "A" in advanced placement calculus has the same numeric contribution to the GPA as does an "A" in family living.

Boyer in College: The Undergraduate Experience in America (1987, p. 34) conducted in-depth interviews with 29 college admission officers regarding the selection criteria they used for admission to their institution. Surprisingly, 28 of the 29 reported that high school GPA or class rank was the primary decision variable for admission.

Achievement test results are generally stated in the form of ACT or SAT scores. Willingham in Predicting College Grades: An Analysis of Institutional Trends Over Two Decades (1990) advocated the joint use of SAT scores and high school records as the basis for admission requirements. However, Crouse and Trusheim (1988) in The Case Against the SAT issued a scathing argument against using the SAT when high school GPA provided a better overall predictor of future student achievement. Spring in Conflict of Interests: The Politics of American Education (1988, p. 139-143) expressed concern regarding the exclusion of certain types of questions on the SAT examination and, in spite of their not-for-profit status, the seemingly over-concern for profit of the ETS management.

Noble (1990) did an exhaustive review of seventeen correlational studies that used high school rank, and high school GPA, versus ACT, or SAT scores. The largest R value was .64 and the majority were less than 0.5. Nobel's (1990) own research reported that ACT scores and high school grades were more reliable predictors of academic performance than could be produced by ACT scores or high school grades alone.

Johnston (1977) found that SAT scores were good predictors of future academic success. Barton (1989) found that high ACT scores indicated not only good grades but also retention in New Mexico colleges. In contrast Zainuba (1985) did not find that ACT scores or high school grade point averages had any significant relationship with GPAs of Texas

Southern University graduating students. Maurovich (1978) reported mixed findings, but did state that total high school GPA could be used as an admission criterion for entering freshman students. However, Maurovich (1978) was very reluctant to generalize the findings of the study and recommended that additional study was warranted.

The School

A number of educators have suggested that the school imparts a wide range of influences on the student. These include diverse areas such as principal's leadership, curriculum content, teacher education/training, number of volumes in the school library, teacher effectiveness, student/faculty ratio, influence of peers, school size, shared values, and the age/condition of the physical plant.

The impact of school size upon student achievement is not clearly understood. Goodlad (1984) reported that "it is not impossible to have a good large school; it is simply more difficult" (p. 309-310). Goodlad (1984) advanced the idea that large schools can be made more effective by creating schools within schools. For example, instead of having a school of 3,000 students, one would create six schools each with 500 students; all contained within the existing buildings. Chubb and Moe (1990, p. 104-105) took the approach that bigger was better since schools could enjoy economies of scale and should be able to provide better services to the

student population. However, they reported no strong findings to support their views. They examined the results of the Administrator and Teacher Survey (ATS) and found that no conclusion could be reached. There were big schools producing good students and likewise big schools producing poor students. The same could also be said for small schools. Conant in The American High School Today: A First Report to Interested Citizens, (1959, p.37) was an advocate of consolidation. He stated that no school should have a graduating class with fewer than 100 students. Boyer in High School: A Report on Secondary Education in America (1983, p. 235) advocated that high schools with more than 1500 should be reorganized into smaller schools. Rutter (1983) reported that student outcomes when compared between schools were not due to school size or to the age of the physical plants.

High school accreditation by SACS is a measure of overall school programs, their quality, and is one of the few direct school variables available for analysis. Accreditation for high schools is accomplished through a process that is specified by the Southern Association of Colleges and Schools (SACS). A uniform set of standards is set for all schools in this region. Data regarding the individual school accreditation status are published in the annual Tennessee Directory of Public Schools (1985, 1986, 1987, 1989). Accreditation covers a five-year period with some degree of annual monitoring.

Recent research has shown that good schools tend to transcend the effects of SES variables. McClean and Sanders (1983) reported that good teachers are effective even when they are located in low-income inner-city schools. Ravitch (1985) advocated that "schools do make a difference, regardless of the family background of students" (p. 101).

The School System

School system inputs tend to be less instructional and more managerial in nature. A school district is charged with carrying out the general state educational policies and programs, staffing, coordinating the educational activities, and the budgeting, monitoring, and disbursements of funds.

The economic consequences of funding upon student achievement is one of the most emotional and perhaps the least understood issue. Research findings have not completely supported the relationship between expenditure levels and student achievement. Childs and Shakeshaft (1986) investigated forty-five major studies on this issue and reported that "nineteen studies reported no relationships, fourteen studies found a positive relationship, and twelve studies indicated a positive relationship under certain conditions" (p. 250). They concluded that the aggregate expenditure per pupil had little effect on academic achievement. However, they did find that if the monies were directly used for instructional purposes, then there were some

increases in student performance (Childs and Shakeshaft, 1986). Peter Coleman (1986) examined the relationship between student achievement and per pupil expenditures for school districts in British Columbia for the years 1981 through 1984. He concluded that effective school districts possessed a type of district ethos that encompassed the classroom, school, and district level consequences. Coleman (1986) reported that effective districts constantly demonstrated high student achievement levels while using very modest levels of per pupil expenditures.

Research on the effects of expenditures per pupil (EPP) on achievement is incomplete and difficult to evaluate. This results not only from the complexity of the number itself but also in what is and should be measured.

The economic well being of the school district has long been used as a barometer of the community's ability to fund education, and also as a measure of social and economic status (SES) of the citizenry. The student is not only a citizen of his/her community, but is also affected by the influence of peers (Chubb & Moe, 1990). Firestone (1987) examined the use of the percentage of students on free or reduced lunches as an SES variable in a study for the Philadelphia school system. He found that SES had a strong impact on student outcomes, but that learning was also greatly influenced by the school principal's leadership skills. The percentage of students on free or reduced lunches is the SES variable that is emphasized

and reported by the Tennessee Department of Education for grades K-12 (Commissioner's Report Card, 1987, 1988, 1989, 1990).

CRITIQUE OF CURRENT STATUS OF RESEARCH

A considerable amount of attention is being given to the examination of the education process. This is particularly true of public education for grades K through 16. Causal factors that increase academic achievement are being sought and advocated by almost every sector of society. However, the research base does not appear to have advanced a great deal in the past twenty years. The prediction of academic achievement at the college level remains based upon high school GPA and achievement scores such as ACT and SAT.

CONTRIBUTION OF RESEARCH TO THE LITERATURE

Despite the fact that this research was of an institutional nature, there are several pertinent factors that make the study important to the literature base. One such factor is that all the graduates of a state's public high schools over a four-year period who attended the major state land grant institution immediately following high school were included. Another factor is that the research took a multi-variable approach to identify items (beyond the usual high school GPA and assessment test scores) that can be used to predict collegiate academic achievement. But, one of the most

important areas, is that the research looks beyond the traditional first semester freshman GPA and follows the students through the end of their sophomore year.

CHAPTER 3

METHODOLOGY

INTRODUCTION

The statistical procedures consisted of the following steps. There were three distinct groups of research questions which encompassed six separate questions. Group one was directed at data and relationships centered around first semester cumulative GPA. The second group of questions involved the data and their relationships regarding fourth semester (end of sophomore year) cumulative GPA. The third and final group was aimed at determining how groups one and two differ, and if they do differ, then what makes up the difference.

For research questions one through four the same methodology was used. First, descriptive statistics were produced for each variable. Second, correlation coefficients were used to show association between any of the individual independent variables and the response variable. The third and final step was to generate a multiple linear regression equation based upon the response variable and the appropriate independent variables. Research question five was aimed at showing any difference in the two individual response variables (first semester cumulative GPA and fourth semester cumulative GPA). A statistical test of difference (correlated

t test) was used for this. Research question six required the use of statistical tests to determine if there was a statistically significant difference in any of the independent variables used to predict first semester cumulative GPA and the independent variables used to predict fourth semester GPA. Here the students who persisted until the end of their sophomore year were compared to those who dropped out for some reason. A test for determining differences was applied in this case. A t-test or chi-square test (depending upon whether the variables were interval or nominal level) was utilized for comparing the survivors versus non-survivors.

Research questions two and four required the use of prediction models. A predictive model is not intended to be used for functional and/or control purposes, even though it may explain a portion of the effects of the individual variables upon the response variables. Draper and Smith (1966) identified three aspects of the usefulness of predictive models. They stated that a predictive model "can and does provide guidelines for further experimentation, it pinpoints important variables, and it is a very useful variable screening device" (Draper & Smith, 1966, p. 235).

Multiple variable linear regression was used to produce the two prediction equations. Separate equations were developed for predicting first semester cumulative GPA and fourth semester cumulative GPA. The models were constructed beginning with all the identified independent variables in the

initial functions. The two initial response functions can be expressed as follows:

$$\begin{aligned} \text{UFCGPA} &= \beta_0 + \beta_1 \text{ACTCOMP} + \beta_2 \text{HSGPA} + \beta_3 \text{SYSFORL} + \beta_4 \text{SYSAEPP} \\ &\quad + \beta_5 \text{HSACRD} + \beta_6 \text{SYSE} + \beta_7 \text{SS12GPCT} + \beta_8 \text{UFFAID} + \\ &\quad \epsilon_{\text{UFCGPA}} \end{aligned}$$

$$\begin{aligned} \text{USCGPA} &= \beta_0 + \beta_1 \text{ACTCOMP} + \beta_2 \text{HSGPA} + \beta_3 \text{SYSFORL} + \beta_4 \text{SYSAEPP} \\ &\quad + \beta_5 \text{HSACRD} + \beta_6 \text{SYSE} + \beta_7 \text{SS12GPCT} + \beta_8 \text{USFAID} + \\ &\quad \epsilon_{\text{USCGPA}} \end{aligned}$$

β_0 = regression intercept term

$\beta_1 \dots \beta_8$ = regression parameter estimates for each of the independent variables

DEPENDENT VARIABLES

UFCGPA = university first semester freshman cumulative GPA

USCGPA = university second semester sophomore cumulative GPA

INDEPENDENT VARIABLES

ACTCOMP = ACT composite score

HSGPA = High school GPA

HSACRD = High school accreditation by SACS (yes or no)

SYSFORL = school system percentage of students on free or reduced lunch

SYSAEPP = school system adjusted expenditure per pupil

SYSE = school system enrollment

SYS12GPCT = twelfth grade graduation percentage for school system

UFFAID = university freshman financial aid indicator

USFAID = university sophomore financial aid indicator

ϵ_{UFCGPA} = regression error term for UFCGPA

ϵ_{USCGPA} = regression error term for USCGPA

STATISTICAL ANALYSIS

Descriptive Statistics

Univariate statistical indicators such as the mean, standard deviation, maximum, minimum, and range were used for the interval data types. These tests were useful for the conventional measures of central tendency, variation about the mean, and for the detection of outliers.

The nominal level variables such as HSACRD, UFFAID, and USFAID utilized statistics such as the number in each classification, total number of observations, maximum, minimum, and mode. The maximum and minimum statistics were used only to test for the extreme values of known data values. For example, high school GPA must be greater than zero and less than four.

Statistics Used to Show Relationships or Associations

The Pearson product moment correlation coefficient measures the linear relationship between two interval or ratio level variables. The letter r was used to signify the correlation coefficient and its value ranged from -1 to $+1$. An r of -1 represented perfect negative correlation and signifies that when one variable increases the other variable will decrease at the same rate. The converse was true if one

variable decreased the other variable will increase. An r value of 0 (zero) indicates that there is no relationship between the two variables. A value of +1 represents perfect positive correlation. When one variable changes the other variable will change correspondingly at the same rate and in the same direction. The p -value associated with r is used to test the null hypothesis that the true correlation is zero. The correlation coefficient is a very useful tool, but it is restricted by the fact that it is used to show a linear relationship and does not indicate cause and affect. Often r^2 will be used since it is always positive and can be expressed as a percentage.

