UNITED STATES MILITARY ACADEMY

WEST POINT, NEW YORK

HONORS THESIS

MODIFYING THE ACEER

by

CDT James C. Palmer

May 2019

Thesis Advisor:	COL Joseph Lindquist
Thesis Advisor:	LTC Kristin Arney
Thesis Advisor:	MAJ Bryan Adams

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE May 2019	3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE Modifying the ACEER 6. AUTHOR(S) CDT James C. Palmer	5. FUNDING NUMBERS
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 Department of Admissions	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)

Each year, thousands of hopeful candidates begin the United States Military Academy's tedious application process. The Department of Admissions possesses the daunting task of selecting the deserving candidates who will gain admission into the Corps of Cadets. As a tool to assess a candidate's academic potential, the Department of Admissions uses an algorithm, formally known as the ACEER, which includes high school class rank and ACT scores for Math, Reading, English, and Science Reasoning. The current ACEER has not been updated since 1992. In the last twenty-seven years, the ACT and the high school curriculum have changed dramatically. Consequently, multiple stakeholders are concerned about the validity of the ACEER. The purpose of this report is to recreate and modify the current ACEER to ensure the USMA Admissions Department is admitting candidates who have the highest potential for academic success. This study conducts thorough research about the USMA Admissions process, Title X of US Code, and implementation of the ACT. With up to 40% of high schools not reporting class rank, the Admissions Department is relying heavily on standardized tests to predict the success of candidates. This study searches for short-term and long-term predictors of academic success through methods of linear regression, logistic regression, and other data analysis techniques. This study culminates in more accurate, simpler, and more authentic models for the Department of Admissions to use when admitting candidates. Our team will present a recommendation for a metric that the Admissions Directorate could use to better predict the success after plebe year with a certain level of accuracy.

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

MODIFYING THE ACEER

CDT James C. Palmer
CDT, United States 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: CDT James C. Palmer

Advisory Team: COL Joseph Lindquist
Thesis Advisor

LTC Kristin Arney
Thesis Advisor

MAJ Bryan Adams
Thesis Advisor

Colonel Tina Hartley
Chair, Department of Mathematical Sciences

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Each year, thousands of hopeful candidates begin the United States Military Academy's tedious application process. The Department of Admissions possesses the daunting task of selecting the deserving candidates who will gain admission into the Corps of Cadets. As a tool to assess a candidate's academic potential, the Department of Admissions uses an algorithm, formally known as the ACEER, which includes high school class rank and ACT scores for Math, Reading, English, and Science Reasoning. The current ACEER has not been updated since 1992. In the last twenty-seven years, the ACT and the high school curriculum have changed dramatically. Consequently, multiple stakeholders are concerned about the validity of the ACEER. The purpose of this report is to recreate and modify the current ACEER to ensure the USMA Admissions Department is admitting candidates who have the highest potential for academic success. This study conducts thorough research about the USMA Admissions process, Title X of US Code, and implementation of the ACT. With up to 40% of high schools not reporting class rank, the Admissions Department is relying heavily on standardized tests to predict the success of candidates. This study searches for short-term and long-term predictors of academic success through methods of linear regression, logistic regression, and other data analysis techniques. This study culminates in more accurate, simpler, and more authentic models for the Department of Admissions to use when admitting candidates. Our team will present a recommendation for a metric that the Admissions Directorate could use to better predict the success after plebe year with a certain level of accuracy.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank COL Joseph Lindquist, LTC Kristin Arney, and MAJ Bryan Adams for their extraordinary commitment to the success of this project. A special thank you to CDT Hailey Conger and CDT Caroline McCusker, who worked on similar problems. Thank you to MAJ William Corson, MAJ David Delcuardro-Zimmerman, and MAJ Dusty Turner. Working in our group was by far the best professional and academic experience I have had at West Point. Thank you to the USMA Department of Admissions, the Department of English and Philosophy, and the Department of Math for your hard work selecting and preparing Cadets to become Army Officers.

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The CEER score is a mathematical model which uses a candidate's high school class rank and standardized test scores to create a metric that aims to predict a candidate's potential for academic success at USMA. The ACEER is the sole academic predictor for candidates applying to West Point who only took the ACT. Our team has identified four major problems with the current ACEER score.

1. The ACEER was created in 1992 and has not been modified since. High school curriculum and standardized tests have both changed significantly since the early 90's, suggesting the ACEER may be outdated.

2. For candidates whose high schools do not publish class rank, USMA uses their standardized test scores to estimate a high school class rank score. This means that up to 40% of ACEER scores are unauthentic, meaning the ACT tests essentially double count. This is extremely problematic as these scores fail to account for any of the experiences or accolades a candidate worked for in his or her four years of high school.

3. ACEER scores are steadily increasing with no rise in CQPA. If ACEER scores are increasing, we should see a rise in CQPA.

4. The ACEER and the SAT CEER are not equated. On average, a student who performs the same on both exams would receive a 30 point increase on the ACEER as opposed to the CEER.

Our team sought to study the current model and create a new ACEER that would solve the problems listed above. Our team performed linear regression and other data analysis techniques to determine the best combination of predictors of success at the end of plebe year. In order to mitigate problems with authenticity, concordance, and complexity, our team recommends the Department of Admissions adopt the following equation:

$$COLLABACEER19 = -273 + 14.97 * ACTM + 8.39 * ACTE + 198 * RATIO \quad \text{(-1)}$$

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND

In 2017, a diverse pool of 13,827 people opened up an application with the hopes of an appointment to West Point. The group was comprised of people from all walks of life - all with a dream to attend the United States Military Academy and serve their country [4]. Of the 13,827 to apply, 4,120 received a nomination, 2,360 were deemed qualified, and 1,257 received an appointment [4].

The United States relies on USMA to commission leaders of character for service in the military. The candidates who receive appointments to USMA will later serve and lead America's sons and daughters. For this reason, it is critical the Admissions Process is as fair and accurate as possible. Each year, the USMA Admissions Department shares the burden of selecting the candidates who receive appointments. This process is completed through a computation of a Whole Candidate Score (WCS) for each candidate. A large portion of the WCS is the College Entrance Examination Rating (CEER) [5]. The composition of the WCS can be found below.

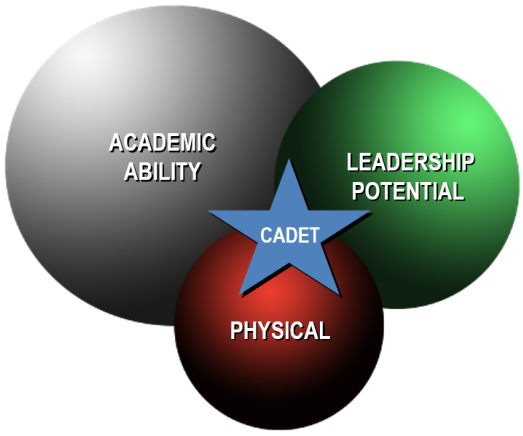

Figure 1. This figure depicts USMA's whole candidate concept. 60% academic, 30% character, and 10% physical.

B. THE ACEER

The CEER score is a mathematical model which uses a candidate's high school class rank and standardized test scores to create a metric that aims to predict a candidate's potential for academic success at USMA. "CEER" is a blanket term covering both the CEER score, which uses the SAT, and the ACEER score, which uses the ACT. The two models produce scores between 200 and 800. A candidate's ACEER/CEER score is a crucial metric for admission to USMA. The metric comprises 60% of the Whole Candidate Score (WCS), which is used to create order of merit lists for competing Congressional nominations and heavily factored into a candidate's application. For this reason, it is vital to the Admissions Department that this score is an accurate predictor of academic success.

1. Composition of the ACEER

The algorithm for the ACEER is described by the following equation:

$$ACEER = 0.219*HSR+9.43*ACT.M+0.45*ACT.SR+4.62*ACT.E+4.01*ACT.R-41.5 \quad \text{(I-1)}$$

Where:

- HSR - High School Class Rank
- ACT.M - ACT Math Score
- ACT.SR - ACT Science Reasoning Score
- ACT.E - ACT English Score
- ACT.R - ACT Reading Score

C. THE ACT

The ACT is the preferred college readiness standardized test in the central United States. The ACT is composed of four major components and a writing component. The four components include: Math, Reading, English, and Science Reasoning. These "big four" components are scored from 0 to 36. The writing exam is scored on a scale from 0 to 12. The Department of Admissions at USMA takes super-scores of the ACT. This means a candidate who takes the ACT two or more times can focus on a specific component and USMA will take the highest score in the category from any of the tests. The ACT refers to

itself as "the leading US college admissions test that measures what you learn in high school to determine your academic readiness for college." [1]. This test, while vastly accepted by colleges and universities as a standard of knowledge, ultimately does not account for the long-term performance of high school students. Figure 2 is a sample score report from the ACT:

D. PROBLEMS WITH THE ACEER

The ACEER is the sole academic predictor for candidates applying to West Point who only took the ACT. Our team has identified four major problems with the current ACEER score.

