3-21-2019

Using Simulation to Model Reserve Officer Training Corps Cadet Flow

Marydell V. Westman

Using Simulation to Model Reserve Officer Training Corps Cadet Flow

THESIS

Marydell V. Westman, Captain, USA

AFIT-ENS-MS-19-M-155

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-19-M-155

USING SIMULATION TO MODEL RESERVE OFFICER TRAINING CORPS CADET FLOW

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Marydell V. Westman, BS

Captain, USA

February 2019

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-19-M-155

USING SIMULATION TO MODEL RESERVE OFFICER TRAINING CORPS CADET FLOW

Marydell V. Westman, BS

Captain, USA

Committee Membership:

Dr. J.O. Miller
Faculty Advisor

Dr. Lance Champagne
Reader

Abstract

The Army Cadet Command oversees the Army Reserve Officer Training. It receives mission goals every year along with a budget. It is their responsibility to recruit enough cadets to fulfill mission goal and remain under budget, while striving to increase racial and gender diversity, and to meet academic and technical specialty goals. This research provides a tool that can change the numbers and percentages of incoming cadets, what year of school they enter, as well as their gender, race, and STEM or non-STEM parameters, in order to explore the composition of graduated commissioned cadets. This study finds that changing the percentage of cadets entering each year results in significant changes in the STEM and non-STEM demographics of the commissioning class. Modeling race and gender disenrollment, as well as specific scholarships, can improve this study.

Acknowledgements

I would like to take this opportunity to thank my thesis committee and my sponsor for their assistance in the completion of my thesis. First, I thank my advisor, Dr. J.O. Miller, for his encouragement, patience, and direction. Next, I would like to thank my reader, Dr. Lance Champagne, for his assistance and for reviewing my thesis. Finally I would like to thank the Army Cadet Command for their help throughout the process.

Marydell V. Westman

Table of Contents

List of Figures

List of Tables

USING SIMULATION TO MODEL RESERVE OFFICER TRAINING CORPS

I. Introduction

General Issue

As the total strength of the Army changes, so does the number of its officers. The Reserve Officer Training Corps (ROTC) accounts for the largest acquisition of new officers in the Army. The Cadet Command is in charge of administering the Army ROTC. ROTC controls both the officers entering active duty, the Army National Guard, and the U.S. Army Reserve. The Cadet Command uses scholarships to entice cadets to sign up. It is part of the Cadet Command's mission to project how many scholarships it should issue in order to meet future Army officer needs.

Problem Statement

Cadet Command wants to better understand the impact of specific policy decisions and to focus on recruitment to produce the number of officers with the desired demographics, backgrounds, and specialties that the Army requires to lead its troops into combat.

Research Objectives/Questions/Hypotheses

The objective of this research is to effectively model the demographic characteristics of individuals throughout the ROTC pipeline so that the Cadet Command can better manage student flow to meet its officer acquisition goals.

Along with modeling how cadets enter the pipeline, the departure of cadets from the ROTC program also needs to be captured. Demographic and specialty information needs to be provided for cadets who have withdrawn to ensure that recruitment can make up for those who are not expected to finish, in order to maintain the desired end-number of commissioned officers as well as an appropriate balance across specialties.

Research Focus

The focus of this research is to create a simulation to explore an appropriate mix of potential cadets to recruit at entry each phase in order to produce the numbers, demographics, and specialties that the Army requires at graduation.

Thesis Organization

This thesis is organized into five chapters. This chapter introduces the material. Chapter two provides a background and literature review that introduces the policies governing ROTC and its cadet recruitment, gives a quick overview of the data available for this research, summarizes how simulations model human behavior, and introduces some methods used to model retention and human behavior. Chapter three discusses the methodology utilized to address the research problem. This discussion includes methods of cleaning the data and developing a simulation to model the cadet flow in ROTC. Chapter four reviews results and presents findings. Chapter five provides concluding comments and suggestions for future research.

II. Background and Literature Review

Chapter Overview

This chapter provides an overview of the Reserve Officer Training Corps (ROTC) scholarship process, along with details on cadet demographics, recruitment, and reasons for leaving ROTC. Subsequently, the chapter discusses human behavior modeling and reviews previous research related to the general problem area and to the approach employed here using simulation.

