UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

HONORS THESIS

USMA PREP SCHOOL: PREDICTING SUCCESS

by

Caroline McCusker

May 2019

Thesis Advisor: Major David Delcuadro-Zimmerman
Second Reader: Colonel Joseph Lindquist
Second Reader: Lieutenant Colonel Kristin Arney

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, 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 this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** May 2019	**3. REPORT TYPE AND DATES COVERED**	
4. TITLE AND SUBTITLE USMA Prep School: Predicting Success			**5. FUNDING NUMBERS**
6. AUTHOR(S) Caroline McCusker			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) United States Military Academy West Point, NY 10996			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) United States Military Academy Preparatory School, Department of Admissions, USMA			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

The purpose of this project is to assist in optimizing candidate admissions to the United States Military Academy (USMA) at West Point from the United States MilitaryAcademy Preparatory School (USMAPS). There are many aspects within admissions to USMA. Candidates who find that they are disqualified in some way, but still show potential to the Admissions Committee, are given the opportunity to attend USMAPS. Currently,the metric to assess academic potential of all candidates is the College Entrance Exam Rank (CEER) Score, which considers High School Class Rank, SAT Verbal score, and SAT Math score. Once a student attends USMAPS, however, their CEER score is not necessarily updated to reflect the work they accomplished and the academic growth they achieved. Therefore, the goal of this project is to create a CEER score that reflects a candidates updated academic potential following their year at USMAPS. Through analysis, we concluded that the most significant defining factors in determining the updated PCEER (Prep School CEER) score are a students SAT Math score, SAT Verbal score, and cumulative Grade Point Average (GPA) at USMAPS. Thislast value is called the Cumulative Academic Performance Score (CAPS), and essentially replaces High School Class Rank from the original CEER score with a metric that effectively captures a students work in between High School and USMA. This allows the Admissions Department to assess a normalized snapshot of a students High School academic record through their SAT scores, and their recent academic record through their USMAPS GPA. The PCEER score can be used by admissions to assess a USMAPS cadets academic potential and can show a candidates growth from their original CEER score. The PCEER accounts for 57.19% more variance than the CEER, meaning it a much better predictor than the CEER score. This paper shows how the PCEER score was obtained and why each factor of it is considered important in assessing academic potential.

14. SUBJECT TERMS			**15. NUMBER OF PAGES**
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT OF** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

USMA PREP SCHOOL: PREDICTING SUCCESS

Caroline McCusker
Cadet, Army
B.S., United States Military Academy, 2019

Submitted in partial fulfillment of the
requirements for the degree of
BACHELOR OF SCIENCE
in **OPERATIONS RESEARCH**
with Honors
from the
UNITED STATES MILITARY ACADEMY
May 2019

Author: Caroline McCusker

Advisory Team: Major David Delcuadro-Zimmerman
Thesis Advisor

Colonel Joseph Lindquist
Second Reader

Lieutenant Colonel Kristin Arney
Second Reader

Colonel Tina Hartley
Chair, Department of Mathematical Sciences

USMA Prep School: Predicting Success

Cadet Caroline McCusker
Department of Mathematical Sciences
United States Military Academy
West Point, New York

Advisors: MAJ David Delcuadro-Zimmerman, COL Joseph Lindquist, COL Kristin Arney
Department of Mathematical Sciences
United States Military Academy
West Point, New York

ABSTRACT:

The purpose of this project is to assist in optimizing candidate admissions to the United States Military Academy (USMA) at West Point from the United States Military Academy Preparatory School (USMAPS). There are many aspects within admissions to USMA. Candidates who find that they are disqualified in some way, but still show potential to the Admissions Committee, are given the opportunity to attend USMAPS. Currently, the metric to assess academic potential of all candidates is the College Entrance Exam Rank (CEER) Score, which considers High School Class Rank, SAT Verbal score, and SAT Math score. Once a student attends USMAPS, however, their CEER score is not necessarily updated to reflect the work they accomplished and the academic growth they achieved. Therefore, the goal of this project is to create a CEER score that reflects a candidate's updated academic potential following their year at USMAPS.

Through analysis, we concluded that the most significant defining factors in determining the updated PCEER (Prep School CEER) score are a student's SAT Math score, SAT Verbal score, and cumulative Grade Point Average (GPA) at USMAPS. This last value is called the Cumulative Academic Performance Score (CAPS), and essentially replaces High School Class Rank from the original CEER score with a metric that effectively captures a student's work in between High School and USMA. This allows the Admissions Department to assess a normalized snapshot of a student's High School academic record through their SAT scores, and their recent academic record through their USMAPS GPA. The PCEER score can be used by admissions to assess a USMAPS cadet's academic potential and can show a candidate's growth from their original CEER score. The PCEER accounts for 57.19% more variance than the CEER, meaning it a much better predictor than the CEER score. This paper shows how the PCEER score was obtained and why each factor of it is considered important in assessing academic potential.

CONTACT: caroline.mccusker@westpoint.edu
david.delcuadro-zimmerman@westpoint.edu

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Colleges in America face the same dilemma every year of deciding which candidates to admit to their institution. The United States Military Academy (USMA) is no different. The difference with USMA, however, is that USMA admissions looks at a whole candidate score. This means that an applicant's academics, physical activities, and leadership skills are all taken into account in the application process. This is known as the "Whole Candidate Concept", which states that "West Point seeks well-rounded young people who demonstrate excellent academic ability, leadership potential, and overall fitness ("New Admissions Committee Training")." The "Whole Candidate Concept" makes the admissions process very lengthy. Candidates are required to get a medical screening to determine whether or not they meet the Army's commissioning and retention standards. They must receive a nomination to the Academy from one of the following sources: a congressperson, senator, the Vice President, the President, or the Superintendent to USMA. Then, the candidate must complete the Candidate Fitness Assessment, a physical test that tests their strength, endurance, and agility. Finally, the candidate must provide their academic information, to include their GPA, their grades, their SAT or ACT scores, and the activities they participated in in High School ("New Admissions Committee Training"). In addition to achieving the "Whole Candidate Concept", USMA Admissions has class composition goals, to include admitting students of different sexes, different ethnicities, and different talent backgrounds ("New Admissions Committee Training").

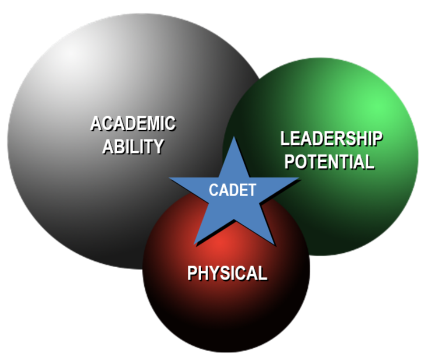

Figure 1. The Whole Candidate Concept

These specific Admissions goals sometimes makes it difficult to find candidates that meet all of the requirements for USMA. Many applicants are "at risk" when they apply, meaning that they are either disqualified in some way from attending USMA. This does not necessarily mean that these candidates are rejected from admission. Some are instead recommended for the United States Military Preparatory School (USMAPS) by the Director of Admissions. This process is shown in Figure 2. The blue arrow represents all of the students that are directly admitted to USMA. The orange arrow represents all of the students that are admitted to USMAPS and then go to USMA.

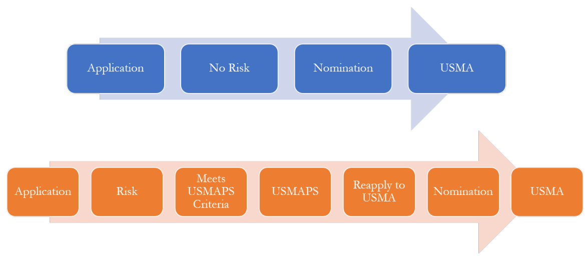

Figure 2. Common Progression to USMA

USMAPS serves as a bridge for these disqualified candidates to get to USMA by addressing specific deficiencies and preparing students for success. The program is specifically tailored to help students strengthen their academic or physical deficiencies to meet the standards that USMA requires. One issue with this system is a lack of granularity for the USMAPS acceptance process because there are so many different factors that affect a candidate's risk profile. The Admissions Department considers the following criteria when considering admitting a student to USMAPS ("New Admissions Committee Training"):

- Is the candidate at risk based on USMA risk levels? - CEER (520), SAT Reading (560), SAT Math (580), ACT English (23), ACT Math (24), ACT Reading (24), ACT Science (23)
- Does the candidate demonstrate strong potential to meet USMAPS program challenges?
- Does the candidate demonstrate exceptional leader potential?
- Does the candidate contribute to USMA class demographics? (i.e. Is the candidate a leader? Athlete? Woman? Minority? Soldier?)
- What is the candidate's CEER score?

This criteria is very subjective; it does not necessarily evaluate a candidate. In an interview, the Dean of USMAPS stated that this is a strength of the program at USMAPS, because it allows for students to be represented both qualitatively and quantitatively (Wong-Dodge Interview). All candidates at USMAPS must reapply to USMA for reconsideration.

Unfortunately, however, the current system for bench-marking students does not directly incorporate progress made while at USMAPS. For most candidates, their risk profile will appear almost identical to the previous year (when they were disqualified) - despite the developmental experience at USMAPS. This paper will attempt to quantify the likelihood of USMAPS Graduate success at USMA. Our research looks at quantifying the risk of admitting USMAPS graduates to ensure that the most qualified candidates that meet the the mission of USMA are admitted. Because of this, quantifying risk of candidate success and looking at determining factors of success at USMA in terms of USMAPS students, who are already disqualified from admissions, becomes extremely important. Specifically, we look at defining a metric to assess USMAPS Cadet Candidates (CCs) in a similar manner to their direct-admit counterparts by incorporating their year at USMAPS, and predicting their likelihood of success at USMA.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

This chapter will provide a brief history of the Prep School, an overview of academics at USMAPS, an overview of research done on the Prep School and success rates from its graduates, as well as an overview of admission to and the definition of success at USMA. The purpose of this chapter is to provide background to help guide our research to find a better way of assessing Prep School candidates before they are admitted into USMA. It will serve as an introduction to the research by assessing and integrating previous research into our own.

A. HISTORY OF USMAPS

On May 4th, 1916, President Woodrow Wilson enacted legislation authorizing the appointment of enlisted soldiers to the United States Military Academy in preparation for WWI by integrating them into the Corps of Cadets (Jebb 4). USMA viewed the soldiers' experience in the Army as an asset to help train and integrate new officers into military lifestyle. Unfortunately, many of these soldiers lacked the academic background needed to succeed at West Point. This resulted in the need for them to attend preparatory schools to bolster their academic skill set. These preparatory schools were controlled by the Army, but were civilian run (USMAPS Website).