1. The ACEER was created in 1992 and has not been modified since. High school curriculum and standardized tests have both changed significantly since the early 90's, suggesting the ACEER may be outdated.

2. For candidates whose high schools do not publish class rank, USMA uses their standardized test scores to estimate a high school class rank score [6]. This means that up to 40% of ACEER scores are unauthentic, meaning the ACT tests essentially double count. This is extremely problematic as these scores fail to account for any of the experiences or accolades a candidate worked for in his or her four years of high school.

3. ACEER scores are steadily increasing with no rise in CQPA. If ACEER scores are increasing, we should see a rise in CQPA. This will be discussed more in the data exploration chapter.

4. The ACEER and the SAT CEER are not equated [7]. On average, a student who performs the same on both exams would receive a 30 point increase on the ACEER as opposed to the CEER. This can be visualized in Figure 3.

E. ENGLISH SCORES AND THE STANDARDIZED TEST WRITING

Another problem this study will investigate is whether or not USMA should be collecting writing samples from standardized tests. The ACT Writing exam is an optional exam performed at the conclusion of the other four tests. Candidates have to opt in to take the exam, which requires additional fees and the additional time discussed above. According to the Director of the Department of Admissions, COL McDonald, USMA is the last school in the Patriot League to collect writing samples from the ACT and the

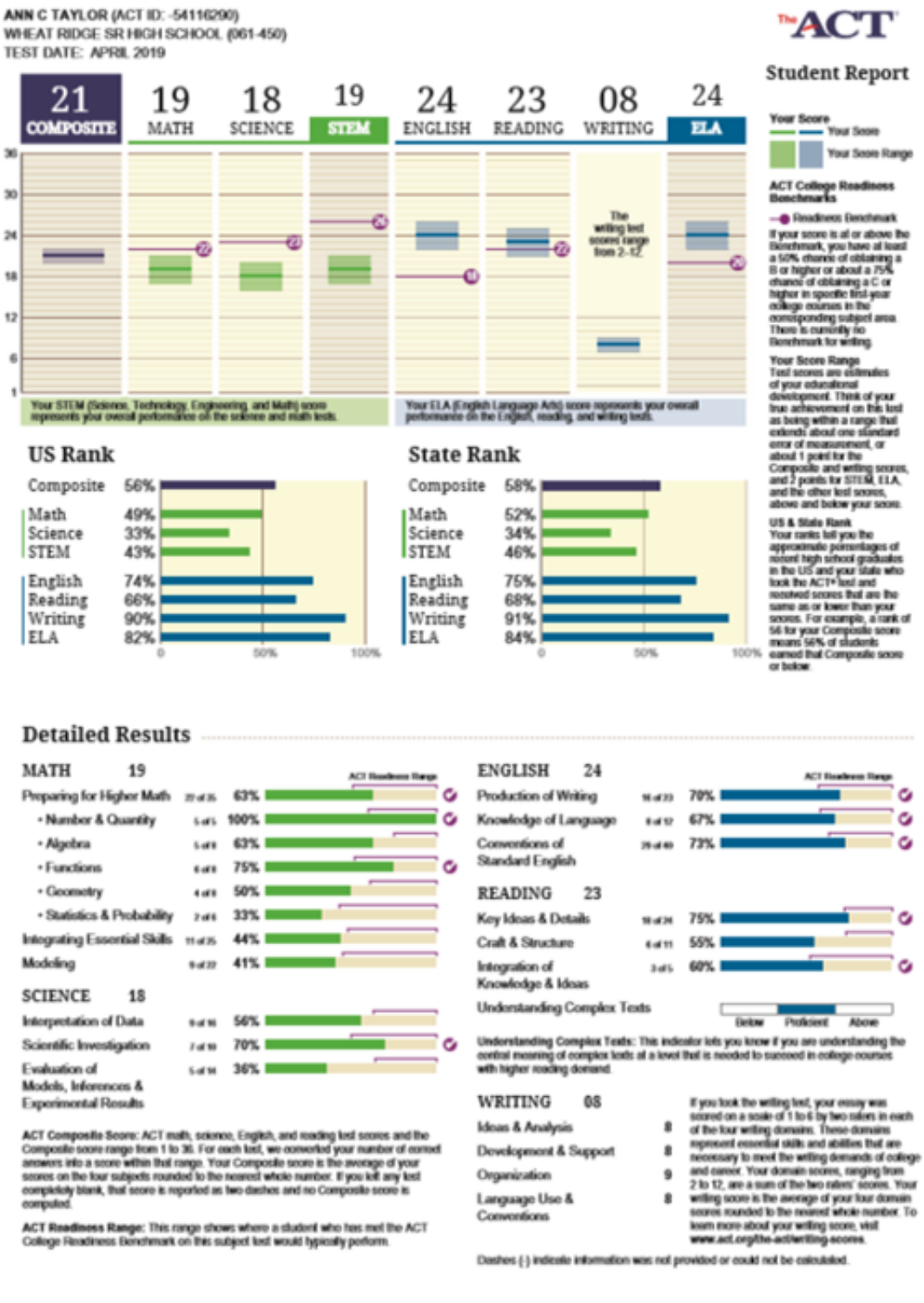

Figure 2. This figure depicts a sample score report for the ACT. Please note that all of the subtests are out of 36. The ACT Writing Exam is out of 12. [1].

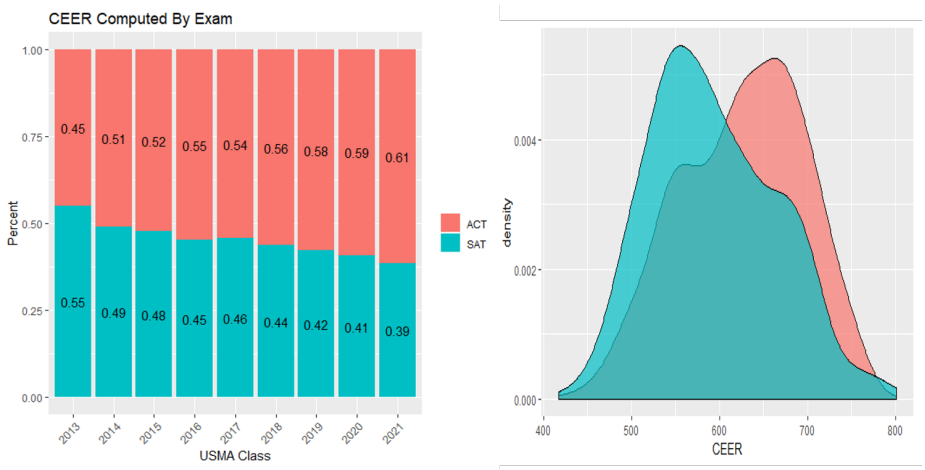

Figure 3. This figure describes the rise in popularity in the ACT and highlights the discrepancy between the scores [2].

SAT. USMA is questioning the validity of the ACT Writing exam. Specifically, USMA is concerned the ACT Writing exam cannot predict success in English courses. If the ACT Writing is a poor predictor, USMA would like to discontinue the requirement. This study will use logistic regression to see if ACT writing scores are good predictors of success in entry level English courses at West Point. EN101 is a course every freshman takes in his/her first semester at the United States Military Academy. This supplementary research question is prompted by recent, poor performance in EN101 and will explore if the ACT Writing exam can explain a Cadet's performance in the course.

F. DEFINING A SUCCESSFUL MODEL

Before our team completed any analysis, we had to define the characteristics that make a linear model successful. In deliberation, our team determined a successful model is simple, easy to interpret, and mathematically accurate. For these reasons, our team took a two-fold approach to assessing a model's performance. For most of our preliminary analysis, our team focuses on the coefficient of determination (R^2 value) to differentiate poor models from good ones. As models become more competitive, our team uses R^2, Akaike information criterion (AIC), and Bayesian information criterion (BIC) to distinguish successful models.

To ensure decision makers could understand and communicate the concept of a model's

predictive power, our team developed a way to clearly assess a model's accuracy. This metric is simply the percentage of observations a model could predict within a given confidence interval. Specifically, our team was concerned with predicting Cadet Quantitative Performance Average (CQPA), at the end of Plebe year, within the sign of a letter grade and within two signs of a letter grade. Below is a table which explains each metric:

Metric	Bounds	What is more accurate?	Explanation
R^2	0 to 1	Closer to 1.	Proportion of variance explained by IVs.
AIC	$\geq$0	Closer to 0.	The information lost by the model.
BIC	$\geq$0	Closer to 0.	The information lost by the model.
.33 CONF.	0-100%	Higher %	% of obs. predicted within the sign of a letter grade.
.66 CONF.	0-100%	Higher %	% of obs. predicted within two signs of a letter grade.

II. LITERATURE REVIEW

The purpose of this literature review is to first gain a deeper understanding of the ACT, standardized testing, the USMA Admissions process, and the ACEER score. Our team examined past studies, methodology, and implementation of the ACEER to further our knowledge and ensure our proposed model builds on results from past studies.

