3-2002

Admissions Criteria and Academic Performance in the AFIT Graduate Cost Analysis Program

Kenneth R. Garwood

ADMISSIONS CRITERIA AND ACADEMIC PERFORMANCE IN THE AFIT GRADUATE COST ANALYSIS PROGRAM

THESIS

Kenneth R. Garwood, Captain, USAF

AFIT/GAQ/ENV/02M-08

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GAQ/ENV/02M-08

ADMISSIONS CRITERIA AND ACADEMIC PERFORMANCE IN THE AFIT GRADUATE COST ANALYSIS PROGRAM

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Acquisition Management

Kenneth R. Garwood, BS

Captain, USAF

March 2002

AFIT/GAQ/ENV/02M-08

ADMISSIONS CRITERIA AND ACADEMIC PERFORMANCE IN THE AFIT GRADUATE COST ANALYSIS PROGRAM

Kenneth R. Garwood, B.S.
Captain, USAF

Approved:

//SIGNED//
Mark A. Ward, Major, USAF (Chairman) date

//SIGNED//
William K. Stockman, Lt Col, USAF (Member) date

//SIGNED//
Daniel E. Reynolds (Member) date

Acknowledgements

I want to thank the staff of the Registrars office for their patient and professional assistance with the collection of data for this research. I would like to express my sincere appreciation to Professor Dan Reynolds for his enthusiasm and expert advice.

Most importantly, I am forever grateful to my family for their support and patience through the many long nights and work-shortened weekends.

Kenneth R. Garwood

Table of Contents

Page

List of Figures

List of Tables

AFIT/GAQ/ENV/02M-08

Abstract

This research examined the criteria used by the Air Force Institute of Technology (AFIT) to determine an applicant's academic eligibility to attend the in-residence Graduate Cost Analysis (GCA) program. The objectives were to evaluate the predictive capability of the current criteria, evaluate other potential predictors, and determine an optimal set of predictors. Academic performance in the GCA program was criterion variable and was measured by the cumulative graduate grade point average (GGPA). Predictive models were developed using stepwise linear regression.

Current AFIT eligibility criteria consists of: the cumulative grade point average of all undergraduate coursework (UGPA), the cumulative GPA of all undergraduate math courses, and minimum scores on either the Graduate Management Admissions Test (GMAT), or the Graduate Record Examination (GRE) verbal and quantitative sections. Other potential predictors considered in this study included other subtests scores of the GRE and GMAT, age, gender, rating of undergraduate school's admissions competitiveness, undergraduate degree type, and various measures of applicant's time in military service.

This research found the GMAT is more useful than the GRE as a predictor of academic performance in the AFIT GCA program. UGPA is also a dependable, though not particularly strong, predictor. The optimal model accounted for up to 45% of the variance in GGPA, and included the GMAT Verbal section score, UGPA, and a dichotomous indicator of prior service as a member of the military enlisted corps.

ADMISSIONS CRITERIA AND ACADEMIC PERFORMANCE IN THE AFIT GRADUATE COST ANALYSIS PROGRAM

I. Introduction

Background

The Air Force Institute of Technology (AFIT), located at Wright-Patterson Air Force Base, Ohio, is the U.S. Air Force's graduate school and its premier professional continuing education institution. Among a variety of programs, AFIT offers in-resident Masters and Doctoral degree programs through its Graduate School of Engineering and Management. “The mission of the Graduate School of Engineering and Management is to produce graduates and engage in research activities that enable the Air Force (AF) to maintain its scientific and technological dominance.” (AFIT Catalog, 2000: 3). A key factor in “producing graduates” is the initial selection and admission of students from among all applicants. The selection process should choose students who are likely to be successful and likely to provide the most benefit to the Air Force.

Air Force personnel desiring to attend AFIT’s Graduate School of Engineering and Management (which will be from here forward referred to simply as AFIT) are first deemed academically eligible by AFIT, and then the AF personnel system selects from the pool of eligible applicants. During this process, both objective and subjective criteria are used to determine eligibility and make selections. It is important that, when possible, the criteria are genuine predictors of the outcome the Air Force desires; which is assumed to be an officer / civilian employee who is better educated and capable of improved service to the Air Force.

This research will examine the criteria used, as well as other criteria that could be used, when selecting students to attend AFIT's Graduate Cost Analysis (GCA) program within the School of Engineering and Management. The factor used to judge academic success will be the cumulative graduate grade point average (GGPA). Statistical analysis will be used in an attempt to identify a predictive relationship between pre-AFIT variables and the criteria of academic success.

Importance of Selection

One might assume that as long as there has been graduate level education, the schools offering it must have used a process to select students for their program. An appropriate selection process provides benefit to the school and student. Schools want to prevent the admission of less-than-qualified individuals because that could diminish both the education provided and academic reputation of the institution. Students prefer to attend the schools with the best reputation possible, because in addition to receiving a quality education, earning a degree from a highly respected graduate degree program can provide a competitive edge when seeking employment. (Ragothaman and Davies, 1998:126). Further, as Kuncel and others (2001: 162) put it, "[a]dmission of poorly qualified students misuses the resources of students, faculty and schools."

Selection of the most appropriate individuals to attend the GCA program is important for many reasons related to the success of the students and the Air Force.

Though a good reputation is important, AFIT differs from civilian institutions in that it has a more direct link to, and interest in, what a student learns and a student's success, both while attending and after graduation. This is because AFIT is part of the same larger organization in which its graduates will work – the U. S. Air Force for most, or on a larger scale, the Department of Defense for all except the foreign military students. AFIT's mission is "to produce graduates and engage in research activities that enable the Air Force to maintain its scientific and technological dominance." (AFIT Catalog 2001-2002: 3) A civilian school does not usually have such a direct link to the future work of its graduates.

Also important to the Air Force is the time and money expended on the student and the benefits not received, should a student fail to graduate from the program. Not only does the follow-on assignment the student was slated to fill remain vacant, but another officer who would have succeeded may have been denied the opportunity to attend AFIT. According to Air University's Financial Management Directorate, the average direct cost per student, for an 18-month graduate degree program, is $101, 495, not including base operating support or the student's pay and allowances. (AU/FM spreadsheet, September 2001)

As shown above, the negative consequences of selecting an individual who delays or fails to complete the program on time are significant. The selection process should utilize the most appropriate and accurate predictors of academic success to prevent the wasted time, money, and opportunity an unsuccessful student consumes.

Eligibility and Selection Process

As mentioned previously, AFIT does not *select* the students who attend, AFIT establishes their academic eligibility. A prospective student submits to AFIT their educational transcripts and a Request for Evaluation, asking for a review of eligibility for the degree programs of interest to the applicant. The Admission and Registrar Directorate (AFIT/RRE) replies with a letter noting eligibility for some or all programs requested. If ineligible, suggestions on how to become eligible may or may not be included.

AFIT/RRE determines eligibility by first using the following criteria, as described in the AFIT newsletter. All M.S. programs require: 1. a baccalaureate degree from an accredited college or university in an appropriate discipline; 2. an overall undergraduate grade point average (UGPA) of at least 3.00; 3. Graduate Record Exam (GRE) scores of at least 500 (verbal) test and 600 (quantitative), or for certain programs, a Graduate Management Admissions Test (GMAT) score of at least 550. The CGA program accepts either GRE or GMAT scores, but also requires a GPA of 3.0 for all undergraduate math courses, up to differential calculus. (AFIT Newsletter, 2000).

If these initial criteria are not met, but other factors indicate academic potential, AFIT/RRE may forward the request for evaluation to the department(s) whose programs are involved. The departments then review the record and consider factors they believe are additional indicators of ability to succeed in the program. (Stockman, 2002; AFIT Catalog, 2000) If the minimum criteria are waived and academic eligibility is granted, the Air Personnel Center (AFPC) is informed. AFPC supervises and directs the overall management and distribution of military and civilian personnel.

On a yearly basis, the Air Force Education Requirements Board (AFERB) determines how many student positions, will be available in the degree programs at AFIT. (AFIT Newsletter, 2000) The prospective student then applies for an advertised opening with the appropriate assignment team at AFPC. The assignment team reviews the officer's entire record and approves/disapproves assignment to AFIT after considering: the applicability of the degree program to the applicant's Core ID (job classification); date of commissioning as an officer; at least 2 years time-on-station (TOS) before leaving for AFIT, and applicant's military record. (Monson, 2001; AFPC website, 2002) Time-on-station is considered because the AF resists moving personnel until they have spent at least 2 years at a single location.

If the number of approved applicants does not exceed the number of available positions, then the approved applicants are assigned to AFIT without convening a selection board. If a selection board is required, it will consist of three voting members who will assess and compare applicants based on, but not limited to, "strength of record, Officer professional development, timing, [undergraduate] GPA" and demonstrated leadership potential. (AFPC website, 2002; Monson, 2001).

For example, Air Force wide there may be 15 predicted vacancies in positions requiring a Cost Analysis Masters degree for the future year 200X. AFERB determines AFIT can admit 10 people to earn a GCA degree and graduate in year 200X. The assignment teams at AFPC then work to fill these 10 slots from the list of academically eligible applicants.

As should be apparent, the admission and selection process relies primarily on objective / quantitative measures and is augmented by subjective / qualitative criteria when the objective measures do not provide a clear decision. Waivers to grant academic eligibility to applicants who do not meet the initial criteria are based largely on subjective judgments by the reviewing official(s). There is no set policy on what additional factors to consider. Rather it is up to the official to review the record and decide if there exists sufficient compensatory evidence to indicate the applicant would succeed in the program despite the low test scores and/or UGPA. This evidence can include, but is not limited to, factors such as: source of degree and perception of that school's admissions and grading policies; the undergraduate major; number of technical and math courses, and the grades earned in these courses; career relation to degree; and any graduate level work. (Stockman, 2002) And, as described previously, the AFPC selection board uses largely subjective factors to rank the applicants for selection.

Ideally, all criteria considered should be predictors of success or predictors of benefit to the Air Force, and most of the criteria mentioned are *believed* to be just that. Additionally, the subjective criteria are, by nature, not easily measured, and their interpretation can vary greatly among the AFIT and AFPC staff. Thus, it may be advantageous to the Air Force if the actual worth of these perceived predictors is quantified in some way, thus improving consistency and reducing subjectivity where appropriate.

Research Objectives

The three primary objectives of this research are:

1. investigate the ability of the current admission criteria to predict student academic performance;

2. investigate the ability of additional individual variables (quantitative and qualitative) to predict student academic performance; and

3. select an 'optimal' set of eligibility/selection indicators having the potential of predicting student academic success.

Summary

This chapter introduced the basic reasons selection of students for a graduate degree program is important, both to the school and the student. The process for the granting of academic eligibility and selection to attend AFIT was described, and the role of objective and subjective criteria discussed. Finally, the research objectives were presented. It is hoped the achievement of these objectives will provide increased insight into which pre-admission factors have a useful predictive relationship with students' academic performance in the GCA program, and which factors do not.

II. Literature Review

Chapter Overview

This chapter begins with a discussion of some of the basic concepts of testing and prediction, describes student selection for graduate education, and then considers what constitutes academic success in general. This is followed by a review of the use of the GRE, GMAT, and other variables as predictors of academic success and their use in admissions decisions. Research methodologies are then reviewed and this chapter concludes by reviewing past works examining AFIT's use of tests and other measures in making eligibility decisions.

Measurement Reliability and Validity

The purpose of this section is to briefly review topics related to statistics and educational and psychological measurement. A full and complete discussion of these topics is beyond the scope of this paper and will not be attempted.

Tests and other forms of measurement are judged according to their reliability and validity, two concepts central to the theory and practice of educational and psychological measurement and evaluation. *Reliability* of a measurement is how accurately and consistently it measures a particular construct. A measurement is reliable if produces the same results over and over again, assuming that what it is measuring is not changing. (Hopkins, 1998:108) Validity can be thought of as the usefulness of the measure, or how well it fulfills the purpose for which it is being used, and how correct are the inferences made from the results of the measure. (Hopkins, 1998:73) A measure can be reliable but not valid, but to have validity it must be at least moderately reliable. For example, an early study to predict academic success used student reaction time as the predictor variable. While reaction time could be measured accurately (i.e., with reliability), the validity coefficient of only -0.02 indicated no useable predictive ability (i.e., no validity). (Hopkins, 1998:109).

The concept of validity can de divided into 3 types of validity: content, criterion-related, and construct. *Content validity* for a measure of academic achievement describes how well a test measures the content and topics, as well as the cognitive processes and abilities objectives, of a given unit, course, or program. Determining if a measure has content validity is primarily a process of logical analysis. If a calculus final exam consisted of simple addition problems and essay questions on American History, it is logical that the test score might not be a true indication of the student's understanding of calculus. Such a tests would be considered to have low content validity for the calculus class. (Hopkins, 1998:73-77)

Criterion-related validity has two sub-classes: concurrent and predictive validity. *Concurrent validity* describes the relationship between one measurement (e.g. a new test) and another measurement (an established test). If the new test is simpler and cheaper and correlates highly with the established test (has concurrent validity), then it may be a viable alternative. Establishing concurrent validity may be the first step to establishing predictive validity. (Hopkins, 1998:96)

Predictive validity describes the ability of a measurement to predict subsequent performance on a criterion. If two measurements, factors, or traits are related, or vary together, they are said to be correlated. A quantitative description of the degree of that relationship is a *correlation coefficient*. The Pearson coefficient of correlation, r, is widely used and summarizes both the magnitude and direction of the relationship, provided it is a linear relationship. It ranges from –1 to 1, where –1 is a perfect inverse relationship, zero indicates no correlation, and 1 is a perfect positive relationship. The accuracy of the predictions, the predictive validity, is determined empirically, and described by the correlation coefficient between the measurement and predicted criterion. (Hopkins, 1998:77-102)

Construct validity describes the degree to which certain abstract psychological traits or abilities are represented by the measurement/test. Psychological constructs are unobservable, theoretical variables such as intelligence, anxiety, motivation, or mathematical aptitude. Determining construct validity requires both logical and empirical means. If a measurement or test is determined to have construct validity, then content and criterion-related validity are assumed, since the content and correlations should have been considered to establish construct validity. (Hopkins, 1998:99-102)

Admission / Selection practices and problems

One might assume that as long as there has been graduate level education, the schools offering it must have used a process to select students for their program. After all, an appropriate selection process provides benefit to the school and student. Schools want to prevent the admission of less-than-qualified individuals because a student's poor performance – in school and/or in post-graduation employment – could diminish both the education provided and the academic reputation of the institution. Students prefer to attend the schools with the best reputation possible, because in addition to receiving a quality education, earning a degree from a highly respected graduate degree program can provide a competitive edge when seeking employment. (Ragothaman and Davies, 1998: 126; Wilson & Hardgrave, 1995: 186).

A prime objective of the graduate admissions process is to select students based on their potential for achieving the school's desired level of academic performance – which can vary by school – while balancing other objectives such as student diversity and professional potential. (Hoefer & Gould, 2000: 225) Toward this end, institutions providing graduate education strive to develop admissions processes and criteria that prevent the admission of individuals unable to complete the program while not denying admission to individuals who would succeed at the desired level.

Most institutions base their decisions on a combination of qualitative and quantitative criteria, most often including standardized test scores and some consideration of past performance, with the relative weight given to each factor based on the schools preference and experience. (Nilsson, 1995:637; Wright and Palmer, 1994:344) Some of the more commonly used criteria include the applicant's: undergraduate GPA, class rank, major, and degree awarding institution; scores on standardized test such as the GRE and GMAT; prior graduate level coursework; professional references; work experience; biographical information such as race, age, and gender; goals statements; and even personal interviews. (Ragothaman and Davies, 1998: 126; Bowman, 1988: 869). Just as the criteria range from the very objective/quantitative to subjective/qualitative, so to do the methods of evaluation.

Examples of graduate education selection processes described in the literature covered the range from subjective to objective. A few are described here. At school A, applications are reviewed by 2-3 faculty members and rated on a scale of 0–5. The faculty then meets and makes final decisions based on cumulative and comparative rankings. At school B, the faculty review the applications and provide to the department head their yes/no vote and a brief explanation. The department head then makes the final decisions. (Bowman, 1998: 870-871) Finally, school C has two methods for admittance to its MBA program – an index system and a petition process. The index combines UGPA and GMAT scores according to a mathematical formula. If the combination score meets the minimum requirement, the applicant is admitted. If not, the applicant may appeal by petitioning the Graduate Council, which accepts or denies the petition at their discretion – less than 10% admitted by petition. (Ahmadi and others, 1997)

The first two processes (schools A & B) were largely subjective, though the individual faculty evaluations may have been objective, subjective, or anywhere in between. School C uses a quantitative criterion, which in this case is also a cut-off criterion, but allows for a subjective evaluation if the applicant who fails to meet that minimum cut-off score makes the extra effort to petition. School C helps illustrates a situation common in graduate selection processes known as compensatory selection. Compensatory selection occurs when individuals who do not meet a minimum requirement are granted academic eligibility because of compensating levels of performance on other factors believed to be predictors of academic achievement. (Dunlap, Henry & Fraser, 1998). This may also occur in schools A and B, but is most clearly illustrated in school C.

To ensure that an institution's selection criteria and processes are providing the type of student desired and not wrongly excluding students who would perform in the desired manner, the admissions decision makers should validate the system. According to the 1991 accreditation standards of the International Association to Advance Collegiate Schools of Business (AACSB) (formerly named American Assembly of Collegiate Schools of Business), "each school must be able to demonstrate empirical documentation that its admission practices and policies are contributing to the realization of its mission." (Hoefer and Gould, 2000:226) The guidance from Educational Testing Service (ETS) – the administrator of the two tests used by the admissions departments of over 1700 graduate institutions, the GRE and GMAT – on the use of scores from either test says an institution should consider not just the test scores, but all pertinent information about an applicant, and that the institution should conduct a validity study to verify the processes and criteria used are accurate predictors of the academic performance the institution desires. (Graduate Record Examinations Board, 2000; and Educational Testing Service, 2001).

Many schools have not heeded this advice. Bowman (1988: 871) found that 65.6% of the 157 schools surveyed had not performed local validity studies on their admissions criteria, and even more felt that faculty were often forced to rely on personal judgment when making selection decisions, due to lack of clarity of the criteria. A lack of clear standards or the inconsistent application of established standards makes validation of the admission/selection process difficult.

Measures of Academic Success

When investigating a relationship between admissions criteria and student performance, the concept of student performance must be defined and an appropriate measure established. What one school, or even one department, considers a successful graduate may not be the same as another school or department. Is the desired success limited to the student's performance while enrolled in the program, or is performance after graduation equally or even more important? To determine if a selection process is admitting the most appropriate students, a researcher or school should establish some student performance factor(s) that can be measured, either directly or by a surrogate. Past researchers have used a variety surrogates, singly or in combinations, to quantify (or operationalize?) the concept of academic success.

GGPA According to Kuncel et al (2001), the most widely used measures of graduate academic performance are cumulative graduate GPA (GGPA) and first year graduate GPA (FYGGPA). This assertion is supported by the research reviewed in this document, much of which used GGPA or FYGGPA as either the sole measure of academic success, or as an element of the success measure. (Abedi, 1991; Ahmadi and others, 1997; Beiker, 1996; Graham, 1991; Hoefer and Gould, 2000; Nilsson, 1995; Arnold and Chakravarty, 1996; Matthews and Martin, 1992; Morrison and Morrison, 1995; and others).

Some advantages to using GGPA to measure academic performance is that grades are supposed to be indicators of a student's understanding and performance in a class. They are derived over time from multiple performances that involve a broad set of skills and attributes. (Cole, 1998) The GGPA "measures long-term work, knowledge acquisition, effort, persistence, and ability." (Kuncel, Hezlett and Ones, 2001:165) Additionally, though the results are inconsistent, GGPA has also been related to post-school job performance and success. Meta-analytical studies by Roth and others (1996) as well as work by Colarelli, Dean, and Konstans (1987) found positive correlation between grades and job performance. On the other hand, Bretz (1989) and Hunter and Hunter (1984) suggest a low predictive validity of grades to job success. Consideration of post-school success is important because this can effect a school's reputation, which has been shown to be a factor in admission decisions. ETS uses FYGGPA as the criterion to validate the GRE and GMAT because taking the average of grades from several professors is a good estimate of individual academic ability and because FYGGPA is readily available. (Goldberg and Alliger, 1992:1025)

GGPA does have its limitations as a measure of academic performance. Grading policies and standards are not consistent between schools, departments, or even teachers. "Grades tend to be scaled within a class regardless of differences in the students' aptitude levels." (Hopkins, 1998:314) Grades in graduate school tend to be A or B and this narrow range makes differentiation between superior and inadequate students more difficult. (Goldberg and Alliger, 1992: 1026; Wilson and Hardgrave, 1995: 193)

Some researchers also used variations on GGPA. One study treated first year GPA as both a continuous and categorical variable (GPA < 3.00, high risk for academic success; $3.00 < \text{GPA} < 3.30$, questionable risk; GPA > 3.30, no risk). The categorical variable allowed the use of non-linear analysis methods and provided predictions more useful to the admissions decision maker, since they are concerned with predicting relative success or failure, not exact GGPA values. (Wilson and Hardgrave, 1995)

Degree attainment Degree attainment is the successful completion of a graduate degree program, and is the simplest measure of success in the program. Use of this dichotomous measure of success – 1 if attained, 0 otherwise – allows the researcher to use some forms of statistical analysis, such as discriminant analysis and logistic regression, which do not work with a continuous criterion variable. While predicting failure is a goal of the admissions decision makers, failure to attain a degree may be due to factors beyond the student's control and unrelated to his/her ability to perform, making it an imperfect measure of success. (Kuncel, Hezlett, and Ones, 2001: 165)

Keith used degree attainment in his 1977 thesis – described later in this document – and found that the only reliable predictor was the volunteer/non-volunteer variable he used as a surrogate for motivation. At that time not all AFIT students had volunteered to attend AFIT and the rate of failure for the non-volunteers was higher (~10%).

Time to Completion Closely related to degree attainment is time to completion, i.e., the amount of time elapsed from starting a degree program until the degree requirements are complete. The idea that less competent students may require more time to complete the degree requirements is logical, but like degree attainment, factors beyond the student's control and unrelated to his/her ability to perform may have greater effect on time to complete than inherent ability. (Kuncel, Hezlett, and Ones, 2001: 165; Fenster and others, 2001:340)

Faculty ratings In some research, the faculty rated the students on factors considered to indicate either achievement or ability as a measure of student performance. Sternberg and Willams (1997) asked the faculty in their study on predicting graduate school success to rate students abilities in five areas – research, analytical, creative, practical and teaching – using a 7-point scale. Critics of Sternberg and William's work highlighted the unreliability of such ratings, citing unreliability due to passage of time, raters' personal biases, and the Halo effect. (Ruscio, 1998: 569) Kuncel, Hezlett and Ones (2001:165) also mentioned these potential problems with reliability of faculty ratings. They went even further and discussed the additional problems of central tendency and how obtaining rating for a large number of students may create such a burden for faculty that the results have poor discriminant validity.

Comprehensive Exam Scores Not all institutions or programs use comprehensive examinations, but those that do, use them to assess the level of job knowledge retained by the graduate student. The exams are usually taken near the end of a degree program and a minimum score is often a graduation requirement. (Kuncel, Hezlett, and Ones, 2001: 165). Comprehensive exam scores have greater variability than grades – giving them better properties for statistical analysis – and are acknowledged as a "summative measure of educational outcomes." (Dunlap, Henley, and Fraser, 1998:458) However, like grades, their value as measures can vary across programs and schools due to aspects such as difficulty, grading policies, relevance to degree. (Kuncel, Hezlett, and Ones, 2001: 165)

Research Productivity This factor is a direct measure of the number of publications or conference papers produced by the student, both during and after graduate school. While this may measure scientific productivity, many students may be training as practitioners not scientists, and have no intention of publishing future works. In addition, quantity of publications is not a measure of their quality. (Kuncel, Hezlett, and Ones, 2001: 165)

Predicting Performance

There has been an abundance of research in the area of identifying the best predictors of graduate school success. Many studies sought to validate current criteria while others looked for new or additional criteria that could improve the selection process. The most popular measures used in admissions, GRE and/or GMAT scores and UGPA, were also the subjects of much of the literature. The reliability and validity of these and other potential measures for use in selection decisions is reviewed below.

Graduate Record Examination (GRE) To clarify any potential misunderstandings, in this document, any reference to GRE or GRE scores refers to the GRE General Test or scores earned on the General Test. This is to distinguish this abbreviation from the GRE Subject Tests that are available. The GRE General Test consists of three separate sub-tests designed to assess knowledge, skills, and abilities that have been acquired over a long period of time, and are relevant to graduate level study. The verbal portion (GRE-V) measures "the ability to analyze and evaluate written material and synthesize information obtained from it, to analyze relationships among component parts of sentences, and to recognize relationships between words and concepts." (Graduate Record Examination Board (GREB), 2000:5) The quantitative portion (GRE-Q) measures "basic math skills and understanding of elementary mathematical concepts, as well as the ability to reason quantitatively and to solve problems in a quantitative setting." (GREB, 2000:5) The analytical portion (GRE-A) measures "ability to understand structured sets of relationships, analyze and evaluate arguments, identify central issues and hypotheses, draw sound inferences, and identify plausible causal explanations." (GREB, 2000:5) Though the scores on each sub-test are reported on the same scale – from 200 to 800 – the test administrator, Educational Testing Service (ETS), cautions against comparing the scores because each measure is scaled separately. (GREB, 2000:11)

ETS has performed its own studies to demonstrate the test's reliability and validity as a predictor of graduate school performance, using the first year GGPA (FYGGPA) as the measure of academic success. They report the reliability of the three portions of the GRE General Test to be above 0.90. Based on data covering the years 1986 to 1990, the predictive validity, expressed as average estimated correlation coefficients, is shown in Table 1. The two last columns, VQA and VQAU, are combinations of the Verbal, Quantitative, and Analytical scores, and UGPA values. Each predictor variable within the combination is multiplied by a unique numerical coefficient, and then all are summed. These models (i.e., the coefficients) were developed using Empirical Bayes regression. (GREB, 2000:24) This table includes only the results for the full sample and the business departments (i.e., student who attained master's degrees in business), though information for all departments is available. These correlation coefficients show a slight to moderate correlation between GRE scores and FYGGPA, and the best predictive validity is achieved when both GRE test scores and UGPA are considered. (GREB, 2000:11)

Table 1. Average Estimated Correlations of GRE Scores and UGPA with FYGGPA

		Predictors					
Departments	N	GRE-V	GRE-Q	GRE-A	UGPA	VQA	VQAU
All Depts	12,013	.30	.29	.28	.37	.34	.46
Business	196	.28	.28	.25	.39	.31	.47

Source: (Graduate Record Examination Board, 2000: 24)

Kuncel, Hezlett and Ones (2001) used meta-analysis to examine GRE and UGPA as predictors of graduate school success. Using data from 1,753 samples yielding 82,659 graduate students, they concluded GRE and UGPA are "generalizably valid" predictors of GGPA, FYGGPA, comprehensive exam scores, and other measures of graduate school performance, but did not have significant predictive validity of degree attainment.

Other studies supported GRE as a valid predictor of graduate school performance. Nilsson (1995) compared GRE and GMAT as predictors of GGPA and found the GRE to have the stronger correlation ($r = 0.449$). House (1998) examined the records of 5,047 graduate students and found GRE-V, GRE-Q and GRE-T (GRE-V + GRE-Q) were all significantly correlated with GGPA, though the GRE consistently over-predicted GGPA for students under age 24, and under predicted GGPA for older students. Fenster and others' (2001) study on students in a MA program in forensic psychology is of particular interest because GRE scores were not part of the selection criteria. This reduced the problem of range restriction, and may account for the comparatively strong correlations (0.63) they found between GGPA and a linear combination of GRE-V, GRE-Q, and UGPA. They also found moderate correlation (0.31) between time-to-complete and a similar linear combination.

Thornell and McCoy (1985) examined the relationship of GRE-V, GRE-Q and GRE-T scores to GGPA for 582 students divided into four subgroups of graduate degree disciplines; education, humanities, fine arts, and math/science. Correlation coefficients for the total sample were $r_{GRE\text{-}V} = 0.47$, $r_{GRE\text{-}Q} = 0.29$, and $r_{GRE\text{-}T} = 0.43$, and the correlations for the subgroups ranged from 0.22 to 0.49 and demonstrated that the GRE sub-test scores had different predictive ability for different degree disciplines. (Thornell and McCoy, 1985). Other researchers seeking to establish predictive validity of the GRE for their programs also noted this variation among degree programs. Results varied widely, with GRE to GGPA correlations for individual programs ranging between –0.62

to 0.81. (Kuncel, Hezlett and Ones, 2001:163) Such widely variable results emphasize the fact that different degree programs may have different valid predictors.