In 1946, the Superintendent of USMA, General Maxwell Taylor, established the Army Prep School at Stewart Army Air Field in Newburgh, NY, to consolidate and standardize the Prep School education for all incoming soldiers ("Overview Brief to COL Higgins). In 1965, USMAPS opened its doors to civilian candidates, offering another avenue for candidates to enter USMA (USMAPS Website). This shows a shift in the mission of USMAPS from bringing in soldiers with prior service experience to assisting "at-risk" candidates in preparing for the academic rigor of USMA. The Prep School also helps USMA meet demographic goals by admitting minorities and recruited athletes that may need more time to prepare for USMA. For example, the USMAPS class of 2018 admitted 108 African-American candidates, 10 Hispanic candidate, 48 female candidates, and 117 recruited athletes out of 227 total students ("Overview Brief to COL Higgins").

Since its inception, USMAPS has relocated to Fort Belvoir, VA, and then to Fort Monmouth, NJ, before finally moving to West Point, NY in 2011. The proximity to USMA has helped USMAPS align its curriculum with that of USMA, meaning that there has been a growth in similarity between the two schools, particularly regarding academic coursework (USMAPS Website).

B. ADMISSION TO USMAPS

Applying to USMA is a long process, consisting of medical qualifications, physical qualifications, a congressional or senatorial nomination, a fitness examination, and standardized testing (USMA Admissions Website). While these factors go in to a whole candidate score that helps the admissions committee determine if a candidate is right for USMA, the committee uses a different metric to measure a candidate's academic potential. This metric, the College Entrance Exam Rank (CEER), takes into account a candidate's high school rank, their SAT verbal score, and their SAT Math score. The most recent CEER score equation is shown in Equation II-1. Additionally, ACT scores can be used in a different calculation to find their ACEER score, shown in Equation II-2("Class of 2020 Calculations").

$$CEER = 0.364\ (High\ School\ Rank) + 0.269\ (SAT\ Verbal) + 0.432\ (SAT\ Math) - 48 \tag{II-1}$$

$$\begin{aligned} ACEER = 0.219\ (HS\ Rank) + 9.43\ (ACT\ Math) + 4.62\ (ACT\ English) \\ +0.45\ (ACT Science/Reasoning) + 4.01\ (ACT\ Reading) - 41.5 \end{aligned} \tag{II-2}$$

The CEER score was first used by West Point for the Class of 1965, and has been a part of the admissions process ever since. Slight changes have been made over time to meet various goals and policies of the Academy, but its fundamentals remain unchanged. This metric is considered a good predictor of academic performance and drives the admissions committee to make their decisions as to whether or not to admit a student to USMA. The minimum risk threshold for a student to be admitted directly to USMA (or to be a "Direct Admit") is a CEER score of 520. However, if a candidate has a CEER or ACEER score below 520, they are flagged and are considered to attend USMAPS (New Admissions Committee Training 24). This threshold was created through historical analysis, which shows that students with a CEER score below 520 were found to be academically at risk during their freshman year (McDonald Interview).

The CEER score is used as a metric to compare candidates, however it might not be the best metric to measure CCs' performance. Data from 2009-2016 shows that the CEER score is not necessarily a good predictor of freshman year academic performance (usmapsData analysis - see Appendix with code). Additionally, the candidate CEER score does not necessarily change after attendance at USMAPS unless a student improves in one of the areas assessed in the CEER score (or ACEER score). This is unlikely because the USMAPS experience does not specifically address standardized test preparation, so there is not a large opportunity for students to improve their ACT or SAT scores. Because of this, most CCs compete for USMA admission after USMAPS with a similarCEER score, which is not a good measure of their academic ability following an additional year of instruction. In general, CCs' CEER scores do not benefit from their year at USMAPS, and thus their candidate profile does not improve.

Another thing to consider is that the CEER takes into account SAT or ACT score, which used to be a focus at USMAPS; however, the focus has shifted to preparing candidates specifically for USMA classes such as English and Math (Jebb Interview). Because of this, the CEER score is potentially an outdated metric that may not be the best way to quantify the risk associated with admitting a student with preparatory experience.

With this information in mind, the focus of this research is going to be to use the current CEER score as a starting point, and find a way to improve it in order to objectively assess the academic ability of students after completing a year at USMAPS. This helps the Admissions Department make informed decisions about the students that they admit to USMA, with a metric that can say how likely a student from USMAPS is to succeed at USMA. This literature review contributes to the efforts of this research by providing information on the academics at USMA and USMAPS, how the current system for admitting students works, and where there is room to make the system better.

C. ACADEMICS AT USMAPS

USMAPS breaks the year down into four quarters in which the faculty are continuously preparing and assessing candidates readiness and qualification for USMA. Candidates begin their USMAPS journey with Cadet Candidate Basic Training, a three-week period of summer training where they are familiarized with the military lifestyle (Jebb 43). Upon completion of Cadet Candidate Basic Training, the candidates enter their first quarter of academics. Each quarter, candidates that are deemed at risk of succeeding are reviewed for their potential by the staff and faculty at USMAPS. “At risk candidates meet any of the following criteria: a GPA below 2.5, a C- or below in any class, a failure of an APFT, a CFA failure, a height/weight failure, or a character mentorship program in-completion (The Road to USMA).

Finally, at the end of the last quarter, the Commandant of USMAPS makes a determination about each candidate’s performance at USMAPS and decides whether to provide his endorsement to attend USMA. The commandant has a series of criteria from which to make his decision. This criteria includes an academic, military, physical, and character evaluation from specific departments within USMAPS (“USMAPS Commandant’s endorsement criteria”).

USMAPS originally focused solely on English and Math for academic courses. The intent of this was to prepare candidates for the heavy emphasis on engineering at USMA, as well as to help candidates improve their SAT scores and make their admission file more desirable (Jebb 50). In 2014, USMAPS added a Science course to its curriculum to help candidates meet the core requirements at USMA (Jebb 51). Additionally, in 2013, the English program was improved to create a more rigorous course that mirrored the kind of academic demands of the freshman year English class at USMA (Jebb interview). It is for this reason that the metrics for assessing CC performance are likely no longer useful: the old metric represent data from USMAPS before all of these curriculum changes were made. A candidate with a 2.5 GPA at USMAPS in 2000 is drastically different than a candidate with a 2.5 GPA at USMAPS in 2017. Because of this, there is need to assess the current metrics and modify or improve them as needed to make a useful model for the Admissions Department. This new model may need to include the Science class grades and the new GPAs that account for the curriculum changes in order to capture a modern view of CC performance.

D. PERFORMANCE PREDICTORS AT THE PREP SCHOOL

In 2016, Dr. Joel Jebb, the director of the English Department at USMAPS, wrote a dissertation looking into the effect of the USMAPS English department on success at USMA. Specifically, Dr. Jebb looked at the performance of USMAPS cadets in their freshman year English class at USMA. He found that on average, USMAPS students had consistently lower GPAs in their freshman year English class than their direct admit counterparts (Jebb 240). He also found that the overall GPA for USMAPS students tended to be approximately 0.4 points lower than their direct admit counterparts, indicating a deficit in USMAPS candidates academic ability at USMA (Jebb 272).

Based on these findings, it seems that USMAPS does not necessarily make a cadet successful at USMA. In response to this, Dr. Jebb discusses many underlying factors that could be effecting these statistics of USMAPS cadets. For example, he considers that USMAPS cadets are already at an academic deficit, since they were disqualified from USMA during their first attempt at admission. He also notes that USMAPS cadets go from taking three to four courses at USMAPS to taking a full five to six courses at USMA. This increase in workload for cadets that are already at an academic disadvantage could be one explanation for lower GPAs of USMAPS cadets (Jebb Interview). Thus, while a year at USMAPS might not make a cadet at an equivalent academic caliber as a Direct Admit, Dr. Jebb believes that the year spent at USMAPS improves and refines academic skills in CCs that give them a better chance of being successful.

While USMAPS might not seem to make a CC equivalent to a Direct Admit, it can be argued that the extra year prepares a candidate well. USMAPS is modeled after USMA; the structure of the day, the focus on academic, military, and physical pillars of success, and the commitment to leadership are mirrored between USMA and USMAPS. Even the courses are similar. At USMAPS, candidates take Math, English, and Science. The faulty at USMAPS are consistently communicating with their faculty counterparts at USMA, and are working toward preparing their students for the challenges of USMA. Therefore, the overall GPA of a cadet that went to USMAPS compared to the overall GPA of a direct admit might not be the best way to predict the success of a preparatory school admit at USMA. Our model will need to look into all measurable aspects of a candidates academics at USMAPS to include grades in individual classes in addition to GPAs.

The other aspect of cadet success that Dr. Jebb mentions is the inability to measure attitude and effort. He firmly believes that USMAPS cadets must have a level of grit to get through the academy that leads to a solid work ethic and success. Dr. Jebb notes that USMAPS is not an accredited university, so those candidates that decide to come to USMAPS must have some sort of high motivation to get through it and get to USMA, considering they cannot transfer their grades to any other university (271). This brings into focus a whole new dimension to the problem of defining success at USMA. The new issue becomes measuring attitude and effort, or grit, which are principles that will drive candidates and direct admits alike to be successful at USMA that cannot be measured by grades or GPAs (Jebb Interview).

This "grit" component can be assessed by faculty, who closely interact with students. School Official Evaluations (SOEs) are filled out by instructors at USMAPS to capture the whole candidate as opposed to just grades. Instructors rate a student's work ethic, influence, communication skills, and ability to perform under pressure on an SOE, which gives USMA qualitative data to help assess student performance. These evaluations are important to consider in the admissions process, however because of the quantitative nature of this research, they will not be used.

In terms of this project, Dr. Jebbs' research will be valuable as a framework to begin a statistical analysis of success at USMA. He used a mixed-methods approach to the problem, meaning he assessed both quantitative and qualitative data, while the approach used for this research will be strictly quantitatively focused. However, in the quantitative analysis it will be important to consider the underlying qualitative factors that could affect the data or create outliers.

These intangible aspects of a candidate cannot be observed through the data alone, but can be a significant contributor to academic performance and success. Thus, the variability and error of the model that we create to predict CC success at USMA will partially be attributed to these aspects of each individual candidate.

In 1979, the Office of the Director of Institutional Research did a study to identify the effect of attending USMAPS on College Board scores and performance at USMA. The researchers found that USMAPS cadets, while still having adequate performance in class, have lower GPAs and class grades than direct admit cadets (Office of the Director 3). However, this might not be attributed solely to USMAPS but instead to the type of student that goes through USMAPS. If a student was academically disqualified at the beginning of their USMAPS career, then perhaps they would be expected to struggle more academically compared to those who were directly admitted to the Academy. Thus, it would make sense for the academic GPAs of CCs to be lower than those of direct admits.

One other thing that this research team looked into is the attrition rates of CCs. It was discovered that CCs have a high attrition rate, meaning a proportionally smaller amount of them graduate compared to direct admits (Office of the Director 6). What it does not indicate, however, is the percentage of the CCs that left the academy for academic reasons. Again, this is important to consider in our research considering there are so many qualitative factors that effect a students performance in school that might not be measurable in a statistical model. These factors will thus not be captured directly in the model, but will be represented in our error values, since they are not quantifiable and are not able to be measured mathematically.