A. INTERVIEW WITH COL MCDONALD 13SEP2018

Our team met with COL Deborah McDonald and her team consisting of LTC Timothy Viles and CPT Owen Tolson. Our team scheduled the interview in order to complete stakeholder analysis, refine our understanding of the problem, and inquire about the deliverables our team will need to create for the project.

COL McDonald gave our team a major assumption during the meeting. COL McDonald stated that for this problem, success at USMA is defined by academic GPA. For the remainder of the study, success will be defined by academic GPA at the end of plebe year. The CEER score strictly comprises the academic prowess of a candidate and is intended to predict academic success and only academic success at USMA after plebe year. This definition of success helps our team define our problem and gives us a major assumption to operate under. Any work or courses completed at USMA after plebe year are assumed to be better predictors of follow on success at USMA than high school data[6].

USMA Admissions also mentioned a Cadet is at risk if he/she has a CEER score below 520. Our team wants to determine if there is any mathematical backing on why 520 is the cut point for a Cadet to be deemed at risk. We learned from the Admissions Department that any class entering USMA can be comprised of up to 25% at risk Cadets. A major question our team will address is whether at risk Cadets more prone to failure at USMA. In other words, is the ACEER an accurate predictor of achieving over a 2.0 GPA plebe year?

The major problem the Admissions Department currently has is the absence of high school rank. COL McDonald stated that 40% of high schools do not publish class rank. Of the remaining 60%, there are little strength of schedule or geographic considerations in place to normalize high school class rank or GPAs. For example, a valedictorian candidate from ru-

ral Wyoming (a class size of 10) may receive a greater benefit from the ACEER calculation than a candidate who is from a large school district such as Fairfax, VA with a class size of 900. To account for this, the Admissions Department, in some cases, is using standardized tests as a replacement for class rank.

B. USMA ADMISSIONS TRAINING BRIEF

This USMA Admissions Training brief is given to incoming officers assigned to the Admissions Department. The presentation is over 150 slides and includes information pertaining to class composition goals, recruiting athletes, admitting Soldiers, the United States Military Academy Preparatory School, and the nomination process for candidates.

Each Class has talent and diversity goals which USMA must attempt to meet. The Class of 2020 goal was scheduled to be composed of greater than 30% Scholars, 25% Leaders, 23% Athletes, and 5% Soldiers [5]. In 2020, USMA shot 6.4% over their Scholar goal while missing both the Leader and Athlete goals. The Admissions Department also attempts to meet diversity goals. Each class is scheduled to admit greater than 20% women, 5% Asians, 11% Hispanics, and 14% African Americans [5]. Admitting and graduating demographics comparable to the population of the United States is incredibly important to USMA.

The brief explains the possible nominations sources for all candidates. For the scope of our project, while very interesting, the nomination process does not appear to play a factor in the formulation of the CEER score. On the other hand, the type of nomination could be considered a predictor of success, warranting further analysis.

The biggest takeaway from this brief is the composition of the Whole Candidate Score (WCS). The WCS is 60% academic ability, suggesting the CEER score is the most important predictor of overall success at USMA. If the Admissions Department does not manipulate the CEER score to reach talent and diversity goals, the processes to reach the quotas are not within the scope of this project.

C. PREVIOUS ATTEMPTS AT THE ACEER AND OTHER RESEARCH OF PREDICTORS

1. Presentation for Director of Admissions: Report on the Validation of the ACEER for Class of 1995 End of 4CL Year

This study was completed on 30 September 1992 by the USMA Office of Institutional Research (OIR). It is an abbreviated version of the brief given to the 1992 Directior of Admissions, COL Rushton. The study's objectives were "to validate the ACEER equation using Class of 1995 End of 4CL Year APS [Academic Performance Score] and to recommend an ACEER for admitting the Class of 1997"[8].

The summary of this study makes it seem like an incredible resource for our team. Unfortunately, the abbreviated nature of the study omits any analysis or methodology from the document. The document contains some demographic data, a few equations, and a couple of graphs. The ACEER used to admit the Class of 1994 was:

$$ACEER = 0.233HSR + 11.97ACTM + 3.02ACTE + 4.91ACTSR - 1.14.ACTR - 45.l \quad \text{(II-1)}$$

Where:

- HSR - High School Class Rank
- ACTM - ACT Math Score
- ACTSR - ACT Science Reasoning Score
- ACTE - ACT English Score
- ACTR - ACT Reading Score

Interestingly, this version of the ACEER punishes candidates for scoring higher on the Reading section of the ACT. A one point increase on the ACT Reading exam would lower a candidate's ACEER by 1.14[8]. For this reason, this model is inadequate for admitting candidates.

Another takeaway from this study is the use of the potential academic performance score (PAPSC). This is an equation used to predict the academic GPA (APSC) of a Cadet at the

end of his/her plebe year using the same predictors as the ACEER. This leads our team to believe OIR worked backwards to create the new ACEER (NACEER). It appears OIR used linear regression to predict Cadets' APSCs and then converted the GPA equation to create the NACEER. The following equations are presented below:

$$PAPSC95 = 0.001266HSR + 0.05462ACTM + 0.02677ACTE + 0.002582ACTSR + 0.02324ACTR - 1.1015 \quad \text{(II-2)}$$

$$NACEER95 = 172.7PAPSC95 + 148.7 \quad \text{(II-3)}$$

$$NACEER = 0.219HSR + 9.43ACTM + 4.62ACTE + 0.45ACTSR + 4.01ACTR - 41.5 \quad \text{(II-4)}$$

The system of equations suggests that PAPSC was the the response that OIR sought to predict then normalize into the NACEER. Regardless, the lack of methodology calls for our team to conduct our own attempt at replicating this study to better understand the orgins of the NACEER. This is neccessary because the NACEER, created in 1992, is still the ACEER scored used today in a vastly different environment.

2. Lack of Equated and Validated CEER and ACEER Formulas

COL Oldaker from the Department of Physics conducted this study in 2000. This study points our problems with the current CEER and ACEER. Most notably, the current scores have not been equated since 1973 [9]. OIR validated the ACEER using the Class of 1995s grades, however the CEER (SAT version) has not yet been validated.

COL Oldaker suggested candidates in the top of their class are penalized if they submit ACT and no SAT scores. He concluded this by creating equations from a series of concordance tables. COL Oldaker calls these "correspondence formulas". Using the same method, COL Oldaker makes another claim, stating that candidates who have greater aptitude in math than verbal skills would be advantaged by submitting the SAT without the ACT[9]. In a system that is supposed to be fair for all parties involved, these problems are

certainly of great concern.

The study goes on to define the correlation coefficient and its role in multiple linear regression. COL Oldaker explains that a R^2 of .7 is needed to explain over half of the variance in the observed data. In simple regression, the correlation coefficient is directly interpretable, but in multi-linear regression, the square of coefficient is interpretable. The current CEER score has a correlation coefficient of .64. COL Oldaker believes the rest of the variance can be explained through differences in: curriculum, textbooks, teachers, tests, and culture.

High School Rank is pinpointed as the most important factor due to its large standard deviation. The coefficients do not necessarily determine which predictor is the most important. The coefficients have not changed much since the 1960 model. The current models are more or less the same models used to admit the classes who fought in Korea and Vietnam.

In summary, COL Oldaker lists potential problems with the current models. As noted above, candidates of higher class rank who submit only the ACT are penalized. Secondly, the difference in coefficients may indicate there are two different pools of Cadets CEER and ACEER. CEER group may be stronger in math while the ACEER group is stronger verbally.

This study recommends the CEER and ACEER should be subject to statistical analysis to ensure they are equally weighted and fair. It also recommends there be an audit of the models to describe why the SAT gives the math advantage and the ACT gives the verbal advantage. This study is particularly interesting to our team because we assumed the CEER and ACEER were fairly well equated. This does not seem to be the case.

a. Review of the CEER Metric v2.3

A multitude of officers from D/MATH completed a review of the current CEER and ACEER. They created a metric, CEER.12, which incorporates the standardized tests writing components and the quality of the school. The CEER.12 uses the best reported percentile from among test score categories.

The study led to the following recommendations: replacement of the current CEER score for the adoption of the CEER.12, implementing the CEER and ACEER along with the

CEER.12 for AY2013, and conducting reviews of the metric every four years. The CEER.12 also changes the cut point at which a student is deemed at risk[10].

This study used five approaches to analyze the problem. The approaches were: (1) updating the current ACEER and CEER models, (2) using modified standardized tests and HSR (CEER.12), (3) using the Patriot League Academic Index, (4) using additional variables, and (5) a black box predictive model.[10]

Each approach was analyzed using the four conditions set by the Department of Admissions. All possible solutions had to have the ability to predict, be considered a simple model, be easily interpretable, and be easy to implement. The winning approach was the creation of the CEER.12. The CEER.12 standardizes each test score before it inputs it into the model. This ensures the model predicts fairly as long as the prior equalities between ACT and SAT are correct.