Description of Cadet Flow

Students enter ROTC though multiple gates. The first input of cadets comes from the High School Scholarship Program (HSSP). High school seniors receive offers for these scholarships as part of a national effort. Active Duty enlisted Soldiers can also receive a Green to Gold scholarship or a Green to Gold Hip Pocket scholarship to attend college, which can range in duration—i.e., 4-year, 3-year, and 2-year scholarships. Once students enter college, they can participate in campus-based recruitment and can apply for a scholarship though their school in the first or second year of study. After the first year or even the second year, those who wish to join the ROTC program can do so by attending the basic camp that covers the initial military science classes they missed. Not every cadet has to receive a scholarship to be in ROTC; they can also contract without a scholarship. Those with green to gold scholarship, prior service, or alternate entry can join in the third year and attend advanced camp in order to graduate after their fourth year. With permission, cadets can do five years of college if they miss some of the required military training or need more time for their academics. A graduate level ROTC program is also available if the student has not previously attended ROTC. Figure 1 graphically depicts the cadets' ROTC entry points.

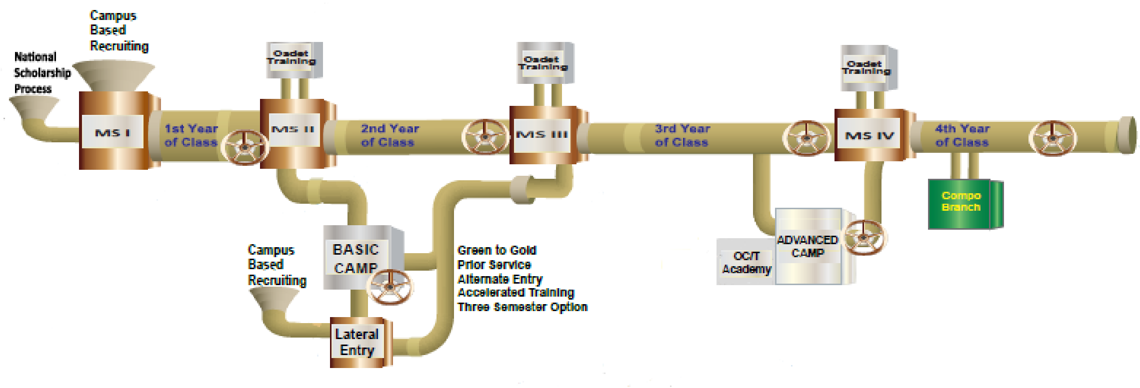

Figure 1. ROTC cadet entry points (Cadet Command, 2018)

ROTC Scholarships

Cadet Command issues national scholarships and works with civilian universities in order to recruit and commission officers for the U.S. Army. The largest recruitment method of ROTC is college scholarships. There are multiple ROTC scholarship types: High School Scholarship Program (HSSP), Campus-Based Scholarship Program, and Green to Gold Scholarship Program.

HSSP consists of 4-year and 3-year Advanced Designee scholarships, given at the national level. The Military Junior College (MJC) 2+2 Program (at designated MJCs) is a subset of the HSSP. This program is open to high school seniors, requiring four years to complete degree requirements, and to current participating MJC freshman enrolled in ROTC. Students attend a designated MJC for the first two years of study and a designated 4-year college, associated with the MJC under this program, for the last two years of study.

The College Scholarship Program consists of: 4-year, 3.5-year, and 3-year Advanced Designees scholarships; 3-year, 2.5-year, 2-year Cadet Initial Entry Training (CIET) scholarships; 2-year Advanced Designees scholarships; 2-year undergraduate degree scholarships; and 2-year graduate degree scholarships.

There are four Green to Gold scholarship options: Active Duty Option, Scholarship, Hip Pocket, and Non-Scholarship. The Active Duty Option (ADO) allows Soldiers to receive their first baccalaureate or advanced degree, provided that they have no more than two years remaining to receive their degree. A 2-year, 3-year, and 4-year scholarship option allows Soldiers to be discharged early in order to pursue their bachelor's or graduate degree. The Hip Pocket option allows participating Division and Corps Commanders to award 2-year scholarships to deserving and qualified Soldiers so that they may be discharged early to pursue a bachelor's or master's degree. The Non-Scholarship option allows Soldiers to be discharged early to pursue their first baccalaureate or advanced degree (USACC Regulation 145-1, 2016). Cadet Command also has a non-scholarship incentive. They can offer non-scholarship cadets a onetime monetary bonus for attending the basic camp (U.S. Army, 2018).