The Comptroller General of the United States did a study in 1976 looking into the attrition rate at the service academies in general. The research draws a link between individual and institutional reasons that cadets quit. The researchers find that there is no one reason that can explain the resignation of any particular cadet. Additionally, they found that there is a higher tendency for those with lower ability at entry to leave the academy, specifically for academic reasons ("Class of 2020" 85). The research suggests that those who drop out of an academy for academic reasons might not drop out because of their academic abilities. However, they instead drop out due to the challenges of adapting to a military lifestyle, which is not by any means a product of individual academic abilities ("Class of 2020" 4). The study also found that mathematical ability was more strongly related to academic attrition than verbal ability in military academies ("Class of 2020" 13).

Overall, this research will be valuable in deciding how to set up a model for USMAPS's effect on success at USMA. Our research can also model the likelihood of a student passing a Math course at USMA based on their Math grade at USMAPS. Additionally, we can add a piece to the model that predicts likelihood that a student will resign from the Academy based on their grades and GPA. All of these aspects of our research will be valuable in optimizing the model to promote admission of candidates that are most likely to succeed.

E. DEFINING SUCCESS

There are many different ways to define collegiate success (Figure 2.1). One way is through academic GPA. The Academy requires a student to achieve at least a cumulative GPA of 2.00 to graduate. The Academy also has "peg point" metrics, which are GPA measurements that a cadet must hit or else they will be put on probation. For example, during freshman year, the peg point is a GPA of 1.67. Another metric for success is individual course grades. USMA requires cadets to achieve a grade of "D" or better in all required academic courses (Office of the Dean 21).

Colonel Deborah McDonald, the director of Admissions at USMA, defines success through GPA. Specifically, she and the Admissions Department are only concerned with cadet performance during plebe year (McDonald Interview). While this might not be applicable to the cadets overall success at USMA, focus on plebe year is important to consider since this is the year that most mirrors the year that CCs have at USMAPS. The classes taken during plebe year are most similar to those at USMAPS, and the environment

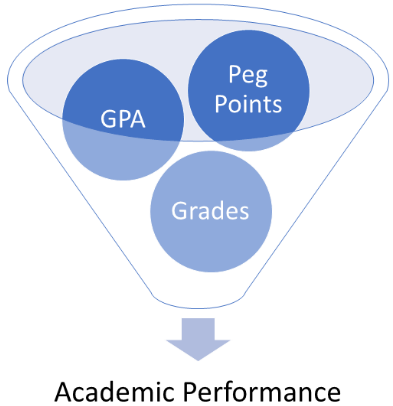

Figure 3. Factors to Determine Academic Performance

and newness of USMA is still a factor in whether or not a CC will quit during plebe year. Additionally, attrition is very high Plebe year. Therefore this research will focus mainly on plebe year as the scope for our problem.

In terms of this project, success can be hard to define. Thus, a combination of passing classes and meeting the peg point GPA for freshman year will be used in this research to get a more holistic view of the cadets in the study. The response variable will be the most important in our model, since it will tell us whether or not a student is more likely to be successful at USMA upon graduation from USMAPS. It will reflect a student's academic performance at USMAPS and predict their likelihood of success at USMA on the same scale as the CEER score, so that it is comparable to the scale that students who are directly admitted to USMA are using.

III. THE DATA

We requested data after reviewing the protocol with the USMA Human Subjects Research Protection Program. Because this project is solely designed to assess the admission process at USMA, it is excluded under DODI3216.02. The data was deidentified by the G5 and the subjects are not readily ascertained through identifiers linked to the subjects. As researchers, we did not attempt to re-identify any subjects, nor did we contact any subjects. The data or results from this study may not be used for any other purpose.

The data for this project includes many different measures of performance of a cadet while at USMAPS and at USMA. It consists of data from the years 2014 to 2016, (USMA graduating classes of 2019-2021) because 2014 is the year that Science was added to the curriculum, and 2016 is the most current and complete data from USMAPS that is available. It includes information on students before they are admitted into USMA or USMAPS and on their performance once admitted.

This data is organized by a cadet ID number, which is unique to each cadet that begins at USMA or USMAPS. Different fields in the data set include what the student's Cumulative Academic Performance Score (CAPS) at USMAPS was, what English class they took each quarter, what their grades were each quarter in all of their classes, what their original CEER score was, and many other measures of performance that can be used to discuss predicting academic success. The fields of the dataframe are in Figure 4.

In order to be more consistent in the data, we omit data points where the CQPA of a student is 0, assuming that this means that the student was never admitted to USMA from USMAPS and therefore never earned a CQPA. Because of this, these students will not help us to determine what predictive factors at USMAPS lead to success at USMA.

Grades in this data are quantified on the scale that USMA uses in determining CQPA from letter grades. This scale is depicted in Figure 5. Each letter grade has a corresponding point score.

A. STATISTICS

This research will rely on a basic foundation of statistics in order to explain the analysis of data. Specifically, this research will use linear regression to explain the correlation

CDT ID
CAPS
English Term Letter Grades (Terms 1-4)
Math Term Letter Grades (Terms 1-4)
English Course Description (Terms 1-4)
Math Course Description (Terms 1-4)
CEER Score
ACT English Score
ACT Math Score
ACT Reading Score
ACT Science and Reasoning Score
SAT Verbal Score
SAT Math Score
SAT Writing Score

Figure 4. Snapshot of Data

between the data and a student's performance. This section provides a brief background on linear regression and how to interpret results.

Linear regression is a method that uses predictor variables to make predictions about response variables (variables that are predicted). In terms of this research, we are trying to find the best predictor variables to assess a student's performance at USMA (the response variable). Using the data that describes a cadet's performance at USMAPS and at USMA, we can use a computer program (R^{TM}) to find a best-fitting line to represent the relationship between the predictor variables and cadet performance (Lane 462).

The basic equation for a regression line is below. $\hat{Y}$ is the response variable, or in this case cadet performance. x is a predictor variable. β is the weight associated with that predictor variable. α represents the intercept (Lane 465).

$$\hat{Y} = \beta x + \alpha \qquad \text{(III-1)}$$

We used multiple linear regression to evaluate different predictor variables, using R^{TM} to decide whether a variable has predictive power or not. The basic equation for this is noted below.

$$\hat{Y} = \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \alpha \qquad \text{(III-2)}$$

Grade Scale	
A+	4.33
A	4
A-	3.67
B+	3.33
B	3
B-	2.67
C+	2.33
C	2
C-	1.67
D+	1.33
D	1
D-	0.67
F	0.33

Figure 5. Quantified Grade Scale

To decide which predictor variables are important (or statistically significant) to the model, we use R™ to calculate a $p-value$. In linear regression, the p-value is the probability of observing a test statistic as extreme or more extreme than the one we observe under the null hypothesis. When we create a linear regression model, we adopt a null-hypothesis that the beta values (seen in Equation III-2) will be 0, meaning that there is an absence of a relationship between predictor variables and response variables (Ronald L. Wasserstein). If a p-value is low, we can reject this null hypothesis and make the conclusion that the predictor should remain in the model. Therefore a low p-value (closer to 0) is representative of a significant predictor variable, and a high p-value (closer to 1) is representative of an insignificant predictor (Lane 680).

To evaluate a completed model, we use the $Adjusted\ R - Squared$ metric. This number represents the percent of variance in the response variable accounted for in the model. Therefore a large Adjusted R-Squared value (closer to 1) is representative of a better model, and a low Adjusted R-Squared value (closer to 0) is representative of a worse model (Adjusted R2).

THIS PAGE INTENTIONALLY LEFT BLANK

IV. METHODOLOGY AND RESULTS

This paper will work to evaluate the validity of the CEER score as a metric for CC's performance and, if necessary, find a better way to quantify risk for CCs different from the CEER score. We will first model USMAPS students' success at USMA as a factor of the candidate CEER score to see if there is a direct correlation between the two. Then, we will build a new model based on information regarding performance at USMAPS to see if the model is more useful. Our model should outperform the CEER score, since the CEER score only takes into account SAT/ACT and high school class rank. The goal is to help the Admissions Department to have a more complete view of the candidate before offering them admittance.

Our first steps are to assess the current metric: the CEER score. Is the score actually a good metric for CC success? If so, how can we improve it? Next, we look at specific parts of measured performance throughout USMAPS as determining factors for success. We look at the CAPS score (cumulative academic performance score) from USMAPS as a predictor of Plebe year CQPA. Additionally, we look at grades in both Math and English classes as predictor variables. Finally, we create a model that is the most predictive and the best model for the Admissions department to consider implamenting.

A. DEFINING SUCCESS

To conduct this research, we began by defining "success" at USMA. Since the head of the Admissions Department's priority is success during Freshman (Plebe) year, we will look specifically at the CQPA (Cumulative Quarter Point Average) at the end of Plebe year as the metric of success. In order to graduate, a cadet must have a CQPA of 2.0. Therefore, we make the assumption that if a cadet maintains a CQPA of 2.0 yearly, they will graduate. Thus, the definition of success is having a CQPA of 2.0 or higher at the end of Plebe year. This is contrary to the peg point CQPA of 1.67, because a cadet can still pass Plebe year with a 1.67, however a cadet cannot graduate without an overall CQPA of 2.00 or higher. Thus, the research team will specifically look at a CQPA of 2.00 or higher as the definition of success.

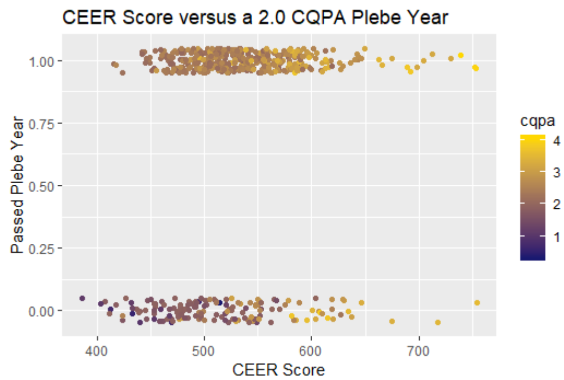

Figure 6. This figure shows whether or not a Plebe had a 2.0 GPA at the end of plebe year (1 is passing, 0 is failing) based on their CEER score. The color of the data point is the Plebe's CQPA at the end of Plebe Year.

B. PREDICTIVE POWER OF THE CEER SCORE

To start, we looked at measuring the CC CEER score as an effective predictor of CQPA. Figure 6 shows a graphic where each data point represents a CC. This graphic describes the issue with the CEER score. As we can see, there are many students that have a low CEER score that pass Plebe year, and many students with high CEER scores that do not pass Plebe year. Therefore, there is some inconsistency that we are working to solve. First, we formulated a linear model to evaluate the predictive power of the CEER score. When we first analyzed the predictive power of the CEER score as a linear model, we found the following relationship between CQPA and CEER score.

$$CQPA = 0.00581\ (CEER\ Score) - 0.634 \tag{IV-1}$$

We found that this model had a very low p-value for CEER Score, meaning that the CEER score is a significant predictor of success at USMA. However, the Adjusted R-squared value of this model is 0.392, which means that 39.2% of the variance is accounted for in this model, and that the model has a lot of room for improvement in predictive power. Additionally, when we look at a plot of whether or not a cadet passed Plebe year with a 2.0 CQPA or higher versus their CEER score (Figure 6), we find that many cadets with very low CEER scores still pass Plebe year. Therefore, evidence indicates that the CEER school is not a good metric for cadets that were CCs, considering many CCs with low CEER scores still pass Plebe year.