Many studies found little value to using the GRE in selection decisions. A meta-analytic assessment by Morrison and Morrison (1995) of the GRE's verbal and quantitative tests' predictive validity for GGPA concluded that those two measures accounted for such a small amount of variance (less than 6%) in the criterion as to be "virtually useless from a prediction standpoint." (Morrison and Morrison, 1995: 313) These results were similar to Goldberg and Alliger's (1992) earlier meta-analysis that found GRE accounted for less than 9% of variance in GGPA. Sternberg (1996) contends the GRE does measure some intellectual abilities, but these are not adequate predictors of graduate school performance. Maybe biologist Doug Bennett of Reed College said it best, as quoted in an article in Science magazine, "…`the GRE can never be expected to predict traits critical to graduate school success such as commitment and ability to work autonomously. After all, says Bennett, often 'the student doesn't even know.'" (Science, 1993: 494)

Graduate Management Admission Test (GMAT) A group of schools seeking to improve their admissions processes met in the early 1950s and formed what later became the Graduate Management Admission Council (GMAC), the organization that now supervises the Graduate Management Aptitude Test (GMAT). ETS administers the test under policies set forth by the GMAC. As an indication of its perceived utility, the GMAT is now used in the admissions processes of over 1700 schools. (Hoefer and Gould, 2000:225) The GMAT is a test of general developed abilities associated with graduate school success, and is intended to provide admission decision makers with one indicator of first year academic performance in a graduate management program. Similar to the GRE, the GMAT consists of three sections: verbal, quantitative, and (unlike the GRE) analytical writing. The verbal section (GMAT-V) "measures the ability to understand and evaluate what is read and to recognize basic conventions of standard written English." The quantitative section (GMAT-Q) "tests basic mathematical skills and understanding of elementary concepts as well as the ability to reason quantitatively, to solve quantitative problems, and to interpret data given in graphs, charts, or tables." The analytical writing sections (AWA) "measure the ability to think critically and communicate complex ideas through writing." The GMAT yields four scores, one for each test and a total score. Each score is reported on a fixed scale. Scores on the verbal and quantitative sections range from 0 to 60, though scores above 44 on the verbal section and above 50 on the quantitative section occur less than 3% of the time. The analytical writing score is scaled between 0 and 9. The total score (GMAT-T) ranges from 200 to 800, with approximately two-thirds of the scores falling between 400 and 600. Average reliability of the GMAT-T is 0.92. (Educational Testing Service, 2001)

ETS has performed studies to determine and monitor the predictive validity of the GMAT, but limits this validity to prediction of FYGGPA in an MBA or similar program. In its most recent study, ETS compiled results of 101 validity studies from 1996 to 1999. Though reported correlations ranged from 0.13 to 0.60, the average of combined GMAT-V, -Q, -T and AWA correlation to FYGGPA was 0.41. The inclusion of UGPA in the combination improved average correlation to 0.47. (Educational Testing Service, 2001)

With over 1700 graduate management programs using the GMAT in their admissions process, and ETS confirming that predictive validity varies by program (ETS, 2001), investigating the GMAT's validity is popular and the results are decidedly mixed.

Many studies support ETS's findings of the GMAT's moderate predictive validity, which can be improved by including UGPA in the analysis. Graham (1991) found the GMAT with UGPA to be the best predictor of the ten considered, but suggested using a UGPA based on only the junior and senior year, versus all undergraduate coursework. Beiker (1996) found GMAT-T to be the strongest predictor of GGPA, and including a GPA based on 11 core undergrad courses improved the linear regression model enough to account for 49% of the variation in GGPA. Similar results were found by Arnold and Chakravarty (1996); Hoefer and Gould (2000); Rothstein, Paunonen, Rush and King (1994); and Swayze (2001).

As mentioned previously, many researchers found the GMAT to have less predictive validity than the average reported by ETS. The analysis by Ahmadi and others (1997) found the GMAT and UGPA did not adequately predict graduate academic success at the study institution and might unfairly deny admission to some qualified students. They recommended inclusion of more qualitative measures into the selection process to more accurately predict academic success. (Ahmadi and others, 1997) It is interesting to note that Ahmadi and others (1997) do not clarify what level of correlation would be considered adequate. Their results showed a GMAT score to GGPA correlation of $r = 0.433$. Other researchers considered this level of correlation acceptable, considering the extensive list of other factors that can affect performance in a graduate education program.

Wright and Palmer (1994) classified UGPA, GMAT-V, -Q, and –T scores into quartiles from high to low scores and investigated the correlations to GGPA. They found the most predictive power in the extreme quartiles, but insignificant differences between adjacent groups. So while the highest scores predicted higher performance, only the very lowest scores predicted poor performance, and even then, not to a level of accuracy sufficient to establish cut-off scores. Wright and Palmer also recommended using additional applicant screening devices, such as letters of recommendation and personal interviews, when making selection decisions. Similar results, recommending less emphasis on GMAT scores and more emphasis on qualitative or biographical factors, were reported by Nilsson (1995) and Wilson and Hardgrave (1995).

Undergraduate Grade Point Average (UGPA) UGPA was probably the most often examined variable in the research reviewed here. This is not surprising, since ETS recommends using UGPA in conjunction with the GRE and GMAT scores when making selection decisions. Based on the commonalities between undergraduate and graduate coursework, it seems reasonable to assume graduate work has a similar variance to undergraduate work, and consider "previous academic performance to be a natural indicator of future academic performance." (Fenster and others, 2001:339) The perception of UGPA as a natural predictor of academic performance is supported by Bowman's 1988 survey of admission practices in Master of Public Administration programs. He found that, when forced to choose, 51% of respondents felt UGPA was the single best indicator of student success, while only 10% chose the GRE (GMAT was not an option). (Bowman, 1988:869-870) The meta-analysis by Kuncel and others (2001) conflicts with this choice. When GRE-V, -Q, -A and UGPA were considered individually, UGPA had the lowest correlation to GGPA. (Kuncel, Hezlet, and Ones, 2001:168).

Many studies had similar findings, that UGPA was positively correlated to academic performance, but the GRE or GMAT had greater correlation to graduate academic performance. (Dunlap, Henley, and Fraser, 1998; Graham, 1991; Kuncel, Hezlett and Ones, 2001) In slight contrast, Ahmadi and others (1997) found UGPA had *better* predictive validity than the GMAT-T score ($r = 0.521$ and $r = 0.433$ respectively) for predicting graduate academic performance, as measured by GGPA. In somewhat of a split decision, Mathews and Martin (1992) found UGPA predicts FYGGPA better for students under age 30 ($r = .438$), than for students age 30 and older ($r = .251$).

Abedi (1991) examined UGPA as a predictor of graduate academic success and compared it with other predictors including age, gender, field of study, source of baccalaureate degree, and whether the student had done any graduate level work. His results "indicated that undergraduate GPA was not a good predictor of graduate academic success" and "has virtually no relationship with any of the measures of graduate academic success." (Abedi, 1991:151,158) Abedi attributed this low predictive power of UGPA to some poor psychometric characteristics of this index: (a) lack of comparability – differences in institutions' quality and grading style; (b) lack of variability – though UGPAs should range from 0 to 4, the majority fall between 2 and 3.5, skewing the distribution; and (c) non-normality of the GPA distribution – the skewed distribution from UGPAs becomes even more skewed for graduate education applicants because students with low GPAs assume they would not be selected and therefore tend not to apply. (Abedi, 1991: 158).

Undergraduate School Rating As mentioned numerous times, different schools and even different departments within a school, have different levels of quality and grading practices, and expect different levels of academic performance from their students. Abedi (1991) mentioned this as a problem with using UGPA as a predictor. In an effort to compensate for this disparity, Hoefer and Gould (2000) included a factor indicating the tier (1 – 4, best to worst) of the undergraduate degree-granting institution, based on ratings by U.S. News. This factor was considered significant in one of the models develop to predict GGPA. (Hoefer and Gould, 2000).

A school rating was included in two previous AFIT theses examining admissions criteria and academic performance. Prokopyk (1988) included a 'Quality of Schools' factor in his regression, though when included individually, this factor was not significant enough to be included in the final model. Building on Prokopyk's quality of schools indicator, Spangler (1989) included the same factor (though renamed to RATE, and described as degree of admissions competitiveness) in his analysis. However, Spangler also combined the RATE factor with the UGPA to form RATGPA, and this combined factor proved a better predictor of GGPA than either individual factor ($r_{RATGPA} = 0.4800$ vs. $r_{RATE} = 0.3707$ and $r_{UGPA} = 0.2437$).

Degree Type Another potential factor to consider in the admissions process is the type of undergraduate degree – Bachelor of Science (B.S.) or Bachelor of Arts (B.A.). In general, a B.S. degree program is considered more quantitatively based than a B.A. program, though exceptions to this generality are common. To see if there was any predictive advantage to this factor, Graham (1991) included a B.S. / B.A. dichotomous variable in his analysis of predictors of academic success in an MBA program. The resulting correlation value of $r = 0.168$, shows little predictive value on its own, and the factor was not significant in the stepwise regression models Graham developed.

Age Numerous researchers examined student age for its effect, if any, on prediction of graduate academic performance. Older students are likely to differ from younger students in time since undergraduate degree, work experience, and family obligations. (Kuncel, Hezlett, and Ones, 2001: 166) Age was looked at in terms of direct predictive ability as well as its effects as a moderator of other variables. Using age as a proxy for work experience, Bieker (1996) included it as both a continuous and dichotomous variable (0 if age < 30, 1 if age ≥ 30) in his analysis of factors affecting academic achievement in a MBA program. "The specification of age as a dichotomous variable is predicated on the hypothesis confirmed by Gayle and Jones (1973) that the performance of younger students is significantly different from that of older students, other things being equal." (Bieker, 1996:43-44). Beiker found age was not a statistically significant predictor variable, regardless of the way it was specified. This finding agrees with research by Dunlap, Henley, and Fraser (1998) and Graham (1991). Hoefer and Gould (2000) also found weak correlation ($r = -.05$) between age as a continuous variable and GGPA, but as a dichotomous variable (old / young), the old indicator was found to be significant in a neural network developed model, but not a stepwise regression model.

Of the examples reviewed here, only Hoefer and Gould (2000) used a non-linear analysis method – neural networks – and only this method included age in the models developed. They suggest that age as a qualitative variable be considered in selection decisions. Matthews and Martin (1992) found age acts as a reciprocal suppressor variable when the cross products with UGPA, GRE-Q or GRE-A are used in a regression equation. In other words, as age increases, it reduces the predictive validity of UGPA, GRE-Q and GRE-A. (Matthews and Martin, 1992: 456). Similarly, House (1998) found significant differences between younger students –age 24 or less – and older students – 25 or older – in the mean error of prediction of GGPA from GRE-T, GRE-V, or GRE-Q. The GRE scores over-predicted the performance of the younger students and under-predicted performance of the older students.

Gender Much of the research that examined gender as a possible predictor variable came to conclusions similar to those for age; i.e., that as a linear predictor of graduate performance, gender shows no statistically significant correlation (Ahmadi and others, 1997, 1997; Bieker, 1996; Graham, 1991; Hoefer and Gould, 2000; Wilson and Hardgrave, 1995), but it may moderate other predictors, such as GRE or GMAT scores, to improve or reduce their predictive validity.

One study found significant differences in GRE predictive validity due to gender and race. Controlling for these biographical factors improved the correlation of GRE-T to GGPA. (Dunlap, Henley, and Fraser, 1998: 461) Hancock (1999) looked at 269 students in an MBA program and found that though there was no discernable difference in academic performance in the program, "the 149 males outscored the 120 females on the GMAT 540.3 to 506.9, a magnitude with far less than 1% chance of occurring if the GMAT is truly gender blind." (Hancock, 1999:93)

Time since Undergraduate Degree Based on the belief that the passage of time since being in an academic environment may effect an individuals performance in future academic endeavors, some researchers included a variable to denote the amount of time from completion of the students bachelor degree (or last academic experience) to entry into the graduate degree program. These studies found only very weak correlation between the measure of time and graduate academic performance, and insufficient statistical significance for this variable to be used in any of the regression based models developed in these studies. (Arnold and Chakravarty, 1996; Graham, 1991; Hoefer and Gould, 2000)

Work Experience is considered by many admission decision-makers to be an important factor in the selection decision, based on the belief that the ability to relate work experiences to concepts presented in the classroom will reinforce those concepts. (Wooten and McCullough, 1991) Despite this, none of the institutions in Wooten and McCullough's (1991) survey required prior work experience for admission. Peiperl and Trevelyan (1997) examined work experience along with other predictors and found insufficient predictive validity to justify its use in a predictive model.

Analysis Methods

As discussed previously, correlation describes a linear relationship between two variables, and the Pearson correlation coefficient is a quantitative measure of that relationship and indicates both the magnitude and direction. If one variable, (the criterion variable) is correlated to one or more other variables (predictors), those other variables may be used individually, or combined in some form, to attempt to predict the criterion variable. The criterion variable is usually referred to as the dependent variable, and the one or more predictors are the independent variables. The dependent variable, Y, is said to be a function of the independent variables X_1, X_2, $\ldots X_n$. The nature – e.g., nominal, ordinal, continuous, interval, etc. – of the predictor variable(s) and the criterion variable limits the type of statistical analysis and predictive models that can be developed from the data. (Glass and Hopkins, 1996)

Linear Regression Per prior discussion, GGPA is the most common criterion variable used to measure academic success. When criterion variable values can be ordered along a continuum, such as grades or performance ratings, regression analysis is considered the 'method of choice' for analysis. (Glass and Hopkins, 1996: 152) It is no surprise then that most prior research on admission / selection criteria for graduate degree programs, used linear regression models to predict academic performance from pre-admission factors. (Ragothaman and Davies, 1998: 126; Wilson and Hardgrave, 1995: 187) The purpose of the regression equation is to predict factors of a new sample based on the qualities of a previous sample. (Glass and Hopkins, 1996: 152) Simple linear regression is used when there is one predictor and one criterion variable. Multiple linear regression is used when more than one predictor variable is to be included in the model. Stepwise linear regression improves upon standard multiple linear regression lets the researcher add or remove the predictors to the model one at a time. This allows the researcher to better understand the contributions to the model each predictor is making, and should prevent the inclusion of insignificant and/or redundant predictor variables. (Glass and Hopkins, 1996)

A few studies contended admissions decision makers were most interested in predicting graduate academic performance categorically – such as exceptional / acceptable / unacceptable or the dichotomous graduate / not-graduate – versus a specific numerical GGPA (which is what linear regression methods predict). (Wilson and

Hardgrave, 1995; Mitchelson and Hoy, 1984) In these cases, logistic regression or discriminant analysis are more appropriate techniques.

Logistic Regression is one method capable of predicting a success / fail type of criterion. The underlying basis for logistic regression is similar to that of linear regression, though logistic regression produces a dichotomous prediction, instead of values along a continuum. Logistic regression dos not require the variables to be normally distributed or have a linear relationship between predictors and the criterion. (Glass and Hopkins, 1996: 182-185)

Discriminant Analysis is used when the criterion variable falls on a nominal scale with two or more categories. This classification technique analyzes the differences between categories and provides a method to classify any set of independent variables into the category it most closely resembles. (Glass and Hopkins, 1996: 182-185; Wilson and Hardgrave, 1995: 187).

Neural Networks are a form of artificial intelligence gaining popularity in prediction scenarios where the relationships between the independent and dependent variables are non-linear and complex. Neural networks are software based models, with different software makers using different techniques for various applications. As such, only general statements are possible because how the programs analyze and manipulate the data may vary greatly from one software package to another, and thus some may be more applicable to admissions criteria studies than others. In general, neural networks are capable of recognizing patterns regardless of the functional form of the relationship, and therefore should be able to enhance predictive validity of the quantitative criteria normally used (e.g., UGPA, GRE scores) by adjusting for the effects of subjective factors (e.g., work experience, degree type and school reputation). Neural networks may be used to predict continuous or categorical criterion. (Arnold and Chakravarty, 1996; Wilson and Hardgrave, 1995)

Three studies examining prediction of performance in graduate programs utilized more than one analysis method in their search for the most appropriate prediction models. Wilson and Hardgrave (1995) compared multiple linear regression (MLR), logistic regression, discriminant analysis (DA), and a neural network (NN) system for ability to predict academic success as measured by FYGGPA. To do the comparison, FYGGPA was treated as both a continuous and categorical variable (GPA < 3.00, high risk for academic success; 3.00 < GPA < 3.30, questionable risk; GPA > 3.30, no risk). None of the models developed could accurately predict the high-risk category, a flaw they considered serious. All three categorical models performed better than the linear regression model, though only marginally. Arnold and Chakravarty (1996) used MLR, DA, and a different NN system than Wilson and Hardgrave. The DA models had classification errors rates as high as 40%, with a best of 27%, which was considered too high to be useful by the researchers. The NN model had only an 18% classification error rate, which was still too high to allow it to be used as a sole selection tool, but accurate enough to provide meaningful input to decision makers. The NN model also provided a 33% reduction over the linear regression model's standard error. Hoefer and Gould (2000) compared stepwise regression and neural networks and found nearly identical predictive validity for the two methods. Their conclusions were similar to that of the two previous studies. Neural networks provide at least as much or more predictive validity as linear regression model, and do so while incorporating qualitative factors. This supports the contention that admissions decision-makers should consider qualitative factors along with the more popular quantitative factors.

AFIT Specific Research

At least six previous theses examined AFIT's admission criteria and policies. Three looked at one or two specific graduate education programs while the other three looked across all of AFIT's in-residence master's degree programs in order to evaluate the validity of admissions criteria in use then. All three of the 'across AFIT' studies reached similar conclusions; 1) that the use of GRE, GMAT and UGPA is valid but better methods are available, and 2) different degree programs should use different combinations of factors to predict academic performance. (Buckley, 1989: 46; Sny, 1991: 60; VanScotter 1983: 74)

In his 1989 thesis, Buckley sought to evaluate the effectiveness of criteria AFIT used as predictors of academic performance. (Buckley,1989: 4) He looked at the UGPA, GRE, and GMAT scores of all students – civilian, U.S. officers, and foreign officers (N = 4170) – who attended AFIT in-resident master's degree programs from 1977 to 1987. With GGPA as the criterion variable, Buckley demonstrated that different predictor variables have greatly different predictive validity among the degree programs. For example, UGPA was found to be a significant predictor in only 2 of 9 System & Logistics programs (though neither was GCA), yet it was significant in all 12 of the School of Engineering programs which had at least one significant predictor (Buckley,1989:31). Buckley theorized this is because most AFIT graduate level engineering programs require students to have an engineering degree, and thus the UGPA is from a program subject similar to the master's program. In contrast, the School of Systems and Logistics required only a 4 year college degree, and thus the UGPA may not be in the same subject as the master's degree. (Buckley,1989: 43)

In 1991, Sny's thesis addressed nearly the same issue as Buckley (1989), though Sny went farther back in time to include students from 1975 to 1987 (N = 4507), and he included predictor variables based on students' age, enlisted years of service, and commissioned years of service. Similar to Buckley, Sny concluded no two programs had the same set of valid predictors. (Sny, 1991:60-61). Both Buckley and Sny developed regression models for the GCA program, and these results are summarized in Table 2 It should be noted that the correlation reported by Buckley and Sny may be unrealistically high due to the inappropriate use of Thorndike's correction for restriction of range. Thorndike's range restriction correction is not applicable to this situation because it assumes that selection was based on only one variable (Thorndike, 1949: 175). At AFIT, eligibility is based on at least two variables, and possibly more if the minimum criteria are not met, at which point the departments review the records and can grant eligibility based on subjective and objective factors, or even conversations with the individual.

Prior to Sny and Buckley, Van Scotter also sought to examine the criterion-related validity of the admissions criteria for AFIT resident master's degree programs. He examined students from 1977 to 1982, (N = 2170) and found the GRE, GMAT and UGPA to be valid for only some programs, and the predictive validity varied widely.

Table 2: Comparison of Results of Previous Studies of the GCA program

Author	Buckley r	Buckley N	Buckley p	Sny r	Sny N	Sny p
GMAT-V	0.7885	33	<0.0001	0.5071	18	0.0317
GMAT-Q	0.5623	33	0.0015			
GMAT-T	0.7535	33	<0.0001	0.5712	18	0.0133
GRE-V				0.996	4	0.0040
GRE-Q	0.763	24	<0.0001	0.996	4	0.0040
GRE-A				0.996	4	0.0040
GRE-T	0.6958	24	<0.0001	0.996	4	0.0040

Three additional AFIT theses – by Keith (1997), Prokopyk (1988) , and Spangler (1989) – sought to find the most accurate predictors of academic success for specific master's degree programs. All three studies added variables to the more common set of GRE and/or GMAT scores and UGPA. These additional variables were often less obviously related to academic success, but their inclusion provided some interesting results. The models developed by these three authors support the conclusion of Van Scotter (1983), Buckley (1989), and Sny (1991), that each program has its own unique set of predictors that most accurately predict academic performance.

In 1977, Keith examined the Graduate Systems Management (GSM) and Graduate Operations Research (GOR) programs, using data covering 1973-1976 (N = 216) and found some interesting correlations. He used degree receipt and GGPA as criterion variables, and performed both multivariate regression and discriminant analysis. At that time, selection of students by the assignment system could be on either a voluntary or non-voluntary basis; i.e., some students were assigned to AFIT without ever submitting an application or otherwise requesting the opportunity. Keith found a much higher failure rate for these non-volunteer students, especially if they were unmarried. He suggested non-volunteer status was a surrogate for a measure of motivation. Of 21 unmarried non-volunteers, 10 failed to graduate, compared to no failures among the 11 single volunteers. Other results indicated the GMAT-Q score to be the best predictor of GGPA for the GLM and GOR programs. (Keith, 1977: 41)

In 1988, Prokopyk analyzed eighteen predictor variables to determine their relationship with final GGPA for students in the Graduate Operations Research (GOR) and Graduate Strategy and Tactics (GST) programs. Prokopyk included a 'Quality of Schools' factor in his regression, based on the belief that the schools with higher admission standards will attract better students. This factor had an overall correlation to GGPA of 0.1956. His conclusions include: a) UGPA is the single most significant predictor of GGPA, and b) each variable's exact significance and contribution varies according to the program (Prokopyk, 1988).

The Graduate Logistics Management (GLM) program was the subject of M. E. Spangler's thesis in 1989. He considered 29 predictor variables in an attempt to develop a statistical model for prediction of GGPA using pre-admission information. Building on Prokopyk's (1988) quality of schools indicator, Spangler included a similar factor (called RATE), both singly and as a cross product with UGPA. This combined factor proved a better predictor of GGPA than either individual factor ($r_{RATGPA} = 0.4800$ vs. $r_{RATE} = 0.3707$ and $r_{UGPA} = 0.2437$).

While this improved correlation is noteworthy, Spangler noted that the method he used for collecting the UGPA data, made UGPA "suspected for low reliability."

(Spangler, 1989:24). He used the UGPA as reported on transcripts from the college that granted the undergraduate degree. This is not the UGPA that AFIT uses when making eligibility decisions. GPA calculation methods and treatment of grades – such as pass, fail, incomplete, or withdrawal, as well as grades for classes accepted as transfer credit – vary from school to school. (Spangler, 1989:23-24) A high GPA from a respected college may mask previously sub-standard performance at other schools if the degree granting school considers only grades earned at its school. The two models Spangler developed based on GRE or GMAT scores had R^2 values of 0.59 and 0.54 respectively. However, the GMAT-based model included a dichotomous variable for whether the student was in the Navy or not. Since only 9 of the 140 member sample were in the Navy, it is not clear why this should be a substantial indicator, or more importantly, how useful this would be in reality.

Summary

The amount of prior work in the area of examining graduate admissions criteria and the prediction of graduate academic performance is quite large, and some of the work most appropriate to this study has been reviewed. The measures of academic performance or success varied greatly though GGPA was the most widely used and accepted. The search for the best predictor variables offered even more variation, as many potential factor from an applicant' past were considered in hope of gaining more insight to the applicants true future performance. No one best predictor was found and one of the most common assertions made throughout the literature was that each graduate program has its own unique set of predictors that would provide the most predictive capability, but it is up to the institution to determine what those are, and revalidate them over time. This study will attempt to establish the most useful criteria for use in selection of students for AFIT's GCA program.

It is also important to note that perfect prediction is realistically unattainable, since academic performance is due to much more than cognitive abilities. Indeed, personality traits such as personal striving for excellence, perseverance, conscientiousness, creativity, organization skills, and sociability can have a greater effect on performance than cognitive ability. (Rothstein, and others,1994)

III. Methodology

Chapter Overview

This chapter describes the methods by which this analysis was conducted. The data collection process is discussed, as well as descriptions of the criterion and predictor variables selected for inclusion in this study. Various techniques of linear regression are discussed, along with other statistical methods for verifying and validating the models developed. Finally, potential problems and known shortcomings are addressed along with the methods of alleviation.

Data Collection

Most of the data for this analysis was obtained from the AFIT Admissions and Registrar Directorate, AFIT/RRD. This office maintains educational records of all students who have attended AFIT in-residence degree programs. Permission for this researcher to view the educational records was obtained in accordance with Family Educational Rights and Privacy Act of 1974 as Amended, section 34 CFR § 99.31 (a)(1).

A listing of all students participating in the Graduate Cost Analysis program for classes 92S through 01M was obtained from AFIT/RRD. The educational record of each student in the sample population was manually reviewed and the pertinent data entered into a computerized spreadsheet. Additional information not always contained in the educational record – e.g., Total Active Federal Military Service Date (TAFMSD) – was provided by AFIT/RRD via the Air Force Personnel Data System. Once all information on individual students was collected, a randomly generated record ID # was assigned to each individual record and any personal identifying information, such as name or birth date, was removed from the data set. All personal data was handled in accordance with the provisions of the Privacy Act of 1974.

Population Studied

Data was gathered on all 109 students who attended the AFIT in-residence Graduate Cost Analysis program in classes 92S (started in 1991) through 01M (completed courses not later than April 2001). This total included 100 USAF officers, one foreign officer, six US Army civilian employees, and two US Air Force civilian employees. However, three individuals withdrew from the program; one due to medical problems, one for personal reasons, and one who was released from active duty. Because these three individuals completed only 2 quarters or less of the program before withdrawing, they were excluded from the analysis. Therefore total sample size (N) is 106.

The GCA Program in which these students participated has undergone changes during the time period covered by this study. This study involves 9 graduating classes, 92S through 01M. The first eight classes occurred while the GCA program was part of AFIT's Graduate School of Logistics and Acquisition Management. During that time, the program was 15 months long and graduates earned 66 quarter hours of credit, 8 hours of which were due to thesis work. In October 1999, the Graduate School of Logistics and Acquisition Management merged with the School of Engineering to form the current organization, the Graduate School of Engineering and Management. Under the new organization, the GCA program was lengthened to 18 months, and graduates now earn 72 quarter hours of credit, 12 hours of that due to thesis work. (Air Force Institute of Technology, 2001).

In general, the GCA curriculum has included courses in statistics, organizational behavior, quantitative decision-making, economics, and project management. Academic eligibility requirements have remained constant throughout the study period, and are as follows:

1. an earned baccalaureate degree from an accredited college or university in an appropriate discipline;
2. an overall undergraduate grade point average of at least 3.00 on a 4.0 scale;
3. minimum GRE scores of 500 on the verbal portion and 600 on the quantitative portion, or a GMAT total score of at least 550; and
4. completion of courses in calculus up to (but not necessarily including) differential equations, with a minimum undergraduate math GPA of 3.00. (Air Force Institute of Technology, 2001: 11,169-170) The department responsible for the GCA program may grant waivers to certain admission criteria on an individual basis.

Criterion Variable

The criterion variable in this study is academic performance, as measured by GGPA. Despite the shortcoming of limited range, GGPA is the most appropriate measure available. In this study, many of the disadvantages of GGPA cited in chapter 2, such as inconsistencies of grading policy and attitudes among institutions, are not applicable since this study is looking at a specific program at only one institution. Of course, inconsistencies among faculty may still play a role, but ETS feels this averaging across many teachers and subjects provides a good measure of true ability. (Goldberg and Alliger, 1992: 1025)

The other potential measures of performance addressed in Chapter 2 are not appropriate for this study. Degree attainment and time-to-completion are not appropriate measures of academic performance in this study because none of the 106 students who completed at least half the program failed to graduate, and only two were late. The rarity of failure or time extensions makes statistical analysis practically pointless. If these were the criterion, almost any model could be at least 98% accurate by predicting success for every individual.

Research productivity is not applicable because most GCA graduates are military officers whose primary duty following graduation is usually as a practitioner of cost analysis, not a researcher. Comprehensive exam score is not an applicable variable because GCA students do not take a comprehensive exam as a requirement of graduation. The reliability of faculty ratings is not high enough to justify the effort that would be required to compile the data. Faculty changes and the passage of time would further hamper the reliability of the data.