The phi coefficient (ϕ) was used to show association between the nominal and interval level variables. Phi takes on values similar to the interval level correlation coefficient of -1 to +1 for a 2x2 table and from 0 to 1 for larger tables (SAS/STAT, 1989, p. 866). The phi coefficient was derived from the chi-square statistic and is reported to be popular when used in the behavioral sciences (Fleiss, 1981, p. 60). Fleiss (1981) also stated that the values of .30 or .35 indicated no statistically significant association.

The multiple correlation coefficient R , was used when examining the correlation between more than two variables. This was very useful when examining the relationships of the various independent variables with the model response variable. Like r , R is most useful when expressed as squared

(R^2). Model R^2 is the ratio of the model sum of squares divided by the total sum of squares. This represents the total variation in the dependent variable that is explained by the contributions of the independent variables (Freund & Littell, 1986, p. 20).

Two different uses were made of Student's t-test. The first application was to test the null hypothesis (H_0) to determine if any of the coefficient estimates in the regression equations were different from zero. (The t value is the ratio of the coefficient estimate divided by its respective standard error.) The second use of the t-test was to look at any differences in freshman cumulative GPA and cumulative GPA measured at the end of the sophomore year. A correlated t-test was used since each individual student's grades must be matched or paired. The difference was calculated between the freshman cumulative GPA and the sophomore cumulative GPA. The t-test was used to test if the value was statistically different from zero.

Research question six required looking at some of the independent variables by classifying the students at the end of their sophomore year as survivors or non-survivors (determining persistence). Survivor classes take on the values of yes and no and therefore are nominal level data. A chi-square test was used to determine if there was a statistically significant difference between the survivor classes (yes or no) and each of the independent variables that

were nominal data types. A nonparametric analysis of variance was performed if the independent variables were interval level data.

Prediction Models

Stepwise regression techniques as described by Freund and Littell (1986, p. 82-92) were used to determine the most accurate combination of independent variables. The stepwise technique begins by calculating a least squares best fit model utilizing the one independent variable that makes the greatest contribution (the largest R^2) to explaining the variation in the response variable. The algorithm then proceeds to select the next variable that can make the greatest additional contribution among the remaining un-selected variables. The procedure is very sophisticated and can determine if a better regression equation can be obtained by "tossing out" a variable that had been previously included in the regression equation. A liberal default value of $p \leq .15$ for the coefficient significance was used.

The t-test was used to test the null hypothesis (H_0) to determine if an individual regression coefficient estimate was different from zero. (The t value in this case is the ratio of the coefficient estimate divided by its respective standard error.)

A one-way analysis of variance (ANOVA) was produced for each regression equation. This section provides information

regarding how the resulting linear equation fits the data. The three primary statistics are $C(p)$, R^2 and the F value.

R^2 and F statistics indicate how much of the variation in the response variable are explained by all of the independent variables (Freund, Littell, & Spector, 1986, p. 51). Adding additional parameters to a model will always increase R^2 . Naturally, the error term will increase in size when non-required variables enter the equation. The $C(p)$ statistic guards against adding unnecessary variables to a model by providing a metric that is a function of not only the number of parameters in the model, but also to the mean square error (MSE) term. $C(p)$ should be approximately equal to the number of independent variables in the regression equation (Walpole & Myers, 1989, p. 447).

The F statistic is the ratio of the mean sum of squares due to the regression and the mean sum of squares due to the residual or error term. The F statistic is also a function of the number of parameters in the model and the number of observations and the pre-determined acceptable probability level. The calculated F statistic should exceed the F value for the F distribution if the regression is statistically significant (Kleinbaum & Kupper, 1978, p. 140).

In addition, the F statistic was used to test the null hypothesis H_0 that each of the parameter estimates (coefficients) were equal zero:

$(H_0 = \beta_0 = \beta_1 = \beta_2 = \dots \beta_9 = 0)$. If the p value ($\text{PROB} > F$) is

small, such as 0.0001, then this indicates that some of the coefficients are not equal to zero and that H_0 should be rejected.

ASSUMPTIONS

Some system wide data were used as if they applied evenly to a student or graduating class. For example, the number of students on free or reduced lunches and expenditures per pupil are reported as a percentage of the entire local school system.

The response variables UFCGPA and USCGPA and the independent variables ACTCOMP, HSGPA, SYSFORL, SYSAEPP, SYSE, and SYSGPCT are treated as interval level data. HSACRD, UFFAID, and USFAID take on values of 0 (no) or 1 (yes) and are treated as categorical or nominal level data.

All values used for expenditures per pupil were expressed in constant 1982 dollars. This allowed any yearly effects to be compared evenly. Implicit price deflators are universally used by researchers doing economic studies, and analysis involving U.S. dollars as their base and the deflators were taken from the Economic Report of the President (1991). Since academic years and calendar years differ, an average of two indicators was used. For example, the 1986 school year included portions of 1985 and 1986. The 1986 deflator as shown in Table 3.1 resulted from the arithmetic average of 109.9 and 112.5.

Table 3.1

GNP Annual Price Deflators and Adjusted
Academic Year Price Deflators

Year	Implicit Price Deflators	Adjusted Implicit Deflators
1985	* 109.9	
1986	112.5	** 111.20
1987	116.8	114.65
1988	120.9	118.85
1989	125.8	123.35

* Source: Economic Report of the President (1991)

** This column is the arithmetic average of the current and prior year values from the implicit price deflators column.

CHAPTER 4

RESULTS AND ANALYSIS OF THE DATA

DATA PREPARATION AND VALIDATION

Data were supplied (11,693 records) by the Student Services Division of the University of Tennessee, Knoxville (UTK) in the form of magnetic tape. Social Security Numbers and any other personal identification data were removed by UTK from the data set. Other data files were constructed from public documents and loaded to a personal computer.

The data supplied by UTK consisted of all first time freshmen from Tennessee private and public high schools for 1986 through 1989. The merged data dropped all schools not listed in the 1989-90 Tennessee Directory of Public Schools (1989). Some small problems arose due to the incompatibility of the high school numbering systems used by UTK and the State Department of Education. In addition, the two organizations often used different names for the same institution. However, these anomalies were resolved by manually verifying names and identification codes in official publications and reports. The final data set consisted of 10,327 student records. In some instances there were missing values for particular data items. For example, 174 students did not report an ACT Composite score.

Tables 4.1 and 4.2 illustrate how closely the research data agreed with data previously published by the

Table 4.1

High School GPA by Year for First-Time Freshmen as Reported by the University of Tennessee as Compared to High School GPA for First-Time Freshmen Who are Graduates of Tennessee Public High Schools

ALL FIRST-TIME FRESHMEN *			TENNESSEE PUBLIC HIGH SCHOOL FIRST-TIME FRESHMEN	
YEAR	GPA	N	GPA	N
1986	3.04	3660	3.08	2562
1987	3.06	3797	3.11	2752
1988	3.10	3756	3.14	2701
1989	3.08	3346 **	3.14	2312

Note: * source: the University of Tennessee Fact Book, 1990
 ** personal communications with Donald Cunningham, UTK Institutional Research Department

Table 4.2

Composite ACT Scores by Year for First-Time Freshmen as Reported by the University of Tennessee Compared to High School Composite ACT Scores for First-Time Freshmen Who are Graduates of Tennessee Public High Schools

ALL FIRST-TIME FRESHMEN *			TENNESSEE PUBLIC HIGH SCHOOL FIRST-TIME FRESHMEN	
YEAR	ACT	N	ACT	N
1986	21.3	3249	21.5	2520
1987	21.6	3374	21.7	2705
1988	22.2	3336	22.4	2657
1989	22.0	2937 **	22.1	2271

Note: * source: the University of Tennessee Fact Book, 1990
 ** personal communications with Donald Cunningham, UTK Institutional Research Department

Institutional Research Department at UTK regarding first-time freshmen for the entire freshman class. The percentage of first-time freshmen represented by the research data set varied from 70% in 1986 to 72.5% in 1987. Some minor adjustments were required because of missing data and inconsistencies. The most serious was the merger of the Knox County and Knoxville City School systems in the middle of the time period used for the analysis. The Knoxville City system did not report for the school year 1986-1987 the percentage of students on free or reduced lunches and the expenditures per pupil. The numbers were estimated by using the ratio of prior year values from the two school systems.

Sevier County did not report the free or reduced lunch percentage for 1986. It should be noted that this was the first year that this value was required by the state. The same level was reported for 1987 and 1988 (35%), and this value was also used for 1986. The reported amount for 1989 was 37%, thus it appeared that 35% was a reasonable estimate.

A total of 117 Tennessee public school systems were represented in the study. This included the merging of the Knox County and Knoxville City systems as one entity. Of the total 117 systems the number of individual school systems represented per freshman class varied from 100 in 1986 to 106 in 1987 and 1988. A total of 257 high schools were included and varied from a low of 203 schools in 1989 to 219 schools in

1987. Table 4.3 illustrates the Tennessee public school population for the respective freshman classes.

FINDINGS

The FINDINGS sections contains the statistical output from the research aimed at addressing each of the respective research questions. Each research question is stated, results are presented in a narrative form and supportive table(s) are provided.

Research Question 1

Is there a statistically significant relationship among any of the selected independent variables and students' first semester collegiate GPA?

The relationship between each independent variable and the response variable UFGPA was examined to determine if there were any worthwhile combinations. The first step was to look at the cumulative data and then for the individual years 1986 through 1989. Table 4.4 shows the descriptive statistics of the first semester cumulative GPA for the years 1986 through 1989. Freshmen cumulative GPA showed a general upward trend in the mean. The percent increases of the means over the base year were 2.07%, 2.12%, and 5.75%. The reason for the large jump in 1989 is unknown. The standard deviation is fairly large as plus or minus one standard deviation is approximately one quality point.

Table 4.3

Number of First-Time Freshmen and Their Respective High Schools and School Systems by Year

	1986	1987	1988	1989	86-89
Students in study	2562	2752	2701	2312	10327
High schools represented	206	219	204	203	257
School systems represented	100	106	106	103	117

Note: 121 of Tennessee's 139 school systems provide programs at the high school level.

Table 4.4

Descriptive Statistics for Freshman First Semester GPA by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	2.34	2.38	2.39	2.50	2.39
% Change		2.07	2.12	5.75	
Median	2.42	2.50	2.50	2.59	2.50
Std Dev	0.95	0.97	1.00	0.97	0.97

Table 4.5 provides an interesting look at the number of first time freshmen who graduated from high schools that were accredited by the Southern Association of Colleges and Schools (SACS). High school accreditation (HSACRD) takes on discrete values of 0 (no) and 1 (yes). Accreditation by SACS is valid for a five-year period barring any unusual circumstances. The percent value for HSACRD varied from 87% to 90%. Thus, 87% to 90% of the incoming freshmen in the respective years graduated from SACS-accredited high schools. There was a general decreasing trend in the percent of incoming freshmen each year coming from schools that were not accredited. The respective percent of students from high schools that were not accredited for each year was 13%, 11%, 10%, and 11%. The data from year-to-year appeared consistent and stable, indicating no major variations over the four-year timespan.

ACT composite scores (ACTCOMP) was one of the variables where the results were of the most interest. More research has been done on assessment scores and high school GPA than any of the other variables of interest. Table 4.6 shows the ACT Composite Scores for incoming freshmen by their high school graduation year. ACTCOMP appeared to be very stable over time with the same median value for each year. The mean increased each year up to 1989 and then reflected a slight decrease over the prior year. The decreasing standard deviation indicated that the values were increasingly consistent and had less variability over the four years.