While a CC attends USMAPS, they retake their SAT and/or ACT to see if they can

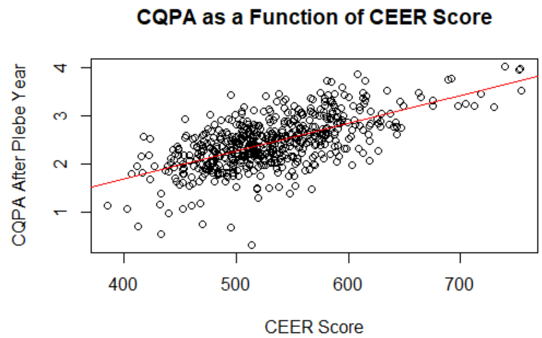

Figure 7. This plot shows the linear relationship between CQPA and CEER score for the USMAPS classes of 2014-2016.

improve theri scores. Thus, their CEER scores change after USMAPS to reflect them taking the SAT/ACT at USMAPS. However, these changes are so small that they don't necessarily show the improvements made during a year at USMAPS. This is shown in Figure 8 and 9. Thus, the CEER score is not a good representation of a student's likelihood of success at USMA.

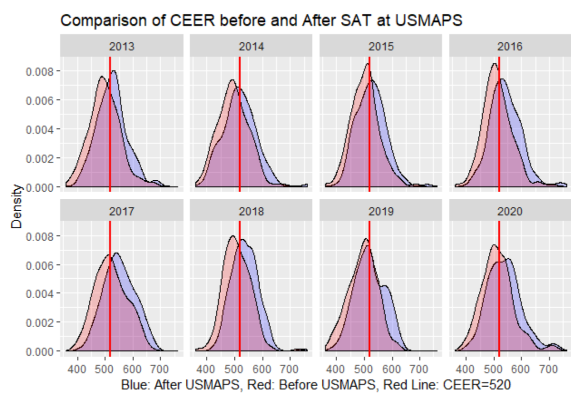

Figure 8. This plot shows a comparison of CEER scores before a student enters USMAPS, and their updated score based on their performance at USMAPS. Blue represents the original CEER score, red represents the updated score.

This problem is the basis of our research, and serves as the baseline model that we will compare our results to. It is important to consider that the current CEER score's Adjusted R-Squared value is 0.392, and any improvement will help the Admissions Department to better predict academic success through Plebe year for CCs. We know that

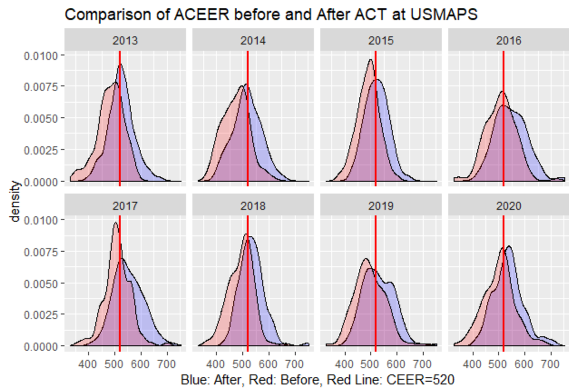

Figure 9. This plot shows a comparison of ACEER scores before a student enters USMAPS, and their updated score based on their performance at USMAPS. Blue represents the original ACEER score, red represents the updated score.

there is a way to get a better model. The question is, what inputs should be considered in this model to make it most useful?

C. CAPS AS A PREDICTOR

The next thing that we looked into is the predictive power of Cumulative Academic Performance Score (CAPS) that a CC receives at USMAPS. This is the USMAPS version of a GPA. We constructed a linear model to show this relationship with Equation IV-2 below.

$$CQPA = 0.891\, CAPS - 0.00477 \quad \text{(IV-2)}$$

This model is much more significant than that with the CEER score as the predictor of CQPA. It has an Adjusted R-squared value of 0.583, meaning it accounts for 58.3% of the variance. This is better than the model with the CEER score. CAPS has a p-value of $2 * 10^{-16}$, which is extremely low. This means that we reject the null-hypothesis that the coefficient value for CAPS is 0, supporting our alternative hypothesis that the beta value for CAPS is not 0.

This model is also helpful when we look at the intercept and slope. If we use this model to predict CQPA after Plebe year, we can see that if a CC had a 0 CAPS at USMAPS, they will achieve a “negative” CQPA, meaning there is almost no chance of success at USMA. Additionally, for every point earned in CAPS, there is almost a point mirrored in the CQPA, since the slope is so close to 1. We can see a plot of our data containing a trend line in Figure 10.

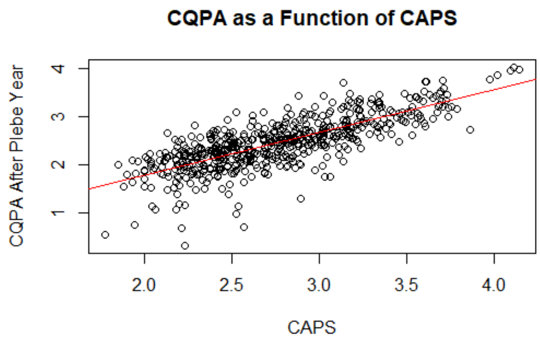

Figure 10. This shows the relatively linear relationship between CAPS and CQPA

D. COMBINING CEER AND CAPS

The next step in analyzing these models is to see what happens should we combine the CEER score and CAPS into one model. The multiple-linear relationship between CEER, CAPS, and CQPA is as follows:

$$CQPA = 0.736\, CAPS + 0.00180\, CEER - 0.534 \qquad \text{(IV-3)}$$

Both CAPS and CEER Score have extremely low p-values ($2*10^{-16}$ and $1.29*10^{-7}$ respectively). This means that they can both be considered significant in determining the CQPA of a CC after Plebe year. The Adjusted R-squared value of this model increases to 0.602, which shows the increasing improvement of our model. Already, we have increased the reliability of predicting CC success, simply by adding their CAPS at USMAPS to the model for predicting their CQPA.

1. Combining CAPS with Raw Data

The next step that we took is to use the raw data to predict CQPA using CEER Score and CAPS. To do this, we created a linear model of CQPA predicted by a candidate's SAT verbal score, SAT Math score, high school class rank (as a number, not a percentage), and CAPS, shown below. The model that is created has low p-values for only the SAT Math score (p-value of $3.19 * 10{-}6$) and CAPS (p-value of $2 - 10^{-16}$. The rest all have p-values above 0.05, suggesting that they are insignificant in predicting CQPA. The Adjusted R-

Squared value for this model is 0.596, which is not a terrible model, however it is less than the Adjusted R-Squared value using only CEER score and CAPS (0.602). The issue with this is that we lose granularity when we convert this data into the CEER score before using it in a model to predict CQPA. Conversely, if we use the raw data, we do not risk losing this granularity from converting the data. Therefore, for the rest of this paper we will be using raw data scores instead of normalized or adjusted scores.

$$\begin{aligned} CQPA = -0.531 + 0.0000502\ (SAT\ Verbal) + 0.00125\ (SAT\ Math) \\ -0.0000180\ (Class\ Rank) + 0.818\ (CAPS) \end{aligned} \tag{IV-4}$$

After this analysis, it is beneficial to look at another model where CQPA is predicted only by the factors that were found to be significant in the previous model. Specifically, a model where the SAT Math score and CAPS are the only predictors of CQPA. The model is below. It has an Adjusted R-Squared value of 0.597. This model still is not as good as the model with CEER score and CAPS as far as Adjusted R-Squared values are concerned.

$$CQPA = -0.501 + 0.00121\ (SAT\ Math\ Score) + 0.822\ (CAPS) \tag{IV-5}$$

E. MATH GRADES

Next we look at the grades that each student earned in their respective Math classes. USMAPS has many different classes that a student can take based on their demonstrated ability. These classes are Algebra & Trigonometry, AP Calculus, Pre-calculus, and Math Modeling/Intro to Differential Equations (USMA MA153). Students are initially placed in a class and then are moved if their skills either don't meet those needed for the class or well exceed the expectations for the it. Students are evaluated on their math ability twice a year, once before entering USMAPS and again after the first semester. They are then placed in an appropriate Math class based on their level of Mathematical expertise. In order to analyze this data, we need to figure out what to use. To do this, we first looked at averaging the grades across all four quarters as a predictor for CQPA.

When we looked at the linear relationship between the average Math grade and CQPA after Plebe year, we found that the Math grade average is significant, as it has a low p-value. The Adjusted R-squared value is 0.356, which is not indicative of a well-performing model. The model is depicted in Equation IV-6. To improve this model, we next look at individual quarter's math grades.

$$CQPA = 1.0522 + 0.516\,(Average\ Math\ Grade) \tag{IV-6}$$

In order to determine which quarter to use, we completed an interaction plot shown in Figure 11, which shows the correlation between Math grades across the quarters. What we expect is that if a student has an A+ in quarter one, they are likely to have an A+ in other quarters as well. This suggests a direct linear relationship, which is what we end up finding.

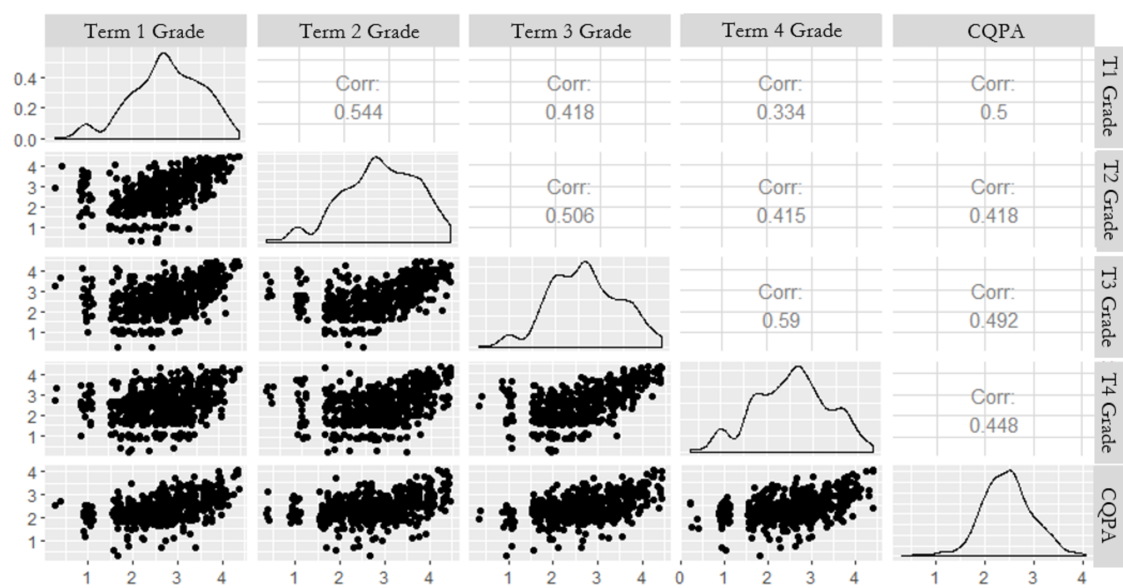

Figure 11. Interaction Plot of Math Grades per Quarter

To interpret this plot, we notice the relationships between the letter grade that a student earns in each quarter, as well as the relationship between the letter grade that a student earns per quarter compared to their CQPA after Plebe year. The correlation is represented in a value located in the grid squares opposite to the plots. We can see that over time, the letter grades seem to move more to the right, or towards higher grades, which makes sense given that the student is spending more time at USMAPS and in Math class and should improve over time. Additionally, we see a tighter grouping of data points, meaning more correlation and therefore more interaction between consecutive quarters. For example, the correlation between a student's letter grade in Quarter 1 and Quarter 2 is

0.573, which is higher than the 0.431 correlation between Quarter 1 and Quarter 3. This makes sense because the student is likely showing compounding improvements across these courses, which will cause the grades earned in consecutive quarters to be more similar.