For all records, GGPA was obtained from an AFIT transcript included in each educational record examined. Not all courses completed at AFIT are included in the GGPA, and courses for which transfer credit was accepted are also not included. The credit hours used for GGPA calculation are shown under the heading 'QHrs' on the AFIT transcript, and total hours of credit are under the heading 'Hrs.' According to the AFIT Graduate Catalog for 2001-2002, academic achievement is indicated by the letter grades and points – used in calculating the grade point averages – as shown in Table 3 (Air Force Institute of Technology, 2001: 22). A review of prior AFIT catalogs confirmed this same grade and point system was in place throughout the study period.

Table 3. AFIT Grade and Points System

Grade	Points	Grade	Points
A	4.0	C+	2.3
A-	3.7	C	2.0
B+	3.3	C-	1.7
B	3.0	D	1.0
B-	2.7	F	0

In addition to the GGPA, data on thesis grade, thesis credit hours, and total quarter hours used to calculate the GGPA, was also recorded during data collection. Thesis grade had been a potential criterion variable, however preliminary data analysis showed this ordinal variable to be limited to only 4 values and extremely skewed toward higher grades, as shown in and Figure 1. The skewness of the distribution may be because of the extensive review and editing process a thesis undergoes before final grades are awarded. Since it is the advisor who reviews/edits the document as well as assigns the grade, a thesis pleasing to the advisor may be more likely to result, even if the student lacks the ability to produce that quality of document on his/her own.

Table 4. Thesis Grade Frequencies

Grade	Points	Count	Prob
B	3.0	7	0.06604
B+	3.3	4	0.03774
A-	3.7	19	0.17925
A	4.0	76	0.71698
	Total	106	1.00000

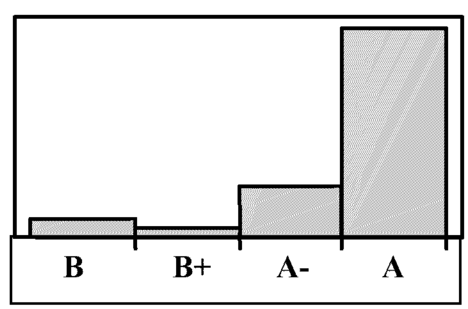


Figure 1. Histogram of Thesis Grades

To remove the effect of the skewed thesis grades on overall GGPA, an adjusted GPA (ADJGPA) was calculated based on all grades other than thesis grade. The calculation is shown below.

$$\text{ADJGPA} = \left[(Qhrs * GGPA) - (ThGrd * ThHrs)\right] \div (Qhrs - ThHrs)$$

where: Qhrs = quarter hours used in GGPA calculation
ThGrd = point value of the thesis grade, per Table 5
ThHrs = quarter hours of credit awarded for thesis work
GGPA = the cumulative grade point average as described above

Separate analysis will be performed on ADJGPA and GGPA, and the results compared to verify whether ADJGPA provides a significant improvement in the validity of the models developed. It is worth noting that due to the program changes that accompanied the organizational changes within AFIT in 1999, credit hours for thesis work increased from 8 hours to 12, and total quarter hours also increased from a standard 66 quarter hours to 72, which increased the percentage of GGPA due to thesis grade form 12.1% to 16.7%. Occasional variations to these totals occurred due to students taking more credit hours than required, or transferring in classes for credit.

Predictor Variables

This section provides definitions, descriptions, and justification of all potential predictor variables included in this analysis.

Scores from GRE and GMAT tests will be included in this study because they are part of the AFIT admission requirements in effect throughout the period covered by this study, and because prior research has shown they have moderate predictive validity in some graduate degree disciplines and programs. All GRE and GMAT scores were obtained from the students' education records maintained by AFIT/RRD. Most education records contained an official score report from ETS, though some contained copies. All information contained in the student educational record was assumed to be authentic and correct. It either test was taken more than once, only the most recent score that occurred before date of entry to AFIT was recorded. If a student took both the GRE and GMAT, both sets of scores were recorded.

As discussed previously, the GRE and GMAT scores are scaled scores. Table 5 provides the scale ranges for each test, as well as the mean, standard deviation, and standard error of measurement, of each score, as reported by ETS. Most tests are not perfect measures of ability, and the standard error of measurement (SEM) is an index of the variation in test scores due to measurement imprecision. For a group of examinees, the SEM estimates the average difference between the observed scores and the true scores. True score is what an examinee would hypothetically achieve if the were no error in the measurement. Roughly 95% of GRE General Test and GMAT test takers should achieve a score within two standard errors above or below their true scores. (GREB, 2000: 13; ETS, 2001: 10)

GRE-V is the scaled score on the Verbal portion of the GRE General Test. AFIT requires a minimum GRE-V score of 500.

GRE-Q is the scaled score on the Quantitative portion of the GRE General Test. AFIT requires a minimum GRE-Q score of 600.

GRE-A is the scaled score on the Analytical portion of the GRE General Test.

GRE-VQ is the sum of the scaled scores of the Verbal and Quantitative portions of the GRE General Test.

GRE-T is the sum of the scaled scores of the Verbal, Quantitative, and Analytic portions of the GRE General Test.

GMAT-V is the scaled score on the Verbal portion of the GMAT.

GMAT-Q is the scaled score on the Quantitative portion of the GMAT.

GMAT-T is the scaled score for the test as a whole. AFIT requires a minimum GMAT-T score of 550.

Table 5. GRE and GMAT Score Statistics

	Range of Scaled Score		Mean	Standard Deviation	Standard Error of Measurement
	Lower	Upper			
GRE-V[a]	200	800	471	114	32
GRE-Q[a]	200	800	569	142	41
GRE-A[a]	200	800	547	131	42
GMAT-V[b]	0	60	28	9	2.8
GMAT-Q[b]	0	60	35	10	3.0
GMAT-T[b]	200	800	528	111	29

[a] source: (Graduate Record Examinations Board, 2000: 14-22)
[b] source: (Educational Testing Service, 2001: 8-13)

UGPA is the Undergraduate Grade Point Average on a scale of 0 to 4.0, as shown in Table 4. In most cases this data was copied from the AFIT Form 95 – a one page form used by the admissions department as a summary sheet of past academic achievements - found in the student's education record. If a UGPA value was not on the Form 95, it was calculated in the same manner used by the admissions department. The UGPA is an average based on all undergraduate course work for which a grade was received, as it appears on the students' academic transcripts. Pass / fail grades are excluded, but failures and repeated classes are included. If an undergraduate institution used a grading system other than as shown in Table 3, AFIT admissions counselors have systems established to convert the non-standard measures to the AFIT letter grade system. Common grading systems requiring conversion are a 0 to 5 point scale, or when an applicant's prior schools included both semester and quarter scheduling systems. Schools using only solid letter grades – e.g., only A, or B, without the differentiation of A- or B+ – are not converted to the +/- system. The UGPA is based on the whole number values associated with the whole letter grades. (AFIT Catalog, 2001; Evans, 2001)

UGPA is included in this study because a minimum UGPA of 3.0 is a requirement of admission to the GCA program.

MGPA is the math GPA. This value is calculated in the same manner as the UGPA, except it contains only grades earned in all undergraduate math and statistics courses. This is included because a math GPA of at 3.00 is an admissions requirement. If this value was not present on the AFIT Form 95, or if the value on the AFIT Form 95 appeared incomplete, it was calculated manually.

RATE is the rating of the admission competitiveness of the undergraduate degree-granting institution of the student. The ratings are on a 6 point scale where 6 is most competitive and 1 the least. Barron's bases their ratings on factors such as percent of applicants admitted, high school class rank and median SAT and ACT scores of incoming freshman. The standards for each rating are provided in Appendix A, and the listing of all schools applicable to this study and their relative ratings are listed in Appendix B. (Profiles of American Colleges 2001, 2000) This variable is included based on

Spangler's finding of improved correlation when school rating and UGPA were combined. (Spangler, 1989: 48-49)

DEGREE is a nominal variable indicating degree type. Variable values are: 1 for all Bachelor of Science (BS) degrees except Bachelor of Science-Business Administration, 2 for Bachelor of Arts (BA), and 3 for a Bachelor of Business Administration (BBA) or Bachelor of Science-Business Administration (BSBA). Graham (1991) considered a similar degree type indicator, and found a correlation to GGPA of $r = 0.168$. Though not a strong relation, it indicates enough positive correlation to warrant inclusion in this analysis.

TIME is a continuous variable to indicate the amount of time – measured in years – between award of the undergraduate degree and entry to the AFIT GCA program. This value is found by subtracting the undergraduate degree completion date from AFIT entry date and dividing by 365. The database for this study was created using Microsoft Excel. In MS Excel, subtracting one date from another returns the number of days between the dates. The number of days is divided by 365 to find the number of years, which is expressed to one decimal place. This variable is included based on the belief that the passage of time since being in an academic environment may affect an individual's performance in future academic endeavors. (Arnold and Chakravarty, 1996; Graham, 1991; Hoefer and Gould, 2000)

AGE is a continuous variable, expressed in years, denoting the age of the student on the date of entry to the AFIT degree program. The value for AGE is found by subtracting the student's date of birth from the date of entry to the AFIT degree program and dividing by 365. Age was included in many prior studies, with mixed results.

(Bieker, 1996; Dunlap, Henley, and Fraser 1998; Graham 1991; Hoefer and Gould, 2000; House, 1998; Kuncel, Hezlett, and Ones, 2001; Matthews and Martin, 1992) Based on these prior studies, age will be investigated for its own predictive ability as well as moderation effect on other variables.

GENDER is a dichotomous nominal variable indicating the gender of the student, where 1 indicates Male, and 2 indicates Female. Based on the many previous researchers who considered gender when attempting to construct a predictive model, (Ahmadi and others, 1997, 1997; Bieker, 1996; Dunlap, Henley, and Fraser, 1998; Graham, 1991; Hoefer and Gould, 2000; Wilson and Hardgrave, 1995), gender will be examined in this study also. Considering Hancock's findings indicating possible gender based bias of the GMAT, this study will also look for any moderation effects gender may have on other predictors. (Hancock, 1999)

PrGGPA is the GPA earned in any prior graduate level coursework. Abedi considered a dichotomous variable to indicate the presence or lack of prior graduate work, but did not find it a significant factor in the models developed.   In this study, it will be looked at only for its correlation to GGPA, but not as part of a predictive model, since applicants are not expected to have previous graduate level experience. In this way, its value as a compensatory factor – one to consider when the basic admissions criteria are not met – will be examined.

Work Experience Though Peiperl and Trevelyan (1997) found only insignificant correlation between GGPA and work experience, Wooten and McCullough (1991) showed that admission decision-makers believed work experience was important. Work experience is included in this study to investigate its validity as a predictor of academic performance in the GCA program. Since time spent in the military service, as an enlisted member or a commissioned officer, can be a measure of employment time, this study will consider the total time in service TMTHS as a proxy for the predictor variable work experience.

TMTHS is the total months of active duty military service, including both enlisted and commissioned time. This value is the sum of EMTHS + CMTHS. EMTHS is a measure of the time in months a student served as an enlisted member of one of the armed services prior to being commissioned an officer. This value was found by subtracting the Total Active Federal Commissioned Service Date (TAFCSD) from the Total Active Federal Military Service Date (TAFMSD) and dividing by 30. Values of 12 or below were converted to 0 for this analysis because it was assumed these short time periods did not represent actual time as an enlisted service member. These EMTHS values of less than 12 were assumed to be due to delayed enlistment, delayed reporting following graduation from a Reserve Officer Training Corps, ROTC program, and/or attendance at Officer Training School. In all cases, a member may be in active duty status for up to a year before being commissioned as an officer. (DoDFMR 7000.14-R, 2002; AFI36-2604, 1999; AFI36-2009, 1999; AFI36-2013, 1994)). CMTHS is a measure of time, in months, of service while a commissioned officer in the US Air Force, before entering the AFIT degree program. CMTHS was calculated by subtracting the TAFCSD from the AFIT date of entry, and dividing by 30.

Van Scotter (1983) and Sny (1991) both found negative correlations for years of enlisted service to GGPA and mild positive correlation between years of commissioned service and GGPA. Both used years as the unit of measure, and neither mentioned the possibility that values near 1 may have been due to the situation described above.

ENLST To investigate whether being prior-enlisted has some predictive ability, a dummy variable (ENLST) was created, where 1 = prior active military service as an enlisted member of any branch of the US Military, 0 = no enlisted service. This variable is not present for the civilian members of the sample population.

Selection of Analysis Method

Unfortunately, the subject of statistics can be very involved and a full explanation of all concepts presented is beyond the scope of this document. However, general descriptions are provided in an attempt to ensure the reader will understand the general ideas and methods presented

Multiple linear regression (MLR) and Stepwise MLR will be used to build the statistical models for this study. Of the four broad methods of statistical analysis discussed in Chapter 2, logistic regression is not applicable to this particular study, discriminant analysis has shown only marginally better performance under limited circumstances, and artificial neural networks have not shown sufficient superiority to justify their added cost, risk, and effort.

Logistic regression predicts a binary response variable, not a continuous one. In this study, the binary variables that could be used for academic performance – such as degree attainment or graduated on time – are not appropriate for the sample data set. This is because no member of the data set failed to graduate, and only four did not graduate on time, of which two graduated late – for reasons that are not known to this researcher – and two graduated early. The early graduates would be considered successes, so only 2 of 106 graduated late. Based on these criteria, the current selection system is already 100% accurate in predicting graduation, and better than 98% accurate in predicting on-time graduation, and needs no improvement.

Previous research using discriminant analysis (DA) in examining graduate student selection has shown mixed results. Arnold and Chakravarty (1996), and Wilson and Hardgrave (1995), both showed marginally better results using DA compared to linear regression techniques, when predicting graduate student success. However, Arnold and Chakravarty (1996) considered an unsuccessful student to be a student who earned one or more grades equivalent to a C or below, on a 4-point scale. This criterion is questionable as a measure of performance because it fails to consider overall success, allowing what may be limited performance problems to overshadow other successes.

Wilson and Hardgrave (1995) used two techniques for model construction. The first used the whole data set (n = 156) to construct the different models, and the resulting the least squares regression model had 52% prediction accuracy compared to 53% for DA. Then, ten other models for each analysis technique were constructed using 10 data sets of 51 items each, randomly selected from the original data set. When the average classification accuracy of these 10 model-sets were compared, the DA models' average accuracy was better for the high-risk classification (44% vs 10% for the LSR), but only 5 percentage points better (44% vs. 38%) than least squares regression in overall accuracy. The mild advantages of DA over linear regression in this type of application appear to be dependent more on the method of classification than on the statistical process itself. Additionally, classifying GGPA – which is already somewhat restricted in range – into categories would not provide an advantage to this study, since all GGPAs in the data set are above the AFIT requirement of 3.00, i.e., none are considered unsuccessful by AFIT standards. Dividing the one point range would further reduce the differentiation in an already narrow range.

Though artificial neural networks (ANN) did show potential to provide better predictive validity than the other three options, this method will not be used in this study. Each study that used artificial neural networks used a different software package, and this is an example of the problem. Regression and discriminant analysis use commonly accepted statistical practices and proven mathematical formulas. ANNs on the other hand, are complex interconnected structures of mathematical models and algorithms whose information-processing paradigm is inspired by the densely interconnected, parallel structure of the mammalian brain. (Battelle Memorial Institute, 1997) There are multitudes of different types of ANNs, and each software maker may utilize different algorithms, processes, and structures. The research reviewed in this study had results that ranged from statistically equal to a 33% reduction in standard error when compared to the relative regression models. Whether this variation in utility is due to the software or the operators, or the situation, is impossible to determine. Additionally, one researcher described the development process as "…an ambiguous and arduous task and, at present, one has difficulty in identifying those independent variables that are the best predictors--information the traditional techniques [regression] can easily provide." (Wilson and Hardgrave, 1995: 193) The cost of ANN software and the potential to choose software less appropriate than others was deemed not worth the risk for this study, when stepwise regression is a readily available and proven technique.

Linear Regression Analysis

Though there are numerous regression analysis techniques, most prior research on admission / selection criteria for graduate degree programs, used linear regression models to predict academic performance from pre-admission factors. (Ragothaman and Davies, 1998: 126; Wilson and Hardgrave, 1995: 187) In *simple linear regression*, the continuous dependent variable, Y, is predicted from a single independent variable X, where both Y and X are assumed to be independent, normally distributed, and linearly related. The basic linear normal error regression model is:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad (1)$$

Where: Y_i = the observed response in the ith trial
X_i = the observed value of the predictor variable in the *i*th trial
β_0 = the Y-intercept of the true regression line
β_1 = the regression coefficient and also the slope of the true regression line
ε_i = the random error term, assumed to be independent and normally distributed. (Neter and others, 1996: 29)

This equation (1) describes the true relation, which is theoretically undeterminable, but it can be estimated with reasonable certainty with the estimated regression equation. The estimated regression equation (2) describes a line that best fits the scatter plot of the X,Y data pairs. Best fit is determined by the least squares criterion, in which the sum of the squared deviations from the predicted values ($\hat{Y}_i$) of the criterion to the observed values of Y (Y_i) is minimized. (Neter and others, 1996: 1-29)

$$\hat{Y}_i = b_0 + b_1 X_i \tag{2}$$

In the estimated simple linear regression equation (2), $\hat{Y}_i$ is the estimated or fitted value of Y in the *i*th trial, b_0 is the estimated intercept of the regression line (β_0) with the Y axis when X = 0, b_1 is the estimated regression coefficient (β_1) and also the slope of the line, and X_i is the observed value of the predictor variable in the *i*th trial. The regression coefficients (b_0 ,b_1) are calculated from the correlation and standard deviations of the independent and dependent variables. (Neter and others, 1996: 1-23; Glass and Hopkins, 1996: 152-189)

Ideally, a linear model should account for all variance in the criterion variable. In the case of a model using multiple independent variables, each variable should account for a portion of the variance not accounted for by another independent variable. For example, if a study is looking at job performance, and among the independent variables considered are age and time on the job, there will likely be inter-correlation between these two factors, because only older people would have had an opportunity to have greater time on the job. Both may be positively correlated to job performance, but the 15% of variance accounted for by age is also accounted for within the 25% covered by time on the job. Including both in the model would still account for only 25% of the variance because 15% is redundant between the two predictors. (Glass and Hopkins, 1996: 171-177)

When more than one predictor variable is being investigated, and all meet the requirements of being linearly related to the criterion variable and normally distributed, then these additional variables may be incorporated into a model using *multiple linear regression*. The estimated multiple linear regression equation (3) is similar to that for simple linear regression, but incorporates the additional predictors and relative regression coefficients.

$$\hat{Y}_i = b_0 + b_1 X_{1i} + b_2 X_{2i} \ldots + b_m X_{mi} \tag{3}$$

The regression coefficients $b_1 \ldots b_m$ weight the m predictor variables X_1, $X_2 \ldots X_m$, so that they are combined in a manner that most accurately predicts Y. The least squared criterion is again employed in determining regression coefficients that minimize the squared deviations (errors) between $\hat{Y}$ and Y. This also ensures the greatest correlation between $\hat{Y}$ and Y. (Glass and Hopkins, 1996: 171)

Calculation of these coefficients is possible by hand, though tedious. Thankfully, there exist computerized statistical analysis packages – such as JMP® – that can handle these calculations rapidly. However, while these systems will find the best model based on the variables provided, they will not remove a predictor variable that is not significant – e.g., due to redundancy or a lack of predictive validity – or that otherwise decreases the models effectiveness. It is then up to the operator to figure out what combination of independent variables provides the best model. If the number of predictor variables (m) is large, trying all combinations, even with computer assistance, can be laborious as well. Stepwise linear regression alleviates some of this. (Glass and Hopkins, 1996: 175)

Stepwise Multiple Linear Regression is useful when the number of potential predictor variables is large because it allows the researcher to search for the best model by adding or removing variables and monitoring the effects on the model. This process of searching for the best model may be done via methodologies known as forward and backward stepwise regression. In each method, the criterion for the addition or removal of a predictor variable can be any of the following; reduction of the error sum of squares, coefficient of partial correlation, the t* statistic, or the F* statistic. This author will use the F* statistic.

When performing forward stepwise regression, the starting point is a list of the potential predictor variables and the basic linear regression equation $Y = \beta_0$ + error (i.e., the equation includes no predictor variables). When using JMP®, the F* value for each predictor is shown, and the first predictor that should be entered into the model is the one with the greatest F* value. The F* value is actually a partial F test, which is the statistic for testing the hypothesis that $\beta_k = 0$. Since b_k is the coefficient for the variable X_k, in the linear regression equation (3), a value of zero would indicate the predictor X_k is not worth including in the model. After each step in the model building process, the F* values for each predictor variables are recalculated based on the current model, which is considered the reduced model – the model with less predictors variables in it compared to the full model. The full model is the hypothetical model after the next step, i.e., with the next predictor added to it. For each predictor variable not yet in the model, the relative F_k* value provides a comparative measure of the reduction in total variance if variable X_k were added during the next step. For the X variables already in the model, the F* shown is an estimated comparative measure of the amount of variance accounted for by their individual presence in the model, assuming the other predictors are also in the model.

The F* value is found using the following equation (4):

$$F_k^* = \frac{SSE_R - SSE_F}{df_R - df_F} \div \frac{SSE_F}{df_F} \tag{4}$$

where: F_k^* = the partial F statistic
SSE_R = the Sum of Squared Errors of the Reduced model
SSE_F = the Sum of Squared Errors of the Full model
df_R = he degrees of freedom for the Reduced model
df_F = he degrees of freedom for the Full model.

Backward stepwise regression is, as one might expect, very similar to forward stepwise regression. However, the starting point is a model containing all the predictors. Variables are removed from the model starting with those with the smallest F_k^* values, and continue to be removed until an acceptable model is developed. It is important to note that during the model building process, large fluctuations in the F* values or other measures of model validity may warrant return of a previously removed variable to the model. The similar situation applies to forward regression, i.e., a variable previously added may be removed if the F* value is decreased below the established threshold.

As the models are developed, the basis for judging predictive validity will be the adjusted coefficient of multiple determination, also called the adjusted r^2 (Δr^2) while accepting only predictors with significance levels as measured by p-value < 0.05. The Δr^2 is a more appropriate measure than the coefficient of multiple determination, r^2, when building multiple regression models because it takes into account the number of predictor variables. Once the regression model contains at least one predictor, the addition of more predictors will never reduce r^2, since the coefficients are adjusted during each step to continually improve the model's fit. Even though r^2 may be large, that does not mean the model is useful. The Mean Squared Error (MSE) may still be too large to allow inferences to be made if precision is required. In contrast to r^2, Δr^2 may get smaller when a predictor is added. (Neter and others, 1996: 230)

Once a model has been developed, via simple, multiple or stepwise multiple linear regression, it will be checked to ensure it meets the basic assumptions of least squares linear regression. Recall that ε represents the error present in the estimated regression model.

Assumption 1: The mean of the probability distribution of ε is 0.

Assumption 2: The variance (σ^2) of the probability distribution of ε is constant over the range of the independent variables $X_{i,1} \ldots X_{i,k}$.

Assumption 3: ε is normally distributed.

Assumption 4: ε is independent, i.e., ε for one value of Y is not effected by, nor effects, any other ε associated with any other value of Y. (McClave and others, 2001: 473)

Analysis of residuals can be used to verify these assumptions. Residuals are denoted as e_i and are the vertical deviation of Y_i from $\hat{Y}_i$, i.e., the vertical distance from the observed to the predicted value of Y on the estimated regression line. This is how far off the prediction was from the actual value of the criterion variable. In a least squares model, the sum of the residuals is, by definition, zero. (Neter and others, 1996)

Verifying normality and a mean of 0 can be done in JMP® by plotting a histogram and a normal quantile plot of the residuals. Plotting the residuals against the fitted values is useful for assessing constancy of error variance. Residuals should also be plotted against each predictor variable to check for independence and possible non-constancy of error variance in relation to specific predictors. (Neter and others, 1996: 98-

111, 236-241). Other tests are available if these basic tests indicate a possible problem. These will be discussed as needed in chapter 4.

Stepwise linear regression will be used to establish baseline models using the admissions criteria as described in the AFIT catalog. Separate models will be developed based on GRE and GMAT scores. Then stepwise regression will be used to investigate combinations of the remaining potential predictor variables to find an optimal model(s) and possibly provide insight into variables useful in compensatory selection.

Restriction of Range

Restriction of range is a problem that is addressed in nearly every research effort reviewed that evaluated the validity of using cut-off scores in admission decisions. When a test such as the GRE or GMAT is used as a selection instrument, the student body consists of only those who met or exceed the minimum, or cut-off criteria. If the cut-off score was high compared to average performance, any subsequent attempt to investigate the correlation between the test score and some criterion will likely be hampered by the effects of range restriction. The performance of the unselected individuals is not known, and thus cannot be included. This may cause the study to greatly underestimate the test's true predictive validity. (Hopkins, 1998: 97) Figure 2 illustrates this well, showing how "[t]he validity coefficient *within* the selected group underestimates the actual predictive value of the test." (Hopkins, 1998:98)

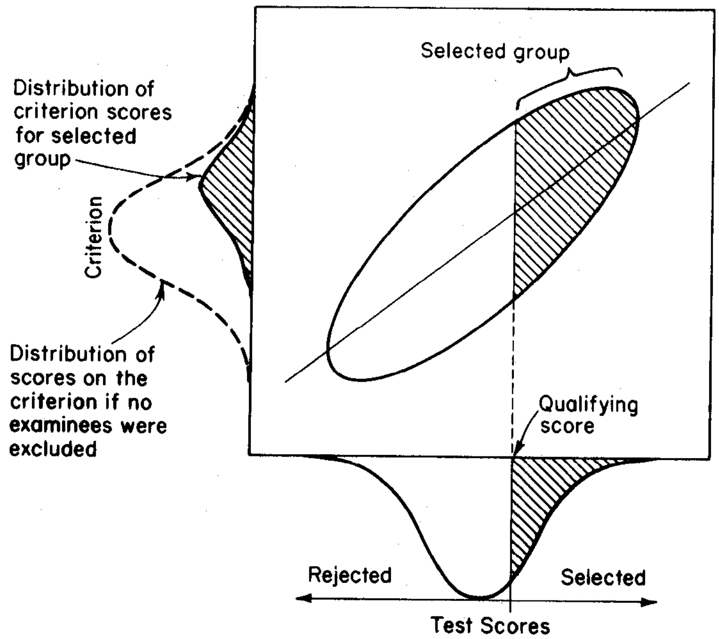


Figure 2. Illustration of Range Restriction on Correlation

(Source: Hopkins, 1998: 98)

Thorndike (1949) demonstrated how drastic an effect this may be. In a study of pilot training, all applicants were admitted no matter how poorly they performed on the selection tests. Biserial correlation coefficients relating pass-fail in training to test scores were computed for the total group (N_T = 1036) and the group that met the selection criteria selected, i.e., were qualified (N_Q=136). Correlation coefficient on the composite test score were 0.64 for the total group, compared to just 0.18 for the qualified group. Other test had similar disparities. (Thorndike, 1949: 169-176)

Some of the reviewed research utilized a correction procedure developed by Thorndike. (Buckley, 1989; Sny, 1991; Van Scotter, 1983) While the procedure Thorndike developed may be valid, it does not apply to studies of AFIT admission policies and processes because it assumes that selection was based strictly on only one variable (Thorndike, 1949:175). At AFIT, eligibility is based on at least two variables (UGPA and GRE or GMAT), and possibly more if the minimum criteria are not met and the waiver process is used. As Thorndike puts it:

> When selection is based, as it often is, on a clinical judgment which combines in an unspecified and inconstant fashion various types of data about the applicant, and when this judgment is not expressed in any type of quantitative score, one is at a loss as to how to estimate the extent to which the validity coefficient for any test procedure has been affected by that screening. (Thorndike, 1949:176)

Data Analysis

Multiple linear regression requires that a value be present for each predictor variable (X_{ij}) in order for the record j to be included in the regression analysis. In other words, if record #17 has values for the GRE scores and UGPA, but is missing the MGPA value, all values in record #17 would be excluded from the regression analysis if MGPA is to be part of it. Most AFIT applicants took either the GRE or the GMAT, not both. Consequently, most records are missing either GRE scores or GMAT scores. If a regression analysis was performed using the entire data set (N = 106) and included both GRE and GMAT scores, all records with just one type of score would be excluded. Due to this requirement, the data set was divided into two subsets, one containing all records with GRE scores ($n_{GRE} = 61$), and the other for all records containing GMAT scores ($n_{GMAT} = 59$). The GRE and GMAT data subsets were each further divided into training and validation data sets.