Table 4.5

Descriptive Statistics for SACS High School
Accreditation by High School Graduation Year

	1986	1987	1988	1989	86-89
(no)	324	301	266	250	1141
(yes)	2238	2451	2435	2062	9186
N	2562	2752	2701	2312	10327
% Acc.	87	89	90	89	89
% Not Acc.	13	11	10	11	11

Table 4.6

Descriptive Statistics for ACT Composite
Scores by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2520	2705	2657	2271	10153
Mean	21.5	21.7	22.4	22.1	21.9
Median	22.0	22.0	22.0	22.0	22.0
Mode	20.0	23.0	24.0	21.0	22.0
Std Dev	4.46	4.33	4.01	4.10	4.25
Maximum	34.0	34.0	33.0	33.0	34.0
Minimum	7.0	7.0	7.0	9.0	7.0
Range	27.0	27.0	26.0	24.0	27.0

Table 4.7 shows that the mean high school GPA (HSGPA) had an increasing trend over the 1986-1989 period. The median and mean are almost the same value with the median about .03 larger for each year. Likewise, the standard deviation remained a constant value of approximately .53 over the entire range. However, the mean HSGPA is very stable over the four-year period with no unusual fluctuations or movement in the scores.

SYSFORL (percentage of students in school system on free or reduced lunches) was very consistent from year-to-year. Table 4.8 shows that the mean value varied very little from year-to-year (from 1986 to 1987, 1.4%; 1987 to 1988, .5%; and from 1988 to 1989 .6%). SYSFORL ranged in value from a maximum of 89% in an inner-city school district to a minimum of 9% in an affluent suburban district. The standard deviation generally decreased over the period by approximately 5 percent. The maximum and minimum (end points of the scale) values changed only in a minor way.

SYSAEPP represents the expenditure per pupil as adjusted for inflation and stated in 1982 dollars for each incoming freshman. The value increased for each year. Using 1986 as the base year, Table 4.9 shows that the mean value for SYSAEPP increases for 1987 through 1989 were 7.27%, 26.8%, and 30.39% respectively. In constant dollars from 1986 through 1989, the average expenditure per pupil for incoming freshmen increased by \$643. The minimum expenditure level increased by \$385 for

Table 4.7

Descriptive Statistics for High School
GPA by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	3.08	3.11	3.14	3.14	3.12
Median	3.10	3.14	3.16	3.15	3.13
Std Dev	0.55	0.53	0.52	0.54	0.53

Table 4.8

Descriptive Statistics for Percentage of Students on
Free or Reduced Lunches by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	35.8	34.4	33.9	34.5	34.6
% Change		1.4	.5	.6	
Median	34.0	34.0	32.0	34.0	33.0
Mode	15.0	14.0	32.0	34.0	32.0
Std Dev	17.31	17.00	12.72	12.76	15.18
Maximum	89.0	88.0	87.0	86.0	89.0
Minimum	10.0	10.0	10.0	9.0	9.0
Range	79.0	78.0	77.0	77.0	80.0

Table 4.9
Descriptive Statistics for Adjusted Expenditures
Per Pupil by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	2118	2272	2685	2761	2541
% Change		7	27	30	
Median	2025	2209	2735	2713	2370
Mode	2025	2209	2735	2713	2209
Std Dev	408	410	436	454	504
Maximum	3336	3393	3898	3965	3965
Minimum	1508	1585	1873	1893	1508
Range	1828	1808	2025	2072	2457

the four years, and the maximum expenditure level increased by \$629. The standard deviation increased in value each year and showed a percentage increase of over 11% from 1986 through 1989. This increase indicated an increasing variability in the expenditure per pupil as experienced by the incoming freshmen.

Table 4.10 depicts the total enrollment of the school systems of the incoming freshman class. The mean value for SYSE (school system enrollment) increased from year-to-year, which indicated that more and more students were coming from larger school systems. However, the standard deviation was larger in 1986 and 1987 than the mean, indicating a large dispersion about the mean value. The coefficient of variation (CV) decreased constantly over the four years. Interestingly, the minimum value for enrollment dropped each year while the maximum system enrollment remained at roughly the same level. The median statistic is perhaps a more meaningful measure of central tendency than the mean for SYSE and the median value changed by less than one hundred students from year-to-year.

As shown in Table 4.11, the mean value for SYS12GPCT (school system twelfth grade enrollment as a percentage of total system enrollment) increased over the entire timespan. In addition, the standard deviation decreased, reflecting less variability about the mean. Overall, the variable was generally consistent and displayed no significant swings in

Table 4.10

Descriptive Statistics for School System
Enrollment by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	21359	21737	28951	30148	25413
Median	16705	17105	16615	15990	16615
Mode	27195	26843	54725	56157	54725
Std Dev	23422	23802	26959	27874	25824
Maximum	112766	113101	112670	111546	113101
Minimum	1097	697	856	395	395
Range	111687	112404	111814	111151	112706

Table 4.11

Descriptive Statistics for School System
Twelfth Grade Percentage Enrollment
by High School Graduation Year

	1986	1987	1988	1989	86-89
N	2562	2752	2701	2312	10327
Mean	6.6	6.9	7.0	7.1	6.9
Median	6.8	7.2	7.2	7.1	7.1
Std Dev	1.06	1.01	0.95	0.92	1.00
Maximum	10.9	10.2	12.0	10.5	12.0
Minimum	2.2	1.7	2.8	4.3	1.7
Range	8.7	8.5	9.2	6.2	10.3

any of the relevant descriptive statistics such as the standard deviation, maximum, median, minimum, and range.

Table 4.12 indicates whether or not university freshmen receive financial aid (UFFAID). The mean value for UFFAID was about 41% over the 1986 through 1989 periods. UFFAID was limited to the values 0 (no) and 1 (yes). The proportions in each of the yes/no cells by years were very consistent and a chi-square of 0.816 indicated that there was no statistically significant difference in UFFAID by year.

Discussion of Simple Correlation Coefficients

Table 4.13 shows the correlations that were developed to indicate the linear relationships between the response variable UFCGPA (freshman cumulative GPA) and the candidate independent variables. To demonstrate the stability over time of each respective r , the coefficients were calculated for each year. The two numbers in each category represented the correlation coefficient and the number of matched pairs that were used to calculate the correlation coefficient. An '*' was used to show statistical significance at the .01 level and an '**' was used to indicate statistical significance at the .05 level.

The r 's between UFCGPA and ACTCOMP (ACT composite scores) were very consistent over time. They varied from a low of .362 in 1986 to a high of .399 in 1989. The value for the

Table 4.12

Descriptive Statistics for Students Receiving
Financial Aid by High School Graduation Year

	1986	1987	1988	1989	86-89
(no)	1487	1630	1578	1353	6048
(yes)	1075	1122	1123	959	4279
N	2562	2752	2701	2312	10327
Percent	42	41	42	41	41

Chi-square = 0.816 df = 3 p = 0.846

Table 4.13

Simple Correlation Coefficients:
Freshman Cumulative GPA Versus Other Independent Variables

	1986 UFCGPA	1987 UFCGPA	1988 UFCGPA	1989 UFCGPA	86-89 UFCGPA
ACTCOMP N	.363* 2520	.396* 2705	.362* 26577	.399* 2271	.379* 10153
HSGPA N	.480* 2562	.476* 2752	.498 2701	.491* 2312	.486* 10327
SYSFORL N	-.024 2562	-.058* 2752	-.042** 2701	-.041** 2312	-.042* 10327
SYSAEPP N	.034 2562	.026 2752	-.018 2701	.028 2312	.034* 10327
HSACRD(1) N	.358 2562	.319 2752	.334 2701	.374 2312	.186 10327
SYSE N	-.065* 2562	-.064* 2752	-.050* 2701	-.025 2312	-.044* 10327
SSGPCT N	-.039** 2562	-.040** 2752	-.030 2701	.002 2312	-.018 10327
SS12GPCT N	.038 2562	.046** 2752	.042** 2701	-.029 2312	.033* 10327
UFFAID(1) N	.372 2562	.357 2752	.344 2701	.398 2312	.256 10327

Note: * significant at $p < .01$, ** significant at $p < .05$
 (1) Phi correlation coefficient used for HSACRD and
 UFFAID

years 1986 through 1989 was .379. This resulted in an r^2 of .144.

The relationship between UFCGPA and HSGPA (high school GPA) was the highest of any of the selected variables. It varied from 0.476 in 1987 to .498 in 1988. The r for the entire time-span was 0.486 resulting in a r^2 of .266.

The SYSFORL (percentage of students in school system on free or reduced lunches) and UFCGPA resulted in consistent negative correlation coefficients. This appears logical as it seems that the larger the percentage of students on free or reduced lunches, the more difficult it would be for a school system to prepare students at a desired readiness level. The r varied from -.058 in 1987 to a high of -.024 in 1986. However, the -.024 was not statistically significant at the .05 level. The value for 1986-1989 was -.042 with an r^2 of .002.

SYSAEPP (school system adjusted expenditure per pupil) failed to correlate with UFCGPA at the statistically significant level of .05 for any of the individual years. The value for the overall timespan resulted in an r of 0.034 and an r^2 of .001 that were statistically significant.

None of the individual years of 1986 through 1989 produced a statistically significant r for HSACRD (high school accreditation) and UFCGPA. The overall years r of .025 was significant at the .05 level. The r^2 was .001.

The r produced by SYSE (school system enrollment) and UFCGPA was negative in all cases. However, the correlation for the year 1989 was not significant at the .05 level. The values varied from a low of $-.065$ in 1986 to $-.050$ in 1988. The all-years r was $-.044$ with an r^2 of $.002$.

The correlation with SS12GPCT (percentage of students in school system that were in twelfth grade) was significant at the .05 level for only 1987 and 1988. The respective values were $.046$ and $.042$. The all-years r was $.03315$ with an r^2 of $.001$.

Summary of Research Question 1

Six of the eight candidate independent variables showed a statistically significant relationship with the response variable freshman cumulative GPA. However, the r values for SYSFORL, SYSAEPP, SYSE, and SYS12GPCT were each close to zero and could be interpreted as having very little, if any, association with UFCGPA. The r value for HSGPA ($.486$) and ACTCOMP ($.379$) did show a relatively strong linear association.

Research Question 2

Is there a regression model that utilizes the stated independent variables that will predict first semester GPA for a student?

Examination of the regression model as displayed in Table 4.14 offered insight into the individual contribution of the coefficients. The intercept value of -1.247 has no physical or logical meaning. The response function would be equal to the intercept when all the coefficients are zero.

The ACT composite score (ACTCOMP) coefficient value of .043 indicated that the cumulative GPA will increase by .043 for each unit of contribution by the ACTCOMP. For example, an ACTCOMP of 20 would add .86 to the GPA value.

The High School Grade Point Average (HSGPA) coefficient contributes the most of any of the parameters. Its value of .79 when coupled with a HSGPA of 3.5 resulted in a value of 2.77 for the cumulative GPA. (It must be kept in mind that the positive contributions of the coefficients must overcome the negative intercept value.) HSGPA and UFCGPA have the same scale, thus this parameter estimate was the easiest to conceptualize. Each unit of change in HSGPA moves the response function .79 units in the same direction.

SYSFORL (percentage of students in school system on free or reduced lunches) had a value of -0.004. The negative value is logical as a high percentage of students on free lunches should represent an economically deprived school district. The values of SYSFORL varied from 9% in an upper income suburban district to 89% for an urban inter-city school system. When the 89% was substituted into the equation for SYSFORL, then the cumulative GPA decreased by .36 points.