When we look at the interaction with CQPA, we find that Quarter 1 has the highest correlation with CQPA. Thus, it would seem that the best predictor for CQPA might be Quarter 1. The next highest correlation is between Quarter 3 and CQPA, and then Quarter 4 and Quarter 2. This suggests that Quarter 1 grades will have the highest predictive power when it comes to CQPA. However, when we consider that each student takes different classes during each of these quarters, there might be different correlational power and value in the model. Therefore, we create a linear model for each quarter. They are depicted below. These models show a linear relationship where if a student took a specific class during a quarter, Algebra & Trigonometry, for example, then we would plug in the value for the grade that they earned in this class into the equation. The value for every other variable would be 0, since the student did not take these classes in that quarter. For example, if a student takes Algebra and Trigonometry in Quarter 1, and scores an A, we would plug a 4.00 in for $(Alg\&Trig)$ in the model, and a 0 for every other variable. The result would be the student's CQPA.

Quarter 1 : Adjusted R-Squared: 0.326

$$CQPA =$$
$$1.615 + 0.188\ (Alg\&Trig) + 0.388\ (APCalc) + 0.297\ (PreCalc) + 0.450\ (USMA153)$$

Quarter 2 : Adjusted R-Squared: 0.371

$$CQPA =$$
$$1.733 + 0.117\ (Alg\&Trig) + 0.280\ (PreCalc) + 0.409\ (APCalc) + 0.411\ (USMA153)$$

Quarter 3 : Adjusted R-Squared = 0.400

$$CQPA =$$
$$1.548 + 0.197\ (Alg\&Trig) + 0.377\ (PreCalc) + 0.438\ (APCalc) + 0.439\ (USMA153)$$

Quarter 4 : Adjusted R-Squared: 0.352

$$CQPA =$$
$$1.786 + 0.117\ (Alg\&Trig) + 0.361\ (APCalc) + 0.318\ (PreCalc) + 0.409\ (USMA153)$$

These results make sense, because the highest beta values are associated with the hardest classes. This means that a student who earns an A in a higher level Math class will have a higher CQPA than a student who earns an A in a lower level Math class. Thus, an A earned from taking USMA 153 at USMAPS will grant a student a larger CQPA than an A earned from taking AP Calculus at USMAPS.

Looking at these results, Quarter 3 holds the most predictive power of CQPA after Plebe year, since it has the highest Adjusted R-Squared value. This differs from our original analysis that Quarter 1 would be the best predictor. There are many reasons that this could be. Quarter 1 is a student's first exposure to collegiate-level classes. For many, it is the first time away from home. Additionally, it will be a CC's introduction into the military lifestyle, which also takes time to get used to. It is also not a necessarily comparable experience to Plebe year, since during Plebe year a Prespter will have had time and experience not only with USMA but also with the combination of a military and collegiate life-style. One other factor that affects this is the Football and Soccer seasons, which start immediately at the beginning of the semester. More than half of the population at USMAPS is part of the Football or Soccer teams, and these students are not only adjusting to a military lifestyle, but are adjusting to the lifestyle of a collegiate athlete as well. Because of these discrepancies, we will not use Quarter 1 as our predictor of CQPA and will instead use Quarter 3. This has the second-highest correlation, the highest Adjusted R-Squared value, and it also makes the most sense. At this point, a student has experienced two quarters of collegiate level classes, and has gotten used to living away from home in a military setting.

It is important to note at this point that the Adjusted R-Squared value from one quarter of Math grades, 0.400, is higher than the Adjusted R-Squared value from the CEER score for CCs (0.392). This is more evidence that the CEER score is not a good metric for CC success at USMA, and a new metric should be used to help the Admissions Department make decisions.

F. ENGLISH GRADES

After completing the analysis of Math grades by term, we looked into English grades as a predictor of CQPA. English grades differ from Math grades because in English, there are only two different classes a student can take: Standard English and Honors English. All students begin in Standard English for the first two quarters, and those that are at the skill level necessary move to Honors English beginning in Quarter 3. Once a student is in Honors English for Quarter 3, they cannot move to Standard English in the next quarter, and the same can be said if a student starts in Standard English.

Similar to the Math Grade data, we started by looking at the average grade a student earned across all four quarters as a linear predictor of CQPA (see equation below). Our p-value for this model is $2 * 10^{-16}$, showing its significance. The Adjusted R-Squared value

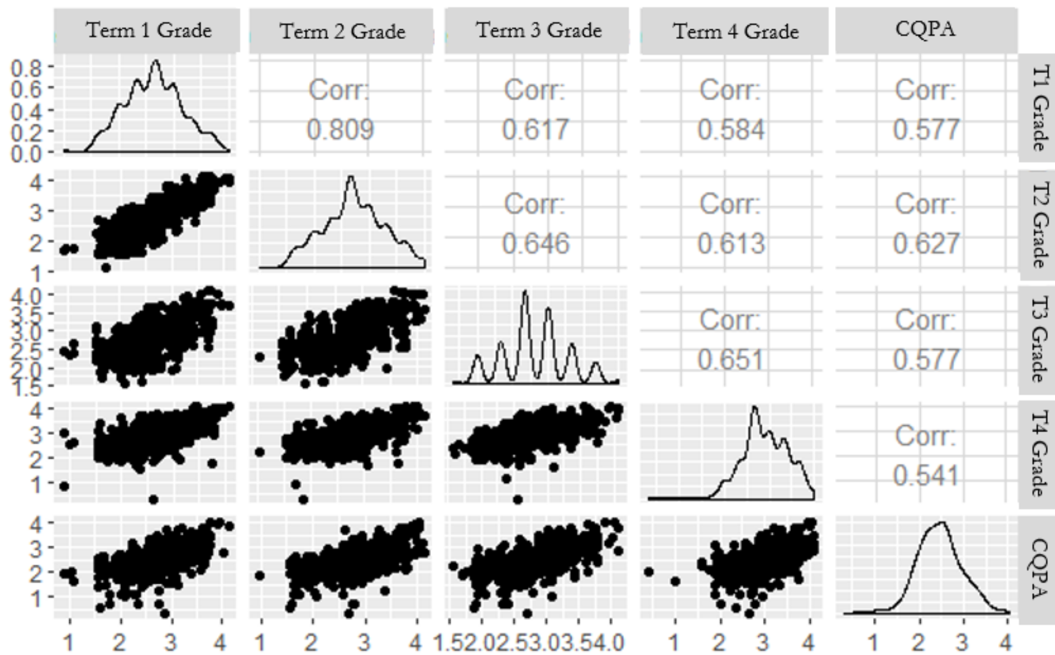

Figure 12. Interaction Plot of English Grades per Quarter

for this model is 0.455, meaning that 45.5% of the variance in CQPA is accounted for in this model. This shows that using average English grade, we account for almost exactly the amount of variance that is accounted for using the average Math grade (Adjusted R-Squared of 0.289).

$$CQPA = 0.226 + 0.802\ (Average\ English\ Grade) \tag{IV-7}$$

What is interesting about this model is that a student seems to gain more from having a higher average English grade in terms of their CQPA. Having a higher average English grade will add 0.802 to a student's CQPA, where as having a higher average Math grade will only earn a student 0.516 in their CQPA. This suggests that if a student does better in English than in Math at USMAPS, they will have a better CQPA at USMA.

Now we look at different quarters of English grades as predictors of CQPA. We began by constructing another interaction plot similar to the one that we constructed for Math grades. The interaction plot is shown in Figure 12. As we can see, Quarter 2 has the highest correlation with CQPA, with a correlational value of 0.627.

The next step is to look at a linear model for each quarter. For Quarters 1 and 2, we will use a standard linear model, but for Quarters 3 and 4, we will use the interaction between a letter grade earned and the class that a student took (Honors or Standard English). The models are shown below:

Quarter 1 : Adjusted R-Squared: 0.332

$$CQPA = 0.948 + 0.572\ (Term\ 1\ Grade)$$

Quarter 2 : Adjusted R-Squared: 0.391

$$CQPA = 0.840 + 0.588\ (Term\ 2\ Grade)$$

Quarter 3 : Adjusted R-Squared = 0.332

$$CQPA = 0.944 + 0.520\ (Standard\ Grade) + 0.605\ (Honors\ Grade)$$

Quarter 4 : Adjusted R-Squared: 0.304

$$CQPA = 1.139 + 0.434\ (Standard\ Grade) + 0.524\ (Honors\ Grade)$$

As we expect, in Quarters 3 and 4, CQPA increases when a student takes the more challenging class - Honors English. However, we also find that the most predictive quarter for CQPA predicted by English grades is Quarter 2. This is contrary to the conclusion from our Math grade data. Additionally, the model from Quarter 2 is significantly less predictive than the Math model (Adjusted R-Squared 0.400).

This information does not mean that English is any less important than Math at USMAPS, however it does beg the question of why Math grades hold more predictive power than English grades. One reason might be because at USMA, the Math courses that students take during Plebe year are worth 4.5 credit hours, whereas English courses are only worth 3.0 credit hours. Therefore students that excel in Math will be more rewarded in their CQPA by higher grades than students that excel in English.

G. CREATING A FINAL MODEL

To begin creating a final model, we used R™ to test many different factors and their predictive power over CQPA as a response variable. Figure 13 shows the different models and their associated Adjusted R-Squared values. The predictor variables in this model were chosen based on the methods discussed above, as well as an analysis of the p-values associated with each variable in each model. The model that we will choose to recommend is highlighted in yellow. It includes the CAPS, SAT Math score, and SAT Verbal score. It is the most valuable model because it captures a USMAPS student's performance at USMAPS (via CAPS), as well as their high school performance (via SAT scores). The reason that we accept this model as the best is because it captures High School and USMAPS performance, and allows for an ease of transition for the Admissions Department, who is looking for a score comparable to the original CEER score in terms of predictors.