Neter and others (1996) recommend that the number of cases in the model-building (i.e., training) data set be 6 to 10 times the number of variables in the predictor pool. (Neter and others, 1996: 437) Since the baseline models will contain 3 to 4 variables, and the improved models may realistically be expected to contain as many as 6 predictors, a minimum size of 48 data records was desired for the training data sets. Based on this, the GRE and GMAT subsets were divided according to the following sizes the GRE data set (n_{GRE} = 61) into a training set of 50 records, reserving 11 for the validation data set; and the GMAT dataset was divided into a training set of size n_{trg} = 48, and a validation data set of size n_{val} = 11.

Once divided according to GRE or GMAT data, each record in the full data sets was assigned an ID # from 1 to n_{GRE} or, as appropriate. These sets were then separated into training and validation sets according to the following random selection method. A random integer generator function (RANDBETWEEN) in Microsoft Excel® was used to generate random numbers between 1 and the n_{full} value for each data set. The cell containing the function was cycled by pressing the F9 key on the keyboard and the resulting numbers recorded until the appropriate n_{val} number of unique integers had been recorded. The records with the matching ID numbers were then removed to form the validation set, and the remaining records become the training set.

The validation data sets will be used to check the predictive ability of the associated models developed. A predictive model developed using regression techniques and a given data set is chosen, at least in part, by how well it fits the given data. Use of a different data set may have lead to the choice of different coefficients (β_0, β_1 etc) or even different predictor variables. The result of this unavoidable flaw in the modeling process is that the mean squared error (MSE) of the model – see equation (5) – will tend to understate the inherent variability of future predictions made with the model. (Neter and others, 1996: 435).

$$MSE = \frac{SSE}{n-p} = \frac{\sum\left(Y_i - \hat{Y}_i\right)^2}{n-p} \qquad (5)$$

where: n = number of observations in the data set
p = the number of predictor variables used in model
SSE = the Sum of Squared Error
Y_i = is the observed response in the *i*th trial
$\hat{Y}$ i = predicted response in the *i*th trial

To measure the predictive capability of the training set model, use it to predict each response in the validation set, and calculate the mean squared prediction error (MSPR):

$$MSPR = \frac{\sum_{i=1}^{n*}(Y_i - \hat{Y}_i)^2}{n*} \qquad (6)$$

where: n* = number of observations in the validation set
Y_i = value of the response in the *i*th validation case
$\hat{Y}_i$ = predicted value of the response in the *i*th validation case based on training data set

If the MSPR is fairly close to the MSE of the training set-based model, then the MSE is not overly biased and can be considered to be an appropriate indicator of the training set-based model's predictive ability. (Neter and others, 1996: 435-436)

Summary

This chapter described the sample population, data collection, and data analysis methods. The criterion variable of this study is academic performance, and GGPA was established as the measure of academic performance. The usefulness of an alternative measure of academic performance, ADJGPA will also be examined in this study. The predictor variables were described, and the processes of simple, multiple and stepwise linear regression were reviewed. The process of model development and comparison was also described, as well as the required division of the data into subsets according to the presence of GRE or GMAT test scores. Chapter 4 discusses the results of the analysis, and chapter 5 will discuss the conclusions drawn form those results.

IV. Results

Chapter Overview

This chapter presents the results of the statistical analysis described in the previous chapter. Two regression models were developed, using only UGPA, MGPA, and either the GRE-V and GRE-Q scores, or the GMAT-T score. These are referred to as baseline models because they represent the predictive capability of the current admissions criteria. Two additional models, referred to as 'improved' models, were developed using all the remaining predictors included in this study. These models are compared and a discussion of other predictors of performance concludes this chapter.

Preliminary Data Analysis

The descriptive statistics shown in Table 6 were developed using all available data. Because not every record contained a value for each variable, the N for all variables is not 106. As an example, most of the records contain scores for either the GRE General Test or the GMAT but not both, since most students did not take both tests.

The correlation matrix was reviewed to identify the potentially strongest predictors of GGPA and ADJGPA, to note possible intercorrelation among predictors, and to identify potential relationships among variables that are unexpected or could aid in understanding the overall situation. Some intercorrelation is expected, such as that among the time related variables of AGE, TIME, CMTHS, and THMHS, as well as intercorrelation among GRE scores. There are also some unexpectedly high intercorrelations, and a few values that seem to contradict each other.

Table 6. Descriptive Statistics

Variable	N	Mean	S.D.	Min	Max	r_{GGPA}	Prob > F
GGPA	106	3.66	0.20	3.012	4		
ADJGPA	106	3.64	0.21	2.972	4	0.986	
GRE-V	61	527	78.75	370	750	-0.032	0.808
GRE-Q	61	667	72.14	480	790	0.205	0.113
GRE-A	61	627	105.41	320	800	0.176	0.175
GRE-T	61	1821	191.89	1390	2220	0.161	0.216
GRE-VQ	61	1194	113.64	1000	1500	0.108	0.406
GMAT-V	53	33.0	4.2	24	43	0.373	**0.006**
GMAT-Q	53	34.4	5.1	24	45	0.281	**0.042**
GMAT-T	59	563	52.67	450	680	0.394	**0.002**
UGPA	105	3.05	0.37	2.35	3.92	0.206	**0.035**
MGPA	104	2.86	0.63	1.51	4	0.096	0.331
RATGPA	102	12.30	4.60	3.2	20.28	0.169	0.090
AGE	106	29.4	5.2	21.9	54.7	-0.137	0.160
TIME	106	6.3	4.2	0.6	23.8	0.023	0.814
EMTHS	19	81.2	33.5	31	137	0.085	0.731
CMTHS	94	57.9	30.4	9	181	-0.123	0.237
TMTHS	94	74.3	46.3	9	200	-0.281	**0.006**

	Category	Frequency
GENDER		
	1 = Males:	91
	0 = Females:	15
ENLST		
	1 = Prior-Enlisted:	19
	0 = Not Prior-Enlisted:	87
DEGREE		
	1 = BS	67
	2 = BA	13
	3 = BSBA/BBA	26

	Category	Frequency
RATE		
	1	7
	2	9
	3	23
	4	23
	5	9
	6	32

EMTHS shows some strong correlation to GRE-Q, and GRE-A, and GRE-T scores, but much less correlation to GMAT-Q, and GMAT-T, however these values are based on only 9 or 10 data points, and had p-values in excess of 0.22 and are thus not considered significant. GRE-V and GRE-A were mildly correlated with AGE and TIME at significance $p<0.05$, though GRE-Q and GMAT scores showed little correlation to AGE and TIME.

Scatter plots of each predictor variable against both GGPA and ADJGPA were created to allow a visual inspection of the relationship between predictor and criterion variables. If the predictor was a categorical variable, a oneway analysis plot was created. The categorical means were tested for statistically significant differences among them using ANOVA and the all pairs studentized t test. Selected scatterplots and oneway analysis plots are included in Appendix E.

Of the four categorical predictor variables – RATE, GENDER, ENLST, and DEGREE – only ENLST had a significant difference in mean GGPA and ADJGPA according to whether the student was prior-enlisted or not. For prior-enlisted students, the mean GGPA and mean ADJGPA are 3.554 and 3.5154 respectively, where as the mean GGPA and mean ADJGPA of non-prior-enlisted students are 3.697 and 3.674 respectively. This agrees with the findings of VanScotter (1983) and Sny (1991) who found negative correlations between years of enlisted service and GGPA.

GRE Baseline Models

The GRE Baseline models were developed using the GRE training data set (n = 50) and included only the predictor variables that are part of AFIT's objective admission criteria, i.e., GRE-V, GRE-Q, UGPA and MGPA.

<u>*GRE Baseline Model for GGPA*</u> The first multiple linear regression run included all four variables and produced a model with the following qualities, $r^2 = 0.127$, $\Delta r^2 = 0.0497$, and prob > F = 0.1805. Seeking to improve on these results, the stepwise linear regression process in the JMP® computer program was used to develop a new the model, with the following results: $r^2 = 0.125$, $\Delta r^2 = 0.0681$, and prob > F = 0.1016. Interactions among the four predictor variables were investigated by including in the stepwise regression process the cross products of all pairwise combinations of the four predictors, along with the four predictors. All but one cross-product variable was eventually eliminated from the model. The final GRE Baseline Model for GGPA produced achieved $r^2 = 0.201$, $\Delta r^2 = 0.130$, and prob > F = 0.0352. The parameter estimates are shown in Table 7, while the actual by predicted plot, ANOVA table, residual plots and other statistical data is provided in Appendix D.

Table 7. Parameter Estimates: GRE Baseline Model for GGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|
Intercept	b_0	2.60377	6.33	<.0001
GRE-Q	b_1	0.00099	2.45	0.0184
UGPA	b_2	0.23609	2.32	0.0247
MGPA	b_3	-0.11552	-2.21	0.0320
$(GRE\text{-}Q - 664)^2$	b_4	0.00001	2.07	0.0441

Examination of the Actual by Predicted plot (Figure 3) identified a possible outlying result – indicated by the arrow.

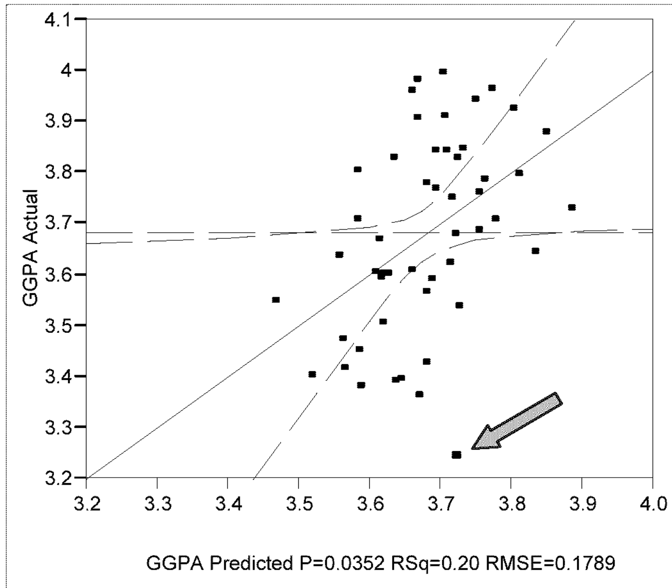


Figure 3. Actual by Predicted Plot: GRE Baseline Model for GGPA

To test the effect this potential outlying record had on the model, the MLR process was repeated with the record excluded. The alternative model had a much-improved ΔR^2 (0.1916) and the p-values remained below the 0.05 threshold for significance. Cook's distance measure (D_i) was used to evaluate the influence this record was having on all the predicted Y values. Interpretation of Cook's distance measure can be done by relating D_i to the F(p, n - p) distribution to find the corresponding percentile value. In this case, D = .118, and corresponding percentile for F(4, 46) is 2.45% – well below the 50% level considered to indicate major influence and justify consideration as an outlier. (Neter and others, 1996: 380-382)

The record was then examined to determine if there was an error or any reason to exclude the record from the study. No errors were found, and though it is the lowest GGPA in the GRE Training data set – and the third lowest in the entire data set – it is within 3 standard deviations of the mean and is not an outlier. No other variables within the record can be considered outlier, and no other justification was found to exclude the record. Regarding outliers, Neter and others (1996: 762) recommend against discarding a variable unless the extreme value is due to some sort of measurement or data recording error. This was not the case, therefore, the GRE Baseline Model for GGPA, as first described, will remain.

GRE Baseline Model for ADJGPA The analogous GRE baseline model for predicting the ADJGPA was developed in nearly the exact same manner as the GRE Baseline Model for GGPA, and the final model contained the same four predictors. The results are slightly better; $r^2 = 0.218$, $\Delta r^2 = 0.148$, and prob > F = 0.0234, with all predictors having p-values < 0.05. Here again, the model parameters are provided below (Table 8), and the full statistical information is provided in Appendix D.

Table 8. Parameter Estimates: GRE Baseline Model for ADJGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|
Intercept	b_0	2.40986	5.56	<.0001
GRE-Q	b_1	0.00111	2.62	0.0119
UGPA	b_2	0.27227	2.55	0.0144
MGPA	b_3	-0.12566	-2.29	0.0270
$(\text{GRE-Q} - 664)^2$	b_4	0.00001	2.10	0.0414

The predictive ability of the GRE Baseline models was then checked according to the method recommended by Neter and others (1996) and discussed in chapter 3, which compares the Mean Squared Error (MSE) of the training model to the Mean Squared Prediction Error (MSPR). For the training set models, $MSE_{GGPA} = 0.0320$ and $MSE_{ADJGPA} = 0.0355$, which compare at an acceptable level to the validation set's $MSPR_{GGPA} = 0.0523$ and $MSPR_{ADJGPA} = 0.0510$.

GMAT Baseline Models

GMAT Baseline Model for GGPA The GMAT Baseline Model was developed using the GMAT training data set (n = 48) and included only the predictor variables that are part of the objective admission criteria, i.e., GMAT-T, UGPA and MGPA. The multiple linear regression process in the JMP® computer program was used and the model produced had the following properties; $r^2 = 0.225$, $\Delta r^2 = 0.173$, and prob > F = 0.0099. However, the prob > |t| exceeded 0.05 for both UGPA and MGPA. Stepwise regression was then tried, and the resulting model ($r^2 = 0.224$, $\Delta r^2 = 0.190$, and prob > F = 0.0033) did not include MGPA, but both remaining predictors were significant.

To be consistent with the process used for the GRE baseline models, interactions were investigated in the same way, though none proved significant enough to warrant inclusion. Thus, the GMAT Baseline Model for GGPA contains only GMAT-T and UGPA, with parameters as shown in Table 9, and full statistical data provided in Appendix D.

Table 9. Parameter Estimates: GMAT Baseline Model for GGPA

Term	Parameter	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	b_0	2.375078	0.362664	6.55	<.0001
GMAT-T	b_1	0.001530	0.000508	3.01	0.0042
UGPA	b_2	0.141234	0.066297	2.13	0.0386

GMAT Baseline Model for ADJGPA A similar analysis for developing the GMAT Baseline Model for ADJGPA produced similar results. Using all three predictors in MLR produced a model that achieved the desired level of significance, but contained variables that did not. Stepwise regression then produced a model, significant in all aspects, with the following properties; $r^2 = 0.235$, $\Delta r^2 = 0.201$, and prob > F = 0.0024. Consideration of interaction effects failed to provide a better model. The GMAT Baseline Model for ADJGPA contains only GMAT-T and UGPA, with parameters as shown in Table 10, and full statistical data provided in Appendix D.

Table 10. Parameter Estimates: GMAT Baseline Model for ADJGPA

Term	Parameter	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	b_0	2.24266	0.38067	5.89	<.0001
GMAT-T	b_1	0.00166	0.00053	3.12	0.0031
UGPA	b_2	0.15193	0.06959	2.18	0.0343

The predictive ability of the GMAT Baseline models was then checked according to the method recommended by Neter and others (1996) and discussed in chapter 3, which compares the Mean Squared Error (MSE) of the training model to the Mean Squared Prediction Error (MSPR). For the training set models, $MSE_{GGPA} = 0.0324$ and $MSE_{ADJGPA} = 0.0357$, which compare at an acceptable level to the validation set's $MSPR_{GGPA} = 0.0512$ and $MSPR_{ADJGPA} = 0.0597$.

For each of the four models developed thus far – GRE- and GMAT-based – the four assumptions required of linear regression – that error (ε) is independent, normally distributed, with a mean of 0, and constant variance – were verified using residual plots, as discussed in chapter 3.

Investigation of Non-Admissions Criteria

The initial task of investigating the potential to improve the predictive ability of the baseline models was to examine the relationships between the non-academic type predictors and GGPA and ADJGPA. This was done in two steps. Having already reviewed the correlation matrix and scatterplots, stepwise MLR was used to build models which did not include standardized test scores (e.g., GRE and GMAT scores) and undergraduate grade point averages. This allowed use of the full sample data set (n = 106), and included as predictors: RATE, GENDER, AGE, ENLST, TMTHS, DEGREE, RATGPA, and TIME.

Two models were developed using the eight predictor variables with GGPA and ADJGAP as the criterion variables. Because these models were for investigation of predictors and not to serve as complete models, a validation set was not reserved, and consequently the model was not validated as were the baseline models.

According to the models, RATE and ENLST were the most useful of the eight predictors of GGPA and ADJGPA. However, RATE was important in an unexpected manner. As shown in Figure 4, the categorical mean GGPAs did not increase along with RATE, as would be expected if RATE was an indicator of student capability. Instead, the means decreased and then increased, as RATE went from 1 to 6. Thus, when predicting

GGPA or ADJGPA, JMP considered it significant if students went to either schools rated 1, 2 and 6, or 3, 4, 5. While the categorical means do vary, a comparison of means using the Studentized t statistic indicated there were no statistically significant differences between the means. A table of these results is provided in Appendix E.

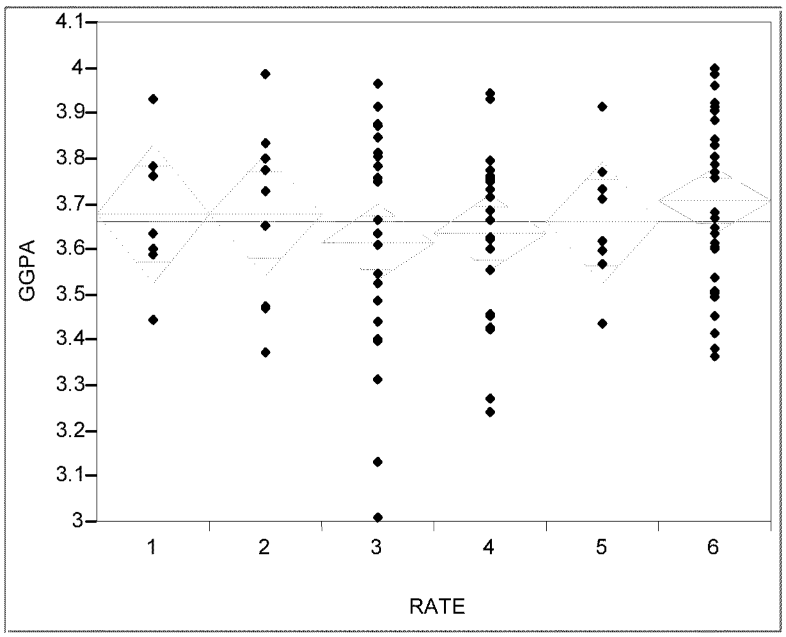


Figure 4. Oneway Analysis of GGPA by RATE

Another unexpected result of this portion of the analysis was the importance of the dichotomous ENLST variable. In the stepwise processes for each model, inclusion of ENLST, and exclusion of TMTHS, provided a better model of this data, despite the fact that only 19 prior-enlisted members were part of the data set, and TMTHS correlation to GGPA and ADJGPA was relatively strong (for this study) r = -0.2808 and –0.2960. The improvement due to ENLST instead of TMTHS varied, but ENLST always provided more improvement in Δr^2 than did TMTHS.

Improved GRE-based Models

Due to the large number of predictor variables, an iterative process of stepwise multiple regression analyses was performed. The first set included GRE-V, GRE-Q, GRE-A, UGPA, MGPA, and then various sets of 3 or 4 of the remaining predictors. The models with the highest Δr^2, while maintaining model and variable significance levels at $p < 0.05$, if possible, were saved and compared to the best model from the next subset analysis. These repeated analyses led to the identification of the variable combinations that provided the most appropriate models. The models were then examined for pairwise interactions among the variables. If interaction effects could improve the model, these cross products were included in the final model.

Improved GRE-based Model for GGPA The most effective linear regression model, based on scores from the GRE General test included the predictors; GRE-A, UGPA, MGPA, GENDER, and ENLST, and achieved the following results: $r^2 = 0.281$, $\Delta r^2 = 0.194$, and prob > F = 0.016. However, four of the five predictor variables exceeded $p < 0.05$, though three are within $p < 0.10$ level of significance. ENLST exceeded even $p < 0.1$, but was kept in the model because removing it reduced Δr^2 to 0.067 and made the model insignificant (Prob > F = .130).

Interactions were investigated next and resulted in an improved model. All cross products were eventually eliminated, except for (ENLST - 0.128)*(GRE-A - 627). The JMP program automatically subtracts the average of each cross product member from itself. This method of transformation prevents the cross product from being a linear combination of two other variables within the model. (Note: The averages used in the transformation are based on the records used in the regression analysis, not the overall means as reported in the descriptive statistics table.)

The final GRE-based Model for GGPA achieved $r^2 = 0.350$, $\Delta r^2 = 0.253$, and prob > F = 0.0061 and was excepted despite some predictors exceeding the desired threshold of significance of prob $|t| < 0.05$. The model parameters are shown in Table 11.

Table 11. Parameter Estimates: Improved GRE-based Model for GGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|	VIF
Intercept	b_0	2.88709	9.51	<.0001	
GRE-A	b_1	0.00071	2.74	0.0091	1.073
UGPA	b_2	0.18455	1.96	0.0568	1.624
MGPA	b_3	-0.10904	-2.02	0.0499	1.705
GENDER	b_4	0.13409	1.90	0.0642	1.060
ENLST	b_5	-0.15078	-1.90	0.0652	1.187
(ENLST - 0.128)*(GRE-A - 627)	b_6	-0.00181	-2.06	0.0463	1.076

Improved GRE-based Model for ADJGPA The analogous improved GRE-based model for predicting the ADJGPA was developed in nearly the exact same manner as the Improved GRE-Based Model for GGPA, and the final model contains the same five predictors. The results are slightly better; $r^2 = 0.327$, $\Delta r^2 = 0.245$, and prob > F = 0.0049, with all but one predictor having p-values < 0.05. Investigation of interactions produced the same basic model as for GGPA, but with improved properties: $r^2 = 0.388$, $\Delta r^2 = 0.296$, and prob > F = 0.0022. Here again, the model parameters are provided in Table 12, and the full statistical information is provided in Appendix F.

Table 12. Parameter Estimates: Improved GRE-based Model for ADJGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|
Intercept	b_0	2.705338	8.65	<.0001
GRE-A	b_1	0.000790	2.95	0.0053
UGPA	b_2	0.214206	2.21	0.0330
MGPA	b_3	-0.108802	-1.96	0.0574
GENDER	b_4	0.160015	2.20	0.0334
ENLST	b_5	-0.191453	-2.34	0.0246
(GRE-A - 627)*(ENLST - 0.128)	b_6	-0.001802	-1.99	0.0533

The predictive ability of the Improved GRE-based models was then checked according to the method recommended by Neter and others (1996) and discussed in chapter 3, which compares the Mean Squared Error (MSE) of the training model to the Mean Squared Prediction Error (MSPR). For the training set models, $MSE_{GGPA} = 0.0279$ and $MSE_{ADJGPA} = 0.0296$, which compare at an acceptable level to the validation set's $MSPR_{GGPA} = 0.0425$ and $MSPR_{ADJGPA} = 0.0424$.

Improved GMAT-based Models

Improved GMAT-based Model for GGPA The most effective linear regression model, based on scores from the GMAT included the predictors; GMAT-V, UGPA, and ENLST, and achieved the following results: $r^2 = 0.498$, $\Delta r^2 = 0.454$, and prob > F = < 0.0001. All predictor variables met the $p < 0.05$ requirement for significance. Interaction effects were investigated, but all cross products were eliminated during the stepwise process due to lack of significance. Therefore, the model witout interactions remains the best model. The parameter estimates are shown in Table 13.

Table 13. Parameter Estimates: Improved GMAT-based Model for GGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|
Intercept	b_0	2.29461	8.02	<.0001
GMAT-V	b_1	0.02348	3.64	0.0009
UGPA	b_2	0.21694	3.06	0.0043
ENLST	b_3	-0.28372	-4.40	0.0001

Improved GMAT-based Model for ADJGPA The analogous improved GMAT-based model for predicting the ADJGPA was developed in nearly the exact same manner as the Improved GMAT-Based Model for GGPA, and the final model contains the same three predictors. The properties are as follows; $r^2 = 0.484$, $\Delta r^2 = 0.438$, and prob > F = <0.0001. Here again, the model parameters are provided in Table 14, and the full statistical information is provided in Appendix F.

Table 14. Parameter Estimates: Improved GMAT-based Model for ADJGPA

Term	Parameter	Estimate	t Ratio	Prob>|t|
Intercept	b_0	2.235403	7.44	<.0001
UGPA	b_1	0.223917	3.00	0.0050
ENLST	b_2	0.023977	3.54	0.0002
GMAT-V	b_3	-0.287655	-4.25	0.0012

The predictive ability of the Improved GMAT-based models was then checked according to the method recommended by Neter and others (1996) and discussed in chapter 3, which compares the Mean Squared Error (MSE) of the training model to the Mean Squared Prediction Error (MSPR). For the training set models, $MSE_{GGPA} = 0.0235$ and $MSE_{ADJGPA} = 0.0260$, which compare at an acceptable level to the validation set's $MSPR_{GGPA} = 0.0139$ and $MSPR_{ADJGPA} = 0.0212$.

Summary

This chapter presented the results of the statistical analysis process. Eight regression models were developed, four of which predicted GGPA, and four that predicted ADJGPA. A summary is provided in Table 15. Summary of Model Statistics. The GRE Baseline Model for GGPA and the GRE Baseline Model for ADJGPA established the relationship between the GCA program admissions criteria (GRE-V, GRE-Q, UGPA and MGPA) and GGPA or ADJGPA that provided the most predictive capability, based on data. The GMAT Baseline Model for GGPA and the GMAT Baseline Model for ADJGPA established the relationship between the GCA program admissions criteria (GMAT-T, UGPA and MGPA) and GGPA or ADJGPA that provided the most predictive capability, based on data. The Improved GRE-based Model for GGPA and the Improved GRE-based Model for ADJGPA each found an optimal set of predictors and defined the relationship between the optimal set of predictors and GGPA, or ADJGPA, which provided the most predictive capability. The Improved GMAT-based Model for GGPA and the Improved GMAT-based Model for ADJGPA each found an optimal set of predictors and defined the relationship between the optimal set of predictors and GGPA, or ADJGPA, which provided the most predictive capability. The importance of these models, as well as conclusions drawn from these and other analyses are considered in chapter 5.

Table 15. Summary of Model Statistics

Model	Predicting GGPA r^2	Δr^2	Predicting ADJGPA r^2	Δr^2
GRE Baseline	0.201*	0.130*	0.218*	0.148*
GMAT Baseline	0.224*	0.190*	0.235*	0.201*
Improved GRE-Based	0.350*	0.253*	0.388*	0.296*
Improved GMAT-Based	0.498**	0.454**	0.484**	0.438**

*$p < 0.05$ **$p < 0.0001$

V. Conclusions

Introduction

This chapter presents the conclusion that can be drawn form the results of this research, and how these satisfy the objectives of this research as described in chapter 1. Additional areas of interest are discussed and suggestions for future research are made.

Achievement of Research Objectives

Research objective 1 was to investigate the ability of the current admission criteria to predict student academic performance. The models developed using only the admissions criteria showed low predictive capability, in general accounting for < 20% of the variance in GGPA and ADJGPA. Of the four baseline models developed, the GMAT-T baseline model for ADJGPA explained the largest portion of the variance ($\Delta r^2 = 0.201$), but still left ~80% unexplained.

The GRE-V score had a correlation to GGPA of only -.0318, and -0.0319 to ADJGPA, and it's contribution to the baseline models was so insignificant it was dropped from the models during the stepwise process. This poor predictive ability calls into question the usefulness of GRE-V scores as a criterion on which to base, even partially, an admission decision.

The GRE-Q scores had, by comparison, much more correlation to GGPA and ADJGPA ($r = 0.2053$ and $r = 0.1944$ respectively). The F ratios showed GRE-Q contributes the most predictive ability of the four factors within the GRE baseline models. The addition of $(\text{GRE-Q} - 664)^2$ also improved the Δr^2 by over 100% – from

0.05 to 0.13 – though accounting for only 13% of the variance in predicted performance is of minimal value when attempting to make a selection decision.

The GMAT-T score had the highest correlations to GGPA and ADJGPA of all predictor variables in this study ($r = 0.3937$ and 0.4202, respectively). Yet, even when combined with UGPA, another comparatively strong predictor, this "best" single predictor could only account for ~20% of the variance in GGPA and ADJGPA.

Undergraduate GPA had correlation to GGPA and ADJGPA values of $r = 0.2064$ and $r = 0.2123$, respectively. On it's own, it has low predictive capability, but adds significantly to every model developed in this study, as indicated by the fact that it had either the second or third highest F Ratio in the effects test tables for each model.

Math GPA is of questionable value to the selection process, due to it's low correlation and inconsistent interactions with and influence on other variables. Sixty of the 104 records with an MGPA value were below the minimum required for admission 3.0, so restriction of range is not a factor in the results of this analysis. MGPA's correlations to GGPA and ADJGPA were very low and not significant ($r = 0.0962$ $p = 0.331$; $r = 0.1021$ $p = 0.303$ respectively). Though a significant predictor in some models, MGPA did not always achieve a prob>|t| < 0.05. It is also notable that for GRE models, including both UGPA and MGPA always improved the model, despite MGPA's negative correlation to GGPA and ADJGPA in GRE models. In contrast, for the GMAT models, MGPA was consistently insignificant, and inclusion always decreased Δr^2, most often by > 0.5.