The national ROTC scholarship program has set goals to ensure that the scholarships are awarded to specific academic discipline groups. The following academic discipline groups receive a percentage of scholarships offered: Engineering 35%, Physical Science (Analytical) 25%, Technical Management 25%, and Generalist 15%. The programs that are not subject to these percentages (USACC Regulation 145-1, 2016) are: 4-year Historically Black College/University (HBCU); 4-year, 3-year, and 2-year Green to Gold; Nurses; 2-year Guaranteed Reserve Forces Duty (GRFD); 2-year Dedicated GRFD, 2-year Basic Camp, and MJC. ROTC candidate requirements are shown in Table 1.

Table 1. ROTC Prospective Candidates' Requirements (USACC Regulation 145-1)

U.S. Citizen or U.S. National
At least 17 years of age at contracting and under 31 years of age in the calendar year in which they are commissioned.
A high school graduate or possessing an equivalent certificate prior to September 1st of the year they receive the ROTC scholarship.
A minimum of 920 on the Scholastic Assessment Test (SAT) or a minimum of 19 on the American College Test (ACT) (waiverable), SAT/ACT writing skills test.
A 2.5 high school grade point average (GPA) for both the four-year applicants and the two-year MJC applicants applying from high school.
A minimum of 2.5 academic, current, and unweighted cumulative grade point average (CGPA) in college at the time of application (waiverable).
A 3.0 ROTC CGPA for cadets already enrolled in ROTC who want to be considered for scholarship.
Pass a medical examination reviewed by the Department of Defense Medical Examination Review Board (DODMERB) successfully or be granted a medical waiver by the Headquarters, U.S. Army Cadet Command.
Pass the height, weight, and body fat percentage standards IAW AR 600-9.
Complete the APFT successfully at the 60/60/60 level prior to contracting.

STEM Incentives

Since STEM majors represent 60% of the recruiting goal, this study explores previous research in identifying potential STEM major candidates. The largest contributor for students to pursue a STEM degree is their belief that they have math skills in 12th grade that were directly impacted by the math skills they obtained in 10th grade Wang (2013). As Figure 2 shows, financial aid also plays a role in a candidate's decision whether or not to apply to a STEM field of study, along with the ability to obtain a graduate degree. An equal sign shows the weights for these factors if the weights are equal across more than one race.