Model	Adjusted R - squared
CQPA ~ Q3 Math Grade : Q3 Math Class + Q3 Eng Grade: Q3 Eng Class + CAPS + SAT Math + SAT Verb	0.578
CQPA ~ SAT Verbal + SAT Math + Q3 Math Grade : Q3 Math Class + Q3 Eng Class : Q3 Eng Grade	0.511
CQPA ~ CAPS + Q3 Math Grade : Q3 Math Class + Q3 Eng Grade : Q3 Eng Class	0.508
CQPA ~ Q3 Math Grade + Q3 Eng Grade	0.42
CQPA ~ CAPS + Average Math Grade + Average English Grade	0.613
CQPA ~ CAPS + Q3 Math Grade + Q3 Eng Grade + SAT Verbal + SAT Writing	0.593
CQPA ~ Average Math Grade + CAPS + SAT Verbal	0.3936
CQPA ~ SAT Math + SAT Verbal	0.607
CQPA ~ CAPS + SAT Math + SAT Verbal	0.597

Figure 13. Models and Associated Adjusted R-Squared Values

In this analysis, we noted that there is a correlational relationship between SAT Verbal and Writing scores. We can see this further in Figure 14. This plot shows the relationships between all aspects of SAT. SAT Writing and Verbal scores have a correlation of 0.574, which is very high. This means that there is colinearity between these two predictor variables. This detracts from the interpretability of the model and therefore will not be useful for this research. The SAT Verbal score was chosen to include in the PCEER because it maximized the Adjusted R-Squared value. Additionally, research has shown that the SAT Writing score is not necessarily a good predictor of success at USMA. It might be a good indicator of how a student commands the English language, but should not be used to predict overall academic success (Conger). Therefore, we will not use the SAT Writing score in our final model and will use the SAT verbal score instead. It is important to note that while the writing score is not necessarily something that should be used in the CEER score, it still sheds light on a student's academic ability and will not necessarily be detrimental to include in an admissions packet.

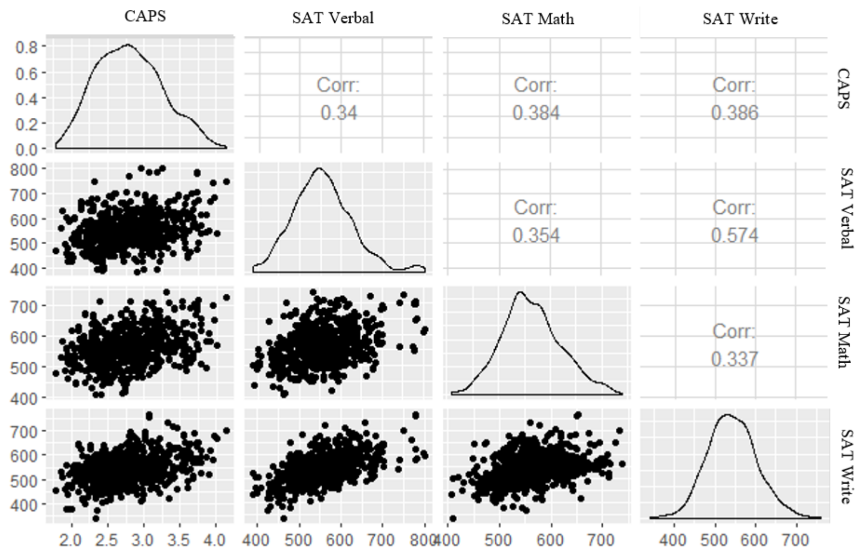

Figure 14. Correlation Plot for SAT

When we look at this model, we can see that it is useful, but we can truly see this usefulness when we look at a histogram of predicting CQPA to see how well we predicted (see Figure 16). This histogram shows the difference between actual and predicted CQPA. It is centered at 0, representing a 0 point change between predicted and actual CQPA. Each bar represents 0.33 in CQPA, or the difference between one letter grade. For example, if a student earned a B as their CQPA after plebe year, but we predicted that they would earn a B+, they would fall in the bar immediately to the left of 0, meaning that their CQPA was 0.33 less than what we predicted. As we can see, the majority of data points falls within 0.33 of the center. Precisely 72.759% of the data points fall in the middle two bars (shown in red on the histogram), meaning that with the PCEER, we can predict a student's CQPA after plebe year within one letter grade 72.759% of the time.

Once we came to this highlighted model noted above, we constructed a similar model using ACT scores. The difference here is that the ACT has test components that do not necessarily mirror the SAT. Specifically, the Science and Reasoning portion of the ACT is not reflected in the SAT. To look at the interaction between the pieces of the ACT, we created another correlation plot, shown in Figure 17. We can see from this plot that there is a high correlation between ACT English, ACT Writing, and ACT Reading. This is to be expected since all of these scores reflect similar parts of the test. In order to more consistently reflect the SAT model, the ACT model will use CAPS, ACT Math, ACT

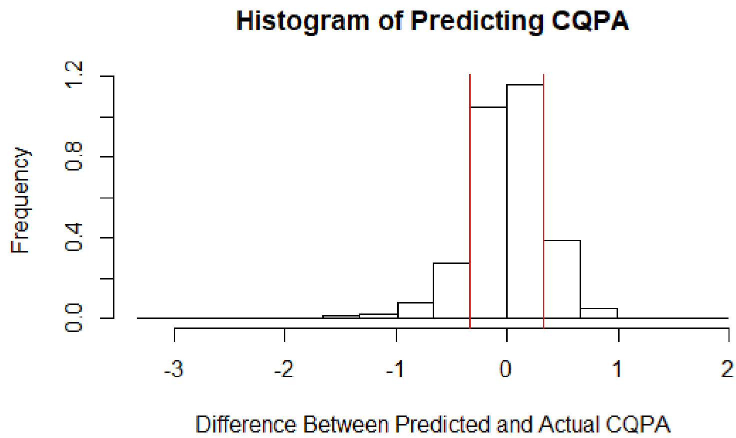

Figure 15. Predicting CQPA using SAT

Reading, and the ACT Science and Reasoning scores.

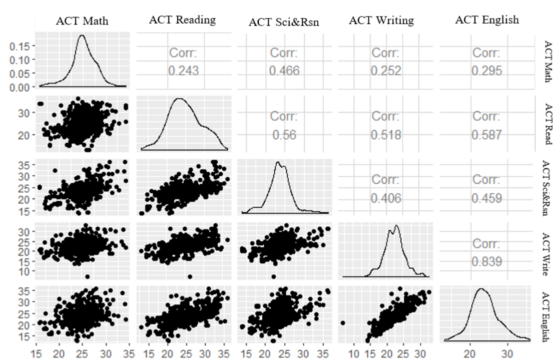

Figure 16. Correlation Plot for ACT

We can do a similar analysis of this model by plotting the actual versus predicted CQPA using this model. This plot and its accompanying histogram are shown below. With this model, we can predict a student's CQPA within one letter grade 72.408% of the time.

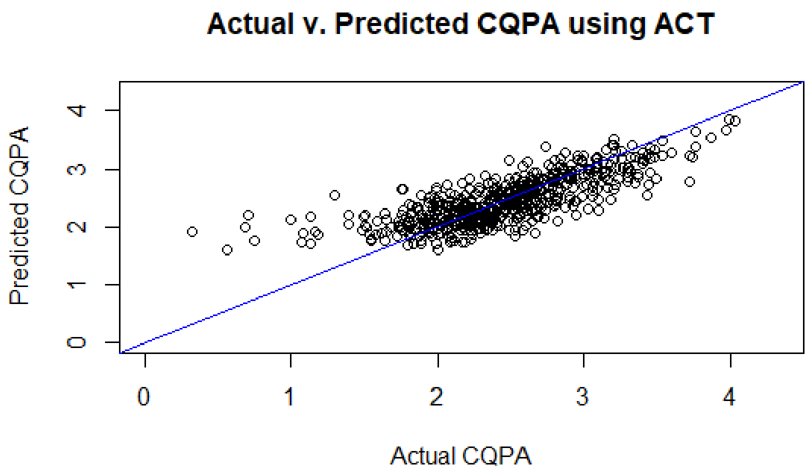

Figure 17. Actual versus Predicted CQPA using ACT

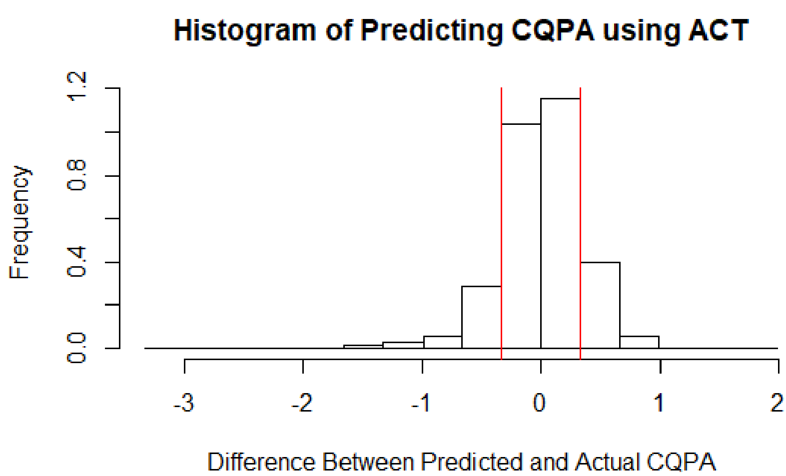

Figure 18. Histogram of Predicting CQPA using ACT

THIS PAGE INTENTIONALLY LEFT BLANK

V. DISCUSSION

The best model incorporates a student's CAPS, SAT Math score, and SAT Verbal score (see below). This model was achieved by testing which factors increased the Adjusted R-Squared value in a linear model that predicts CQPA while keeping the factors significant (i.e. having a low p-value). The Adjusted R-Squared value for this model is 0.597 which is a 57.19% increase from the original CEER score. This means the model accounts for approximately 35% more variance than the original CEER score for CCs.

$$CQPA = -0.534 + (0.819 * CAPS) + 0.00125(SAT\ Math) + (0.0000495 * SAT\ Verb) \tag{V-1}$$

The PCEER equation was created by normalizing the linear equation that predicts CQPA to represent a set of data with a mean of 600 and a maximum of 800, to accurately reflect the original range of the CEER score. The PCEER is shown below. In this model, it is possible to exceed a score of 800, but this is the case in the original model as well and is not a concern for this research.

$$PCEER = 159.0475 + (0.184 * SAT\ Math) + (0.00731 * SAT\ Verbal) + (120.977 * CAPS) \tag{V-2}$$

Once we have this model, we can look at how much of a student's CQPA after plebe year is predicted by each component of the PCEER. See the chart below for a visual representation. We can see that the majority of the PCEER is accounted for by CAPS from USMAPS. The second biggest contributor to the PCEER is the SAT Math Score, and the third and smallest contributor is the SAT Verbal Score. This shows that the time that a student spends at USMAPS is a better metric for success, which again identifies a problem with using the normal CEER score to measure CCs in terms of predicting academic success and performance.