Table 4.14

Regression Analysis for First Semester
Freshman Year Cumulative GPA Model

Dependent Variable: UFCGPA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	5	2751.63499	550.32700	818.815
Error	10147	6819.81491	0.67210	
C Total	10152	9571.44991		
Prob>F	0.00000			
Root MSE	0.81982	R-square	0.2875	
Dep Mean	2.39250	Adj R-sq	0.2871	
C.V.	34.26613	C(p)	3.5653	

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEPT	1	-1.2370	0.0719	-17.214	0.0001
ACTCOMP	1	0.0426	0.0022	19.781	0.0001
HSGPA	1	0.7901	0.0175	45.092	0.0000
SYSFORL	1	-0.0041	0.0006	-7.360	0.0001
SYSAEPP	1	0.0001	0.0000	6.280	0.0001
HSACRD	1	0.1231	0.0266	4.636	0.0001

Variable	Number In	Partial R**2	Model R**2	F	Prob > F
HSGPA	1	0.2459	0.2459	3309.6808	0.0000
ACTCOMP	2	0.0331	0.2790	466.6463	0.0001
SYSFORL	3	0.0032	0.2822	45.3039	0.0001
SYSAEPP	4	0.0037	0.2860	53.2344	0.0001
HSACRD	5	0.0015	0.2875	21.4957	0.0001

Similarly, the 9% value decreased the response function by only .04 points.

The parameter estimate for SYSAEPP (school system adjusted expenditure per pupil) had a value of only .0001. The research data used to develop the regression equation ranged from a maximum of \$3965 to a minimum of \$1508. The range of almost \$2500 represented about .25 quality points difference between a student from a well-funded system than from a system that was funded at a minimum level. (It should be noted that the \$1508 was from 1986 data and that the minimum level of funding in 1989 was \$1893 for a \$385 increase.)

The parameter estimate for HSACRD (high school accreditation) was .123. HSACRD can only assume the values of 0 and 1. Thus, if a student came from a high school that was accredited by SACS, then his/her predicted cumulative GPA would be .123 higher than a student that did not come from an accredited high school.

HSGPA accounted for almost 25% of the variation in UFCGPA (university freshman college grade point average). The remaining variables accounted for less than 5% of the total variation that was explained by the regression model. ACTCOMP with an R^2 of .0331 accounted for the majority of the amount contributed by the last 4 variables.

The parameters were examined over the 1986 through 1989 timespan to determine their stability over time. Table 4.15

shows that all the variables appear to be reasonably stable, particularly those that had the greatest contribution to the response function (HSGPA and ACTCOMP). This should not be surprising since the four-year timespan of the study is probably not long enough to reflect any structural changes in the independent variables.

Summary of Research Question 2

Research question one (1) was an exploration of the relationships between the candidate independent variables and the response variable freshman cumulative GPA (UFCGPA). A regression equation was produced which reflected the "best" unbiased estimator of freshman cumulative GPA.

$$\begin{aligned} \text{UFCGPA} = & -1.236955 + 0.042619(\text{ACTCOMP}) + 0.790071(\text{HSGPA}) \\ & -0.00406(\text{SYSFORL}) + 0.000106(\text{SYSAEPP}) + 0.123126(\text{HSACRD}) \end{aligned}$$

The response function can be utilized by substituting the specified independent variables with values from the students' records and from Tennessee State Department of Education publications. The following example illustrated how the equation was used.

Given: ACTCOMP = 24
 HSGPA = 3.2
 SYSFORL = 30%
 SYSAEPP = 2500
 HSACRD = 1 (yes)

The above values were then substituted into the

Table 4.15

Regression Parameter Estimates
for First Semester Freshmen Model by Year

Variable	----- Parameter Estimates -----				
	1986	1987	1988	1989	86-89
INTERCEPT	-1.2111	-1.3159	-1.3779	-1.4112	-1.2370
ACTCOMP	0.0376	0.0479	0.0413	0.0450	0.0426
HSGPA	0.7744	0.7617	0.8569	0.7846	0.7901
SYSFORL	-0.0030	-0.0048	-0.0045	-0.0061	-0.0041
SYSAEPP	0.0002	0.0002	0.0001	0.0002	0.0001
HSACRD	0.1424	0.1259	0.0743	0.1119	0.1231

prediction equation.

$$\text{UFCGPA} = -1.236955 + 0.042619(24) + 0.790071(3.2) - \\ 0.00406(30) + 0.000106(2500) + 0.123126(1)$$

$$\text{UFCGPA} = 2.58$$

The confidence interval was constructed around the estimated mean value of UFCGPA by using the MSE (mean square error), the inverse of the XX' matrix $(XX')^{-1}$ and the corresponding t statistic.

$$\text{MSE} = 0.81982 \quad t = 1.96$$

The confidence interval was calculated as $\pm .02$. Thus, there is a less than 5% chance that the true value for UFGPA will be less than 2.56 or greater than 2.60 when using the above values for ACTCOMP, HSGPA, SYSFORL, SYSAEPP, and HSACRD.

Research Question 3

Is there a statistically significant relationship among any combination of the independent variables and students' fourth semester collegiate GPA?

Research question 3 was addressed in an identical manner as research question number 1. First, descriptive statistics were produced for the response variable and the candidate independent variables. Next, correlation coefficients were evaluated to determine if linear relationships existed between the response variable and individual independent variables.

Data are those records for individuals who entered UTK as first-time freshmen immediately following high school

graduation and were still enrolled at the end of their sophomore year.

Table 4.16 shows that the mean values for USCGPA (sophomore cumulative GPA) and standard deviation values were relatively stable from year-to-year.

Table 4.17 shows that the percentage of students from SACS-accredited high schools increased by approximately .01 for each year. There was an increase of 90 students from 1986 to 1987 and 73 students from 1987 to 1988.

The fourth semester descriptive statistics for ACT Composite Scores (ACTCOMP) are shown in Table 4.18. The mean values for ACTCOMP increased each year. From 1986 through 1988 the mean increased by .8 or almost one full point. In addition, the standard deviation decreased from 4.45 in 1986 to 4.04 in 1988 or -.41 over the three-year span. This reflected less variability in the ACTCOMP. The median scores remained very constant at 22, 22, and 23 for the three years. The mode increased dramatically from 20 to 24 over the same timespan.

The mean value for high school GPA (HSGPA) increased in a similar fashion to the ACT Composite scores in the previous table. Table 4.19 shows that the mean HSGPA score for the students surviving to the end of their sophomore year increased by .03 from 1986 to 1987 and by .02 from 1987 to 1988. The median value for HSGPA increased at approximately the same rate as the mean value. The standard deviation

Table 4.16

Descriptive Statistics at End of Sophomore Year:
Cumulative GPA by High School Graduation Year

	1986	1987	1988	Total
N	1798	1919	1961	5678
Mean	2.69	2.64	2.73	2.69
Median	2.67	2.64	2.73	2.68
Mode	2.58	2.29	2.83	2.58
Std Dev	0.61	0.65	0.64	0.64
Maximum	4.00	4.00	4.00	4.00
Minimum	.48	.14	.13	.13
Range	3.52	3.86	3.87	3.87

Table 4.17

Descriptive Statistics at End of Sophomore Year:
SACS Accredited High Schools by High School
Graduation Year

	1986	1987	1988	86-88
Yes	1580	1713	1770	5063
No	218	206	191	615
N	1798	1919	1961	5678
Percent	88	89	90	89

Table 4.18

Descriptive Statistics at End of Sophomore Year: ACT
Composite Scores by High School Graduation Year

	1986	1987	1988	86-88
N	1775	1888	1928	5591
Mean	21.8	22.0	22.6	22.2
Median	22.0	22.0	23.0	22.0
Mode	20.0	23.0	24.0	24.0
Std Dev	4.45	4.27	4.04	4.26
Maximum	34.0	34.0	33.0	34.0
Minimum	8.0	8.0	10.0	8.0
Range	26.0	26.0	23.0	26.0

Table 4.19

Descriptive Statistics at End of Sophomore Year:
High School GPA by High School Graduation Year

	1986	1987	1988	86-88
N	1798	1919	1961	5678
Mean	3.16	3.19	3.21	3.18
Median	3.18	3.24	3.24	3.21
Mode	4.00	4.00	4.00	4.00
Std Dev	0.53	0.51	0.49	0.51

decreased by almost .04 over the three years. This reflected less variability in the incoming high school averages. Generally, the statistics for the fourth semester students' high school GPA were stable over the three years.

Table 4.20 shows the descriptive statistics for the school system percentage of students on free or reduced lunches (SYSFORL) as represented by students who remained enrolled until the end of their sophomore year. The mean value for SYSFORL decreased by over two percentage points from 1986 through 1988. The median value was very close to the mean over the three-year span. However, the mode was 15, 14, and 32 for 1986, 1987, and 1988. The standard deviation dropped by almost 5 points over the three years. Generally, the descriptive statistics for SYSFORL were stable over the timespan.

The mean value for school system adjusted expenditure per pupil (SYSAEPP) increased by \$563 from 1986 through 1988. SYSAEPP had a general increase as the minimum funding level rose by \$365, and the maximum level increased by \$562 over the three year span. More importantly, the median value increased by \$710. The standard deviation went down slightly in 1987 and then ended up about \$25 higher in 1988 than in 1986. The descriptive statistics in Table 4.21 shows the stability of SYSAEPP. Table 4.22 was used to show the number of students who are receiving financial aid at the end of their sophomore year (USFAID). USFAID remained relatively constant from 1986

Table 4.20

Descriptive Statistics at End of Sophomore Year:
Percentage of Students on Free or Reduced Lunches
by High School Graduation Year

	1986	1987	1988	86-88
N	1798	1919	1961	5678
Mean	35.8	34.3	33.8	34.6
Median	34.0	34.0	32.0	32.0
Mode	15.0	14.0	32.0	32.0
Std Dev	17.27	17.21	12.46	15.77
Maximum	89.0	88.0	87.0	89.0
Minimum	10.0	10.0	10.0	10.0
Range	79.0	78.0	77.0	79.0

Table 4.21

Descriptive Statistics at End of Sophomore Year:
Adjusted Expenditures Per Pupil
by High School Graduation Year

	1986	1987	1988	86-88
N	1798	1919	1961	5678
Mean	2122	2265	2685	23651
Median	2025	2209	2735	22093
Std Dev	405	402	431	478
Maximum	3336	3393	3898	3898
Minimum	1508	1585	1873	1508
Range	1828	1808	2025	2390

Table 4.22

Descriptive Statistics at End of Sophomore Year:
Students Receiving Financial Aid by High
School Graduation Year

	1986	1987	1988	86-88
Yes	550	554	594	1698
No	1248	1365	1367	3980
N	1798	1919	1961	5678
Percent	31	29	30	30

through 1988. Approximately 30% of the students received some sort of financial assistance during their sophomore year.

Table 4.23 shows the school system total enrollment (SYSE) for UTK students who survived through their sophomore year. The mean value for SYSE increased from year-to-year with the greatest increase from 1987 to 1988 (6,798). However, the standard deviation was larger or almost as large as the mean for the timespan. This indicated a large amount of variability in SYSE and is reflected in the huge swings in the median values of plus or minus 10,000 students each year.

Twelfth grade enrollment as a percentage of total system enrollment (SYS12GPCT) was generally stable from 1986 through 1989 for those students completing their university level sophomore year. Table 4.24 shows a slight increase in mean value for SYS12GPCT during the three years. The standard deviation decreased approximately .12 points indicating less variability. There were no significant swings in the maximum, median, and minimum categories.

Correlation coefficients were produced to show the association between USCGPA (university sophomore cumulative GPA) and the individual independent variables. Pearson's r was produced for all interval-level variables while the phi coefficient was calculated for the interval vs. nominal level variables (HSACRD and USFAID). Table 4.25 shows the individual correlation coefficients and the number of observations used for each calculation.