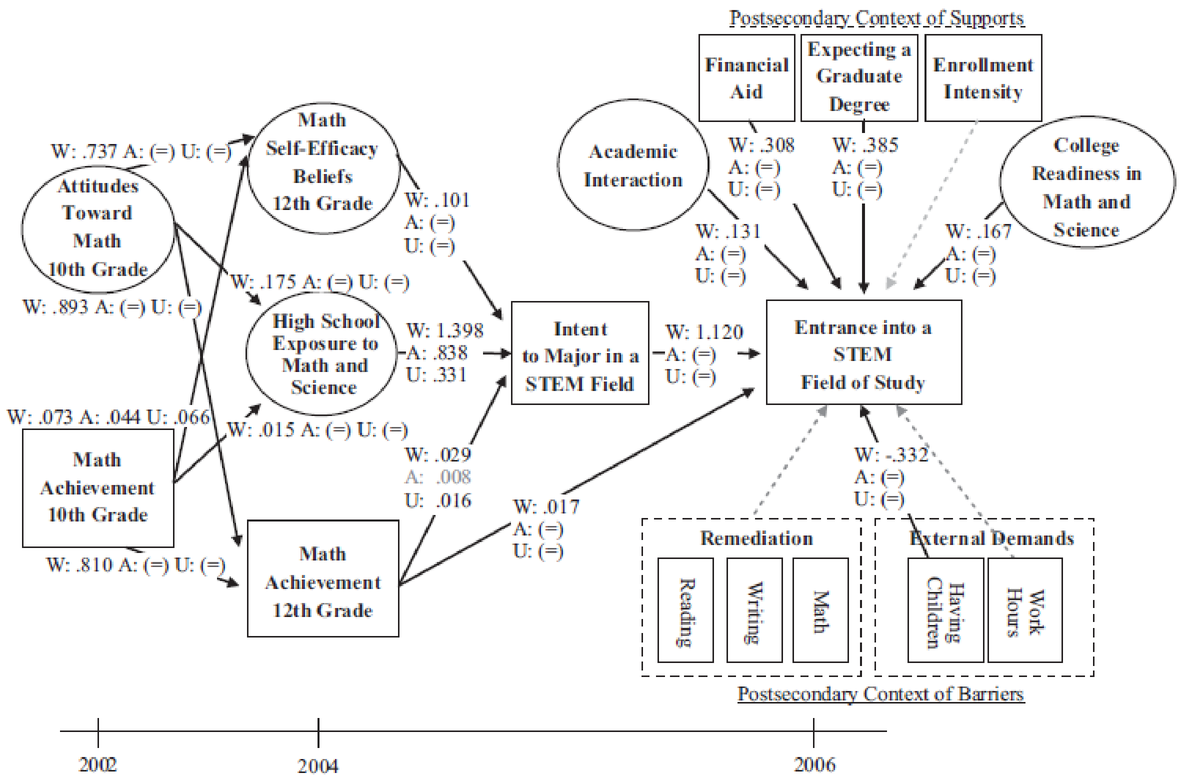

Figure 2. Results of the final multiple-group structural equation modeling (SEM) model based on race (Wang, 2013:1105).

Figure 2 does not break down race beyond white (W), Asian (A), and underrepresented minorities (U). Figure 2 does not show the statistics for women; however, Wang (2013) does find that women differ from men in that they perceived their math skills to be inferior.

Because ROTC entices students to contract and enter the Army based on scholarships, the influence of financial aid and a STEM major degree choice factors into this study. There is a positive correlation between financial aid and choosing a STEM major (Castelman, Long, and Mabel, 2018: 136-66). The analysis shows that some students who start out in a STEM program later change to a non-STEM major. This change may be related to the requirement of maintaining a higher GPA in order to maintain the scholarship. ROTC cadets must maintain a 2.5 GPA to maintain their scholarship, but the scholarship is tied to their original STEM major. The HSSP STEM major scholarship is funded at a national level, not at a college level as with the College Scholarship Program. Furthermore, the Castelman, Long and Marbel study finds that, when provided scholarships, minorities are more likely to attend college thus creating a more diverse demographic (Castelman, Long, and Mabel, 2018:136–166).

Demographics

A Congressional Budget Office (CBO) study in 1989, focusing on the social composition of the U.S. military forces, concluded that the officer corps was proportionally more representative of minorities than the enlisted corps. It is important to recruit officers who adequately represent the country's population to ensure a positive public perception of the social mix within the Army forces representing the nation. This CBO study found that African Americans were overrepresented and females were underrepresented in newly commissioned officers (*Social Representation in the U.S. Military*, 1989*).*

There is a widening racial gap as officers increase in rank. One reason for this is that there is a larger number of General Officer slots for the Combat Arms' specialties. The CBO study showed that minorities tend to choose Combat Support and Combat Service Support specialties, while white cadets tend to prefer Combat Arms. There was a correlation between occupational specialty and how far a Soldier progresses up the military career ladder. It was found that most cadets get their requested branch choice. Minorities had less mentorship once commissioned, which is hypotheses by Hall (2009) to hold them back from promotion. One conclusion of the study indicated that the Cadet Command controls branching at the cadet level and, in order to insure a distribution of minorities at higher levels in the military, the branching methodology needs to be changed (Hall and others, 2009).

A RAND study focused on the use of Standardized Scores in cadet selection. Its findings indicated that the ACT and SAT exams produce varied mean scores based on racial and ethnic differences. White students outperformed every other racial group in all areas of the test, except for Asians on the math portion of the test. This study also found no bias against racial/ethnical minorities on the ACT and SAT tests. Thus, this discrepancy may influence the selection of cadets in order to recruit the desired demographics (Giglio and others, 2012).

Recruitment of Cadets

Understanding what to highlight during the recruitment of Cadets influences the number and type of cadets a program will receive. A survey was conducted to discover what can be done to recruit military nurses. The survey was given to military members, nursing students, and the general public. The findings indicated that there is a need to build public awareness of military nurses, personal hazards of military service are exaggerated. It was also found that job security and economic benefits are a large incentive, and intellectual challenges, along with financial benefits, should be emphasized when recruiting nurses. As with nurses, general education on the job security and financial benefits is needed to help recruit potential cadets (Applebum and others, 2014).