With this model for the PCEER, we can predict a student's CQPA during plebe year. A plot of their predicted CQPA versus their actual CQPA is depicted below. This shows a trend of a generally 45 degree angle, which signifies the accuracy of our model. Most of the variability in the model comes from students who we predicted would have

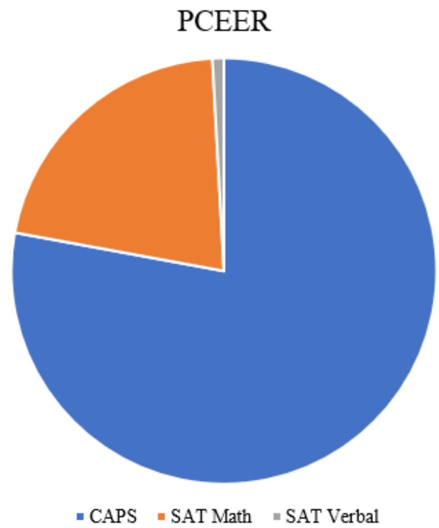

Figure 19. Variable Contributions to PCEER

a passing CQPA around a 2.00, but end up having a much lower CQPA after Plebe year. When we look closely at these data points, it becomes apparent that all of these data points with a CQPA below 1.2 that fall in this group of outliers represent students that were in the bottom half of their class at USMAPS. Specifically, all of these students have a USMAPS class rank of below 90. This is significant because it can help the Admissions department determine which CCs are more at-risk. The CCs that we predict would have an average CQPA that might actually have a low CQPA are likely to be in the bottom half of their class at USMAPS.

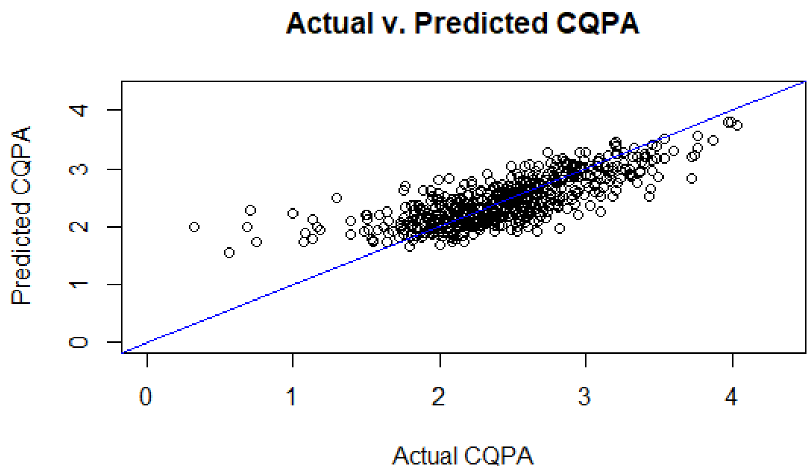

Figure 20. Plot of Actual versus Predicted CQPA

Similarly, if a student took the ACT, the model with the most predictive power includes CAPS, ACT Math score, ACT Reading score, and ACT Science and Reasoning score (see below). This model has an Adjusted R-squared value of 0.5015, meaning that it accounts for 50.15% of the variance when predicting CQPA.

$$CQPA = -0.482 + (0.810 * CAPS) + (0.0141 * ACT\ Math) + (0.00448 * ACT\ Reading) + (0.0101 * ACT\ Sci\ Rsn) \tag{V-3}$$

Like for the SAT, a PCEER score was created from this model with normalized values: a mean of 600 and a maximum of 800. This equation is the PCEERa shown below. A pie-chart showing the contributions of each component of the PCEERa is also shown below. Similar to the PCEER, most of the PCEERa is predicted by CAPS.

$$PCEERa = 169.482 + (114.944 * CAPS) + (2.00745 * ACT\ Math) + (0.635 * ACT\ Reading) + (1.434 * ACT\ Sci\&\ Rsn) \tag{V-4}$$

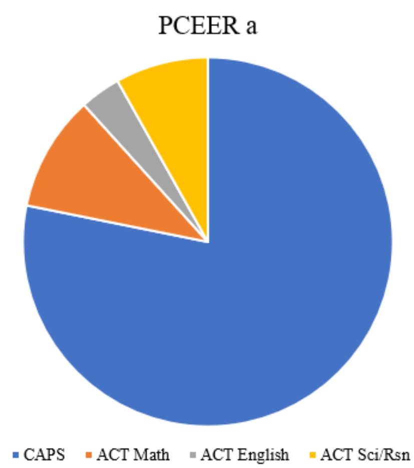

Figure 21. Variable Contributions to PCEERa

VI. CONCLUSION AND RECOMMENDATION FOR FUTURE WORK

This research sheds significant light on the admissions process to USMA from USMAPS and suggests just a few ways to improve the process. The current CEER score is not a good representation of CC performance and should not be used as a metric to assess academic capabilities. Generally speaking, most any model that includes a CC's performance at USMAPS will be better than the current CEER score at predicting success at USMA. The model that this research team created is just one way to improve the current CEER score. It is a simple model with only three inputs that captures a student's academic performance both before and during USMAPS, both of which are important considerations when applying to USMA. Future work should test this model on future classes at USMA to see how it performs. Then, a new risk threshold can be created to help the admissions department better assess candidate files.

Following this research should be a more in-depth analysis on the effects of non-academic related predictors of success. For example, a student's participation in sports, their faculty appraisal score, their military or character score at USMAPS, etc. These might allow the researcher to better capture a candidate's "grit" and resilience.

A more in-depth analysis on how CCs perform at USMA would also be beneficial. How many fail after they get through Plebe year? How many decide to leave USMA or are separated from USMA? How many graduate at the top of their class? How well to students from USMAPS measure up to students from other military preparatory schools such as Marion Military Institute or Valley Forge Military Academy? Research regarding overall CC performance at USMA would be very interesting to see and would shed light on how well Prespters perform throughout their entire time at USMA.

THIS PAGE INTENTIONALLY LEFT BLANK

VII. WORKS CITED

Adjusted R2 / Adjusted R – Squared : What is it used for?. Statistics How To, 2019. Web. 21 March 2019.

Class of 2020 Calculations. Diagram. United States Military Academy. 03 October 2018. Comptroller General of the United States. Student Attrition at the Five Federal Service Academies. Department of Defense, Commerce, and Transportation, 1976. Web.

Conger, Hailey. *Predicting Academic Potential at West Point using SAT Scores and High School Class Rank Scores*. 2019. United States Military Academy. Undergraduate Thesis.

Jebb, Joel. Personal Interview. 26 September 2018.

Jebb, Joel. USMAPS English: Needless Detour or Pathway to Success? Diss. Columbia University, 2016. Web.

Lane, David M. "Introduction to Linear Regression." *Introduction to Statistics*. PDF file.

McDonald, Deborah. Personal Interview. 13 September 2018. New Admissions Committee Training. Slides briefed to Admissions Department. 2016-2017. Microsoft Power-Point File.

New Admissions Committee Training. United States Military Academy, 2016. Web.

Office of the Dean. Academic Program Class of 2019: Curriculum and Course Descriptions. West Point. 2016. Web.

Office of the Director of Institutional Research. Performance of USMAPS Graduates at USMAPS, USMA, and as Army Officers. West Point. 1975. Web.

Overview Brief to COL Higgins USMAPS. Slides briefed to Commandant of USMAPS. 21 NOV 2017. Microsoft Power-Point File.

"The Road to USMA R-Day as of 03 AUG 16." Slides for USMAPS Cadets. 03 AUG 16. Microsoft Power-Point File.

usmapsData. Retrieved from USMA Office of Institutional Research. Microsoft Excel File.

USMAPS (USMA) Website. Accessed on numerous occasions for all information cited as USMAPS.

USMA Website. Accessed on numerous occasions for all information cited as USMA.

Wasserstein, Ronald L. & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108

Wong-Dodge, Melanie. Personal Interview. 29 November 2018.