Table 4.23

Descriptive Statistics at End of Sophomore Year:
School System Enrollment by High School Graduation Year

	1986	1987	1988	86-88
N	1798	1919	1961	5678
Mean	21320	21654	28452	23896
Median	17352	26843	16615	16615
Mode	27195	26843	54725	54725
Std Dev	23536	23674	26590	24895
Maximum	112766	113101	112670	113101
Minimum	1079	1131	1271	1079
Range	111687	111970	111399	112022

Table 4.24

Descriptive Statistics at End of Sophomore Year:
Twelfth Grade Enrollment as a Percentage
of School System Enrollment by
High School Graduation Year

	1986	1987	1988	86-88
N	1798	1919	1961	5678
Mean	6.7	6.9	7.0	6.9
Median	6.8	7.2	7.2	7.0
Mode	7.0	7.5	7.2	7.5
Std Dev	1.06	1.02	0.94	1.02
Maximum	10.9	10.2	12.0	12.0
Minimum	2.2	1.7	2.8	1.7
Range	8.7	8.5	9.2	10.3

Table 4.25

Correlation Coefficients:
Sophomore Cumulative GPA Versus
Other Independent Variables

	1986	1987	1988	86-88
ACTCOMP N	.464* 1775	.457* 1888	.416* 1928	.446* 5591
HSGPA N	.557* 1798	.536* 1919	.581* 1961	.557* 5678
SYSFORL N	-.042 1798	-.061* 1919	-.050** 1961	0.051* 5678
SYSAEPP N	.018 1798	.002 1919	-.046** 1961	.011 5678
HSACRD(1) N	0.369 1798	0.389 1919	0.406 1961	0.247 5678
SYSE N	-.122* 1798	-.117* 1919	-.043 1961	-.083* 5678
SS12GPCT N	.012 1798	.044** 1919	.066 1961	.043* 5678
USFAID(1) N	0.447* 1798	0.423** 1919	0.437* 1961	0.321* 5678

Note: * significant at $p < .01$, ** significant at $p < .05$
(1) Phi coefficient used for HSACRD and USFAID

The r for USCGPA and ACTCOMP (ACT composite scores) declined slightly from year-to-year. The r values for each year and for the total timespan were statistically significant at the .01 level.

HSGPA (high school GPA) correlated with USCGPA with the values higher than any of the other independent variables. The individual r 's were all statistically significant at the .01 level.

The correlations of SYSFORL (school system percentage of students on free or reduced lunches) with USCGPA were negative for each year. All functions in each column in the table were statistically significant to at least the .05 level except for 1986 which was significant at the .10 level.

The r 's for SYSAEPP (school system adjusted expenditure per pupil) were not statistically significant for the total timespan or for any of the individual years.

The phi coefficient was used for HSACRD (high school SACS accreditation) and USCGPA. The highest coefficient was .406 in 1988, but none of the correlations were statistically significant at the .05 level.

The correlations for SYSE (school system enrollment) and USCGPA were negative for each year. This indicated that the larger system enrollments tended to reduce academic performance. The r value became less negative (more positive) from 1986 through 1988. The 1988 value was not statistically significant at the .05 level, but was significant at .10.

SYS12GPCT (school system percent of students in 12th grade) vs. USCGPA increased from year-to-year (.03 and .02). However, neither 1986 or 1987 were statistically significant at the .05 level (1987 was significant at .10 level).

The phi coefficient produced statistically significant values for each year in the all-years categories for the association of the students receiving financial aid during their sophomore year (USFAID) and their sophomore cumulative GPA (USCGPA).

Summary of Research Question 3

SYSAEPP and HSACRD were the only two variables that failed to show a statistically significant relationship with sophomore cumulative GPA. ACTCOMP, HSGPA, and USFAID reflected reasonably high levels of association. Despite being statistically significant, SYSFORL, SYSE, SS12GPCT had correlation coefficients very close to zero which reflected little if any linear association.

Research Question 4

Is there a regression model that utilizes the stated independent variables that will predict fourth-semester cumulative GPA for a student?

The predicting of cumulative GPA at the end of the sophomore year should be more accurate than predicting first semester GPA since many of the poorer students will fall by

the wayside and leave only the more able students in the surviving data base. Table 4.26 illustrates the results from the stepwise regression. Two additional variables, USFAID and SYSE were added to the prediction equation by the stepwise procedure. However, as it was shown in their contribution to the R^2 , neither made a meaningful contribution. HSGPA accounted for 32% of the variation in the dependent variable USCGPA. ACTCOMP was the number two contributor with almost 5% contribution. Whether or not a student was receiving financial aid (USFAID) accounted for almost 1% of the variation in the dependent variable. USFAID was the only variable that represented factors that took place during the college experience.

In general, the parameter estimates for the sophomore-year model were smaller than the parameter estimates for the freshman-year model. This was offset by the more positive intercept term in the sophomore equation. The intercept term was -0.010271, as contrasted by the freshman intercept term of -1.236955. The larger intercept meant that mathematically the student was starting from a higher base and that the larger contribution of each coefficient was not necessary to achieve more respectable GPAs.

The parameter estimate for HSGPA (high school GPA) was the highest of any of the independent variables. The value of .579414 indicated that for each unit of HSGPA the sophomore cumulative GPA changed by .579414 units.

Table 4.26

Regression Analysis for End of Sophomore Year
Cumulative GPA Model

Dependent Variable: USCGPA

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value
Model	7	865.77713	123.68245	497.058
Error	5583	1389.21256	0.24883	
C Total	5590	2254.98969		
Prob>F	0.00000			
Root MSE	0.49883	R-square	0.3839	
Dep Mean	2.68680	Adj R-sq	0.3832	
C.V.	18.56588	C(p)	5.0439	

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > T
INTERCEPT	1	-0.0103	0.0617	-0.167	0.8677
ACTCOMP	1	0.0333	0.0018	18.825	0.0001
HSGPA	1	0.5794	0.0152	38.173	0.0001
SYSFORL	1	-0.0028	0.0005	-6.054	0.0001
SYSAEPP	1	0.586E-4	0.0000	3.870	0.0001
HSACRD	1	0.0523	0.0228	2.292	0.0219
USFAID	1	0.1365	0.0149	9.188	0.0001
SYSE	1	-0.686E-6	0.0000	-2.204	0.0276

Variable	Number In	Partial R**2	Model R**2	F	Prob>F
HSGPA	1	0.3210	0.3210	2642.7710	0.0000
ACTCOMP	2	0.0472	0.3683	138.1051	0.0001
USFAID	3	0.0077	0.3759	68.65	0.0001
SYSFORL	4	0.0055	0.3814	49.6047	0.0001
SYSAEPP	5	0.0016	0.3831	14.6928	0.0001
HSACRD	6	0.0004	0.3834	3.1804	0.0746
SYSE	7	0.0005	0.3839	4.8564	0.0276

ACTCOMP (ACT composite scores) had a parameter estimate of .0333. Thus, each unit change in ACT composite score will increase or decrease the response function by .0333. A student with an ACT composite score of 25 would have had a predicted cumulative score greater than .17 over a student with an ACT composite score of 20.

USFAID (financial aid during sophomore year) had a parameter estimate of .13652. USFAID assumed values of 0 or 1 depending on whether a student received financial aid or not. The prediction equation indicated that students receiving financial aid would have had a cumulative GPA at the end of his/her sophomore year of .14 higher than students not receiving financial aid.

SYSFORL (percentage of students in school system on free or reduced lunches) had a parameter estimate of -0.002845. The negative value indicated that high values of SYSFORL indicated low standard of living conditions in the school district community.

SYSAEPP (school system adjusted expenditures per pupil) had a parameter estimate of .0000585 (It must be kept in mind that the scale for the independent variable is in whole dollars.).

The parameter estimate for HSACRD (high school accreditation) was .052328. HSACRD took on values of 0 for a student's high school not being accredited to a value of 1 for the high school being accredited. The sophomore-year model

indicated that a student coming from a SACS accredited high school would have had a predicted cumulative GPA of .05 higher than a student from a non-accredited high school. However, the coefficient's standard error term of .02283 was almost half the value of the coefficient itself. This indicates that there was a high amount of variability in the usefulness of the HSACRD coefficient.

SYSE (school system enrollment) was a variable that failed to be useful in the freshman-year model but was used in the sophomore year prediction model. It was the last variable entered into the equation and had the least statistical significance of any of the parameter estimates ($\alpha = .0276$). The value of $-.000000686$ indicated that very large school districts could have a small negative impact on the cumulative GPA.

Table 4.27 was used to show the stability of the coefficients if they were calculated for each year rather than a single equation for the entire timespan. The individual coefficient estimates appeared to be reasonably stable over time and did not change drastically from one year to the next.

Summary of Research Question 4

Research question four (4) dealt with the construction of a regression equation that used the candidate independent variables and that could be used for predicting sophomore cumulative GPA (USCGPA). A regression equation produced the

Table 4.27

Regression Parameter Estimates for End of
Sophomore-Year-Model by High School Graduation Year

Variable	----- Parameter Estimates -----			
	1986	1987	1988	Total
INTERCEPT	0.080903	-0.200062	-0.011008	-0.010277
ACTCOMP	0.033887	0.037154	0.026840	0.033323
HSGPA	0.525456	0.576762	0.656838	0.579414
SYSFORL	-0.001898	-0.002530	-0.004852	-0.002845
SYSAEPP	0.000110	0.875E-4	0.373E-4	0.585E-5
HSACRD	0.045129	0.070125	0.031101	0.052328
USFAID	0.107196	0.165887	0.130389	0.136520
SYSE	-0.142E-5	-0.203E-5	0.869E-6	-0.686E-6

"best" unbiased estimator of the response variable. The equation was:

$$\begin{aligned} \text{USCGPA} = & -0.010277 + 0.033323(\text{ACTCOMP}) + 0.579414(\text{HSGPA}) \\ & -0.002845(\text{SYSFORL}) + 0.000058547(\text{SYSAEPP}) + 0.052328(\text{HSACRD}) \\ & +0.136520(\text{USFAID}) -0.000000686(\text{SYSE}) \end{aligned}$$

The response function could be utilized by substituting the specified independent variables with values from the students' records and from Tennessee State Department of Education publications. The following example illustrates how the equation could be used.

Given: ACTCOMP = 24
 HSGPA = 3.2
 SYSFORL = 30%
 SYSAEPP = 2500
 HSACRD = 1 (yes)
 USFAID = 1 (yes)
 SYSE = 5000

The above values were then substituted into the prediction equation.

$$\begin{aligned} \text{USCGPA} = & -0.010277 + 0.033323(24) + 0.579414(3.2) - \\ & 0.002845(30) + 0.000058547(2500) + 0.052328(1) + \\ & 0.136520(1) -0.000000686(5000) \end{aligned}$$

$$\text{USCGPA} = 2.89$$

The confidence interval was constructed around the estimated mean value of UFGPA by using the MSE (mean square

error), the inverse of the XX' matrix $(XX')^{-1}$ and the corresponding t statistic.

$$MSE = 0.49883 \quad t = 1.96$$

The confidence interval was calculated as ± 0.02 . Thus, there was less than 5% chance that the true value for UFGPA will be less than 2.86 or greater than 2.92 when using the above values for ACTCOMP, HSGPA, SYSFORL, SYSAEPP, HSACRD, USFAID, and SYSE.

Research Question 5

Is there any statistically difference in students' fourth semester collegiate GPA and their first semester collegiate GPA?