The CEER.12 is a single model which address many of the problems pointed out in previous studies. It is a great attempt at the current state of the CEER score. Our team chose to borrow methodology from this study.

3. Predicting academic success for students of color within STEM majors

John Gibson conducted this study on behalf of Purdue University in 2013. The study aimed to "predict what pre-college characteristics predict collegiant STEM success using GPA as a metric"[3]. Specifically, the study attempts to predict success for non-white students. The factors Mr. Gibson considered were: high school grade point average, standardized tests, and college prep coursework.

The study first determined "intercorrelations of variables" and then validated "the assumptions of multiple linear regression...via a series of descriptive analyses"[3]. The study then completed multiple linear regression using all of the variables and created the following correlation matrix:
When conducting simple linear regression, the researcher found the most significant predictor of cumulative college GPA was in fact high school core GPA. While the methods used in this study are nothing out of the ordinary, the idea that high school core GPA was

Variable	College GPA	ACT	AP credits	Transfer	HS core	HS GPA	HS Math	HS Sci
College GPA	1							
ACT	0.286*	1						
AP credits	0.275*	0.533*	1					
Transfer	0.029	0.040	−0.040	1				
HS Core	0.370*	0.304*	0.275*	0.097*	1			
HS GPA	0.357*	0.363*	0.325*	0.074	0.804*	1		
HS Math	0.304*	0.260*	0.229*	0.060	0.810*	0.637*	1	
HS Sci	0.339*	0.278*	0.251*	0.082	0.848*	0.697*	0.647*	1

Note: $*p < 0.01$

Figure 4. This figure depicts the correlation matrix used in Mr. Gibson's study [3].

found to be the best predictor is fairly problematic for our team. USMA admits candidates from across the country while many public state schools pull a majority of their students from a specific region, potentially having more similar high schools than the national pool.

When conducting multiple linear regression, the researcher found that AP credits and standardized test scores are the most significant predictors of cumulative college GPA. This gives our team another avenue of approach through the consideration of AP courses. A possible research question could look at the number of AP courses a student took or the aggregate scores of his/her AP exams.

D. CHANGES TO THE ACT AND CONCORDANCE

1. Guide to the 2018 ACT/SAT Concordance

This document describes the CollegeBoard's procedure for comparing the 2018 ACT to the 2018 SAT. These figures are incredibly valuable if the CEER score can be transformed into one model. If the scores can be inputted as one type, using a single model, it is much simpler than having two models. Comparing individual numbers is much easier to do than comparing two multiple linear regression models.

The concordance between individual components of the tests achieved a correlation coefficient of .85 or higher[7]. Because the correlation coefficient is not 1, variance is lost when the scores are compared against each other. The authenticity of the comparison decreases each time it is completed. This means while there are processes to compare and equate, when iterated or done in sequence, only a fraction of the authentic value remains. A key takeaway from this is simplicity. In an ideal world, every candidate would take the same entrance exam. Implementing this would eliminate both the problems of concordance

between scores and equating the CEER and ACEER.

III. DATA COLLECTION & EXPLORATION

A. DATA SAMPLE

Our team began our data collection by completing Collaborative Institutional Training Initiative (CITI) training to ensure we were comfortable with the processes and procedures required to handle human data. CITI strives to protect the privacy of human subjects during experiment or data analysis. To receive the data, our team filed an Institutional Review Board (IRB) exception request. Prior to receiving any data, all of the personally identifiable information was removed. This greatly enhanced the protection of the subjects' privacy for the duration of the study.

1. Classes of 1994-1997

Our team requested this data to replicate the study completed in 1995. This data included the following components:

- ACT/SAT scores.
- CQPA recorded after first semester Plebe year.
- Predicted CQPA using ACEER and CEER.
- ACEER and CEER scores.

2. Classes of 2010-2021

The second group of data included a over ten classes from USMA. This data was used to the remainder of the study and will be explained in later paragraphs. The components collected included:

- ACT/SAT scores.
- CQPA recorded after first semester Plebe year.
- Predicted CQPA using ACEER and CEER.
- ACEER and CEER scores.
- Source of appointment.

- High school and collegiate sports information.
- High school four-year college percentage.
- Letter grades from core classes including: EN101, EN102, MA103, MA104, MA206.

B. DATA EXPLORATION

After receiving the data, our team verified the correctness through summaries and histograms. It was clear that some of the data did not make sense. For example, the given ACT Writing scores ranged from 0 to 36, while the test is scored with a maximum of 12. The data was scanned for inconsistencies and corrected when possible. The series of plots below seeks to explain some of the patterns and abnormalities in the data.

The histogram, Figure 5, contains nine years of computed ACEER scores from admitted candidates. The mean ACEER for this period of time is slightly greater than 620, while the standard deviation is around 70.

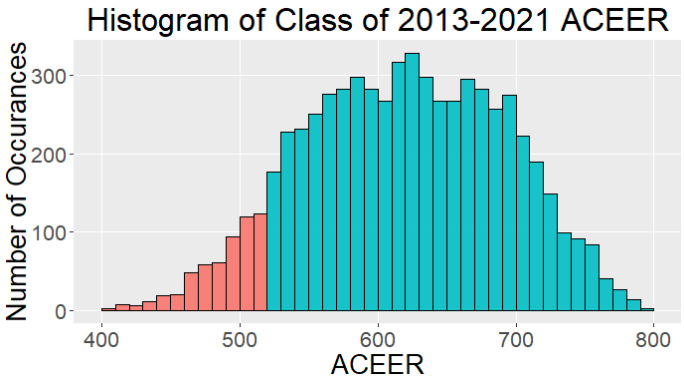

Figure 5. This histogram depicts the ACEER scores from the Class of 2013 to the Class of 2021. An ACEER of 520 or below puts a Cadet at-risk. The at-risk observations are colored in red.

One of the major problems with the ACEER is that it is outdated. Figure 6 depicts the change in average of the ACEER from 2013 to 2021. The blue points on the plot represent the average CQPA, which does not seem to be changing. CQPA has ranged from 2.94 to 3.00, but there is no predictable trend associated with an increase or decrease in CQPA. This is problematic as the average ACEER continues to increase.

The same problem can be visualised by looking at the histograms from 2013 and 2021, respectively. These histograms are displayed in Figure 7. The two distributions fail the eye test. The average has increased, but it also appears the entire distribution has shifted to the

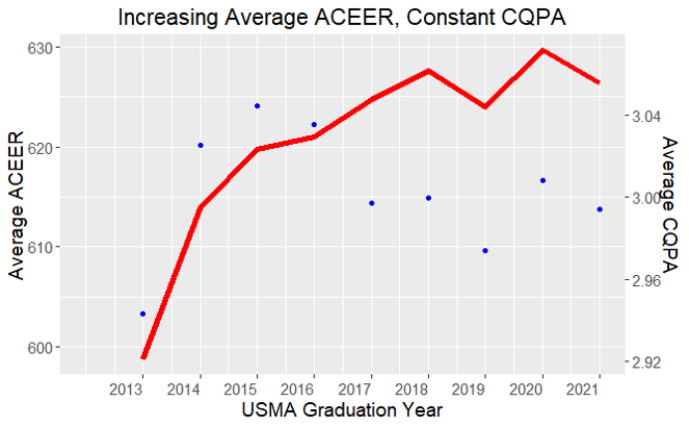

Figure 6. This histogram displays the CQPA of every Cadet from the Classes of 2013-2021.

right. Additionally, there are more observations in 2021 than 2013, suggesting either that more students are taking the ACT or ACEER scores are becoming greater than their SAT counterpart.

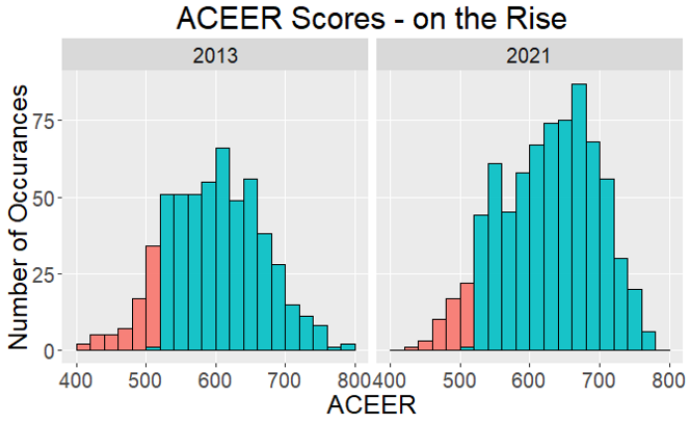

Figure 7. These histograms display the number of occcurances for ACEER scores in two different year groups.

After examining most of the major players in the data frame, our team started to look at some oddities we found in the set. For example, some high schools do not use a 4.0 grading scale. Scales out of 5, 6, and 100 are significantly popular, as seen in Figure 9. This is useful to know, as high school GPA serves as a promising candidate to predict collegiate success.