VIII. APPENDIX: CODE

```
library(tidyverse)
library(dplyr)
library(GGally)
setwd('C:/Users/x97354/Documents/2019/Thesis')

#Data
data <- read_csv("ConsolidatedData.csv")
data2 <- read_csv("usmapsdata2.csv")
data2 <- data2 %>%
  select(CDTID,english_term1_crse_descr,english_term2_crse_descr,english_term3_crse_descr,
        english_term4_crse_descr,math_term1_crse_descr,math_term2_crse_descr,math_term3_crse_descr,
        math_term4_crse_descr)
usmapsdat <- data %>%
  filter(usmaps_rank != "NA") %>% #Filters the data to include only Prepsters
  filter(caps > 0) %>%  #filters out all students that did not complete Prep School
  filter(cqpa >0) %>%  #filters out all students that did not complete Plebe year
  filter(apply_clyr>=2019) %>%
  filter(SAT_M_RAW>0,SAT_V_RAW>0) %>%
  filter(act_math>0, act_read>0,act_sci_rsn>0) %>%
  mutate(atrisk <- ifelse(ceer_scr<=520,1,0)) %>%
  left_join(data2,by = c("CDT_ID"="CDTID"))

usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "A+"]<-4.33
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "A"]<-4.00
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "A-"]<-3.67
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "B+"]<-3.33
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "B"]<-3.00
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "B-"]<-2.67
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "C+"]<-2.33
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "C"]<-2.00
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "C-"]<-1.67
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "D+"]<-1.33
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "D"]<-1.00
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "D-"]<-0.67
usmapsdat$math_term1_ltr_grd[usmapsdat$math_term1_ltr_grd == "F"]<-0.33

usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "A+"]<-4.33
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "A"]<-4.00
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "A-"]<-3.67
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "B+"]<-3.33
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "B"]<-3.00
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "B-"]<-2.67
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "C+"]<-2.33
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "C"]<-2.00
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "C-"]<-1.67
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "D+"]<-1.33
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "D"]<-1.00
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "D-"]<-0.67
usmapsdat$math_term2_ltr_grd[usmapsdat$math_term2_ltr_grd == "F"]<-0.33

usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "A+"]<-4.33
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "A"]<-4.00
```

```
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "A-"]<-3.67
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "B+"]<-3.33
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "B"]<-3.00
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "B-"]<-2.67
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "C+"]<-2.33
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "C"]<-2.00
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "C-"]<-1.67
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "D+"]<-1.33
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "D"]<-1.00
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "D-"]<-0.67
usmapsdat$math_term3_ltr_grd[usmapsdat$math_term3_ltr_grd == "F"]<-0.33

usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "A+"]<-4.33
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "A"]<-4.00
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "A-"]<-3.67
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "B+"]<-3.33
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "B"]<-3.00
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "B-"]<-2.67
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "C+"]<-2.33
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "C"]<-2.00
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "C-"]<-1.67
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "D+"]<-1.33
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "D"]<-1.00
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "D-"]<-0.67
usmapsdat$math_term4_ltr_grd[usmapsdat$math_term4_ltr_grd == "F"]<-0.33

usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "A+"]<-4.33
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "A"]<-4.00
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "A-"]<-3.67
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "B+"]<-3.33
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "B"]<-3.00
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "B-"]<-2.67
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "C+"]<-2.33
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "C"]<-2.00
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "C-"]<-1.67
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "D+"]<-1.33
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "D"]<-1.00
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "D-"]<-0.67
usmapsdat$english_term1_ltr_grd[usmapsdat$english_term1_ltr_grd == "F"]<-0.33

usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "A+"]<-4.33
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "A"]<-4.00
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "A-"]<-3.67
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "B+"]<-3.33
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "B"]<-3.00
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "B-"]<-2.67
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "C+"]<-2.33
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "C"]<-2.00
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "C-"]<-1.67
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "D+"]<-1.33
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "D"]<-1.00
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "D-"]<-0.67
```

```
usmapsdat$english_term2_ltr_grd[usmapsdat$english_term2_ltr_grd == "F"]<-0.33

usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "A+"]<-4.33
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "A"]<-4.00
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "A-"]<-3.67
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "B+"]<-3.33
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "B"]<-3.00
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "B-"]<-2.67
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "C+"]<-2.33
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "C"]<-2.00
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "C-"]<-1.67
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "D+"]<-1.33
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "D"]<-1.00
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "D-"]<-0.67
usmapsdat$english_term3_ltr_grd[usmapsdat$english_term3_ltr_grd == "F"]<-0.33

usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "A+"]<-4.33
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "A"]<-4.00
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "A-"]<-3.67
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "B+"]<-3.33
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "B"]<-3.00
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "B-"]<-2.67
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "C+"]<-2.33
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "C"]<-2.00
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "C-"]<-1.67
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "D+"]<-1.33
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "D"]<-1.00
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "D-"]<-0.67
usmapsdat$english_term4_ltr_grd[usmapsdat$english_term4_ltr_grd == "F"]<-0.33

usmapsdat$math_term1_ltr_grd<-as.numeric(as.character(usmapsdat$math_term1_ltr_grd))
usmapsdat$math_term2_ltr_grd<-as.numeric(as.character(usmapsdat$math_term2_ltr_grd))
usmapsdat$math_term3_ltr_grd<-as.numeric(as.character(usmapsdat$math_term3_ltr_grd))
usmapsdat$math_term4_ltr_grd<-as.numeric(as.character(usmapsdat$math_term4_ltr_grd))
usmapsdat$english_term1_ltr_grd<-as.numeric(as.character(usmapsdat$english_term1_ltr_grd))
usmapsdat$english_term2_ltr_grd<-as.numeric(as.character(usmapsdat$english_term2_ltr_grd))
usmapsdat$english_term3_ltr_grd<-as.numeric(as.character(usmapsdat$english_term3_ltr_grd))
usmapsdat$english_term4_ltr_grd<-as.numeric(as.character(usmapsdat$english_term4_ltr_grd))

usmapsdat<- usmapsdat %>% rowwise() %>%
  mutate(mathgrade =
mean(c(math_term1_ltr_grd,math_term2_ltr_grd,math_term3_ltr_grd,math_term4_ltr_grd))) %>%
  mutate(englishgrade =
mean(c(english_term1_ltr_grd,english_term2_ltr_grd,english_term3_ltr_grd,english_term4_ltr_grd)))



#Evaluation of the CEER Score
ceermod <- lm(cqpa~ceer_scr,data = usmapsdat)
summary(ceermod)

plot(usmapsdat$ceer_scr, usmapsdat$cqpa,
```

```
    xlab = "CEER Score", ylab = "CQPA After Plebe Year",
    main = "CQPA as a Function of CEER Score")
abline(ceermod,col = "red")

options(scipen = 999)
Pass=c(seq(2.000,4.300,0.001))
usmapsdat<-usmapsdat %>%
  mutate(Passed=ifelse(cqpa%in%Pass,1,0))
usmapsdat%>%
  ggplot(aes(x=ceer_scr,y=Passed,color=cqpa))+geom_jitter(height = 0.05)+
  labs(x = "CEER Score", y = "Passed Plebe Year", title = "CEER Score versus a 2.0 CQPA Plebe Year") +
  scale_color_continuous(low = "midnightblue", high = "gold")


#CAPS as a predictor
capsmod <- lm(cqpa~caps, data = usmapsdat)
summary(capsmod)

plot(usmapsdat$caps,usmapsdat$cqpa,xlab = "CAPS", ylab = "CQPA After Plebe Year",
    main = "CQPA as a Function of CAPS")
abline(capsmod, col = "red")

ceercapsmod <- lm(cqpa~caps+ceer_scr, data = usmapsdat)
summary(ceercapsmod)

ceerrawcapsmod <- lm(cqpa~SAT_V_RAW+SAT_M_RAW+class_rank+caps,data = usmapsdat)
summary(ceerrawcapsmod)

strongvars <- lm(cqpa~SAT_M_RAW+caps,data = usmapsdat)
summary(strongvars)


#Math Grades
avgmath = lm(cqpa~mathgrade,data = usmapsdat)
summary(avgmath)

usmapsdat%>%
  select(math_term1_ltr_grd,math_term2_ltr_grd,math_term3_ltr_grd,math_term4_ltr_grd,cqpa) %>%
  ggpairs(lower = list(continuous=wrap("points", position="jitter")))#NEED TO PUT IN PAPER

term1math = lm(cqpa~math_term1_crse_descr:math_term1_ltr_grd,data = usmapsdat)
summary(term1math)

term2math = lm(cqpa~math_term2_crse_descr:math_term2_ltr_grd,data = usmapsdat)
summary(term2math)

term3math = lm(cqpa~math_term3_crse_descr:math_term3_ltr_grd,data = usmapsdat)
summary(term3math)

term4math = lm(cqpa~math_term4_crse_descr:math_term4_ltr_grd,data = usmapsdat)
summary(term4math)
```

```
#English Grades
avgenglish = lm(cqpa~englishgrade,data = usmapsdat)
summary(avgenglish)

usmapsdat%>%
  select(english_term1_ltr_grd,english_term2_ltr_grd,english_term3_ltr_grd,english_term4_ltr_grd,cqpa)
%>%
  ggpairs(lower = list(continuous=wrap("points", position="jitter")))

term1eng = lm(cqpa~english_term1_ltr_grd,data = usmapsdat)
#no interaction because students only take standard english during 1st semester
summary(term1eng)

term2eng = lm(cqpa~english_term2_ltr_grd,data = usmapsdat)
summary(term2eng)

term3eng = lm(cqpa~english_term3_ltr_grd:english_term3_crse_descr, data = usmapsdat)
summary(term3eng)

term4eng = lm(cqpa~english_term4_crse_descr:english_term4_ltr_grd,data = usmapsdat)
summary(term4eng)


#Final SAT Model
SATmodel = lm(cqpa ~ caps+SAT_M_RAW+SAT_V_RAW,data=usmapsdat)
summary(SATmodel)

pred = data.frame(cqpaPred=predict(SATmodel,newdata=usmapsdat))
mS = 200/(max(pred$cqpaPred)-mean(pred$cqpaPred))
bS = 600-mS*mean(pred$cqpaPred)
satCoefs = mS*coef(SATmodel)
satCoefs[1]=satCoefs[1]+bS
usmapsdat <- usmapsdat %>% # Add a colunm that has the PCEER for ACT
  mutate(PCEER_S =
satCoefs[1]+satCoefs[2]*caps+satCoefs[3]*SAT_M_RAW+satCoefs[4]*SAT_V_RAW) %>%
  rowwise %>% mutate(predCQPA =
coef(SATmodel)[1]+coef(SATmodel)[2]*caps+coef(SATmodel)[3]*SAT_M_RAW+coef(SATmodel)[4]*SAT
_V_RAW) %>%
  mutate(diffPred = cqpa-predCQPA) %>%
  mutate(within_ltr_grd = ifelse(abs(diffPred)<=0.33,1,0))

count(usmapsdat,within_ltr_grd)
414/nrow(usmapsdat)

plot(x = usmapsdat$cqpa,y = usmapsdat$predCQPA,xlim=c(0,4.33),ylim=c(0,4.33),main="Actual v.
Predicted CQPA",
     ylab = "Predicted CQPA",
     xlab = "Actual CQPA")
fit = lm(cqpa~predCQPA, data = usmapsdat)
abline(a=coef(fit)[1], b=coef(fit)[2],col="blue")
```

```
hist(usmapsdat$diffPred, breaks = c(-3.33,-2.99,-2.66,-2.33,-1.99,-1.66,-1.33,-0.99,-0.66,-
0.33,0,0.33,0.66,0.99,1.33,1.66,1.99),
     main = "Histogram of Predicting CQPA",xlab = "Difference Between Predicted and Actual CQPA", ylab
= "Frequency")
abline(v=c(-0.33,0.33),col = "red")


#Final ACT Model
ACTmodel = lm(cqpa~caps+act_math+act_read+act_sci_rsn,data = usmapsdat)
summary(ACTmodel)

predACT = data.frame(cqpaPredACT=predict(ACTmodel,newdata=usmapsdat))
mSACT = 200/(max(predACT$cqpaPredACT)-mean(predACT$cqpaPredACT))
bSACT = 600-mS*mean(predACT$cqpaPredACT)
ACTCoefs = mSACT*coef(ACTmodel)
ACTCoefs[1]=ACTCoefs[1]+bSACT
usmapsdat <- usmapsdat %>% # Add a colunm that has the PCEER for ACT
  mutate(PCEER_A =
ACTCoefs[1]+ACTCoefs[2]*caps+ACTCoefs[3]*act_math+ACTCoefs[4]*act_read+ACTCoefs[5]*act_sci_r
sn) %>%
  rowwise %>% mutate(predCQPAACT =
coef(ACTmodel)[1]+coef(ACTmodel)[2]*caps+coef(ACTmodel)[3]*act_math+coef(ACTmodel)[4]*act_read
+coef(ACTmodel)[5]*act_sci_rsn) %>%
  mutate(diffPredACT = cqpa-predCQPAACT) %>%
  mutate(within_ltr_grdACT = ifelse(abs(diffPredACT)<=0.33,1,0))

count(usmapsdat,within_ltr_grdACT)
412/nrow(usmapsdat)

plot(x = usmapsdat$cqpa,y = usmapsdat$predCQPAACT,xlim=c(0,4.33),ylim=c(0,4.33),main="Actual v.
Predicted CQPA using ACT",
     ylab = "Predicted CQPA",
     xlab = "Actual CQPA")
fit = lm(cqpa~predCQPAACT, data = usmapsdat)
abline(a=coef(fit)[1], b=coef(fit)[2],col="blue")

hist(usmapsdat$diffPredACT, breaks = c(-3.33,-2.99,-2.66,-2.33,-1.99,-1.66,-1.33,-0.99,-0.66,-
0.33,0,0.33,0.66,0.99,1.33,1.66,1.99),
     main = "Histogram of Predicting CQPA using ACT",xlab = "Difference Between Predicted and Actual
CQPA", ylab = "Frequency")
abline(v=c(-0.33,0.33),col = "red")
```