A variable was created that was the difference between USCGPA and UFCGPA scores ($DIFFGPA = UFCGPA - USCGPA$). A t -test was used to test the null hypothesis that the population mean was zero (SAS/STAT User's Guide, Volume 2, pp. 1638-1639). Only those students who had grades in both the fall semester of their freshman year and at the end of their sophomore year were included, thus disregarding any dropouts.

Table 4.28 shows that the mean value for the difference in freshman cumulative GPA and sophomore cumulative GPA was negative and indicates that, on the average, the cumulative GPA would be lower at the end of the sophomore year than for the first semester GPA. However, the decrease was not drastic, and was less than one tenth of a quality point.

Table 4.28

Analysis of Statistical Difference
in First Semester Cumulative GPA and
End of Sophomore-Year Cumulative GPA

N	Mean	Std. Error	T	Prob > T
5678	-0.07	0.0071	-9.9003	0.0001

There was a statistically significant difference in the first semester cumulative GPA and the sophomore year fourth semester cumulative GPA.

Research Question 6

Is there any statistical difference in the independent variables used to predict first semester GPA and the independent variables used to predict fourth semester cumulative GPA?

Research questions one and two were used to examine the independent variables and their relationships with first semester collegiate-cumulative GPA. Research questions three and four were directed at the variables associated with the cumulative GPA at the end of the sophomore year. Research question five was used to examine if there was any statistically significant difference in the cumulative GPA at the end of the first semester of the freshman year and the cumulative GPA at the end of the sophomore year. Thus research question six was designed to determine if there were statistically significant differences in the explanatory variables used in the prediction equations for the students that persisted through the end of their sophomore year and those who did not. Tables 4.29 through 4.36 display the various statistical tests for survivor versus non-survivor groups. Each independent variable was examined with the appropriate test depending on whether the variable was nominal

or interval-level data. If the variable was interval level, then a test was done to insure that it was normally distributed and the two groups had equal variances.

HSACRD (high school accreditation by SACS) is a nominal data type and was restricted to the values of 0 (no) and 1 (yes). A 2x2 frequency table was constructed (Table 4.29) to display the cell values for survivor class vs. HSACRD categories. The null hypothesis (H_0) was tested to determine if there was no statistical difference in the high school accreditation by SACS of students who persisted through their sophomore year and those who did not persist. A chi-square statistic was calculated to test the hypothesis. The chi-square value of 1.602 was not large enough to justify rejecting the null hypothesis. Therefore, there is not a statistically significant difference in students who come from SACS accredited high schools and whether or not they persist until the end of their collegiate sophomore year.

ACTCOMP (ACT composite scores) were treated as interval level data and could be tested with parametric tests if the assumption of data independence and normality were satisfied. The variable failed to be normally distributed and a non-parametric analysis of variance test was applied. The students who survived had a mean ACT composite score of 1.14 points higher than those who did not survive. An F value of 117.8 resulted in a probability level of 0.0001. Thus, the null hypothesis (H_0 : there is no statistical difference in the

Table 4.29

Analysis of Persistence by SACS High School Accreditation

	Frequency Percent Row Pct Col Pct	HSACRD		Total
		NO	YES	
SURVIVOR	NO	276	2061	2337
		3.44	25.71	29.16
		11.81	88.19	
		30.98	28.93	
	YES	615	5063	5678
		7.67	63.17	70.84
		10.83	89.17	
		69.02	71.07	
	Total	891	7124	8015
		11.12	88.88	100.00

Chi-square = 1.605 df = 1 p = 0.205

Table 4.30

Analysis of Persistence
by ACT Composite Scores

SURVIVOR	N	Mean	Among MS	Within MS
NO	2291	21.0	2131.73	18.0963
YES	5591	22.2		

F = 117.8 df = (2290,5590) p = 0.0001

Table 4.31

Analysis of Persistence
by High School GPA

SURVIVOR	N	Mean	Std Dev	Std Error
NO	2337	2.94	0.5423	0.0112
YES	5678	3.18	0.5108	0.0068

$F' = 1.13$ $df = (2336, 5677)$ $p = 0.0005$

Table 4.32

Analysis of Persistence
by Percentage of Students in School System
on Free or Reduced Lunches

SURVIVOR	N	Mean	Std Dev	Std Error
NO	2337	34.8	15.93	0.3295
YES	5678	34.6	15.77	0.2093

$F' = 1.02$ $df = (2336, 5677)$ $p = 0.5663$

Table 4.33

Analysis of Persistence
by Adjusted Expenditures Per Pupil

SURVIVOR	N	Mean	Std Dev	Std Error
NO	2337	2354	491	10.2
YES	5678	2365	478	6.3

$F' = 1.05$ $df = (2336, 5677)$ $p = 0.1242$

Table 4.34

Analysis of Persistence
by Students Receiving Financial Aid During
Their Freshman Year

SURVIVOR	Frequency Percent Row Pct Col Pct	UFFAID		Total
		NO	YES	
NO		1406	931	2337
		17.54	11.62	29.16
		60.16	39.84	
		29.95	28.04	
YES		3289	2389	5678
		41.04	29.81	70.84
		57.93	42.07	
		70.05	71.96	
Total		4695	3320	8015
		58.58	41.42	100.00

Chi-square = 3.415 $df = 1$ $p = 0.065$

Table 4.35

Analysis of Persistence
by School System Enrollment

SURVIVOR	N	Mean	Std Dev	Std Error
NO	2337	24415	25383	525.1
YES	5678	23896	24895	330.4

$F' = 1.04$ $df = (2336, 5677)$ $p = 0.2607$

Table 4.36

Analysis of Persistence
by Percentage of Students in Twelfth Grade
as a Percentage of Total School System Enrollment

SURVIVOR	N	Mean	Std Dev	Std Error
NO	2337	6.8	1.02	0.0211
YES	5678	6.9	1.02	0.0135

$F' = 1.00$ $df = (2336, 5677)$ $p = 0.9859$

ACT composite scores of those who persist through their sophomore year and those who did not persist) was rejected and the results indicated that there was a statistically significant difference in the ACT composite scores of those students who survived through the end of their sophomore year and those students who did not survive.

Table 4.31 was used to show any statistical difference in survivor and non-survivor groups regarding their high school GPA (HSGPA). Survivors had a mean value for HSGPA of approximately .24 points higher than those who did not survive. The null hypothesis (H_0) states that there was no statistical difference in the high school GPA for those students who remained in school through the end of their sophomore year and for those who did not. The F test reflected that there was a statistically significant difference in the high school GPA of survivors and non-survivors. The standard error term of the non-survivors was almost twice the magnitude of the error term for the survivors. The standard deviation was only slightly higher for the non-survivors than for those who remained in school through the end of their sophomore year. The results indicated that H_0 was rejected and that there is a statistically significant difference in HSGPA between survivors and non-survivors.

The mean and standard deviation values for SYSFORL (school system percentage of students on free or reduced lunches) were virtually the same in Table 4.32 for both survivor classes. The survivor = no standard error term at .329 was more than 50% larger than the .209 value for the survivor = yes category. The F test results found no statistically significant difference in SYSFORL for the survivor classes. Thus, the null hypothesis (H_0 : there is no statically significant difference in SYSFORL for students who survived through their sophomore year and those students who did not) could not be rejected.

The mean values for SYSAEPP (school system adjusted expenditures per pupil) of survivor classes were different by only a little more than ten dollars or about four tenths of one percent of the mean. The null hypothesis (H_0) states that there is no statistically significant difference in SYSAEPP for students who persist through their sophomore year and those who do not. The F test, as shown in Table 4.33, failed to reveal a statistically significant difference in SYSAEPP between the students who were in school at the end of their sophomore year and those students who were not.

Both survivor class (yes,no) and financial aid indicators (yes,no) were nominal (categorical) data. A frequency table was developed (Table 4.34) and a chi-square test was applied to the data. The resulting chi-square value of 3.415 was significant at the .10 level, however it was not significant

at the pre-determined alpha of .05. Thus, the null hypothesis that there was no statistically significant difference in whether students received financial aid and if they remained enrolled to the end of their sophomore year or not could not be rejected.

SYSE (school system enrollment) when categorized by survivor classes, revealed some interesting results. The mean, standard deviation, and standard error terms, as shown in Table 4.35, were all larger for the survivor = no class than their corresponding values in the survivor = yes category. However, the F test failed to show any statistically significant difference in SYSE and the survivor classes. Thus, the null hypothesis (H_0 : there is no statistically significant difference in SYSE and students who remained enrolled through their sophomore year and for those students who did not survive) could not be rejected.

SS12GPCT (percentage of students in the school system that were enrolled in the twelfth grade) revealed little insight. The mean and standard deviation terms were virtually the same for both survivor classes. The standard error term for the survivor class = no was almost twice the value of the survivor = yes category. The null hypothesis (H_0) was stated that there is no statistically significant difference in SS12GPCT for students who remain in school until the end of their sophomore year and those who do not. However, the F test failed to be statistically significant and the null hypothesis

could not be rejected.

Summary of Research Question 6

HSGPA (high school GPA) and ACTCOMP (ACT Composite scores) were both statistically significant at the .01 level. The mean value for HSGPA was found to be almost a fourth of a quality point (.24) higher for students who persisted through their sophomore year as compared to students who did not survive. Mean ACTCOMP values were approximately 1.14 higher for survivors. HSACRD (high school accreditation by SACS), SYSFORL (school system percentage of students on free or reduced lunches), SYSE (school system enrollment), UFFAID (freshmen receiving financial aid), and SYS12GPCT (percentage of school system enrollment as represented by the 12th grade enrollment) were not statistically significant at the .05 level.

UNANTICIPATED RESULTS

Unanticipated results will often crop up when research questions are addressed. The data revealed that the incoming freshman classes were heavily made up of a few school districts and high schools. For example, it is revealed in Table 4.37 that almost 30% of the incoming freshmen over a four-year period came from ten high schools. Even more

dramatic was that about 45% of the students in this timespan came from five school systems.

Table 4.37
First-Time Freshmen
Supplied by the Top Feeder High Schools
And School Systems by Year

	1986	1987	1988	1989	86-89
No. of students in study	2562	2752	2701	2312	10327
Top 10 feeder schools...No.	700	814	803	752	3069
...%	27.3	29.6	29.7	32.5	29.7
Top 5 feeder systems....No.	1101	1245	1210	1056	4612
....%	43.0	45.2	44.8	47.7	44.7

CHAPTER 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

SUMMARY

Background

Admission into college is based upon the hope and anticipation of academic success. Due primarily to high drop-out rates and the poor performance of many students, the cost to universities of recruiting and retaining students continues to increase. The inability to predict the academic achievement of college bound students is an ongoing problem that can be understood with systematic research.

Purpose

Are there a set of variables that will contain the components of collegiate readiness and will perhaps provide insight into the selection process? This question is one that is of particular interest to those charged with admitting students at the freshman level. At the University of Tennessee the question more specifically is: Are there variables that are readily available through The Tennessee State Department of Education and from the University of Tennessee that can be used for constructing and utilizing academic prediction methodologies without having to resort to lengthy and costly surveys and data collection schemes?

The majority of studies in recent years have emphasized

instructional and curriculum processes and individual student characteristics rather than attempting to identify school system classifications and broad indicators of student performance. Earlier studies have generally restricted their investigations to identifying relationships with college freshmen and a very limited number of secondary school variables. This research provided an opportunity for a longitudinal study over a four-year time period and eliminated short term influences. In this investigation, there was an examination of the entire population of Tennessee high school seniors who entered the University of Tennessee from 1986 through 1989. The students were tracked through the end of their sophomore year. To summarize briefly, the purpose of this research was to construct a mathematically based model that could be used to predict student cumulative grade point average (GPA) at the University of Tennessee for graduates of Tennessee public high schools.