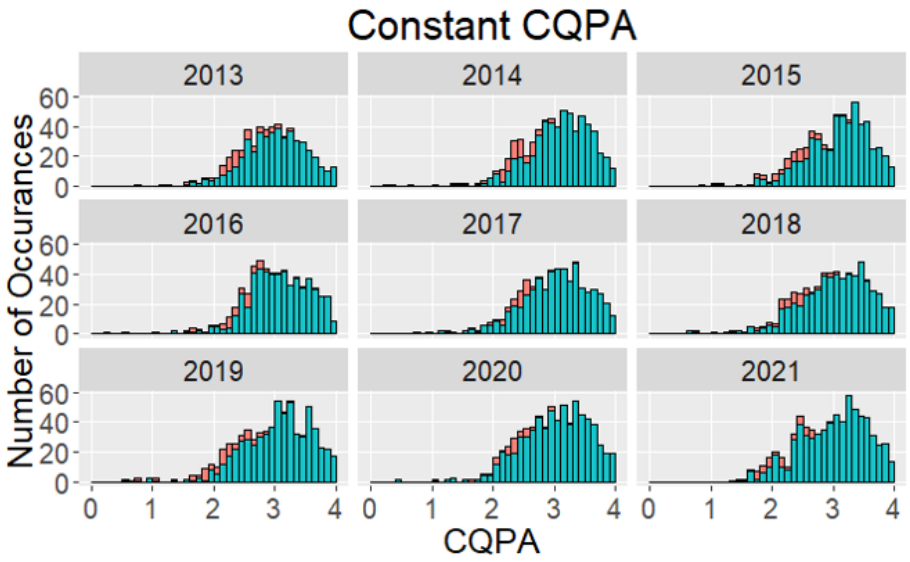

Figure 8. These histograms depict CQPA from the years 2011 to 2019.

Most importantly, the preliminary data exploration made our team experts in our data set. Knowing the ins and outs of the data, the name for each variable, and how to manipulate new columns using software were all invaluable skills moving forward in the project.

1. Tools Used For Data Exploration

Our team started this study with little to no idea how to manipulate a data set. Over the course of the study, our team grew more and more comfortable mutating and filtering the data. This was done using the *tidyverse* package in the software, R[11].

C. DATA EXPLORATION TAKEAWAYS

Exploring the data set made our team experts of the data and increased our confidence using the software *RStudio*. Our brief analysis uncovered a few important findings:

- ACEER scores have changed over time while CQPA has remained fairly constant.
- High schools use different grade scales.
- Some years contain incomplete data - analysis will have to be completed with only year groups that have complete data.

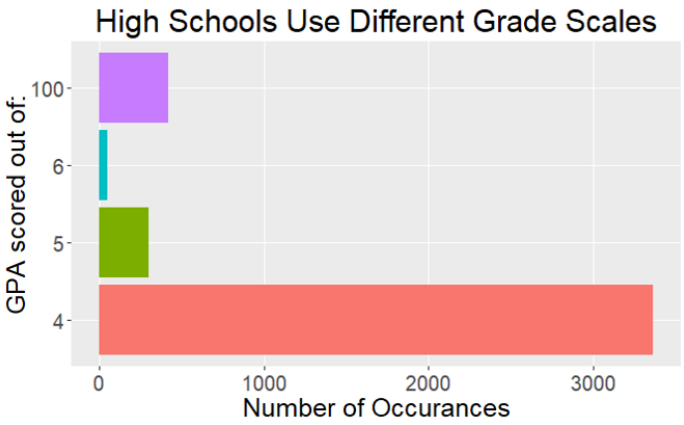

Figure 9. This column graph shows us that many candidates have different high school grading scales.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. REPLICATING THE ACEER

A. THE CLASS OF 1997

The first step in the study was to replicate the ACEER study completed in 1995. Our team completed this by filtering the data set to include only the Class of 1997. The Class of 1997 was chosen because they are the only class in the old data set to have recorded CQPAs. Our team performed a linear regression to see how well the ACEER predicted CQPA. After a brief analysis of this performance, our team sought to make our own "predicted ACEER" using the equation given in the study. The team then performed single linear regression by attempting to predict the Cadets actual CQPA using our "predicted ACEER".

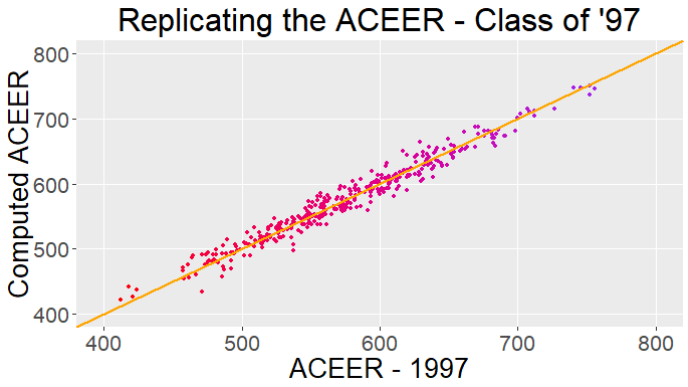

Figure 10. This figure displays the the '95 model replicating the Class of '97 data.

Our team is somewhat uncomfortable with these results. If the '95 equation was used to compute these ACEER scores, the R^2 should be 1. The model displayed above obtained an R^2 value of .95, meaning 5% of the variance went unexplained. This means another process was used for at least some of the observations in computing the ACEER score. The lack of explanation during the previous studies makes it impossible to know why this model is not a perfect match. Our team decided to replicate more current data using the same methodology to see if this discrepancy still exists.

B. REPLICATING THE ACEER USING CURRENT DATA

Our team sought to successfully recreate the ACEER scores in the data by using the associated ACT scores and high school rank for the Class of 2021. Below is a plot of our

replication.

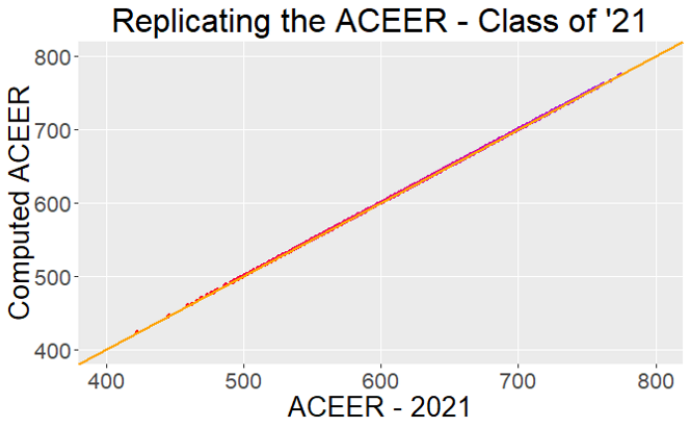

Figure 11. This figure displays the the current model replicating the Class of 2021 data perfectly.

Recreating the ACEER proved to be a great success for our team. Our next step was to test the ACEER for accuracy. Our goal became to assess the current ACEER model's accuracy in predicting the academic success of Cadets after plebe year.

V. UPDATING THE ACEER USING CLASS OF 2021 DATA

Our team began our search for new coefficients for the ACEER by calculating ACEER scores using observations from the Class of 2021. Our team completed this by using the equation of the current ACEER. All of these observations matched the ACEER scores currently in the data set. The performance of this model, its ability to predict CQPA, was recorded in summary format. The next step was to use high school rank and ACT scores to create a new, updated model which would more accurately predict actual CQPA. This was completed by using the linear model function in RStudio. The coefficients were recorded, along with the accuracy of the model. This model's output, predicted CQPA, was placed into the data set as a new column. Our team then converted this predicted CQPA into an ACEER score using a linear transformation. The model was then compared to the baseline. Both models use all the same predictors found in the ACEER, but have different coefficients. The new model serves as an update to the current ACEER to better reflect more current data. The goal of this sub-study was to analyze if the ACEER is an outdated metric. Our team asked the question, "if the coefficients were updated, would the metric have more predictive power?"

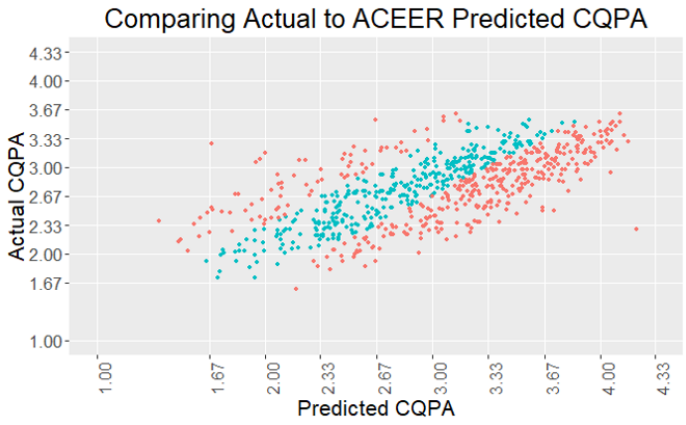

Figure 12. This figure displays how many observations were successfully predicted by the ACEER within .33 of the CQPA the Cadet received.