<u>*Research objective 2*</u> was to investigate the ability of additional individual variables (quantitative and qualitative) to predict student academic performance.

It is notable that, except for a very weak negative correlation between GRE-V and GGPA, the only other negative correlations to GGPA are related to the passage of time, e.g, AGE, CMTHS, and TMTHS. The correlation of TMTHS to GGPA and ADJGPA was relatively strong (for this study) $r = -0.2808$ and -0.2960. This goes against the belief of some researchers and admissions decision makers (Peiperl and Trevelyan, 1997; Wooten and McCullough, 1991), that work experience - which some researchers measured by age or actual years of employment (Kuncel, Hezlett and Ones, 2001) – is positively correlated to academic performance in graduate school.

As discussed previously, despite the comparatively strong negative correlation of TMTHS to GGPA and ADJGPA, the dichotomous ENLST variable consistently prevented the inclusion of TMTHS in the developed models. ENLST provided greater increases in Δr^2 than did TMTHS, and inclusion of both variables decreased Δr^2 substantially. The Means ANOVA test found a significant difference in the means of prior enlisted and non-prior enlisted students (Figure 5). The scatterplot comparison (Figure 6) illustrates another difference between the past performance of non-prior enlisted compared to prior-enlisted students in the GCA program. No prior-enlisted student within this study earned a GGPA above 3.8 or below 3.2.

TIME, with a very low and insignificant correlation to GGPA and ADJGPA, was never found to be a significant factor, even in developmental models where other measures of time or work experience (TMTHS, ENLST, AGE) were significant or at least nearly significant.

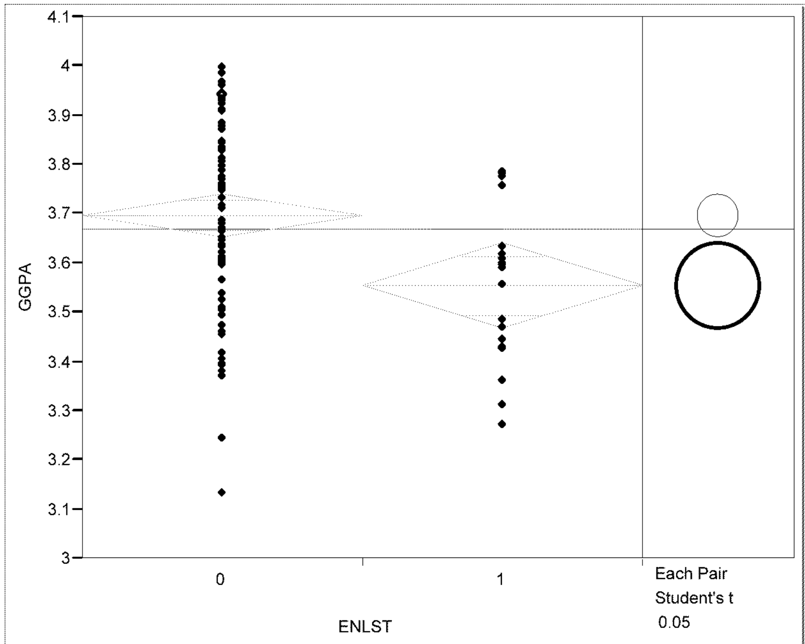


Figure 5. Oneway Analysis of GGPA by ENLST

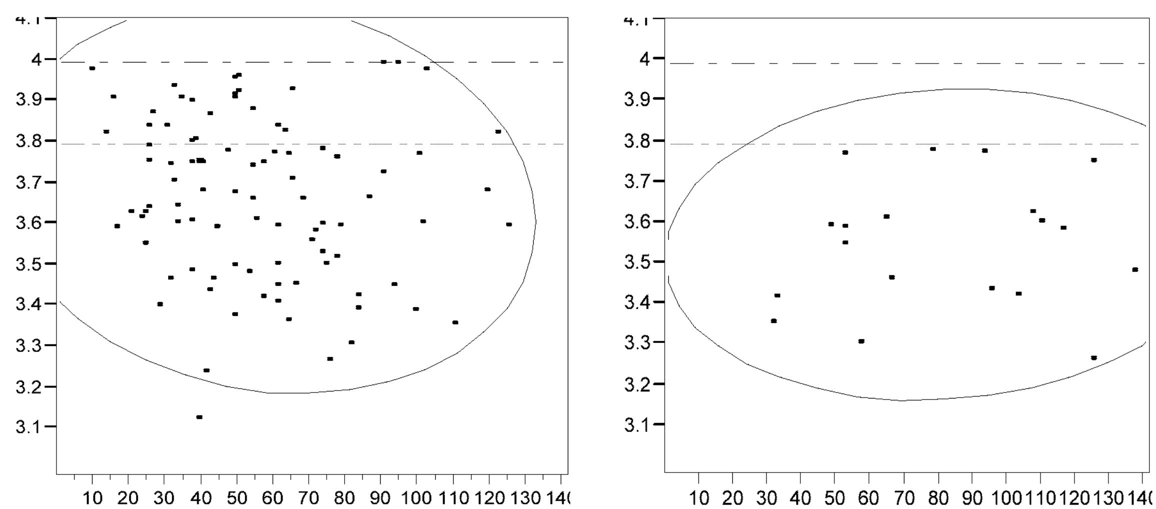


Figure 6. Scatterplots of GGPA by CMTHS (left) and EMTHS (right)

Ratings of the admissions competitiveness of undergraduate schools (RATE) did not provide the expected increase in predictive capability. Neither RATE or RATGPA proved to be useful predictors in this study. This disagrees with the results of Spangler (1989), who found the combination of RATE*UGPA to provide a substantial improvement in correlation to GGPA. As was shown in chapter 4, there was no statistically significant difference in the mean GGPA or ADJGPA between the 6 categories of RATE. In fact the slight pattern that did emerge, did not follow the expected relation that the more competitive the undergraduate school, higher the average levels of graduate academic performance. RATE and RATGPA were not found to be significant contributors to the predictive capability of any of the models developed in this study

Though not included in any regression models, previous graduate level academic performance, as measured by PrGGPA, showed surprisingly little predictive validity for performance in the GCA program. When regressed against GGPA and ADJGPA, the resulting models were insignificant and had R^2 values below 0.02. The average GGPA for individuals with prior graduate work was 3.64; 0.04 below the average for those without prior graduate work. The scatterplot of GGPA vs. PrGGPA is provided in Figure 7. Despite the seemingly obvious value of PrGGPA as a predictor, this study does not support it's use when making a selection decision.

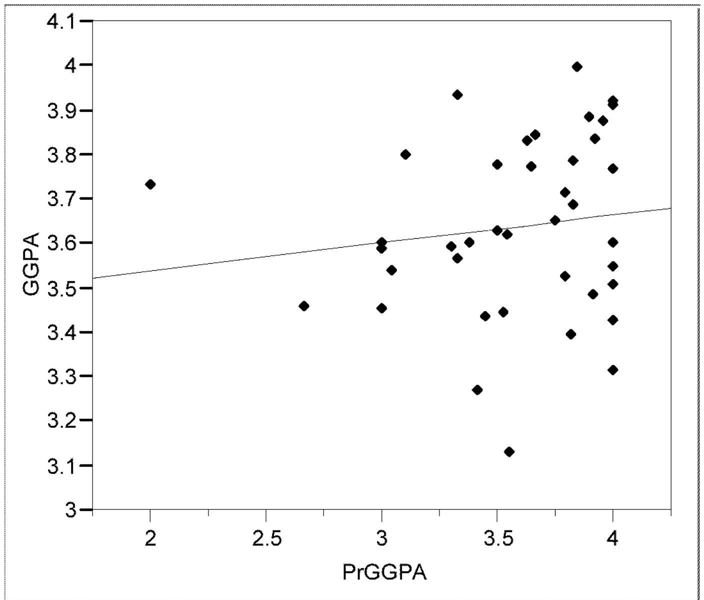


Figure 7. Scatterplot of GGPA by PrGGPA

AGE was not found to be a significant predictor of graduate academic performance, as measured by GGPA and ADJGPA. It's correlation to standardized test scores ranged from 0.3730 for GRE-V to -0.2636 for GRE-Q, and it provided no useable interaction effects with the other variables. The literature discussing age related bias of standardized test scores was not conclusive, so this result is not unexpected.

Degree type, as denoted by DEGREE, did not provide any predictive capability. The categories and their mean GGPA were: #1 (BS degrees), 3.657; #2 (BA degrees), 3.673, and #3 (BBA or BSBA degrees), 3.679. A studentized t test comparing the differences in the means indicated these means are not significantly different.

<u>Research objective 3</u> was to select an 'optimal' set of eligibility/selection indicators having the potential of predicting student academic success. The obvious options for the optimal set of criteria are the four 'Improved' regression models described in chapter 4. The GRE-based model accounted for up to 30% of the variation in GGPA

and ADJGPA, but the GMAT-based model consistently outperformed them. The GMAT Baseline models included GMAT-T where as the Improved GMAT-based Models included GMAT-V instead.

The optimal set of predictors of graduate academic performance, as measured by GGPA (or ADJGPA), in the GCA program is the Improved GMAT-based Model for GGPA as shown in equation (7).

$$GGPA = b_0 + b_1{*}GMAT\text{-}V + b_2{*}UGPA + b_3{*}ENLST \tag{7}$$

where: $b_0 = 2.2946112$
$b_1 = 0.0234784$
$b_2 = 0.2169419$
$b_3 = -0.283723$

Both GMAT and UGPA can be said to have content validity, and this analysis established at least moderate predictive validity, when used in combination with other predictors. The importance of ENLST to the predictive capability of the model is apparent from the development process, i.e., it's predictive validity has been established. However, it's content validity is not fully understood by this researcher. Many reasons can be surmised, but none seem better than another, and all would require extensive research to verify.

This equation (7) may be used to predict an applicant's expected academic performance, as measured by GGPA. The result of this equation is not meant to be used as either a cut-off score or a single criteria on which to base admission. It is meant to provide a more accurate assessment of an applicant's expected academic performance, compared to the use of the current admission criteria.

Discussion

The prediction of academic performance in a program of graduate education is not an easy task. The sources of variation are too numerous to list and nearly impossible to quantify. However, most potential variation must be accepted as part of life, and thus cannot be predicted. This study attempted to account for the variance that is predictable.

That said, it is important to reiterate that 100% of the students in this study graduated on time with a GGPA above the minimum acceptable 3.0. Despite lack of validity in the current admission criteria, the current selection process does work, though the reasons for AFIT's high success rate are most likely not due to the selection process.

One measure of the validity of the admission criteria is to check if the ability to meet those criteria indicates better performance. By classifying all records with the dichotomous variable 1 = met minimum admissions criteria ($n_1 = 26$), 0 = did not meet minimum admission criteria ($n_0 = 80$), a significant difference in the mean GGPA and ADJGPA was found. Though both were respectable GGPAs, the difference indicates that meeting all criteria does predict better performance. However, since not meeting the minimum also predicts acceptable performance, the usefulness of the cut-off scores is questionable. In addition, this study found that 70% of the lowest GGPAs were earned by students who did not meet all criteria. This may not be surprising, but a more illustrative fact is that, of the highest 10 GGPAs, 80% were earned by students who did not meet all admission criteria.

The comparison of the criterion variables GGPA and ADJGPA indicated that ADJGPA may be a better measure of academic performance. In most models where ADJGPA was the criterion variable, the predictive validity was greater than for the comparable GGPA model. Though, these difference were small, the trend was fairly consistent. Unfortunately, the differences are not great enough to make a clear choice.

Suggestions for Future Research

As was mentioned earlier, AFIT underwent a large internal reorganization that took effect in October 1999. As part of the reorganization, the GCA program is now part of the School of Engineering and Management. The program is placing increased emphasis on statistical and other applied quantitative analysis techniques. (Stockman, 2002) This shift has already affected classes 02M and 03M, and will likely continue in future classes. It is suggested that a new study, similar to this one, be performed as soon as enough students have completed the revised curriculum to allow for the results to be considered significant. This study was obviously limited to data on past students, most of whom completed the program when it was only 15 months long and under different management.

Due to the ability of neural networks to include complex relationships and to learn over time, it is suggested that a future study include analysis using neural networks. Once completed, if the model is successful, it may be possible to implement it into the admissions process and allow it to learn over time, and improve and adapt as the program undergoes future shifts of purpose, emphasis and leadership.

Uses of different measure for academic success are another area for future research. The GPA from the first two quarters of instruction may be of more interest to the eligibility decision makers, than overall GGPA. Under the current curriculum, the program has a heavier coursework load in the first two quarters, than in the remaining four quarters. The initial series of quantitative courses is when the students may encounter the most new material, and the differences in pre-admission experience and education may have the largest effect, since later courses usually build on the initial courses.

The reasons ENLST is an important part of the optimal regression equation, i.e., the content validity of ENLST is not fully understood. Investigation into whether prior enlisted members can be truly be expected to perform at lower levels than non-prior's, would be valuable. Given the corroborating results of Van Scotter (1983) and Sny (1991), it seems likely the relationship is valid. The interesting aspect is why. Is it related to attitude, or proximity to retirement? Maybe prior enlisted are, on average, older and their typical family situation creates less time for school-work. The potential reasons are too numerous to discuss further, but discovering them may provide a significant boost to the ability to predict their performance at AFIT.

Summary

The objectives of this research effort were accomplished. The objective criteria currently used by AFIT to make academic eligibility decisions provide little predictive validity, accounting for at most only 20% of the variation in GGPA.

The values of other predictors were examined. Most predictor variables that were measures of time, or related to time and or work experience – e.g., TMTHS and AGE – were negatively correlated to GGPA. The dichotomous variable ENLST, was the most consistent non-academic predictor variable.

The GMAT is more useful than the GRE as a predictor of academic performance in the GCA program. UGPA is also a dependable, though not particularly strong, predictor. The optimal model includes the GMAT-V score, UGPA and ENLST, and accounted for up to 45% of the variance in GGPA. Though this is better than the current criteria, the optimal model is intended to provide improved insight into an applicants expected performance. Other criteria should be included in the overall selection process. It is hoped that the discussion in this thesis of some of these other variables shed some light on their actual usefulness.

Appendix A.

Criteria for Admissions Competitiveness Ratings

Criteria for Ratings of Admissions Competitiveness of Undergraduate Schools

Category	High School Rank	High School GPA	Median SAT I Score	Median ACT Score	Selection Ratio	Rating
Most Competitive	Top 10-20%	A to B+	655-800	29+	< 1/3	6
Highly Competitive	Upper 20-35%	B+ to B	620-654	27-28	< 1/3 to 1/2	5
Very Competitive	Upper 35-50%	No less than B-	573-619	24-26	50 – 75%	4
Competitive	Upper 50-65%	C to B-	500-572	21-23	75 - 85%	3
Less Competitive	Upper 65%	< C	< 500	< 21	85%	2
Non-Competitive	Requires H.S. Diploma or equivalent SAT or ACT not required				98%	1
Special	Schools with specialized programs, e.g. professional music or art degree, and schools oriented toward working adults.				varies	varies

Source: (Profiles of American Colleges 2001, 2000)

Appendix B. Admissions Competitiveness Ratings, by School

Admissions Competitiveness Ratings of Undergraduate Degree-granting Schools

School Rating	Undergraduate School
4	Alfred University, Alfred NY
3	Austin Peay State University
2	Boise State University, Boise ID
5	Boston University, Boston MA
5	Brigham Young University, Provo UT
3	California State University, Sacramento CA
6	Carnegie Mellon University
3	Central Michigan University, Mt Pleasant MI
3	Central Washington University, Ellensburg WA
3	Chicago State University, Chicago IL
3	East Central University, Ada OK
3	Embry-Riddle Aero University, Daytona Beach FL
4	Fordham University, New York NY
1	Fort Hays State College, Hays KS
5	George Washington University
6	Georgetown University, Washington DC
4	Gonzaga University, Spokane WA
4	Hendrix College, Conway AR
4	Illinois Institute of Technology, Chicago IL
3	Illinois State University, Normal IL
NR*	Inter American University of Puerto Rico, San Juan PR
4	Iowa State University of Science and Technology, Ames IA
NR*	Korea Military Academy
2	Lubbock Christian College, Lubbock TX
4	Miami University, Oxford OH
4	Michigan State University
4	Mississippi State University
2	Missouri Southern State College, Joplin MO
3	North Carolina A&T State University, Greensboro NC
2	Northland College, Ashland WI
6	Northwestern University,
2	Norwich University, Northfield VT
4	Penn State University
3	Portland State University, Portland OR
3	Purdue University

School Rating	Undergraduate School
5	Rensselaer Polytechnic Institute, Troy NY
3	San Diego State University, San Diego CA
5	State University College, Geneseo NY
3	Stephen F. Austin State University, Nacodoches, TX
1	The University of Akron, Akron OH
NR*	Toledo University, Toledo OH
3	Troy State, Troy AL
6	United States Air Force Academy
4	University of Alabama
3	University of Arkansas
5	University of California, Irvine CA
3	University of Cincinnati, Cincinnati OH
4	University of Delaware, Newark DE
1**	University of Maryland - University College, College Park MD
5	University of Miami, Coral Gables FL
3	University of Minnesota
4	University of Minnesota, Twin Cities campus
4	University of Mississippi
3	University Of North Carolina at Charlotte, Charlotte NC
6	University of Notre Dame
4	University of Oklahoma, Norman OK
4	University of Portland, Portland OR
5	University of Richmond, Richmond VA
3	University of South Alabama, Mobile AL
4	University of South Carolina, Columbia SC
3	University of Tennessee, Knoxville
5	University of Texas at Austin, Austin TX
1	University of Texas at San Antonio, San Antonio TX
3	University of Vermont, Burlington VT
4	Virginia Polytechnic Institute and State University, Blacksburg VA
1	Wayland Baptist University
3	West Virginia University, Morgantown WV
2	Wright State University, Dayton OH

*NR = Not Rated

** Rating for the University of Maryland - University College, College Park MD was 'Special' due to its stated orientation toward working adults. The written review indicated all applicants that meet basic requirements are accepted, so this author assigned a rank of 1.

Appendix C. Correlation Matrix

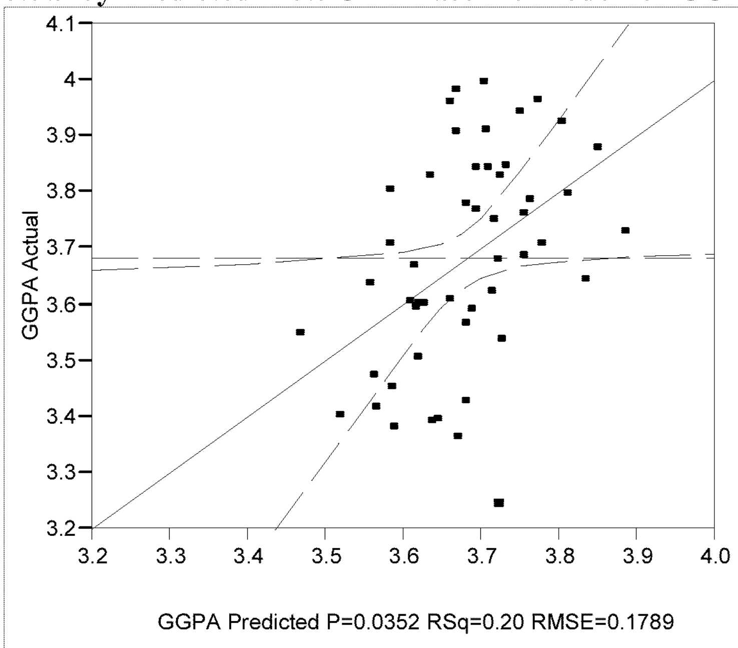
	GGPA	ADJGPA	GRE-V	GRE-Q	GRE-A	GRE-T	GRE-VQ	GMAT-V	GMAT-Q	GMAT-T	UGPA	MGPA	RATGPA	AGE	TIME	EMTHS	CMTHS	TMTHS	GENDER	RATE	DEGREE
GGPA	1.000																				
ADJGPA	0.986**	1.000																			
GRE-V	-0.032	-0.032	1.000																		
GRE-Q	0.205	0.194	0.133	1.000																	
GRE-A	0.176	0.175	0.261*	0.557**	1.000																
GRE-T	0.161	0.156	0.604**	0.736**	0.866**	1.000															
GRE-VQ	0.108	0.101	0.777**	0.727**	0.534**	0.886**	1.000														
GMAT-V	0.373*	0.375*	-0.010	0.283	-0.022	0.124	0.210	1.000													
GMAT-Q	0.281*	0.300*	-0.242	0.687*	0.454	0.476	0.381	0.296*	1.000												
GMAT-T	0.394*	0.420*	-0.154	0.565*	0.270	0.356	0.340	0.758**	0.841**	1.000											
UGPA	0.206*	0.212*	0.050	-0.124	-0.111	-0.087	-0.044	-0.044	-0.127	-0.058	1.000										
MGPA	0.096	0.102	-0.037	0.024	-0.062	-0.040	-0.010	-0.040	0.127	0.091	0.368**	1.000									
RATGPA	0.169	0.148	0.075	-0.032	0.147	0.094	0.031	0.292*	0.204	0.2692*	-0.062	-0.084	1.000								
AGE	-0.137	-0.137	0.373*	-0.043	-0.264*	-0.008	0.231	-0.104	-0.182	0.046	-0.006	0.063	-0.303*	1.000							
TIME	0.023	0.019	0.427*	-0.053	-0.295*	-0.007	0.262	0.114	0.047	-0.191	-0.128	-0.001	-0.138	0.869**	1.000						
EMTHS	0.085	0.133	-0.103	0.570	0.519	0.463	0.362	-0.421	0.189	0.086	0.195	-0.089	-0.287	0.000	-0.327	1.000					
CMTHS	-0.123	-0.139	0.345*	-0.092	-0.236	-0.022	0.167	-0.172	-0.014	-0.124	-0.223	-0.123	0.136	0.550**	0.716**	-0.312	1.000				
TMTHS	-0.281*	-0.296*	0.232	-0.009	-0.135	0.018	0.147	-0.094	-0.110	-0.119	0.137	0.100	-0.117	0.790**	0.490**	0.742*	0.630**	1.000			
GENDER	0.050	0.077	0.049	0.023	-0.083	-0.017	0.049	-0.096	-0.007	0.114	0.050	-0.081	-0.027	0.067	0.012	0.356	0.064	0.171	1.000		
RATE	0.105	0.079	0.031	-0.005	0.148	0.087	0.018	0.303	0.281	0.313	-0.354	-0.270	0.947	-0.299	-0.102	-0.316	0.194	-0.144	0.009	1.000	
DEGREE	0.049	0.073	0.076	-0.268	-0.135	-0.144	-0.117	0.047	-0.111	-0.046	0.200	0.151	-0.363	0.105	0.156	-0.605	-0.095	-0.007	0.076	-0.421	1.000

Appendix C. Correlation Matrix

Appendix D. Statistics for Baseline Models

GRE Baseline Model for GGPA

Actual by Predicted Plot: GRE Baseline Model for GGPA



Summary of Fit: GRE Baseline Model for GGPA

R^2	0.201277
Adj R^2	0.13028
Root MSE	0.178884
Observations	50

Analysis of Variance: GRE Baseline Model for GGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.3628712	0.090718	2.8350
Error	45	1.4399708	0.031999	Prob > F
C. Total	49	1.8028421		0.0352

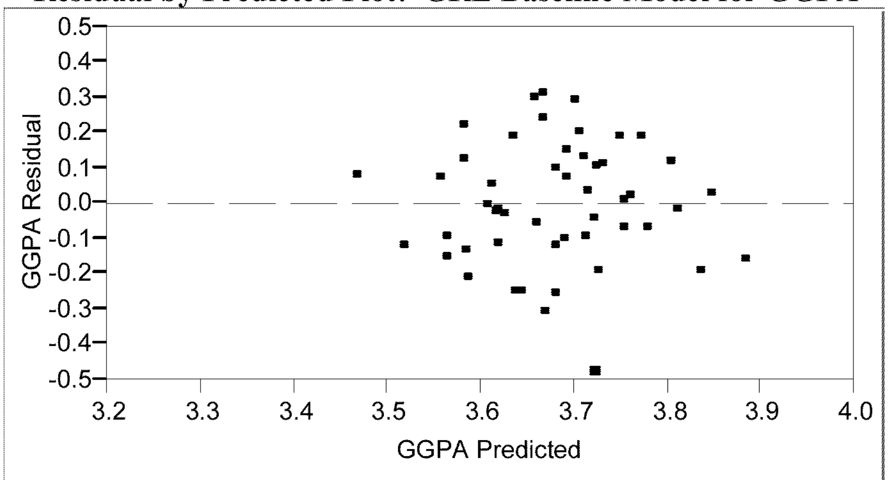
Parameter Estimates: GRE Baseline Model for GGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	2.603776	0.411449	6.33	<.0001	
GRE-Q	0.000986	0.000403	2.45	0.0184	1.1799
UGPA	0.236091	0.101597	2.32	0.0247	1.6601
MGPA	-0.115519	0.052209	-2.21	0.0320	1.5664
$(\text{GRE-Q} - 664)^2$	0.000011	0.000005	2.07	0.0441	1.1629

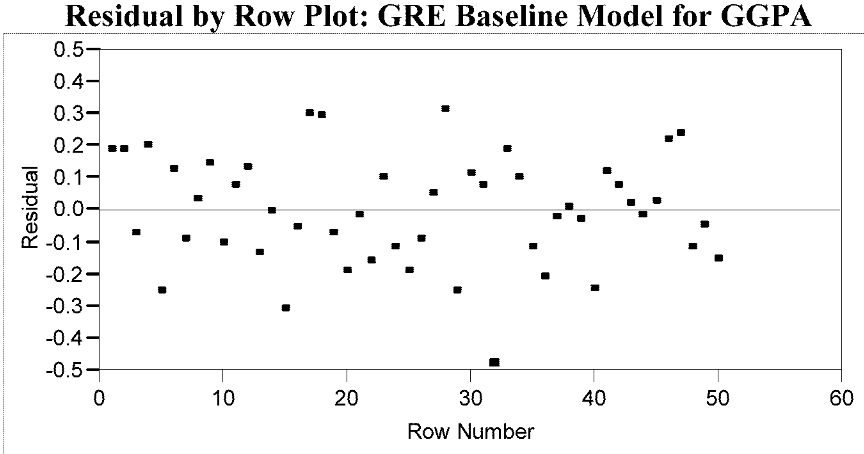
Effect Tests: GRE Baseline Model for GGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GRE-Q	1	1	0.1913792	5.9807	0.0184
UGPA	1	1	0.1727975	5.4000	0.0247
MGPA	1	1	0.1566617	4.8958	0.0320
$(GRE\text{-}Q)^2$	1	1	0.1372377	4.2888	0.0441

Residual by Predicted Plot: GRE Baseline Model for GGPA

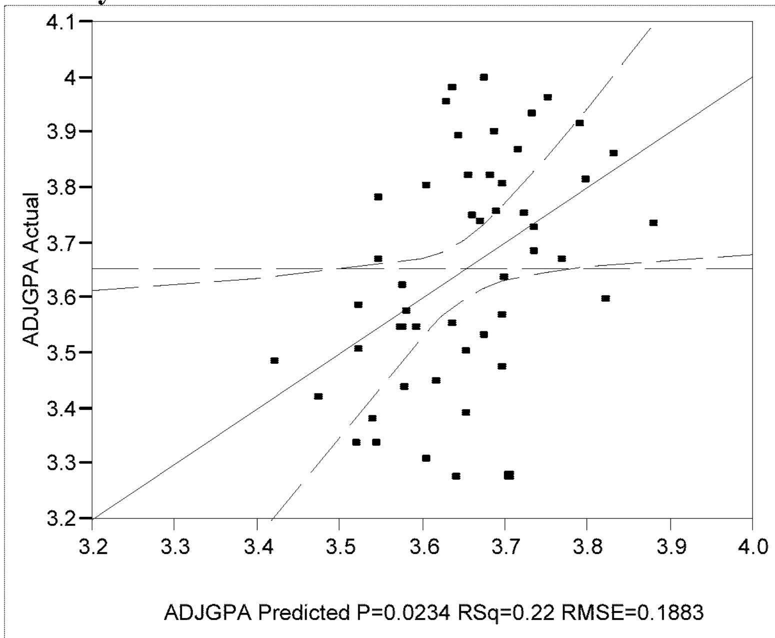


Residual by Row Plot: GRE Baseline Model for GGPA



GRE Baseline Model for ADJGPA

Actual by Predicted Plot: GRE Baseline Model for ADJGPA



Summary of Fit: GRE Baseline Model for ADJGPA

R^2	0.21787
Adj R^2	0.148348
Root MSE	0.188334
Observations	50

Analysis of Variance: GRE Baseline Model for ADJGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	0.4446226	0.111156	3.1338
Error	45	1.5961446	0.035470	Prob > F
C. Total	49	2.0407671		0.0234