Significance

The significance of this research was addressed from the perspectives of the various stakeholders such as students, parents, high schools, and the University of Tennessee. The "freedom to choose" movement and the many variations such as vouchers affect parents, students, and schools. If they ever become reality, freedom to choose and voucher systems will certainly pressure public high schools to demonstrate that

their graduates can be successful in higher education. Conversely, students and their parents must make decisions that assure readiness necessary for collegiate academic success. Representatives of higher education would like to spend their time and resources attracting and retaining students who can successfully complete their chosen undergraduate program.

Methodology

A review of the recent literature regarding the prediction of a student's readiness for college reveals several different paths that could be pursued. However, it is possible to identify a taxonomy of major categories that can support systematic research. The hierarchical categories such as the student, the family, the high school, and the school district have an impact on academic achievement. Recently, researchers have attempted to encapsulate, identify, study, and measure the contributions of these components. (Chubb & Moe, 1990, p. 111).

In the current study, student records were supplied by the UTK Student Services Division and were merged with data obtained from the Tennessee State Department of Education. Each record represented the academic performance of a student who graduated from a Tennessee secondary school from 1986 through 1989 and who attended UTK the following summer or fall. The final data set contained 10,327 student records and

represented 117 Tennessee public school systems and 257 high schools.

Several variables were identified that could be supported by the literature and were available from public sources or from the UTK student records system. The candidate independent variables were ACT Composite scores (ACTCOMP), high school GPA (HSGPA), school system percentage of students on free or reduced lunches (SYSFORL), school system adjusted expenditures per pupil (SYSAEPP), school system enrollment (SYSE), school system twelfth grade enrollment as a percentage of total enrollment (SYS12GPCT), and whether or not a student was receiving financial aid (UFFAID and USFAID). The two dependent variables were first semester freshman cumulative GPA (UFCGPA) and second semester-sophomore cumulative GPA (USCGPA). The following six research questions were developed to be addressed in the research.

1. Is there a statistically significant relationship among any of the above independent variables and students' first semester collegiate cumulative GPA?
2. Is there a regression model that utilizes the stated independent variables that will predict first semester cumulative GPA for a student?
3. Is there a statistically significant relationship among any of the above independent variables and students' fourth semester collegiate cumulative GPA?
4. Is there a regression model that utilizes the stated

independent variables that will predict fourth-semester cumulative GPA for a student?

5. Is there any statistically significant difference in students' fourth semester collegiate GPA and their first-semester collegiate GPA?

6. Is there any statistically significant difference in any of the independent variables used to predict academic performance of students who persisted through their sophomore year and those who did not?

Research questions 1 and 2 pertained to the candidate independent variables and the construction of a regression equation for the prediction of first semester collegiate cumulative GPA.

Results

The investigation regarding research questions 1 and 2 revealed that six of the eight candidate independent variables showed a statistically significant relationship with the response variable freshman cumulative GPA (UFCGPA). However, the r values for SYSFORL, SYSAEPP, SYSE, and SYS12GPCT were each very close to zero and could be interpreted as having very little, if any, association with UFCGPA. The r value for HSGPA (.486) and ACTCOMP (.379) did show a relatively strong linear association.

A stepwise linear regression equation was calculated that produced the "best" unbiased estimator of freshman cumulative

GPA. The equation was:

$$\text{UFGPA} = -1.2370 + 0.0426(\text{ACTCOMP}) + 0.7901(\text{HSGPA}) \\ -0.0041(\text{SYSFORL}) + 0.00011(\text{SYSAEPP}) + 0.1231(\text{HSACRD})$$

$$R^2 = .2875 \quad F = 818.815 \quad \text{Prob}>F = 0.0000$$

$$\text{MSE} = .81982 \quad C(p) = 3.57 \quad (p)\text{arameters} = 5$$

The independent variables in the regression model explained 29% of the variation in the response function. The F statistic was statistically significant, the mean square error term (MSE) was relatively small and the C(p) term was reasonably close to the number of parameters (5) in the model.

Research questions 3 and 4 were the parallel questions to questions 1 and 2 except they were concerned with the second semester-sophomore year cumulative GPA (USCGPA) as the response variable. The candidate independent variables were ACTCOMP, HSGPA, SYSFORL, SYSAEPP, SYSE, SYS12GPCT, and USFAID (whether or not a student was receiving financial aid during his/her sophomore year).

SYSAEPP and HSACRD were the only two variables that failed to show a statistically significant relationship with sophomore cumulative GPA. As in research question 1 that dealt with the freshman cumulative GPA, only ACTCOMP and HSGPA reflected reasonably high r levels. Variables such as SYSFORL, SYSE, SS12GPCT, and USFAID failed to achieve values that indicated a strong linear association.

A regression equation produced the "best" unbiased estimator of sophomore cumulative GPA. The equation was:

$$\begin{aligned} \text{USCGPA} = & -0.0103 + 0.0333(\text{ACTCOMP}) + 0.5794(\text{HSGPA}) \\ & -0.0028(\text{SYSFORL}) + 0.58\text{E-}4(\text{SYSAEPP}) + 0.0523(\text{HSACRD}) \\ & +0.1365(\text{USFAID}) -0.686\text{E-}6(\text{SYSE}) \end{aligned}$$

$$R^2 = .3839 \qquad F = 497.058 \qquad \text{Prob}>F = 0.0000$$

$$\text{MSE} = .49883 \qquad C(p) = 5.04 \qquad (p)\text{arameters} = 7$$

The independent variables in the regression model explained 38% of the variation in the response function. The F statistic was significant at the .05 level while the mean square error term (MSE) was relatively small and the C(p) term was reasonably close to the number of parameters (7) in the model.

Research questions 5 and 6 were designed to determine if anything significant was different between examining first semester results and the fourth semester (end of sophomore year) results.

The two dependent variables (UFCGPA and USCGPA) were analyzed to determine if there was a statistically significant difference in them. An examination revealed that there was a statistically significant difference between the first-semester cumulative GPA and the sophomore-year fourth semester cumulative GPA. The cumulative GPA for students tended to decrease between their freshman and sophomore years. The decrease was not drastic and, in fact, was less than one tenth of a quality point.

Students were identified as those that persisted until the end of their sophomore year and those that did not persist

past their first semester. Each independent variable was examined to determine if there was a statistically significant difference between those who persisted and those who did not.

Results of the statistical tests indicated that students who persisted until the end of their sophomore year had higher ACT composite scores and higher high school GPA's than those who did not persist. Both of these variables were statistically significant at the .01 level. SACS high school accreditation (HSACRD), percentage of students on free or reduced lunches (SYSFORL), adjusted expenditures per pupil (SYSAEPP), freshman financial aid variable (UFFAID), the 12th grade enrollment as a percentage of total system enrollment (SYS12GPCT) and school system enrollment (SYSE) were not statistically significant.

Discussion

The finding that high school GPA (HSGPA) and ACT Composite scores (ACTCOMP) were the two most important variables in the prediction equations for both first-semester cumulative GPA and fourth-semester cumulative GPA was consistent with prior research. The two in combination have been judged to be more accurate than when used alone as reported by Noble (1990) in an exhaustive research effort. Willingham (1990) also advocated the joint use of achievement tests and high school GPA for college-admission requirements.

The fact that students with greater high school GPA's and higher ACT Composite scores were more likely to remain in

school than students with lower scores should not be a big surprise. The overwhelming evidence of the usefulness of HSGPA and ACTCOMP for predicting academic success (at least first semester performance) should add credence to their standing out among the other independent variables.

School-system size effects were represented by school system enrollment (SYSE) and the twelfth grade enrollment as a percentage of total system enrollment (SYS12GPCT). Of the two, only SYSE was able to enter the stepwise regression model and this applied only to the sophomore-year prediction model. The opinions of the effects of high school size and school system size by researchers are mixed. Some report that economics of scale allow for efficiencies that smaller systems can not enjoy (Chubb & Moe, 1990; Conant, 1959). Negative influences were reported by those concerned with academic achievement. The results of this research (i.e., the negative correlation coefficients (UFCGAP versus SYSE and USCGPA versus SYSE) and the extremely small regression coefficient in the sophomore-year prediction model) tend to agree with Goodlad (1984) and Boyer (1983) regarding the undesirable effect of large schools and large school districts. However, it appears that system size as a regression parameter does not affect academic achievement to a proportion of concern.

In the past, a great deal of interest was shown regarding the use of expenditure per pupil as a predictor of academic performance. The results of this research were consistent

with findings as reported by Hanushek (1986), Childs and Shakeshaft (1986) and Peter Coleman (1986). Perhaps too much should not be read into the findings regarding the contribution of expenditures per pupil as a predictor of academic achievement. First of all, a single value for expenditure per pupil is an over simplification of a complex process. The number fails to convey any meaning of the distributions of the funds at the various grade levels, nor how much was spent on individual academic content areas, nor how much was spent on non-academic endeavors such as transportation, cafeterias, or sports programs versus how much was spent directly on academics. The actual time the money was distributed during the school year was not shown. In short, this indicator is subject to much misuse.

Virtually no research was encountered regarding the use of high school accreditation as a measure of school effectiveness. However, high school accreditation by SACS (HSACRD) was present in a minor way in both prediction equations. It was not found to be statistically significant when survivors versus non-survivors were examined. The fact that most of Tennessee's large high schools are accredited by SACS and that a large percentage of incoming freshmen are from large high schools overshadows any contribution of this variable.

This research must be viewed as descriptive in nature and should be used as prescriptive only after very careful

consideration. Regression equations result from fitting equations to data where the data values are within discrete ranges. A model is only valid in predicting a response variable when values for the independent variables are within the range restrictions (bounds) of the fitted data.

Despite the fact that this research was of an institutional nature, there are several pertinent factors that make the study important to the literature base. One such factor is that all the graduates of a state's public high schools over a four-year period who attended the major state land grant institution immediately following high school were included. Another factor is that the research took a multi-variable approach to identify items (beyond the usual high school GPA and assessment test scores) that can be used to predict collegiate academic achievement. However, one of the most important areas, is that the research looked beyond the traditional first semester freshman GPA and followed the students through the end of their sophomore year.

Application of the research findings can be used in several ways. The model can be used to justify higher entrance requirements necessitated by budgetary constraints. This research found that high school GPA and ACT scores are the most useful variables in predicting academic achievement and persistence. Thus, the raising of college entrance requirements could be justified within the model framework.

Another direction could be to examine the unexplained

variation in the prediction equations. More variation is not accounted for than is identified. Thus, intervention schemes, such as identification of the "at risk" students could possibly be very useful in assisting students in overcoming shortcomings and deficiencies due to their backgrounds and experiences. This latter approach would be much more constructive and productive for not only the students, but would increase retention and possibly result in reduced drop out rates.

CONCLUSIONS

The large number of individual student records and the amount of representation from a few large high schools and school districts may have masked some results from smaller schools. However, one of the primary objectives of the research was to construct prediction equations as they applied to incoming freshmen from Tennessee public high schools. Most of the incoming freshmen are represented by several large school districts and the prediction equations reflect these factors.

Education is a cumulative process. High school GPA (HSGPA) is the resulting measurement produced by the product of grades and course hours divided by the sum of the course hours. HSGPA is not a point estimate, rather it is a measurement of performance over a prolonged period. Factors such as family influences, parental economic status, and even

native intelligence contribute to both the high school and college academic experience. Therefore, it should not come as any great surprise that HSGPA is the most important predictor of college GPA. All of the above stated factors are present in HSGPA. They may not be represented in full strength, but an argument can be made that they certainly had a significant input.