Ultimately, this answer was yes! Updating the coefficients of the ACEER increased the model's predictive power. In fact, updating the coefficients accounted for 9% more of the

variance in the data. This suggests the current ACEER might be outdated. The original and updated coefficients can be found in the table below:

Predictor	Original Model	Updated Coefficient Model
High School Rank	.219	.310
ACT Math	9.43	12.232
ACT English	4.62	4.976
ACT Reading	4.01	2.629
ACT Science Reasoning	.45	2.841
Intercept	-41.5	-236

Updating the coefficients also increased the percentage of observations within one sign of a letter grade from 46% to 55%. If the Department of Admissions insisted the five predictors in the current model were correct, they could be more accurate, by all metrics, by simply updating the coefficients.

VI. ACT WRITING AS A PREDICTOR OF SUCCESS

Our team completed separate analysis on short term predictors to explore the possibility of adding them to the ACEER. The ACT Writing exam was the most extensive substudy. Our team wanted to see if the ACT Writing exam could predict success in entry level English courses[12]. Predicting success in EN101 using ACT Writing involved a variety of methods including: linear regression, logistic regression, ROC/cut-point analysis, and cross-validation. Each technique will be described in depth in the following paragraphs.

A. LINEAR REGRESSION

The first approach our team used to analyze EN101 performance was linear regression. The first step was to create a new column in the data set named "number GPA". Our data contained letter grades (A+, A, A-, etc.). These grades were converted into corresponding grade point averages in accordance with the following table. Our team used simple and multiple linear regression to analyze how well ACT Writing and other variables could predict a Cadet's EN101 grade. Our team used R^2 as a metric of success for our models.

A+	4.33
A	4.00
A-	3.67
B+	3.33
B	3.00
B-	2.67
C+	2.33
C	2.00
C-	1.33
D	0.00
F	0.00

A plot that shows the amount of variance in this data can be seen in Figure 13. Based on how far the observations are spread from the line placed on the graph, it becomes doubtful whether the ACT Writing can really predict the grade a student receives in EN101. No model could possibly account for all of the variance shown in the data. When using just ACT Writing as a lone predictor of EN101 grade, our R^2 was .0506. Our best model, which includes long term predictors such as high school class rank, had an R^2 of .2888.

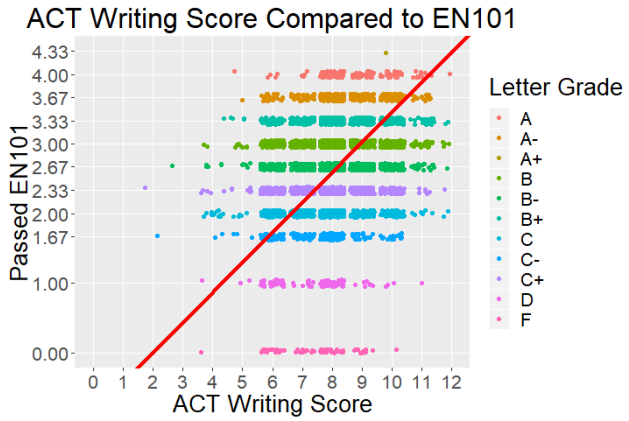

Figure 13. This figure displays how many observations were successfully predicted by the ACEER within .33 of the CQPA the Cadet received.

B. LOGISTIC REGRESSION

Our team used logistic regression to predict whether a Cadet would pass or fail EN101. Our team experimented with ordinal logistic regression, but settled on binary logistic regression to conduct the study. This type of logistic regression uses one or more variables to predict whether an outcome will or will not happen. Our team completed both single-variable and multi-variable logistic regression. Our team analyzed the log odds of the model as a metric of success. The model shown in the plot only uses ACT Writing as a predictor. This figure displays how many observations were successfully predicted by our best logistic regression model. The proportion of the red circle inside the green circle indicates reality. This logistic model, using only ACT Writing, is capable of predicting with 75% accuracy. The naive model, saying everyone will pass, is 96% accurate. The combined model was over 97% accurate, suggesting that long term predictors may be able to predict EN101 success. Figure 15 is an example of a plot depicting the success or failure (pass/fail) of Cadets based on the ACT Writing score. The figure displays Cadets who passed in blue and those who failed in red. In a perfect world, there would be some ACT Writing score in which all Cadets who received the specific score or greater passed and those who did not failed.

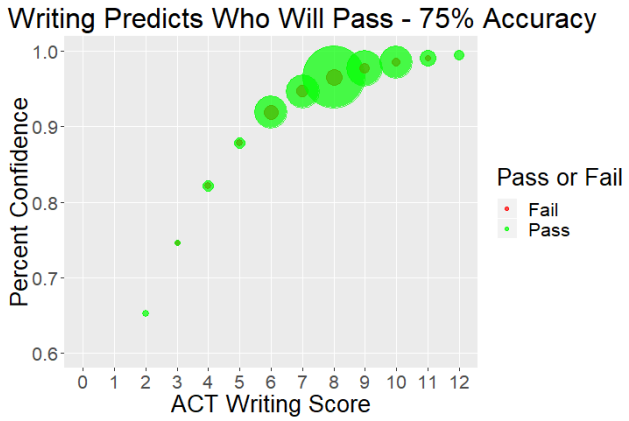

Figure 14. Note: The red circle is oversized for visual purposes - there was no group in which more than 20% of Cadets failed.

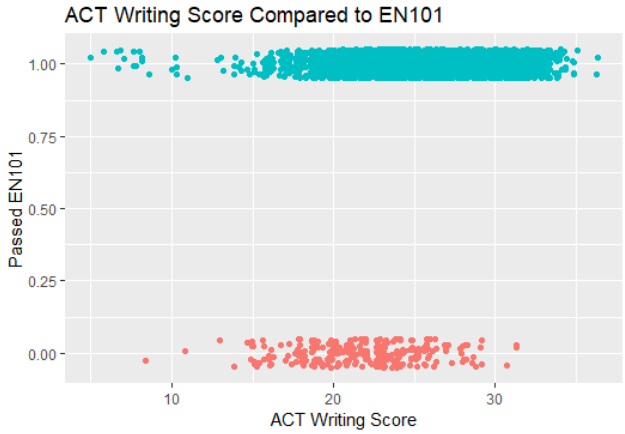

Figure 15. Clearly, this data is not ideal for logistic regression.

1. Cut-Points and ROC

In order to examine and create cut points, hard estimates which predict (with specific confidence) the number of failures expected to occur, our team completed receiver operating characteristic analysis. The ROC curve is created by making a plot of the false positive rate on the x-axis and the true positive rate on the y-axis. The curve displays the accuracy at various thresholds. Our team extrapolated potential cut-points from the curve based on the risk the stakeholder was willing to assume.

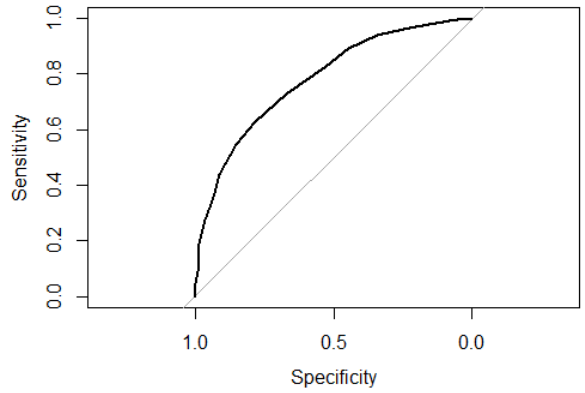

Figure 16. This figure displays the ROC curve for the predicted model.

2. Cross Validation

Our team also cross-validated our relevant models. The process of cross-validation begins by dividing the data set into folds. Our team divided the data into 10 folds. The folds are then used to make training and test sets. A model is created using only some folds of data, and then tested on the remaining folds. This process is repeated until all folds have been trained and tested. This procedure verifies the model and helps explain the accuracy of the model.

C. ACT WRITING RESULTS

Given the weak predictive power of the ACT Writing exam in both linear and logistic models, our team does not believe one can determine whether a Cadet will fail EN101 based on his/her writing score. Additionally, the ACT Writing score should not be added as a predictor to the ACEER.

VII. PREDICTORS OF SUCCESS

A. SHORT TERM PREDICTORS

A short term predictor is a predictor, such as a standardized test, which attempts to predict a candidate's collegiate academic performance. Other examples of potential short term predictors would be AP exam scores. With the ACT Writing score severely underperforming as a predictor of success, our team continued our search for other short term predictors we could include in an improved model.

From previous analysis of the original ACEER model, our team understood the most important short term predictor to be ACT Math. In order to find which short term predictors were more important than others, our team multiplied each predictor by the mean of its column in the data set and normalized it as a percentage of all predictors in the model. The two most important predictors were ACT Math and ACT English.

At this time, our team factored in an important consideration. If the ACT and the SAT had different tests, would only similar predictors be sufficient in a final model? In other words, it would make sense that the two exams should test similar skills. Ultimately, our team decided to make our best models independently and then collaborate on an updated model. The ACT Math and ACT English scores were preferred front runners, as the SAT Math and SAT Verbal sections very closely resemble the skills tested on these exams. By formulating a multitude of linear models, our team tested and explored the possible combinations of short term predictors.