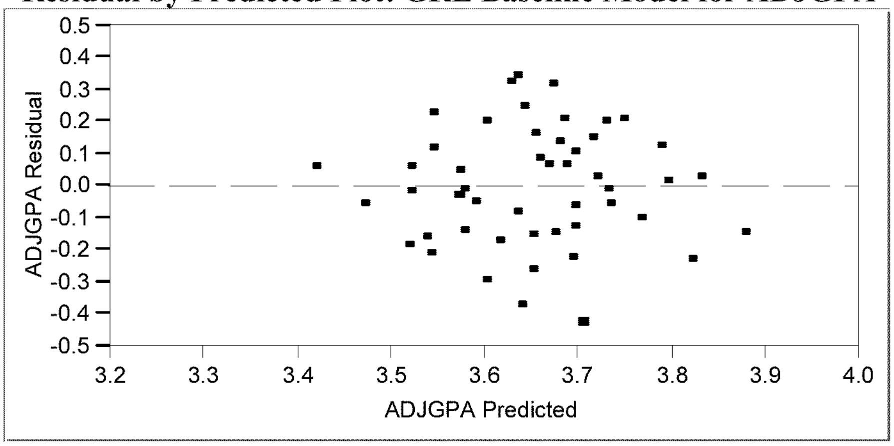
Parameter Estimates: GRE Baseline Model for ADJGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.409860	0.433187	5.56	<.0001
GRE-Q	0.001112	0.000424	2.62	0.0119
UGPA	0.272266	0.106965	2.55	0.0144
MGPA	-0.125661	0.054967	-2.29	0.0270
$(GRE\text{-}Q\text{-}664.2)^2$	0.000011	0.000005	2.10	0.0414

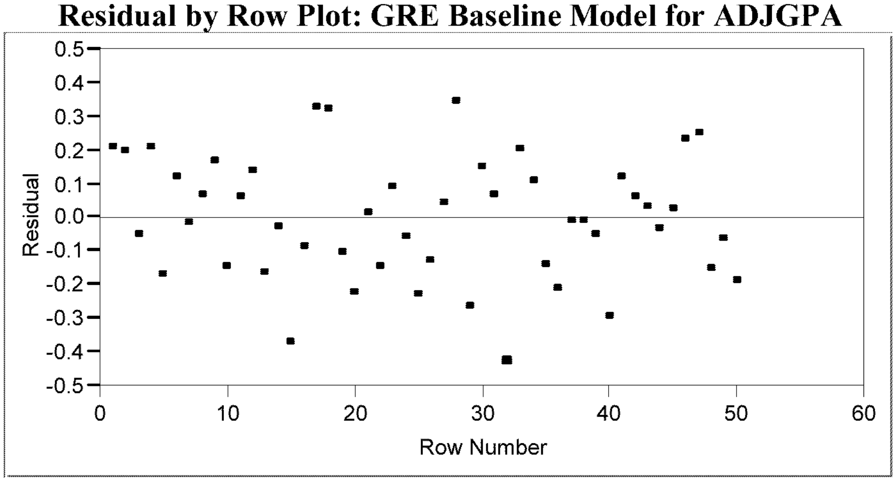
Effect Tests: GRE Baseline Model for ADJGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GRE-Q	1	1	0.24353648	6.8660	0.0119
UGPA	1	1	0.22980748	6.4789	0.0144
MGPA	1	1	0.18537654	5.2263	0.0270
GRE-Q*GRE-Q	1	1	0.15629996	4.4066	0.0414

Residual by Predicted Plot: GRE Baseline Model for ADJGPA

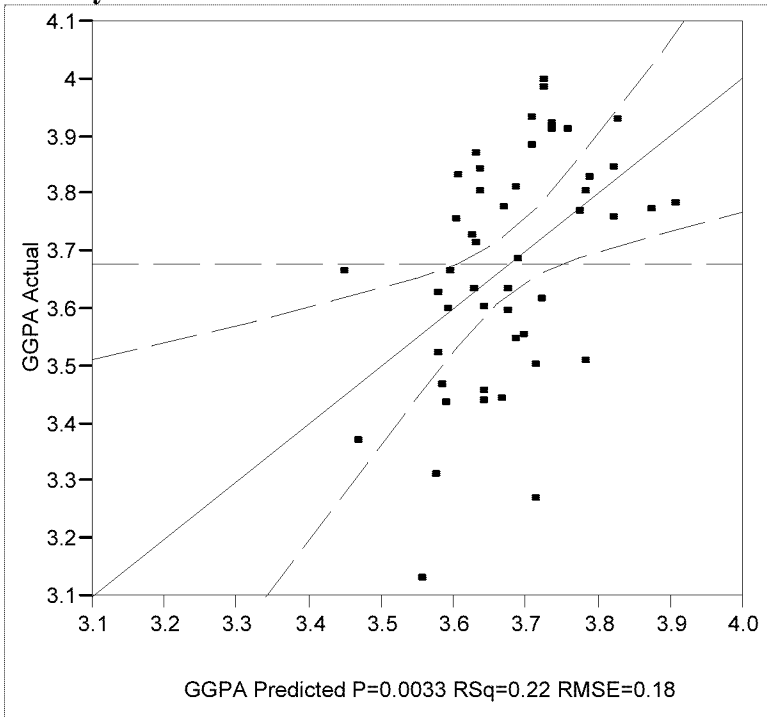


Residual by Row Plot: GRE Baseline Model for ADJGPA



GMAT Baseline Model for GGPA

Actual by Predicted Plot: GMAT Baseline Model for GGPA



Summary of Fit: GMAT Baseline Model for GGPA

R^2	0.224099
Adj R^2	0.189615
Root MSE	0.18
Observations	48

Analysis of Variance: GMAT Baseline Model for GGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.4211053	0.210553	6.4986
Error	45	1.4579962	0.032400	Prob > F
C. Total	47	1.8791015		0.0033

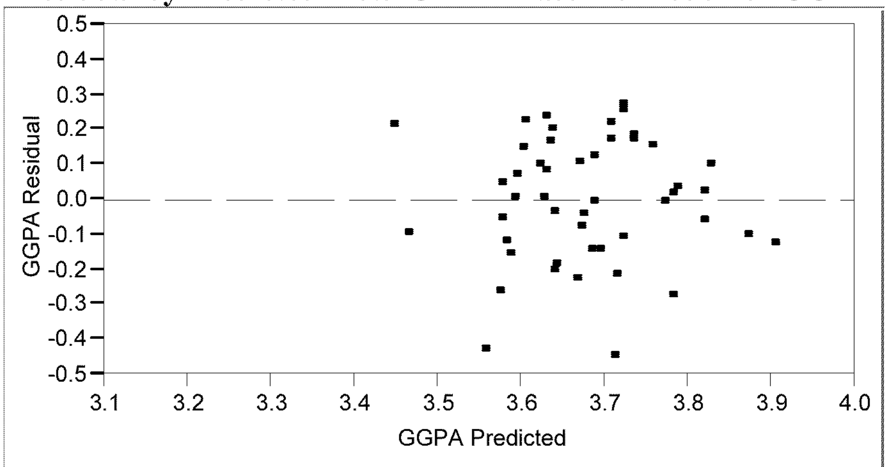
Parameter Estimates: GMAT Baseline Model for GGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.3750777	0.362664	6.55	<.0001
GMAT-T	0.0015299	0.000508	3.01	0.0042
UGPA	0.1412342	0.066297	2.13	0.0386

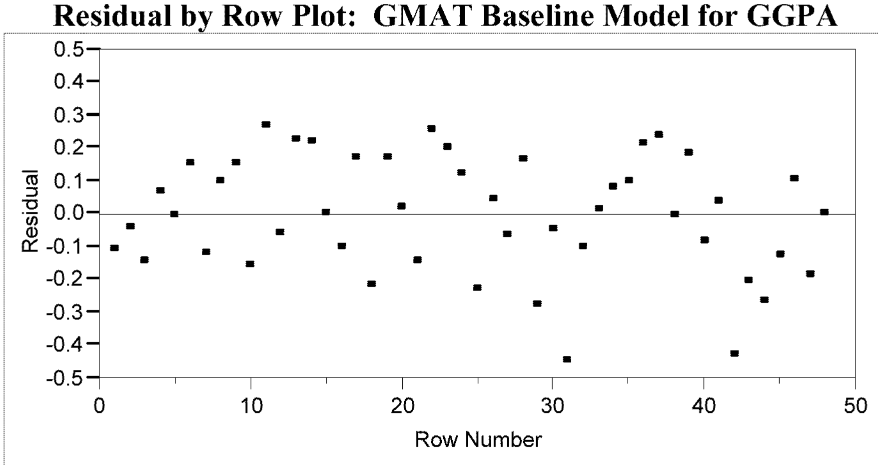
Effect Tests: GMAT Baseline Model for GGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GMAT-T	1	1	0.29409123	9.0769	0.0042
UGPA	1	1	0.14703925	4.5383	0.0386

Residual by Predicted Plot: GMAT Baseline Model for GGPA

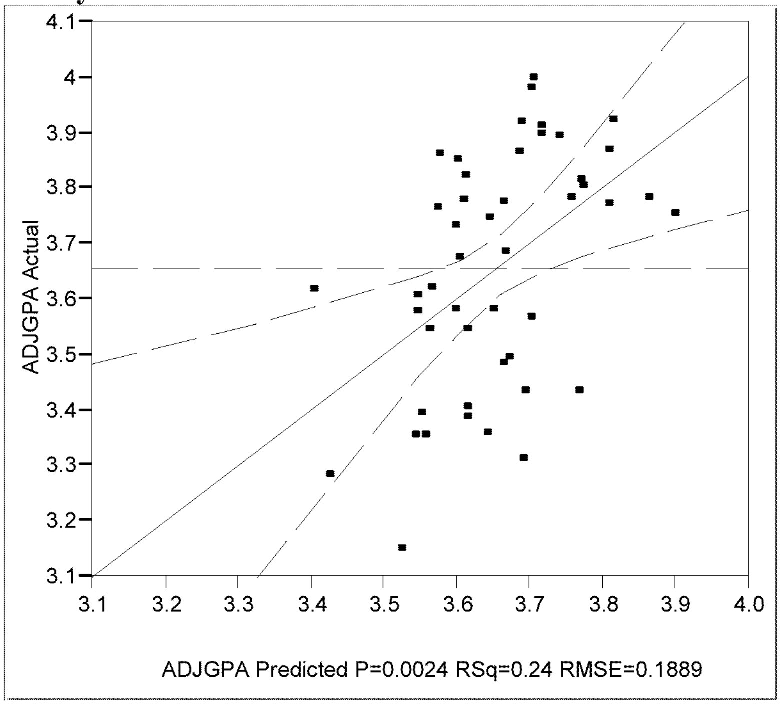


Residual by Row Plot: GMAT Baseline Model for GGPA



GMAT Baseline Model for ADJGPA

Actual by Predicted Plot: GMAT Baseline Model for ADJGPA



Summary of Fit: GMAT Baseline Model for ADJGPA

R^2	0.235371
Adj R^2	0.201388
Root MSE	0.188939
Observations	48

Analysis of Variance: GMAT Baseline Model for ADJGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	2	0.4944888	0.247244	6.9260
Error	45	1.6064005	0.035698	Prob > F
C. Total	47	2.1008893		0.0024

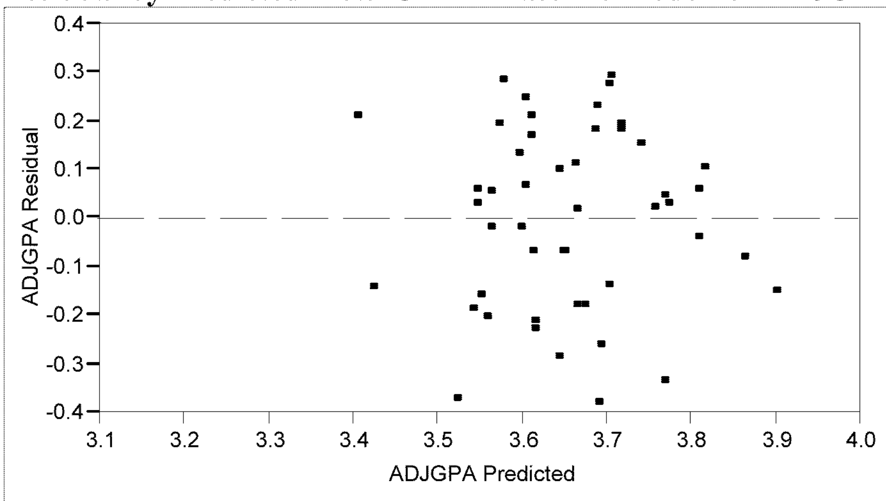
Parameter Estimates: GMAT Baseline Model for ADJGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.24266	0.38067	5.89	<.0001
GMAT-T	0.00166	0.00053	3.12	0.0031
UGPA	0.15193	0.06959	2.18	0.0343

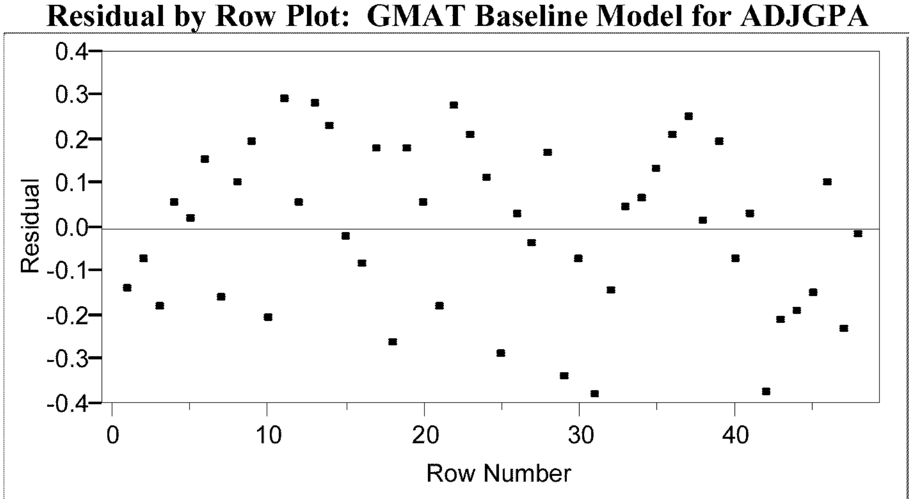
Effect Tests: GMAT Baseline Model for ADJGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GMAT-T	1	1	0.3477482	9.7414	0.0031
UGPA	1	1	0.1701608	4.7667	0.0343

Residual by Predicted Plot: GMAT Baseline Model for ADJGPA

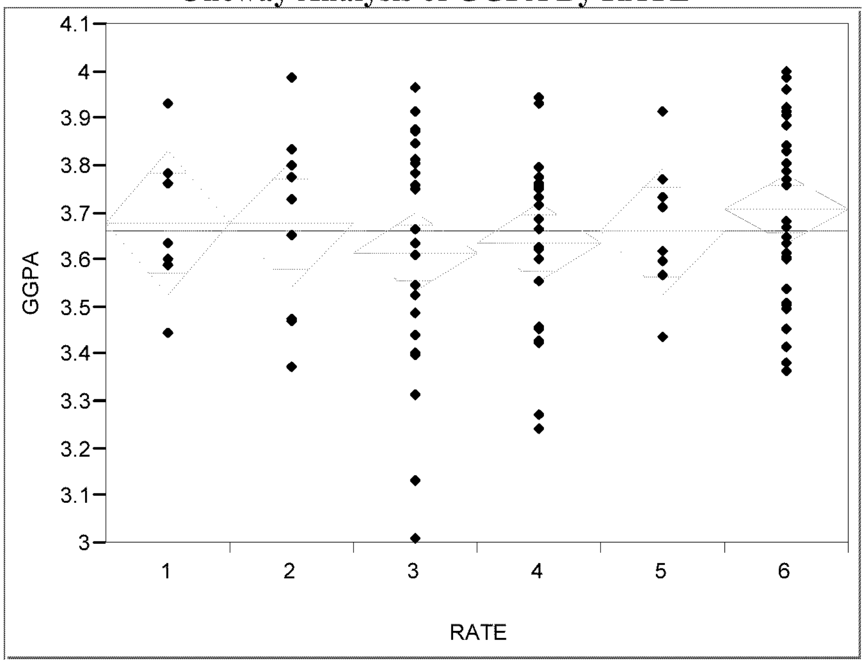


Residual by Row Plot: GMAT Baseline Model for ADJGPA



Appendix E. Selected Scatterplots and Oneway Analysis Plots

Oneway Analysis of GGPA By RATE



Summary of Fit: Oneway Analysis of GGPA By RATE

R^2	0.033138
Adj R^2	-0.0167
Root MSE	0.201602
Observations	103

Analysis of Variance: Oneway Analysis of GGPA By RATE

Source	DF	Sum of Squares	Mean Square	F Ratio
RATE	5	0.1351230	0.027025	0.6649
Error	97	3.9424258	0.040644	Prob > F
C. Total	102	4.0775488		0.6509

Means for Oneway Anova Oneway Analysis of GGPA By RATE

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	7	3.67900	0.07620	3.5278	3.8302
2	9	3.67800	0.06720	3.5446	3.8114
3	23	3.61713	0.04204	3.5337	3.7006
4	23	3.63843	0.04204	3.5550	3.7219
5	9	3.66122	0.06720	3.5278	3.7946
6	32	3.70947	0.03564	3.6387	3.7802

Std Error uses a pooled estimate of error variance

Means Comparisons: Oneway Analysis of GGPA By RATE

Dif = Mean[i]- Mean[j]	6	1	2	5	4	3
6	0.00000	0.03047	0.03147	0.04825	0.07103	0.09234
1	-0.03047	0.00000	0.00100	0.01778	0.04057	0.06187
2	-0.03147	-0.00100	0.00000	0.01678	0.03957	0.06087
5	-0.04825	-0.01778	-0.01678	0.00000	0.02279	0.04409
4	-0.07103	-0.04057	-0.03957	-0.02279	0.00000	0.02130
3	-0.09234	-0.06187	-0.06087	-0.04409	-0.02130	0.00000

Alpha = 0.05

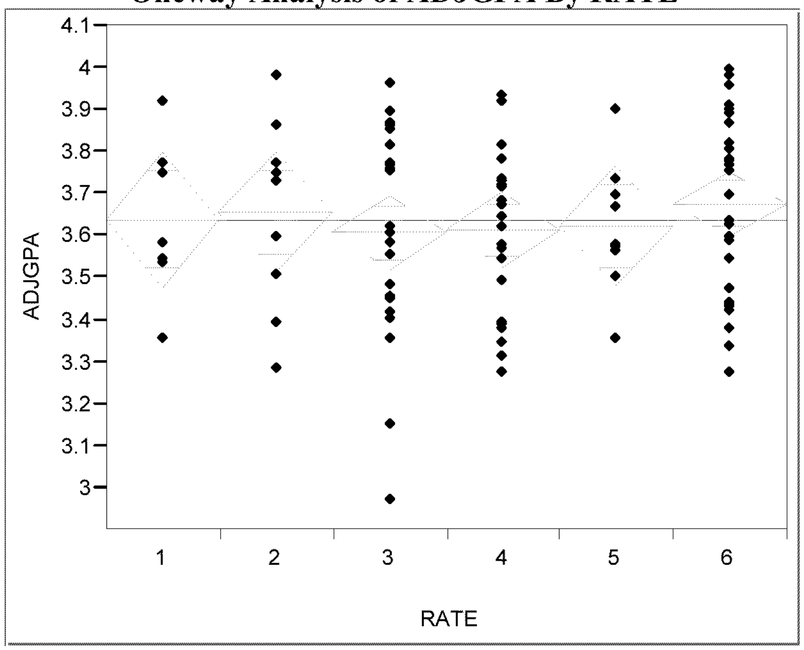
Comparisons for each pair using Student's t

t = 1.98472

Abs(Dif)-LSD	6	1	2	5	4	3
6	-0.10003	-0.13649	-0.11950	-0.10272	-0.03835	-0.01704
1	-0.13649	-0.21388	-0.20064	-0.18387	-0.13216	-0.11085
2	-0.11950	-0.20064	-0.18862	-0.17184	-0.11776	-0.09645
5	-0.10272	-0.18387	-0.17184	-0.18862	-0.13453	-0.11323
4	-0.03835	-0.13216	-0.11776	-0.13453	-0.11799	-0.09669
3	-0.01704	-0.11085	-0.09645	-0.11323	-0.09669	-0.11799

Positive values show pairs of means that are significantly different.

Oneway Analysis of ADJGPA By RATE



Summary of Fit: Oneway Analysis of ADJGPA By RATE

R^2	0.019485
Adj R^2	-0.03106
Root MSE	0.215108
Mean of Response	3.638223
Observations	103

Analysis of Variance: Oneway Analysis of ADJGPA By RATE

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
RATE	5	0.0891948	0.017839	0.3855	0.8577
Error	97	4.4883451	0.046272		
C. Total	102	4.5775399			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	7	3.63971	0.08130	3.4783	3.8011
2	9	3.65544	0.07170	3.5131	3.7978
3	23	3.60726	0.04485	3.5182	3.6963
4	23	3.61400	0.04485	3.5250	3.7030
5	9	3.62222	0.07170	3.4799	3.7645
6	32	3.67722	0.03803	3.6017	3.7527

Std Error uses a pooled estimate of error variance

Means Comparisons: Oneway Analysis of ADJGPA By RATE

Dif= Mean[i]-Mean[j]	6	2	1	5	4	3
6	0.00000	0.02177	0.03750	0.05500	0.06322	0.06996
2	-0.02177	0.00000	0.01573	0.03322	0.04144	0.04818
1	-0.03750	-0.01573	0.00000	0.01749	0.02571	0.03245
5	-0.05500	-0.03322	-0.01749	0.00000	0.00822	0.01496
4	-0.06322	-0.04144	-0.02571	-0.00822	0.00000	0.00674
3	-0.06996	-0.04818	-0.03245	-0.01496	-0.00674	0.00000

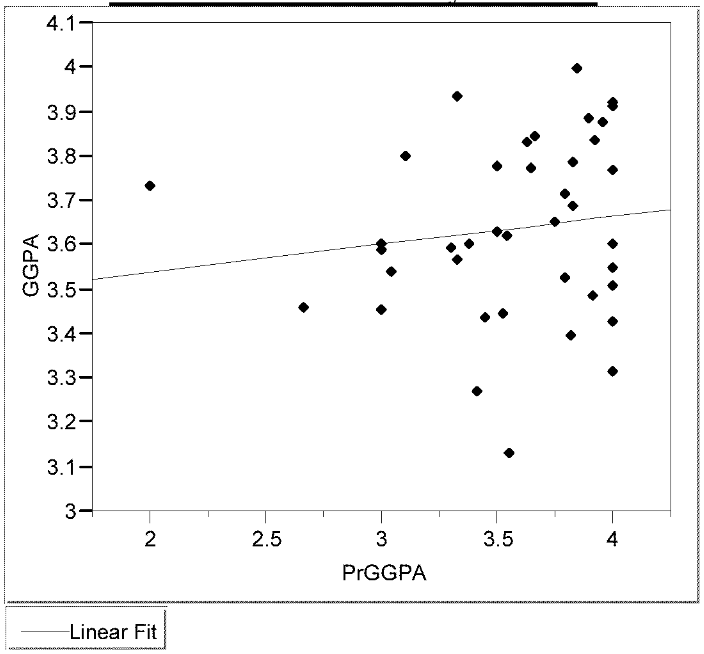
Alpha = 0.05

Comparisons for each pair using Student's t, t = 1.98472

Abs(Dif)-LSD	6	2	1	5	4	3
6	-0.10673	-0.13931	-0.14064	-0.10609	-0.05349	-0.04675
2	-0.13931	-0.20126	-0.19942	-0.16803	-0.12642	-0.11968
1	-0.14064	-0.19942	-0.22820	-0.19766	-0.15858	-0.15184
5	-0.10609	-0.16803	-0.19766	-0.20126	-0.15964	-0.15290
4	-0.05349	-0.12642	-0.15858	-0.15964	-0.12589	-0.11916
3	-0.04675	-0.11968	-0.15184	-0.15290	-0.11916	-0.12589

Positive values show pairs of means that are significantly different.

Bivariate Fit of GGPA By PrGGPA



Linear Fit GGPA = 3.4131435 + 0.0629886 PrGGPA

Summary of Fit: Bivariate Fit of GGPA By PrGGPA

R^2	0.01889
Adj R^2	-0.00692
Root MSE	0.20063
Mean of Response	3.63838
Observations	40

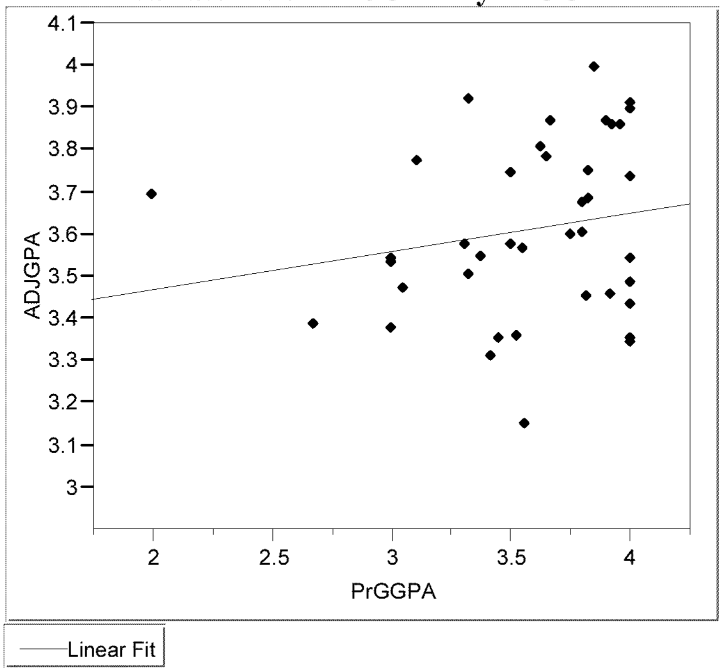
Analysis of Variance: Bivariate Fit of GGPA By PrGGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.0294575	0.029457	0.7318
Error	38	1.5295759	0.040252	Prob > F
C. Total	39	1.5590334		0.3977

Parameter Estimates: Bivariate Fit of GGPA By PrGGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.4131435	0.265188	12.87	<.0001
PrGGPA	0.0629886	0.073631	0.86	0.3977

Bivariate Fit of ADJGPA By PrGGPA



Linear Fit

ADJGPA = 3.290814 + 0.0896766 PrGGPA

Summary of Fit: Bivariate Fit of ADJGPA By PrGGPA

R^2	0.03595
Adj R^2	0.01058
Root MSE	0.205269
Mean of Response	3.611475
Observations	40

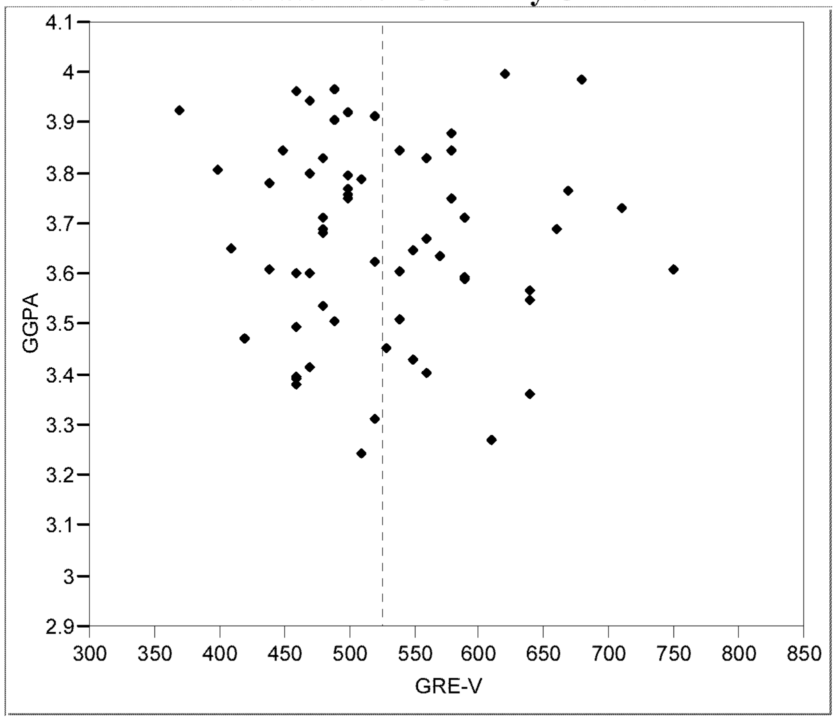
Analysis of Variance: Bivariate Fit of ADJGPA By PrGGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	0.0597076	0.059708	1.4170
Error	38	1.6011403	0.042135	Prob > F
C. Total	39	1.6608480		0.2413