ACT Composite scores are used as an indicator of collegiate readiness. This provides a means of measuring all high school students with the same instrument. ACT Composite scores eliminate different teacher grading practices and other variables that make it difficult to filter out differences in school systems and high school experiences. The other variables found in the regression equations such as school system percentage of students on free or reduced lunches (SYSFORL), adjusted expenditures per pupil (SYSAEPP), and high school accreditation (HSACRD) are useful and do make a contribution to the prediction equation and should be used when possible. However, high school GPA and ACT Composite scores, when used together, are the two primary contributors to the prediction equation.

RECOMMENDATIONS

Several factors that would contribute to better predictor equations were either purposely omitted or have been identified in the research. The high end of the spectrum

includes students who elected to go to colleges with more selective admission policies. Probably very little could be done to include this in the data set. However, the most obvious omission in the predictor model is the exclusion of those students who were either rejected from admission to UTK or elected to enter other state higher education institutions. A broader range of students could be identified in Tennessee by (1) including the remaining undergraduate programs at University of Tennessee locations. (Chattanooga and Martin). (2) including the state supported universities which are not part of the University of Tennessee system. (Austin Peay State University, East Tennessee State University, Tennessee State University, Memphis State University, Middle Tennessee State University, and Tennessee Technological University), and (3) including the state community college system. (Chattanooga State Technical Community College, Cleveland State Community College, Motlow State Community College, Roane State Community College, Nashville Technical Community College, Columbia State Community College, Walters State Community College, Volunteer State Community College, Jackson State Community College, Shelby State Community College, and Dyersburg State Community College). The state supported higher education institutions are coordinated through the Tennessee Higher Education Commission and have somewhat common reporting and curricular practices.

The new enhanced ACT Assessment test should be evaluated

to determine the impact upon its use as a screening device for admissions officers. The revised examination uses different components to measure student achievement and the use of the composite cannot be compared evenly. High school students began taking the examination in 1990, thus a data base is accumulating.

Some gain could be made by comparing the individual ACT content areas with performance on like subjects in college. For example, the ACT math scores could be matched with the grades that students obtained in collegiate mathematics classes.

LIST OF REFERENCES

LIST OF REFERENCES

- Amos, O. M., Jr. (1987). Economics concepts, analysis, and applications. Belmont, CA: Wadsworth Publishing Company.
- Annual statistical report of the department of education for the scholastic year ending June 30, 1986. Nashville: State Department of Education.
- Annual statistical report of the department of education for the scholastic year ending June 30, 1987. Nashville: State Department of Education.
- Annual statistical report of the department of education for the scholastic year ending June 30, 1988. Nashville: State Department of Education.
- Annual statistical report of the department of education for the scholastic year ending June 30, 1989. Nashville: State Department of Education.
- Boyer, E. L. (1987). College: The undergraduate experience in America. New York: Harper and Row.
- Boyer, E. L. (1983). High school: A report on secondary education in America. New York: Harper and Row.
- Barton, G. W. (1989). Freshman retention: An examination of similarities and differences among returning and non-returning students one year after college admission (Doctoral dissertation, New Mexico State University, 1988). Dissertation Abstracts International, 50, 634A.
- Bosserman, D. C. (1985). Idaho public high school size as an indicator of success in graduation from the University of Idaho 1980 - 1984 (Doctoral dissertation, University of Idaho, 1985). Dissertation Abstracts International, 46, 1527A.
- Childs, T. S. & Shakeshaft, C. (1986, Fall). A meta-analysis of research on the relationship between educational expenditures and student achievement. Journal of Education Finance, 12, 249-263.
- 'Choice' must include private/church schools, voucher advocate says. (1990, February). TEA News, p. 1,7.
- Chubb, J. E. & Moe, T. M. (1990). Politics, markets, America's schools. Washington, D.C.: The Brookings Institution.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Frederic, D. W., & York, R. L. (1966). Equality of educational opportunity. Washington: National Center for Educational Statistics.

Coleman, J. S. & Hoffer, T. (1987). Public and private high schools: The impact of communities. New York: Basic Books.

Coleman, P. (1986, Summer). The good school district: A critical examination of the adequacy of student achievement and per pupil expenditures as measures of school district effectiveness. Journal of Education Finance, 12, 71-96.

Commissioner's report card 1985-86: A report about your school system (1987, Jan.). Nashville: Tennessee Department of Education.

Commissioner's report card 1986-87: A report about your school system (1988, Jan.). Nashville: Tennessee Department of Education.

Commissioner's report card 1987-88: A report about your school system (1989, Jan.). Nashville: Tennessee Department of Education.

Commissioner's report card 1988-89: A report about your school system (1990, Jan.). Nashville: Tennessee Department of Education.

Conant, J. B. (1959). The American high school today: A first report to interested citizens. New York: McGraw-Hill.

1986-87 Directory of public schools, approved nonpublic, special state schools and the department of education state of Tennessee (1986, Sept.). Nashville: State Department of Education.

1987-88 Directory of public schools, approved nonpublic, special state schools and the department of education state of Tennessee (1987, Sept.). Nashville: State Department of Education.

1988-89 Directory of public schools, approved nonpublic, special state schools and the department of education state of Tennessee (1988, Sept.). Nashville: State Department of Education.

- 1989-90 Directory of public schools, approved nonpublic, special state schools and the department of education state of Tennessee (1989, Sept.). Nashville: State Department of Education.
- Draper, N. R. & Smith, H. (1966). Applied regression analysis. New York: John Wiley & Sons, Inc.
- Fact book 1988-89 (1989, March). Knoxville: The University of Tennessee.
- Fienberg, S. E. (1978). The analysis of cross-classified categorical data. Cambridge, MA: The MIT Press.
- Firestone, W. A. (1987, Oct.). Meaning in method: The rhetoric of quantitative and qualitative research. Educational Researcher, 16-21.
- Fleiss, J. L. (1981). Statistical methods for rates and proportions. 2nd ed. New York: John Wiley & Sons.
- Freund, R. J. & Littell, R. C. (1986). SAS system for regression. Cary, NC: SAS Institute Inc.
- Freund, R. J., Littell, R. C. & Spector, P. C. (1986). SAS system for linear models. Cary, NC: SAS Institute Inc.
- Gay, L. R. (1987). Educational research: competencies for analysis and application. 3rd ed. Columbus, OH: Merrill Publishing Co.
- Good, T. L. & Brophy, J. E. (1986). School Effects. In M. C. Whittrock (Ed.), Handbook of research on teaching (pp. 571- 602). New York: The American Educational Research Association.
- Goodlad, J. I. (1984). A place called school: prospects for the future. New York: McGraw-Hill.
- Hanushek, E. A. (1986, Sept.). "The economics of schooling: production and efficiency in public schools". Journal of Economic Literature. Vol. XXIV, p. 1141-1177.
- Hess, B. B., Markson, E. W. & Stein, P. J. (1988). Sociology. New York: Macmillan Publishing Company. Third Ed.
- Hickey, S. K. (1988). The degree to which effective schools research is being implemented in a rural Tennessee county as seen by principals and teachers (Doctoral Dissertation, The University of Tennessee, Knoxville, 1987).

- Johnston, J. D. (1977). The university compensatory program student: placement, course performance, and high school basic skill scores (Doctoral dissertation, University of Georgia, 1977). Dissertation Abstracts International, 38, 1252A.
- Kleinbaum, D. G. & Kupper, L. L. (1978). Applied regression analysis and other multivariable methods. North Scituate, MA: Duxbury Press.
- Lieberman, M. (1989). Privatization and educational choice. New York: St. Martin's Press.
- Maurovich, R. L. (1978). The effectiveness of the high school record as a predictor of academic success for freshman students at Michigan State University (Doctoral dissertation, Michigan State University, 1978). Dissertation Abstracts International, 39, 1373A.
- McClellan, R. A. & Sanders, W. L. (1983). Objective component of teacher evaluation: A feasibility study. Knoxville: The University of Tennessee.
- Moline, A. E. (1987). Financial aid and student persistence: An application of causal modeling. Research in Higher Education, 26(2), 130-147.
- Noble, J. P. (1990, April). Improving college grade predictions using expanded high school coursework information and standardized test scores as predictors. Paper presented at the national conference of the American Educational Research Association, Boston, MA.
- Noble, J. P. & Sawyer, R. (1987). Predicting grades in specific college freshman courses from ACT test scores and self-reported high school grades. (ACT Research Report Series 87- 20). Iowa City: The American College Testing Program.
- Proceedings (March-April, 1991). Southern Association of Colleges and Schools, 43(2), Decatur, GA.
- Public schools of choice. (1990). Alexandria, VA: the Association for Supervision and Curriculum Development.
- Ravitch, D. (1985). The schools we deserve: Reflections on the educational crises of our time. New York: Basic Books.
- Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. In L. S. Shulman and G. Sykes (Eds.). Handbook of Teaching and Policy. New York: Longman.

- SAS/IML software usage and reference, (1989). Cary, NC: SAS Institute Inc., Version 6, First Edition, 501 p.
- SAS language guide for personal computers, release 6.03 edition. (1988). Cary, NC: SAS Institute Inc.
- SAS/STAT user's guide, Version 6, 4th ed., Volume 2. (1989). Cary, NC: SAS Institute Inc.
- Sawyer, R. (1985). Using demographic information in predicting college freshman grades. (Research Report No. 87). Iowa City: The American College Testing Program.
- Sawyer, R. & Maxey E. J. (1979). The validity over time of college freshman grade prediction equations. (ACT Research Report 80). Iowa City: The American College Testing Program.
- Siegel, S. (1956). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill.
- Smith, C. E. (1990). 21st Century: Statewide goals and objectives for educational excellence. Nashville: Tennessee State Department of Education.
- Standards of the commission on secondary schools (1991). Decatur, GA: Southern Association of Colleges and Schools.
- Spring, J. (1988). Conflict of interests: The politics of American education. New York: Longman.
- Walpole, R. E. and Myers, R. H. (1989). Probability and statistics for engineers and scientists. New York: Macmillan Publishing Company, Fourth Edition.
- Warren, J., Innes, T., Quarles, D. (1977). Tennessee assessment of schools results report 1975-1976. Knoxville: The University of Tennessee, State Testing and Evaluation Center.
- Willingham, W. W., Lewis, C., Morgan, R., Ramist, L. (1990). Predicting college grades: An analysis of institutional trends over two decades. New York: The College Board and Educational Testing Service.
- Zainuba, M. A. (1985). Relationship of selected admissions criteria and college graduating grade point average as a determinant of student's academic success (Doctoral dissertation, Texas Southern University, 1984) Dissertation Abstracts International, 46, 910A.

VITA

Ken R. Adcock was born in Smithville, Tennessee on April 20, 1941 and graduated from DeKalb County High School in May, 1959. He received a B.S. in Mathematics from Tennessee Technological University in August, 1964 and a MBA from the University of Tennessee at Chattanooga in August, 1976.

Adcock worked one year as a junior high school mathematics teacher in Madison County, Alabama. He then joined Arnold Engineering Development Center (AEDC) in Tullahoma, Tennessee as a Scientific Programmer. He left AEDC in March, 1970 to become a member of the Management Information Services group with Bowater Incorporated in Calhoun, Tennessee. Adcock began his career with Bowater as a programmer and served in management science, systems development, and information center capacities. He left Bowater Incorporated in August, 1989 where he was Manager of the Information Center in Greenville, South Carolina. He joined AGC Corporation in Cleveland, Tennessee as Vice President of Research and Development and is currently serving as President.

Adcock is a member of the American Educational Research Association (AERA), Society for the History of Technology (SHOT), The Institute for Management Sciences (TIMS), the Association for Computing Machinery (ACM), the Mid-Southeast ACM chapter, and the IEEE Computer Society.