B. LONG TERM PREDICTORS

A long term predictor is one which encompasses a candidate's entire high school academic experience in order to predict first year academic success. Examples of long term predictors include high school class rank, high school GPA, the strength of a candidate's high school, and the faculty appraisal score. The pros and cons of each can be found in the table below:

Predictor	Pros	Cons
High School Rank	Function of candidates performance related to peers	Not reported
High School GPA	Objective performance metric	Different scales
% to 4 yr. College	Validates rank and/or GPA	Potentially unfair
Faculty Appraisal Score	Moderate success predicting scholarships	Subjective

As our team analyzed these predictors, a few problems came to our attention:

1. Only 40% of high schools report class rank. This long term predictor is inevitably insufficient. The Department of Admissions is not and should not be comfortable with creating proxy class ranks (using the ACT) for up to 40% of candidates.

2. A myriad of ethical problems accompany factoring in the strength of a candidate's high school. Candidates who do not attend prestigious high schools or hail from low income areas are punished for their socioeconomic status and/or geographic location. The percent of students who move on to attend a four year college might be mathematically significant, but it is not worth the potential unfairness it may bring to the process.

3. The Department of Admissions does not agree with the assessment of faculty appraisal score. USMA believes letters from high school teachers are too subjective to determine the fate of potential candidates.

1. High School GPA as a Ratio

Our team realized that high school grade point average would be the only tangible replacement for high school class rank as a long term predictor in the model. The only problem associated with this replacement is that many high schools do not grade on the same scale. This problem was dealt with using three separate courses of action.

The first course of action involved only using observations from candidates who were graded out of 4.0 in high school. This course of action would require the Department of Admissions to adjust each high school grade point average to be graded on a four point scale. A quick histogram of this approach revealed that many candidates have well over a 4.0 GPA out of 4.0. Observations colored in red have below a 4.0, while blue observations indicate observations above this threshold.

Interestingly, our team found that having above a 4.0 GPA had little to no predictive power. For example, having a 5.0 GPA in high school does not increase a candidate's chances of achieving success versus a candidate who received a 4.0 GPA in high school.

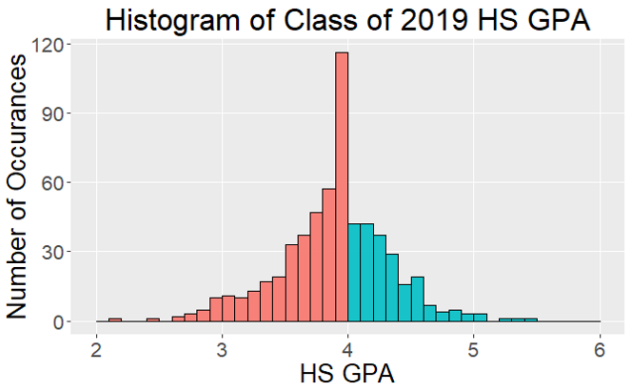

Figure 17. This figure displays the histogram for candidates who were graded off of a 4.0 scale in high school.

Because of this, our team decided to take a different approach.

The second course of action was to use all of our observations and create a new variable, the GPA ratio. The GPA ratio was calculated by dividing a candidate's high school GPA by his/her grade scale. Interestingly, high schools with grade scales other than four appeared to have lower ratio GPAs. This was apparent especially for high schools with grade scales out of 100.

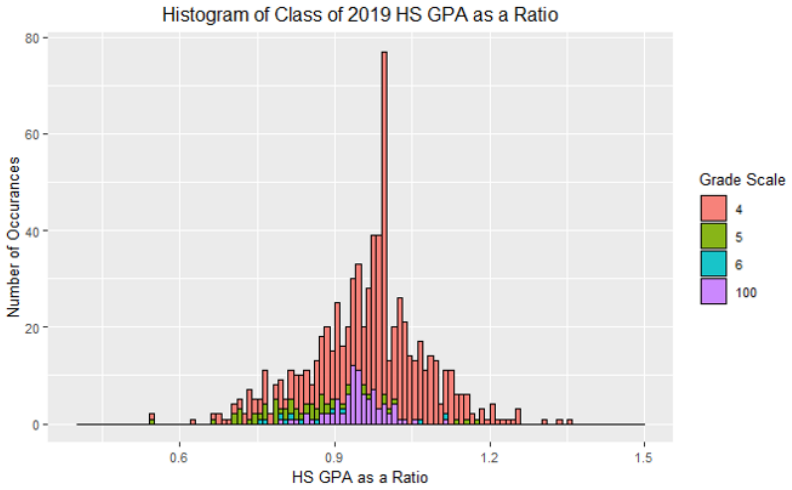

Figure 18. This figure displays the histogram for the ratio GPA of candidates. The histogram is colored based on the grade scale of the high school.

After making various linear models, our team quickly figured out that ratios of GPA greater than 1.0, much like GPAs over 4.0, did not have predictive power. Our team's third

and final approach to harness this predictor was to cap the ratio at 1.0. This approach seemed fair, made sense, and served as the best predictor of academic performance. The plot below will depict whether the candidate received a cap based on the color of the bin.

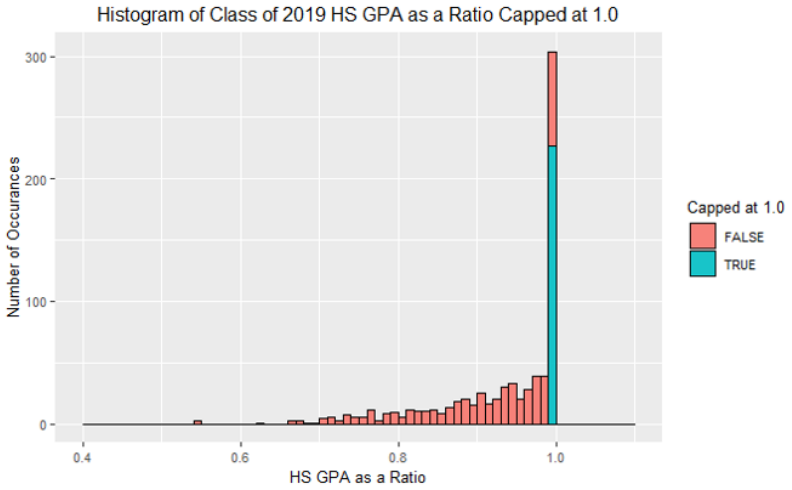

Figure 19. This figure displays the histogram for the ratio GPA of candidates capped at 1.0.

The GPA ratio capped at 1.0 would be the long term predictor included in the final model. This method is arguably the fairest as it mitigates disadvantaged populations. Other methods favor students of greater socioeconomic status, or students who enroll in either many or no AP courses. For this reason, we recommend USMA accepts all versions of weighted GPAs. This method diminishes the effect of the weighted GPA by capping the ratio at 1.0. This cap also protects the populations, such as the students who are graded out of 100, who cannot reach the 1.0 standard. If it were not for the cap, their counterparts would be rewarded for ratios greater than one, which seem to be unattainable on a 100 point grade scale.

VIII. RESULTS

This chapter will outline the results from hours of model building and data analysis.

A. BEST MODEL

Our team's best model was created using the variables:

- ACT Math
- ACT English
- ACT Science Reasoning
- ACT Reading
- High School GPA as a Ratio Capped at 1.0

This model possessed an adjusted R^2 of .461 and could predict a Cadet's CQPA within one sign of a letter grade 56% of the time. It could also predict an observation within two signs of a letter grade 89% of the time. This best model beats out original ACEER model in all metrics except R^2. The original model accounts for 1% more variance.

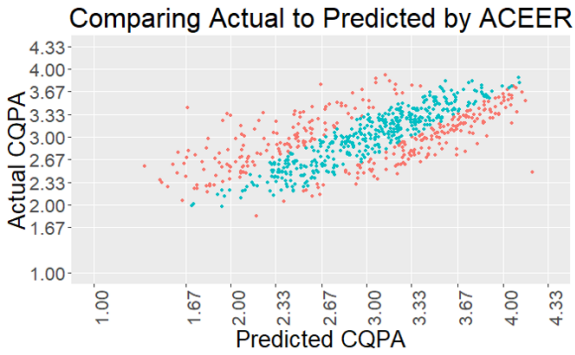

Figure 20. This figure displays the width of observations correctly predicted in blue.