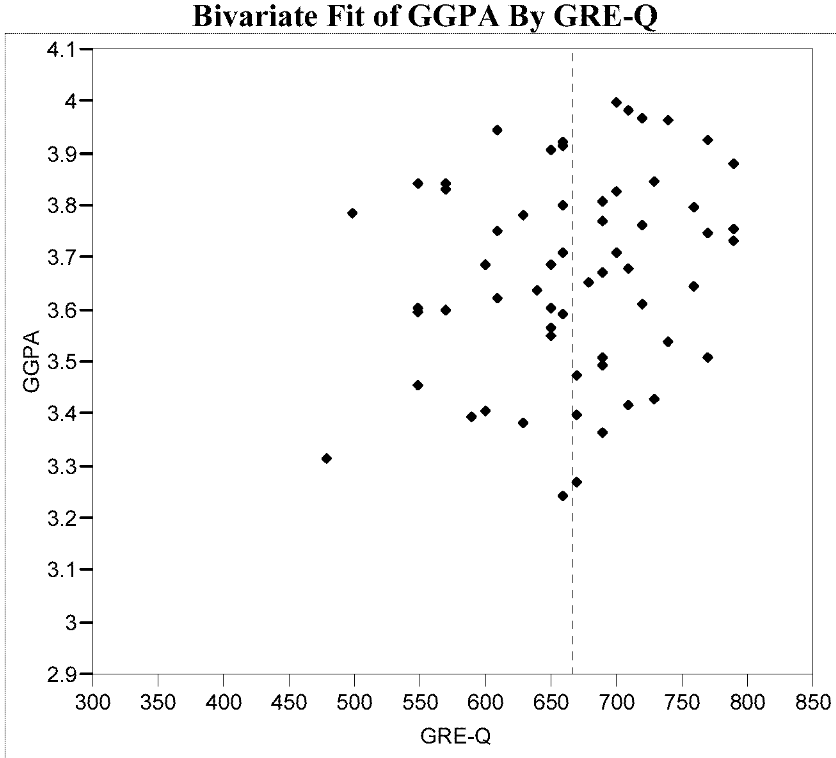
Parameter Estimates: Bivariate Fit of ADJGPA By PrGGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.290814	0.271321	12.13	<.0001
PrGGPA	0.089677	0.075333	1.19	0.2413

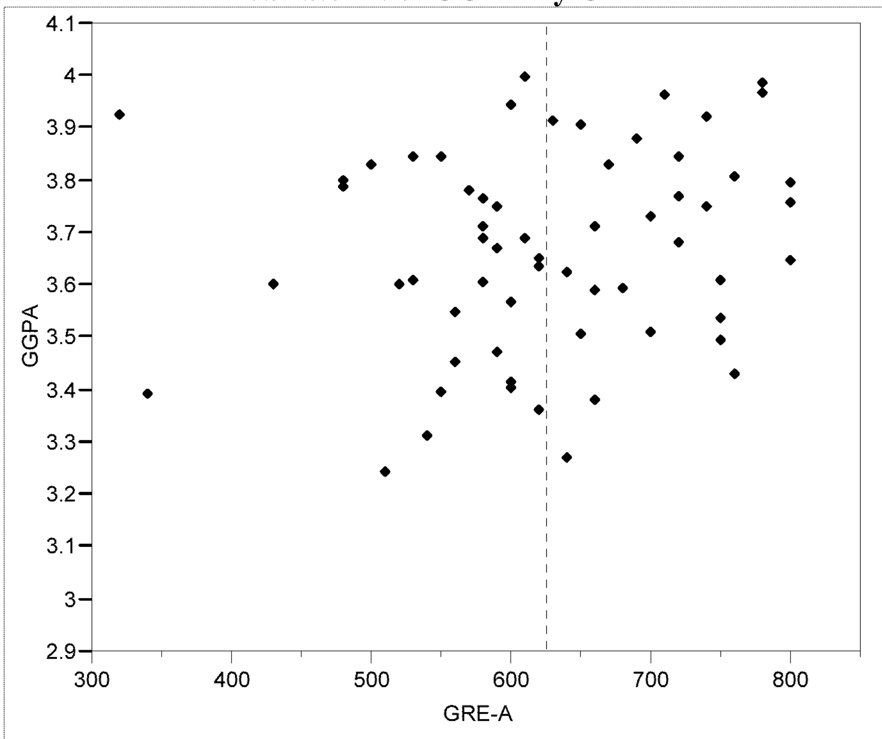
Bivariate Fit of GGPA By GRE-V



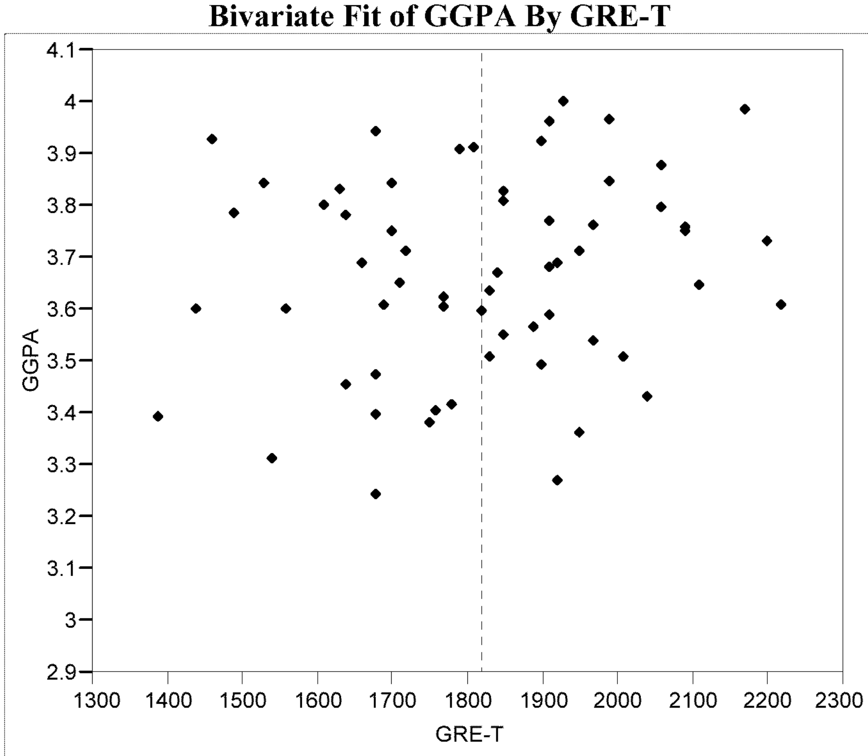
Bivariate Fit of GGPA By GRE-Q



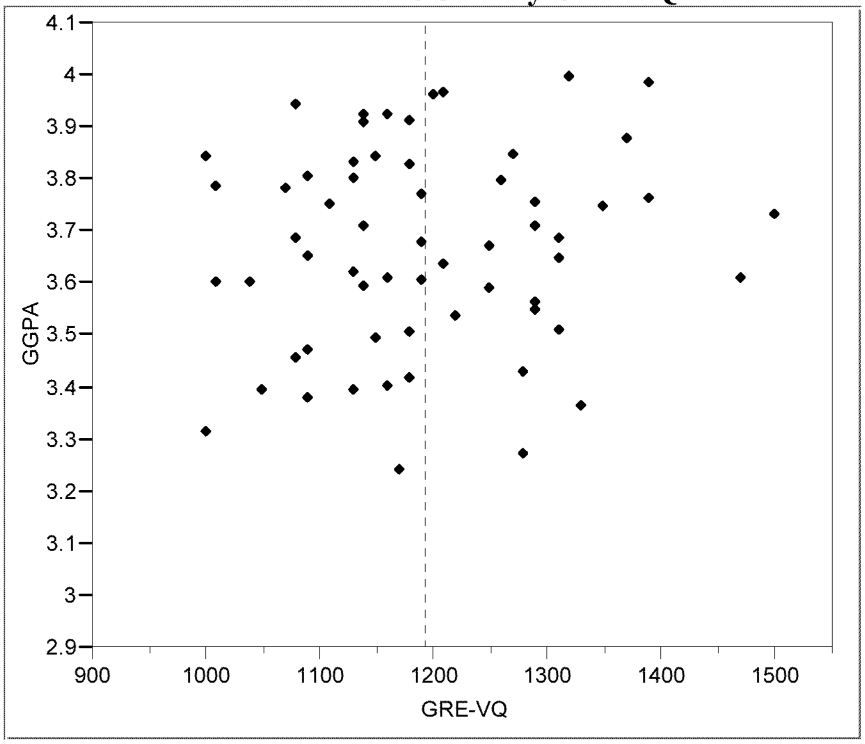
Bivariate Fit of GGPA By GRE-A



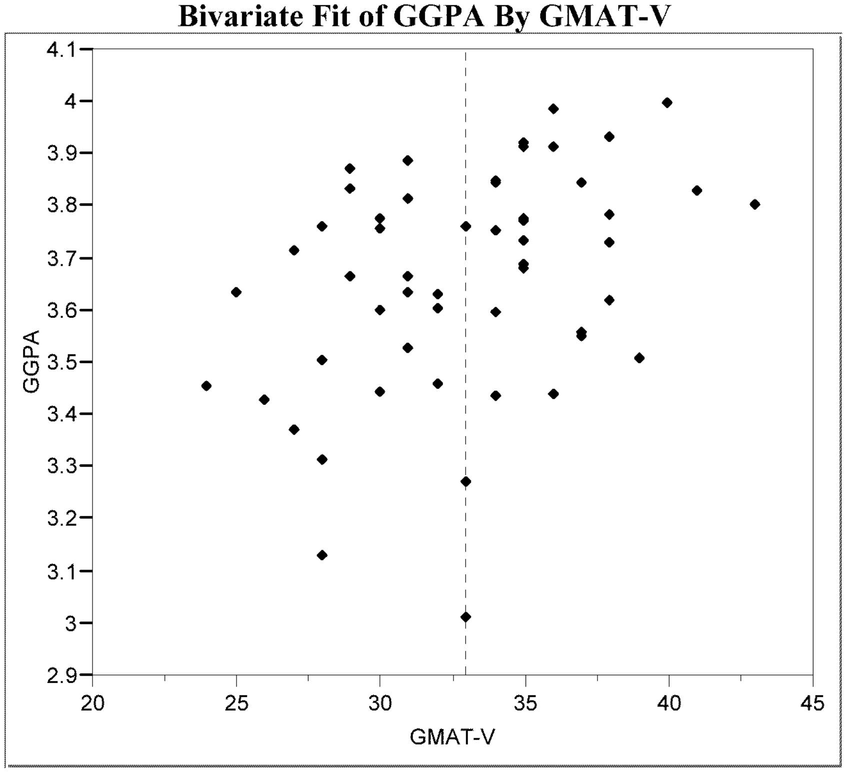
Bivariate Fit of GGPA By GRE-T



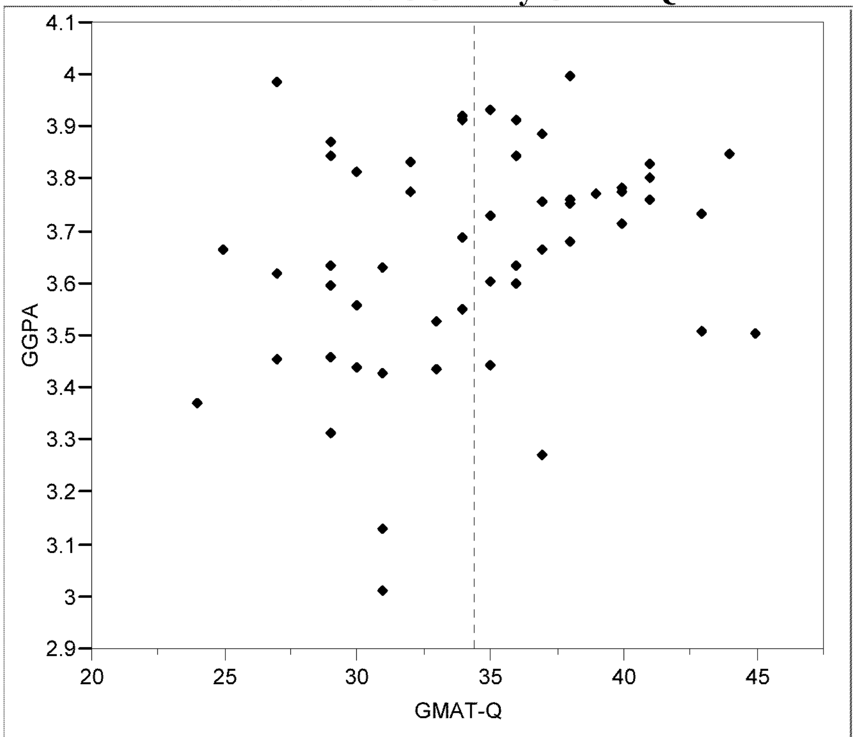
Bivariate Fit of GGPA By GRE-VQ



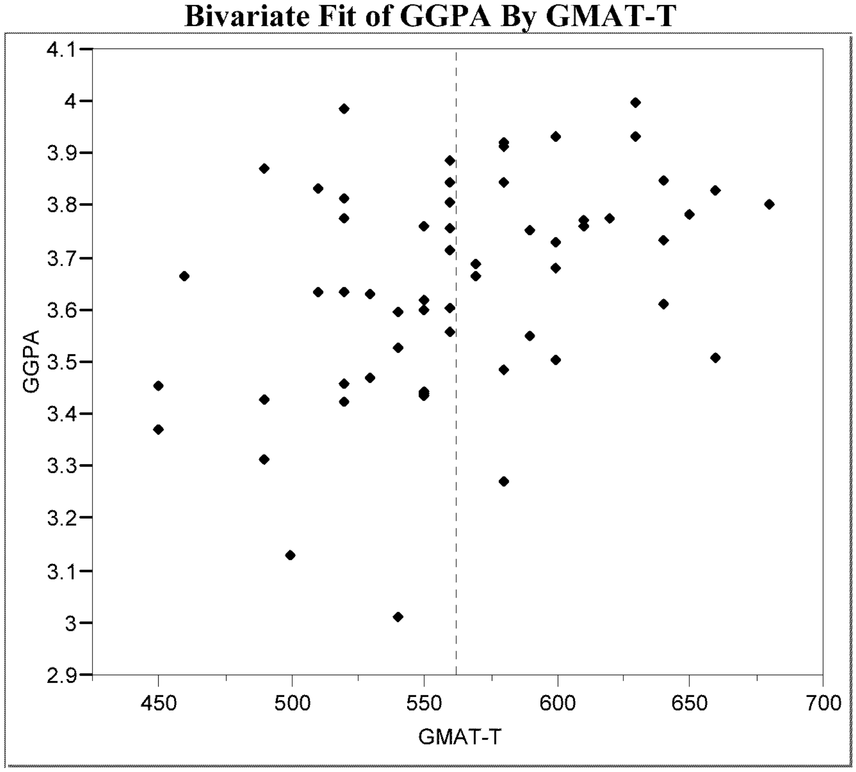
Bivariate Fit of GGPA By GMAT-V



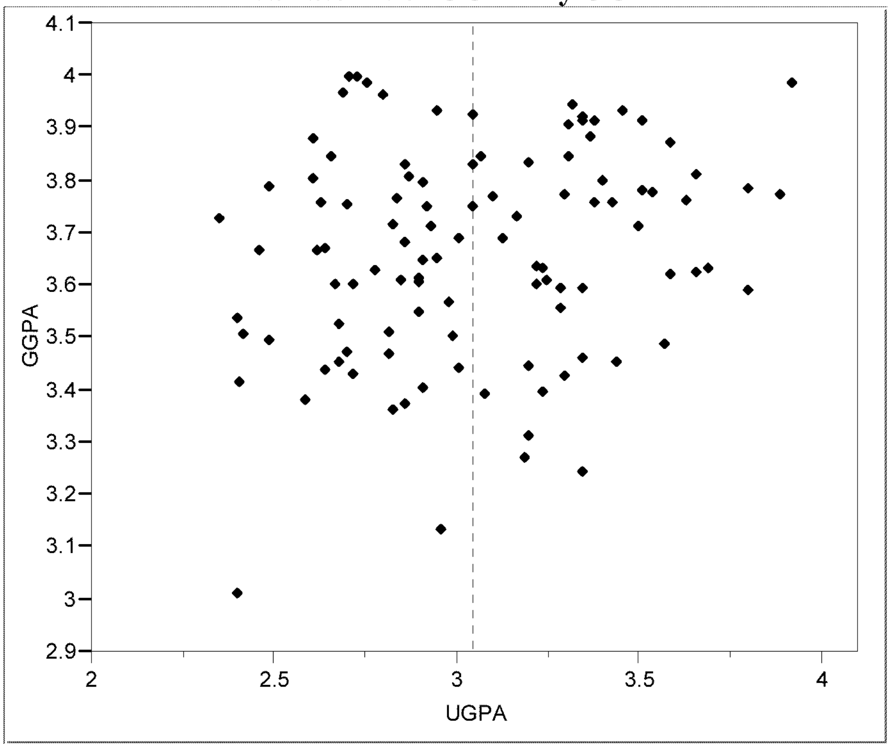
Bivariate Fit of GGPA By GMAT-Q



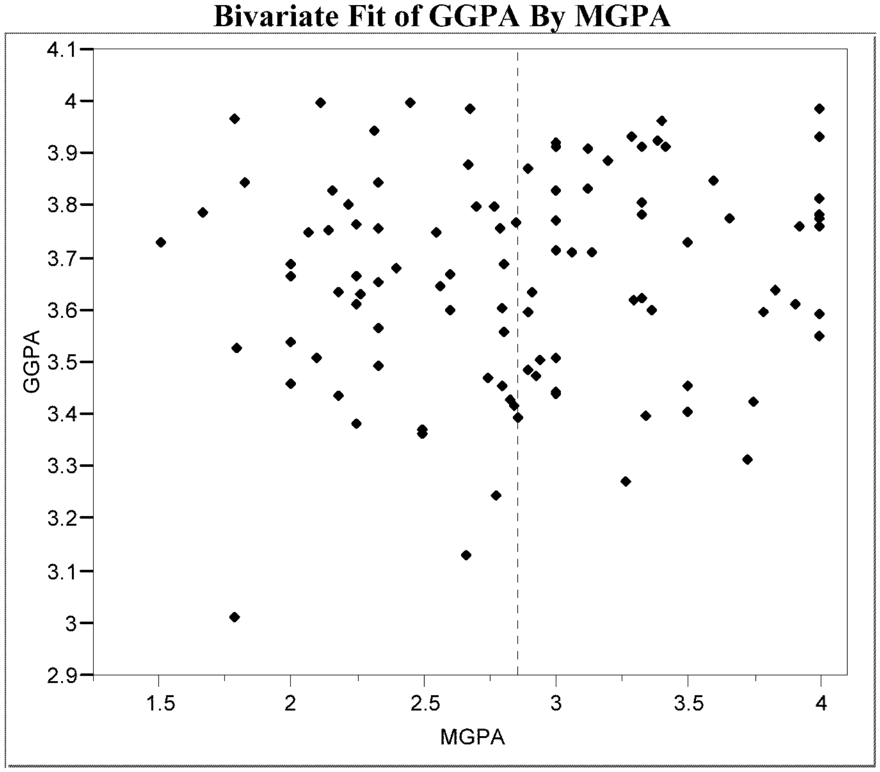
Bivariate Fit of GGPA By GMAT-T



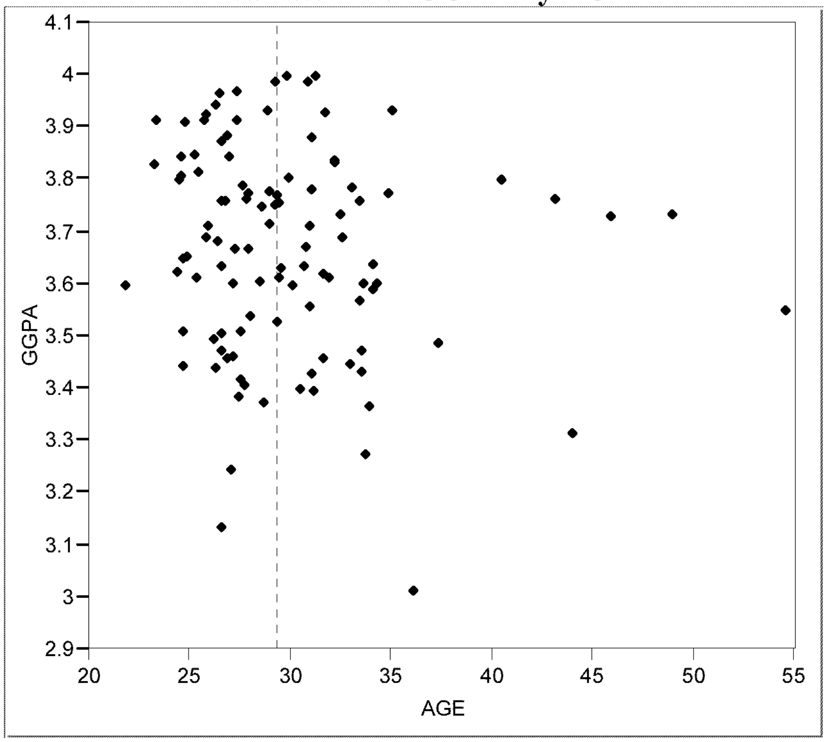
Bivariate Fit of GGPA By UGPA



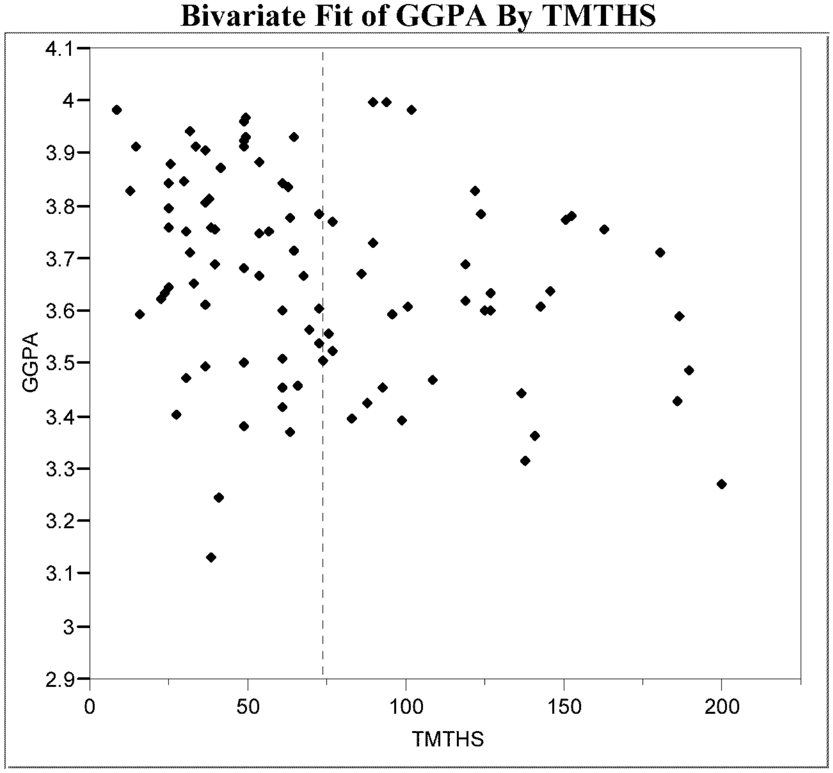
Bivariate Fit of GGPA By MGPA



Bivariate Fit of GGPA By AGE



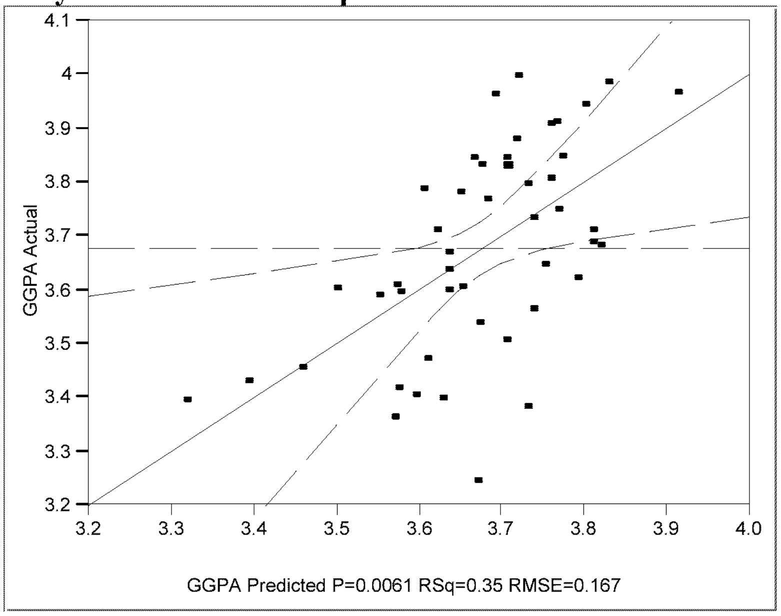
Bivariate Fit of GGPA By TMTHS



Appendix F. Statistics for Improved Models

Improved GRE-based Model for GGPA

Actual by Predicted Plot: Improved GRE-based Model for GGPA



Summary of Fit: Improved GRE-based Model for GGPA

R^2	0.350047
Adj R^2	0.252554
Root MSE	0.167029
Observations	47

Analysis of Variance: Improved GRE-based Model for GGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.6010166	0.100169	3.5905
Error	40	1.1159434	0.027899	Prob > F
C. Total	46	1.7169600		0.0061

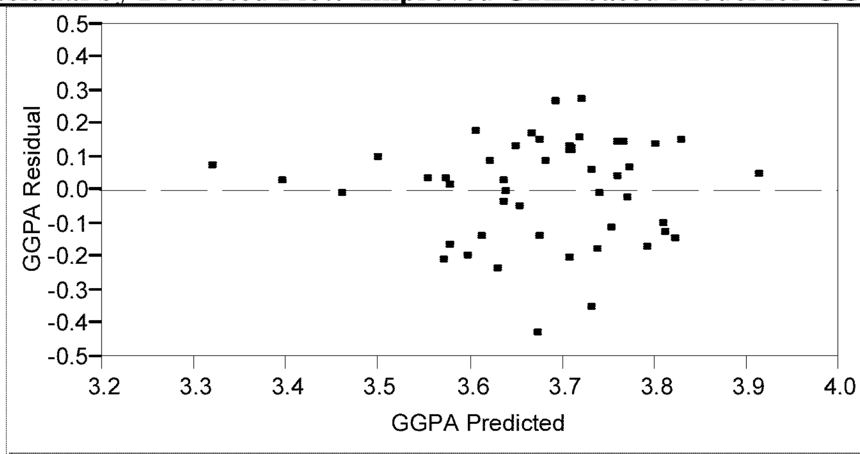
Parameter Estimates: Improved GRE-based Model for GGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.887093	0.30349	9.51	<.0001
GRE-A	0.000713	0.00026	2.74	0.0091
UGPA	0.184548	0.09410	1.96	0.0568
MGPA	-0.109047	0.05393	-2.02	0.0499
GENDER	0.134088	0.07045	1.90	0.0642
ENLST	-0.150781	0.07953	-1.90	0.0652
(GRE-A - 627)*(ENLST - 0.128)	-0.001806	0.00088	-2.06	0.0463

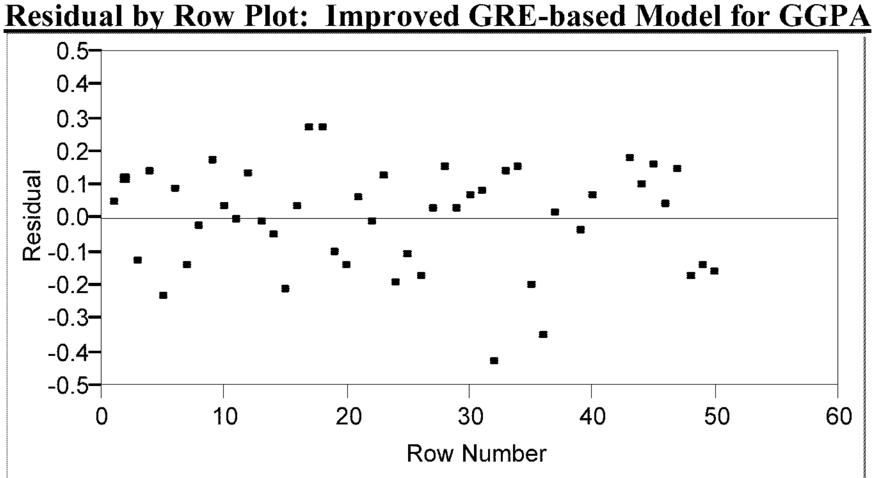
Effect Tests: Improved GRE-based Model for GGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GRE-A	1	1	0.21000438	7.5274	0.0091
UGPA	1	1	0.10730702	3.8463	0.0568
MGPA	1	1	0.11404727	4.0879	0.0499
GENDER	1	1	0.10106258	3.6225	0.0642
ENLST	1	1	0.10029137	3.5949	0.0652
GRE-A*ENLST	1	1	0.11800857	4.2299	0.0463