This model was created in RStudio using the linear model function. Our team set GPA as a ratio (capped at 1.0) and all of the ACT scores (no writing) to predict CQPA. The coefficients for this model were recorded. Our team then used a linear transformation to create a model that would predict a score between 200-800. Finally, we labeled our new

model the ACEER19. The equation for this model is:

$$ACEER19 = -320 + 13.23 * ACTM + 5.93 * ACTE + 3.34 * ACTSR + 2.71 * ACTR + 189 * RATIO \quad \text{(VIII-1)}$$

A histogram displaying the distribution of the scores can be found in Figure 21. The color of the observations corresponds to whether the model predicted the student's CQPA within a sign of a letter grade. Red bins indicate a failure to do so, and blue indicates successful attempts. Finally, our team wanted to see which predictors were the most important in our

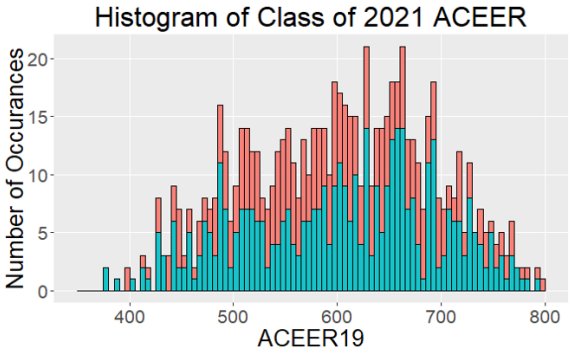

Figure 21. This figure displays a histogram of predicted ACEER scores using our best model.

final solution. After the normalization process, GPA as a ratio and ACT Math are the two most powerful predictors, accounting for 35% and 34% of the model, respectively. There will be more discussion and validation around this model in later chapters.

B. COLLABORATIVE MODEL

Our team also created a collaborative model using the variables:

- ACT Math
- ACT English
- High School GPA as a Ratio Capped at 1.0

These variables were in accordance with the same predictors used for the SAT. This model possessed an adjusted R^2 of .452 and could also predict a Cadet's CQPA within one sign of a letter grade 56% of the time. It could also predict an observation within two signs of a letter grade 90% of the time. This model could be used to assure simplicity and

fairness as the SAT counterpart model would ideally measure the same set of skills. This model was compared with a model for the CEER using GPA as a ratio capped at 1.0, SAT Verbal, and the SAT Math exams.

$$COLLABACEER19 = -273 + 14.97 * ACTM + 8.39 * ACTE + 198 * RATIO \quad \text{(VIII-2)}$$

THIS PAGE INTENTIONALLY LEFT BLANK

IX. AT-RISK ANALYSIS

Previously, a ACEER or CEER below 520 would place a Cadet “at-risk”. USMA is only willing to a certain percentage of ”at-risk” Cadets. USMA did not have an explanation for why the cut point was set at 520, and so our team performed some analysis.

Receiving below a 520 on the previous ACEER would place a candidate in the bottom ten percent of his/her class. This can be visualized in Chapter 3, Figure 5. A major problem associated with creating a new ACEER model is that a score of 520 on the old model does not equate to a 520 in the new model. In fact, our model would place almost a quarter of the Class of 2021 at-risk.

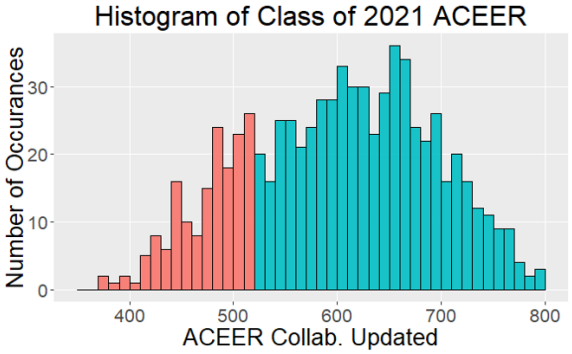

Figure 22. This figure displays the outrageous number of Cadets the new model would flag as “at-risk”.

Our team decided it was best to define the bottom 10% of the class as being “at-risk”. This is also proportional with the percent of Cadets USMA admits who have below a 520 ACEER. After comparing the ACT and SAT model, our team declared the new at-risk cut point to be a CEER/ACEER score of 465.

THIS PAGE INTENTIONALLY LEFT BLANK

X. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSION

The current ACEER metric is outdated and unauthentic. The coefficients being used have not been updated since 1992. In the last 24 years, there have been numerous changes that greatly hinder the current ACEER's ability to predict academic success. The major problem with the ACEER is that high school rank is not wildly reported. For over 40% of the candidates USMA admits, there is no long-term academic predictor factored into the academic third of their admissions process. This is incredibly problematic as our team saw long-term academic predictors consistently appear as more important than short-term predictors. Double counting short-term predictors is an unfair and inadequate approach to admitting candidates.

The easiest way to solve this problem is to utilize the high school GPA of candidates in some form. Our team created multiple models using high school GPA which performed on par with or outperformed the current ACEER. The data for a GPA based model is always readily available, greatly increasing the simplicity and fairness of the metric. The success our team had using GPA as a ratio offers a simple and accurate solution to the problems addressed above.

At this time, the best models available to predict the academic success of candidates are:

$$ACEER19 = -320 + 13.23 * ACTM + 5.93 * ACTE + 3.34 * ACTSR + 2.71 * ACTR + 189 * RATIO \quad \text{(X-1)}$$

$$COLLABACEER19 = -273 + 14.97 * ACTM + 8.39 * ACTE + 198 * RATIO \quad \text{(X-2)}$$

B. RECOMMENDATIONS

The Department of Admissions will have to make a trade-off between the accuracy of the model and the complexity of the model. This definition of complexity includes the model's available data, ease of implementation, level of subjectivity, and inability to represent both short and long term high school success. A useful tool for the Department of Admissions might be a tradespace as seen in Figure 23.

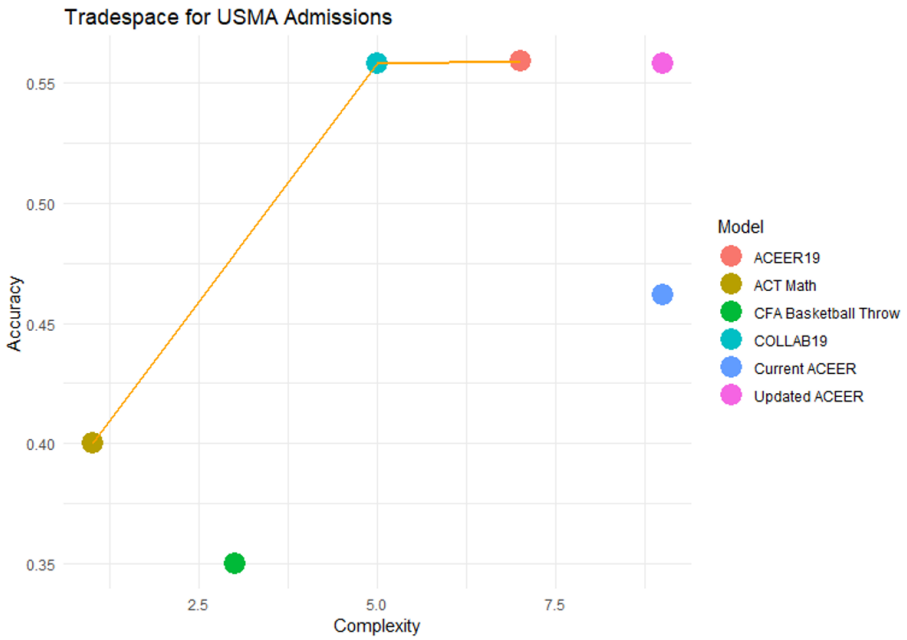

Figure 23. This figure allows the Department of Admissions to view a portfolio of optimal solutions.

In order to foster a fair and objective admissions process, our team strongly recommends the Department of Admissions adopt the equation:

$$COLLABACEER19 = -273 + 14.97 * ACTM + 8.39 * ACTE + 198 * RATIO \quad \text{(X-3)}$$

This model should be implemented for the following reasons:

1. The model is simple. The model uses only three predictors which are all readily available. The only data fields needed to calculate this metric are the ACT Math score, ACT English Score, high school GPA, and high school grading scale.
2. The model is fair. The model is fair for all high school grade scales and mitigates bias between the SAT and the ACT by using similar predictors.
3. The model is just as, if not more, accurate than the original.

LIST OF REFERENCES

[1] American College Testing. Interpreting your results. https://www.act.org/content/act/en/products-and-services/the-act/scores.html.

[2] COL Lindquist. Increased use of aceer, 2019.

[3] John Gibson. Predicting academic success for students of color within stem majors, 2013.

[4] United states military academy at west point. https://www.usma.edu/admissions/SitePages/Class%20Profiles.aspx.

[5] USMA Department of Admissions. Usma admissions training brief, 2017.

[6] COL McDonald. Personal interview, 13 September 2018.

[7] ACT CollegeBoard. Guide to the 2018 act/sat concordance, 2018.

[8] Office of Institutional Research. Report on the validation of the aceer for class of 1995 end of 4cl year.

[9] COL Oldaker. Lack of equated and validated ceer and aceer formulas, 14 November 2000.

[10] Department of Mathematical Sciences. Deans time study ay13-1 report, 16 January 2013.

[11] Hadley Wickham. tidyverse, 2017.

[12] LTC Zupancic. Personal interview, 27 November 2018.