Residual by Predicted Plot: Improved GRE-based Model for GGPA

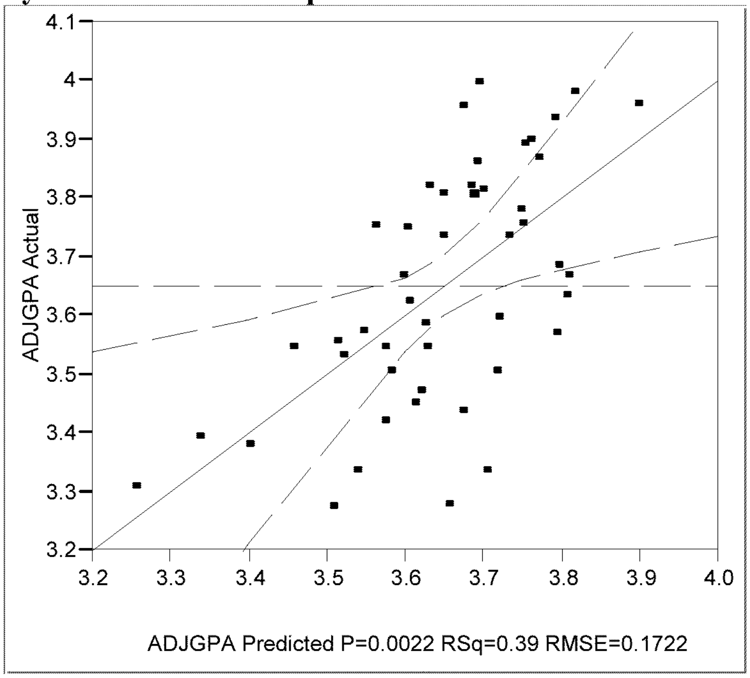


Residual by Row Plot: Improved GRE-based Model for GGPA



Improved GRE-based Model for ADJGPA

Actual by Predicted Plot: Improved GRE-based Model for ADJGPA



Summary of Fit: Improved GRE-based Model for ADJGPA

R^2	0.387724
Adj R^2	0.295882
Root MSE	0.172182
Mean of Response	3.648426
Observations	47

Analysis of Variance: Improved GRE-based Model for ADJGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	6	0.7509458	0.125158	4.2217
Error	40	1.1858617	0.029647	Prob > F
C. Total	46	1.9368075		0.0022

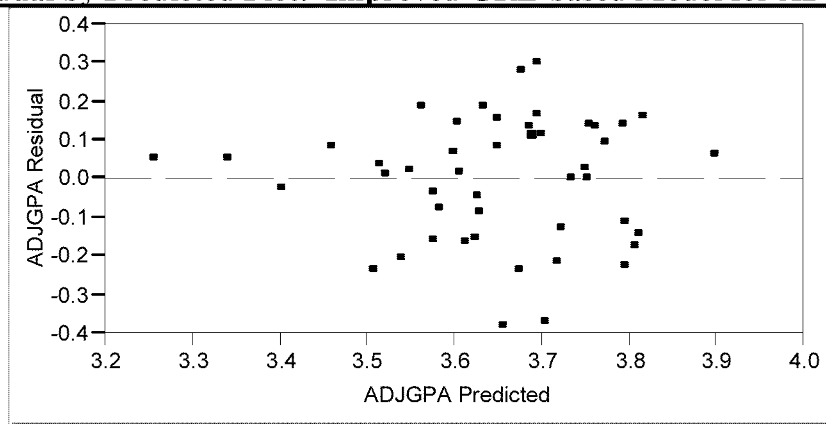
Parameter Estimates: Improved GRE-based Model for ADJGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.705338	0.312856	8.65	<.0001
GRE-A	0.000790	0.000268	2.95	0.0053
UGPA	0.214206	0.097002	2.21	0.0330
MGPA	-0.108802	0.055598	-1.96	0.0574
GENDER	0.160015	0.072624	2.20	0.0334
ENLST	-0.191453	0.081979	-2.34	0.0246
(GRE-A - 627)*(ENLST - 0.128)	-0.001802	0.000905	-1.99	0.0533

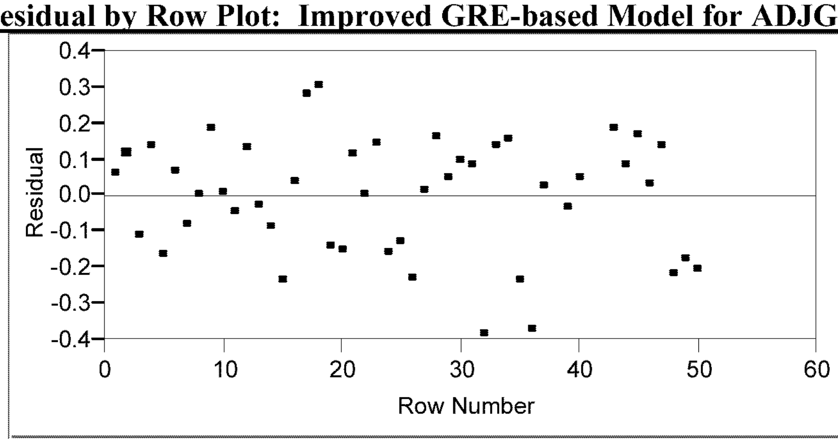
Effect Tests: Improved GRE-based Model for ADJGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GRE-A	1	1	0.25785997	8.6978	0.0053
UGPA	1	1	0.14456812	4.8764	0.0330
MGPA	1	1	0.11353561	3.8296	0.0574
GENDER	1	1	0.14392424	4.8547	0.0334
ENLST	1	1	0.16169389	5.4541	0.0246
GRE-A*ENLST	1	1	0.11752431	3.9642	0.0533

Residual by Predicted Plot: Improved GRE-based Model for ADJGPA

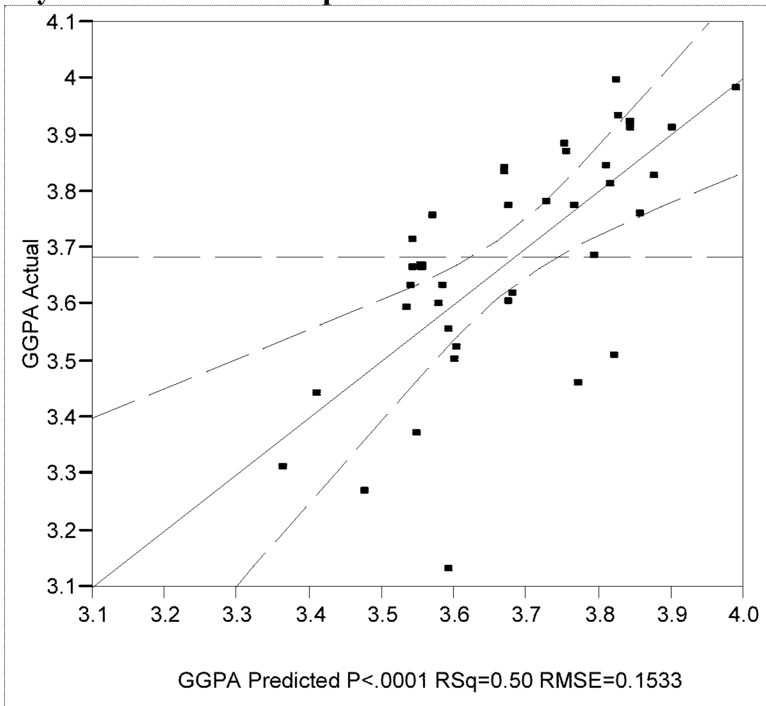


Residual by Row Plot: Improved GRE-based Model for ADJGPA



Improved GMAT-based Model for GGPA

Actual by Predicted Plot: Improved GMAT-based Model for GGPA



Summary of Fit: Improved GMAT-based Model for GGPA

R^2	0.498283
Adj R^2	0.454014
Root MSE	0.153322
Mean of Response	3.684184
Observations	38

Analysis of Variance: Improved GMAT-based Model for GGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.7937940	0.264598	11.2558
Error	34	0.7992637	0.023508	Prob > F
C. Total	37	1.5930577		<.0001

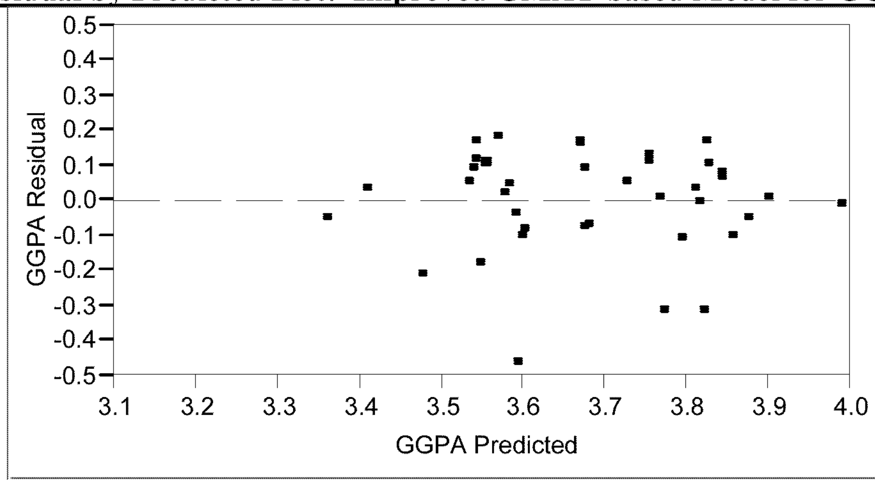
Parameter Estimates: Improved GMAT-based Model for GGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.294611	0.28600	8.02	<.0001
GMAT-V	0.023478	0.00645	3.64	0.0009
UGPA	0.216942	0.07096	3.06	0.0043
ENLST	-0.283723	0.06443	-4.40	0.0001

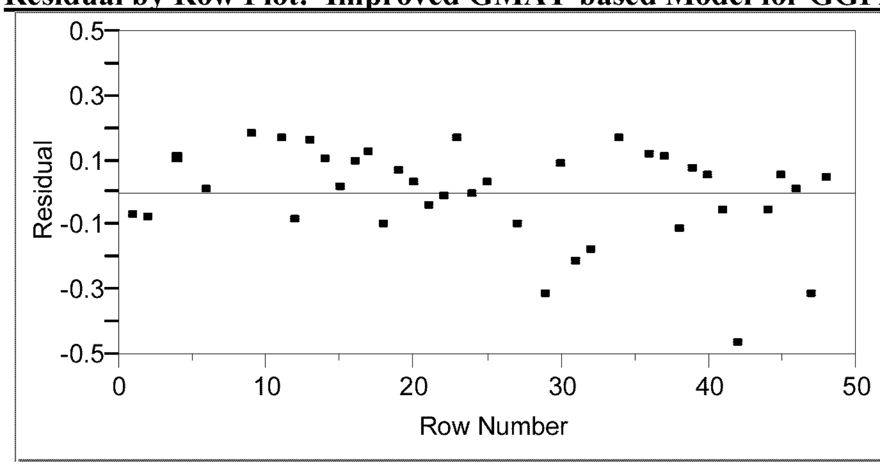
Effect Tests: Improved GMAT-based Model for GGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
GMAT-V	1	1	0.31183371	13.2651	0.0009
UGPA	1	1	0.21972201	9.3468	0.0043
ENLST	1	1	0.45591382	19.3942	0.0001

Residual by Predicted Plot: Improved GMAT-based Model for GGPA

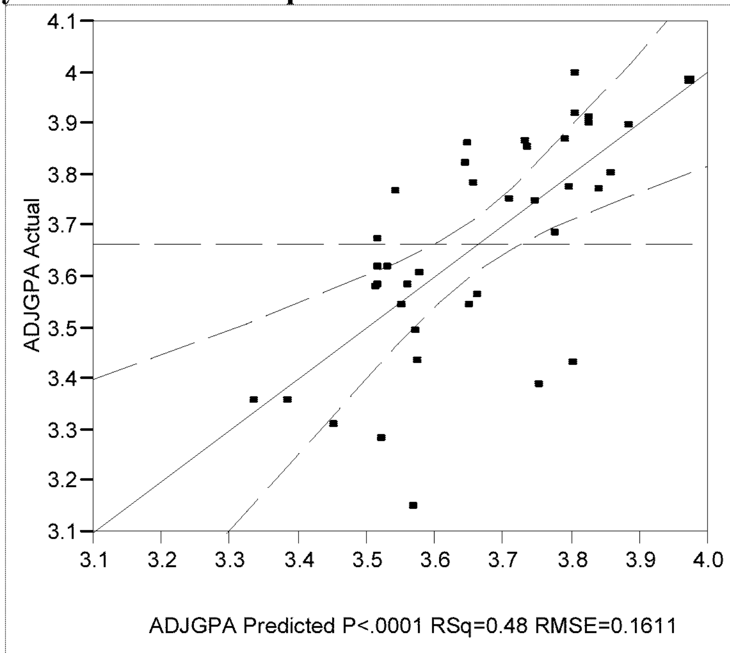


Residual by Row Plot: Improved GMAT-based Model for GGPA



Improved GMAT-based Model for ADJGPA

Actual by Predicted Plot: Improved GMAT-based Model for ADJGPA



Summary of Fit: Improved GMAT-based Model for ADJGPA

R^2	0.483576
Adj R^2	0.438009
Root MSE	0.161131
Mean of Response	3.662526
Observations	38

Analysis of Variance: Improved GMAT-based Model for ADJGPA

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	0.8266016	0.275534	10.6124
Error	34	0.8827519	0.025963	Prob > F
C. Total	37	1.7093535		<.0001

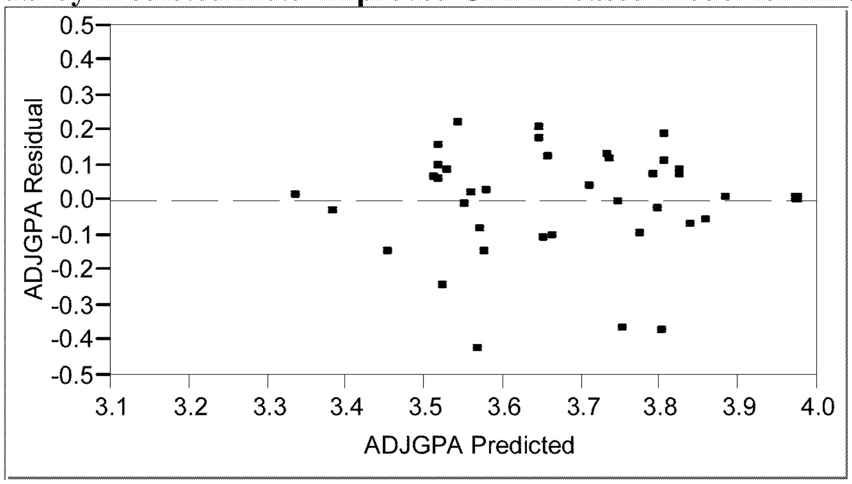
Parameter Estimates: Improved GMAT-based Model for ADJGPA

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	2.235403	0.300566	7.44	<.0001
UGPA	0.223917	0.074574	3.00	0.0050
GMAT-V	0.023977	0.006775	3.54	0.0012
ENLST	-0.287655	0.067707	-4.25	0.0002

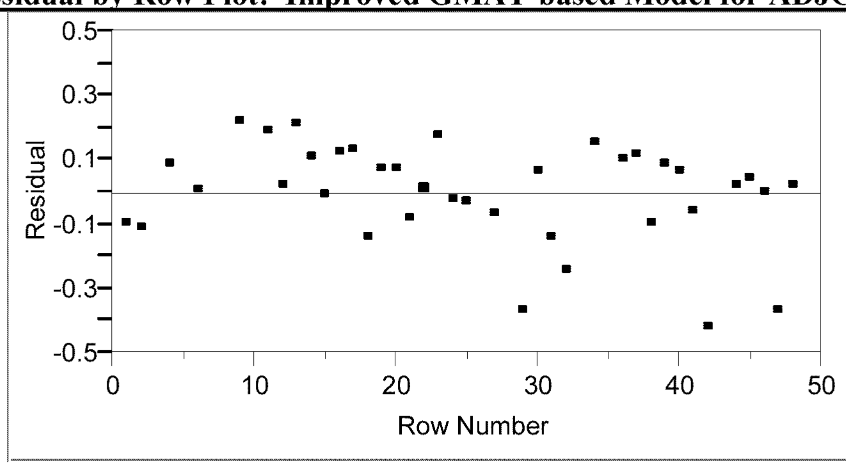
Effect Tests: Improved GMAT-based Model for ADJGPA

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
UGPA	1	1	0.23407886	9.0158	0.0050
GMAT-V	1	1	0.32522288	12.5263	0.0012
ENLST	1	1	0.46863527	18.0499	0.0002

Residual by Predicted Plot: Improved GMAT-based Model for ADJGPA



Residual by Row Plot: Improved GMAT-based Model for ADJGPA



Bibliography

Abedi, Jamal. "Predicting Graduate Academic Success from Undergraduate Academic Performance: A Canonical Correlation Study," *Educational & Psychological Measurement*, 51: 151-160 (Spring 1991).

AFIT Education Newsletter, AFITRP 36-139, V4(1), March 2000.

Ahmadi, Mohammad, Farhad Raiszadeh, and Marilyn Helms. "An Examination of the Admission Criteria for the MBA Programs: A Case Study," *Education*, 117: 40-546 (Summer 1997).

Air Force Institute of Technology, *Academic Year 2001-2002 Catalog*. Academic Catalog of the AFIT Graduate School of Engineering and Management, 2001.

Air Force Personnel Center (AFPC). from the "Career Field Opportunities" page of the AFPC Assignment System website, located at: http://afas.afpc.randolph.af.mil/Acquis/options.html (21 January 2002)

Air University/Financial Management Directorate, Air Education and Training Command. Spreadsheet, "Average and Variable Training Cost per Graduate – AFIT". Maxwell AFB AL. 26 September 2001.

Arnold, Larry R., and Amiya K. Chakravarty. "Applicant Evaluation in an Executive MBA Program," *Journal of Education for Business*, 71: 277-283 (May/June 1996).

Bieker, Richard F. "Factors Affecting Academic Achievement in Graduate Management Education," *Journal of Education for Business*, 72: 42-46 (September/October 1996).

Bowman, James S. "Admission Practices in Master of Public Administration Programs: A Nationwide Study," *Public Administration Review*, 48:867-875 (September/October 1988).

Bretz, Robert. D. "College Grade Point Average as a Predictor of Adult Success: A Meta-analytic Review and Some Additional Evidence," *Public Personnel Management*, 18: 11-22 (1989)

Buckley, Kevin W. *An Evaluation of the Air Force Institute of Technology Student Selection Criteria*. MS thesis, AFIT/GSM/LSR/98S-2. School of Systems and Logistics, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, September 1989 (AD-A215619).

Colarelli, Stephen. M., Roger A. Dean, and Constantine Konstans. "Comparative Effects of Personal and Situational Influences on Job Outcomes of New Professionals," *Journal of Applied Psychology*, 72:558-566 (November 1987).

Cole, Nancy S. "Merit and Opportunity: Testing and Higher Education at the Vortex" *The GRE FAME Report Serie*s Vol 1: 12-18. Princeton NJ: Educational Testing Service, 1998.

Dunlap, Katherine M., H. Carl Henry, and Mark W. Fraser. "The Relationship Between Admissions Criteria and Academic Performance in an MSW Program," *Journal of Social Work Education*, 34:455-463 (Fall 1998).

Educational Testing Service, *Guide to the Use of GMAT Scores*. Effective August 1, 2001. Princeton NJ: Graduate Management Admission Council, 2001.

Evans, Richard D., Admissions Counselor, AFIT/RRE, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH. Personal interviews. October - January 2002.

Fenster, Abe, Keith A. Markus, Carl F. Wiedemann, Marc A. Brackett, and John Fernandez. "Selecting Tomorrow's Forensic Psychologists: A Fresh Look at Some Familiar Predictors," *Educational & Psychological Measurement*, 61: 336-348 (April 2001).

Glass, Gene V., and Kenneth D. Hopkins. *Statistical Methods in Education and Psychology*. Needham Heights MA: Allyn and Bacon, 1998.

Goldberg, Edith L., and George M. Alliger. "Assessing the Validity of the GRE for Students in Psychology: A Validity Generalization Approach," *Educational & Psychological Measurement*, 52: 1019-1027 (1992).

Graduate Record Examinations Board (GREB), *Guide to the Use of Scores*. for the years 2000-2001. Princeton NJ: Educational Testing Service, 2000.

Graham, L. D. "Predicting Academic Success of Students in a Master of Business Administration Program," *Educational & Psychological Measurement*, 51: 721-727 (Fall 1991).

Hancock, Terence. "The Gender Difference: Validity of Standardized Admission Tests in Predicting MBA Performance," Journal of Education for Business, 75: 91-93 (November/December 1999).

Hoefer, Peter and James Gould. "Assessment of Admission Criteria for Predicting Students' Academic Performance in Graduate Business Programs," *Journal of Education for Business*, 75: 225-229 (March/April 2000).

Hopkins, Kenneth D. *Educational and Psychological Measurement and Evaluation*. Needham Heights MA: Allyn and Bacon, 1998.

House, J. Daniel. "Age Differences in Prediction of Student Achievement From Graduate Record Examination Scores," *The Journal of Genetic Psychology*, 159: 379-382 (September 1998).

House, J. Daniel, and Edward J. Keeley. "Differential Prediction of Adult Student Performance from Miller Analogies Test Scores," The Journal of Genetic Psychology, 157: 501-503 (December 1996).

Hunter, J. E., and R. F. Hunter. "Validity and Utility of Alternative Predictors of Job Performance," Psychological Bulletin, 96: 72-98 (1984).

Keith, Robert W. *Success Predictors for Students to Attend AFIT Department of Systems Management Master's Degree Programs.* MS thesis, AFIT/GSM/SM/77S-7. School of Engineering, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, September 1977 (AD-A046110).

Kuncel, Nathan R., Sarah A. Hezlett, and Deniz S. Ones. "A Comprehensive Meta-Analysis of the Predictive Validity of the Graduate Record Examinations: Implications for Graduate Student Selection and Performance," *Psychological Bulletin*, 127: 162-181 (January 2001).

McClave, James T., P. George Benson, and Terry Sincich. *Statistics for Business and Economics* (8th Edition). Upper Saddle River NJ: Prentice Hall, 2001

Mitchelson, Ronald L., and Don R. Hoy. "Problems in Predicting Graduate Student Success," *Journal of Geography*, 83: 54-57 (March/April 1984).

Monson, Michael G., Comptroller Assignment Officer, Acquisition Officer Assignment Branch, Mission Support Officer Assignment Division, Air Force Personnel Center, Randolph AFB TX. Telephone interview. 11 January 2002.

Morrison, Todd, and Melanie Morrison. "A Meta-Analytic Assessment of the Predictive Validity of the Quantitative and Verbal Components of the Graduate Record Examination with Graduate Grade Point Average representing the Criterion of Graduate Success," *Educational and Psychological Measurement*, 55: 309-316 (April 1995).

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models* (4th Edition). Boston MA: The McGraw-Hill Companies, Inc, 1996.

Nilsson, Johanna E., "The GRE and the GMAT: A Comparison of their Correlations to GGPA," *Educational & Psychological Measurement*, 55: 637-640 (August 1995).

Peiperl, Maury A. and Rose Trevelyan. "Predictors of Performance at Business School and Beyond," *Journal of Management Development*, 16:354-367 (1997)

"Predicting Performance in Graduate School," *Science*. 260: 494 (23 April 1993).

Profiles of American Colleges 2001. Hauppage NY: Barron's Educational Series, Inc, 2000.

Prokopyk, William N. *A Study of Academic Success Predictor Variables for Students Enrolled in AFIT's Graduate Operations Research Programs*. MS thesis, AFIT/ENS/GST/88M-10, School of Engineering, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, March 1988 (AD-A190760).

Ragothaman, Srinivasan, and Thomas Davies. "Using Artificial Intelligence Techniques to Predict M.B.A. Admission Decisions," *College Student Journal*, 32: 125-134 (March 1998).

Roth, Philip.L., Craig. A. BeVier, Fred. S. Switzer, and Jeffery. S. Schippmann. "Meta-Analyzing the Relationship Between Grades and Job Performance," *Journal of Applied Psychology*, 81: 548-556 (October 1996).

Rothstein, Mitchell G., Sampo V. Paunonen, James C. Rush, and Gillian A. King. "Personality and Cognitive Ability Predictors of Performance in Graduate Business School," *Journal of Educational Psychology*, 86: 516-530 (December, 1994).

Sny, Daniel R. *A Validity Study on the Air Force Institute of Technology Student Selection Criteria for Resident Master's Degree Programs*. MS thesis, AFIT/GLM/LSR/91S-59. School of Systems and Logistics, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, September 1991 (AD-A246812).

Spangler, Mark E. *A Regression Model For Predicting Academic Success of Prospective Students in the AFIT Graduate Logistics Management Program*. MS thesis, AFIT/GSM/ENS/89D-39. School of Systems and Logistics, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, December 1989 (AD-A218241).

Stockman, William A., Assistant Professor of Acquisition Management, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH. Personal interview. 10 January 2002.

Swayze, Susan. "Study Shows GMAT Predicts Academic Performance in Executive MBA Programs," *Graduate Management News*, Newsletter of the Graduate Management Admissions Council (November/December 2001).

Thornell, John G. and Anthony McCoy. "The Predictive Validity of the Graduate Record Examinations for Subgroups of Students in Different Academic Disciplines," *Educational and Psychological Measurement*, 45: 415-419 (Summer 1985).

Thorndike, Robert L. *Personnel Selection: Test and Measurement Techniques.* New York: John Wiley & Sons, Inc, 1949.

Van Scotter, James R. *Predictors of Success in Air Force Institute of Technology Residents Master's Degree Programs: A Validity Study*. MS thesis, AFIT/GSM/LSR/89S-2. School of Systems and Logistics, Air Fore Institute of Technology (AU), Wright-Patterson AFB OH, September 1983 (AD-A134391).

Wilson, Rick L. and Bill C. Hardgrave. " Predicting Graduate Student Success in an MBA Program: Regression versus Classification," *Educational & Psychological Measurement*, 55: 186-195 (April 1995).

Wooten, Bob E., and Charles D. McCullough. "Does Work Experience Help or Harm Students in MBA Programs?," *Business Forum*, 16(4): 12-14 (Fall 1991).

Wright, Robert E. and John C. Palmer. "GMAT Scores and Undergraduate GPAs as Predictors of Performance in Graduate Business Programs," *Journal of Education for Business*, 69: 344-348 (July 1994).

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
26-03-2002	Master's Thesis	August 2000 – March 2002

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
ADMISSIONS CRITERIA AND ACADEMIC PERFORMANCE IN THE AFIT GRADUATE COST ANALYSIS PROGRAM	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Garwood, Kenneth R., Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Building 640 WPAFB OH 45433-7765 DSN: 785-4539	AFIT/GAQ/ENV/02M-08
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
N/A	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research examined the admissions criteria used by the Air Force Institute of Technology (AFIT) to determine academic eligibility for the in-residence Graduate Cost Analysis (GCA) master's degree program. Using the cumulative graduate grade point average earned at AFIT as the measure of academic performance, comparisons were made between the predictive capability of the current criteria and other potential factors. Factors considered included scores from the Graduate Record Examination (GRE) and the Graduate Management Admissions Test (GMAT), undergraduate cumulative and math grade point averages, applicant's age and gender, ratings of the undergraduate school's admissions competitiveness, undergraduate degree type, and various measures of the applicant's time in military service. Predictive models were developed using stepwise linear regression. This research found the GMAT is more useful than the GRE as a predictor of GGPA in the AFIT GCA program. The optimal model accounted for up to 45% of the variance in GGPA, and included the GMAT Verbal score, UGPA, and a dichotomous indicator of prior service as a member of the military enlisted corps.

15. SUBJECT TERMS

Mathematical Prediction, Personnel Selection, Regression Analysis, Screening, Academic Performance, Graduate Education, Admissions Criteria

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					Mark A. Ward, Major, USAF, AFIT/ENV
a. REPORT	b. ABSTRACT	c. THIS PAGE			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	140	(937) 255-3636 ext 4742

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18