A study by the U.S. Army Research Institute for the Behavioral and Social Sciences found that ROTC cadets had more knowledge of the Army and closer ties to the military than non-cadets in college. Information about ROTC reaches ROTC candidates though interpersonal and media-based communication. Friends, ROTC personnel on campus, and recruiters all play a role in informing personnel of what ROTC is and what the benefits are. Pamphlets, radio/T.V., magazine, and newspaper recruiting methods are also influential in enlightening students about ROTC (Gilbert and Wilson, 1983: 413–418).

Surveys given to undergraduate students and ROTC cadets found that ROTC cadets were more likely to respond positively to questions regarding self-sacrifice, attraction to public service, and commitment to public values. At the same time, ROTC cadets responded lower to compassion than non-ROTC undergraduate students. There was also a strong positive correlation between athletics and ROTC cadets. In addition, the study also found that focusing on occupational military motivations would change what is currently attracting cadets to the military (Clerkin, 2014:442–458).

Reasons for Leaving ROTC

When looking at how to commission a desired number of officers in the Army, it is necessary to consider the cadets who choose to leave ROTC. The Cadet Command published some of the reasons why cadets choose to quit the program: financial concerns about attending college or not having the time to work a part time job; skepticism about life as a commissioned officer; misconception about what the Army demands; dissatisfaction with an aspect of the ROTC program or with what life will be like as an officer; competing requirements, such as those of athletics, fraternities or sororities, and student governments; and campus perception of ROTC (CC Pam 145-37, 2008). Recorded reasons why cadets chose to disenroll included: academic work plans do not support the required ROTC classes, personnel fitness, drug or alcohol offense, medical issues, enlisting in the Army, and legal issues (CCIMM, 2018). The possible penalties include of disenrollment are the repayment of financial support or a possible involuntary call to active enlisted duty in fulfillment of contractual obligations.

Available Data

The Cadet Command provided three separate databases pulled from the CCIMM that included information about ROTC cadets from 1981 until the present. For its purposes, this research paper considers the data from only the last six years.

Human Behavior Modeling

An important factor in simulating the ROTC process flow is modeling the behavior of individual cadets. Multiple regression and factor analysis are two well-known methods for modeling but there are many others. Table 2 shows a list of model construction methods.

Table 2. Model Construction Methods (Vickers, 2004:395)

Measurement Models			
	Dimensional models		
		Exploratory	
			Exploratory factor analysis (EFA)
			Multidimensional scaling (MDS)
		Confirmatory	
			Confirmatory factor analysis (CFA)
	Categorical models		
		Exploratory	
			Exploratory cluster analysis (ECA)
			Latent class analysis (LCA)
		Confirmatory	
			Expectation-maximization mixture analysis
			Taximetrics
Path Models			
	Dimensional models		
		Exploratory	
			Regression, including multiple regression
			Analysis of variance (ANOVA)
			Hierarchical linear models (HLM)
		Confirmatory	
			Structural equation modeling
	Categorical models		
		Exploratory	
			Categorical and limited dependent variables (CLDV)
		Confirmatory	
			Taximetrics
			Latent class analysis (LCA)

One of the issues with traditional EFA, such as principle components analysis (PCA), is that it takes all factors into consideration, which can increase the complexity of the analysis (Vickers, 2014:398). An important consideration for theory formation and testing is the number of factors that should be retained. Interpretability with PCA, on its own, is not a good guide for factor reduction; however, testing methods, such as retaining components with an eigenvalue greater or equal to one, may be a good guide(Vickers, 2014). To avoid overfitting, which does not have any major issues besides wasted effort in analyzing additional factors, an upper limit should be set for how many factors should be extracted. Vickers (2014) believes that the common use of PCA should be replaced by principal factors analysis and that oblique rotation should replace orthogonal rotation. The reason that the author suggests oblique rotation is because it provides a simple structure and produces results that are easier to interpret than orthogonal rotation. The fit of the model should also be measured using the root mean square error of approximation (RMSEA) (Vickers, 2014:395). When dealing with human behavior, confirmatory factor analysis (CFA) can be better than EFA because it provides greater flexibility in model construction along with stronger testing of the models. The analyst can also impose a specific set of constraints that the CFA program then uses to estimate parameters. The drawback of using CFA is that the analyst needs to know about data before evaluating them. They need to know the number of dimensions and to make informed guesses to support each of the three basic factor analysis decisions and the indicator variables defined between latent traits. The three basic factor analysis decisions are: know the number of latent traits to measure, designate which indicator variables define each latent trait, and specify a pattern of correlations between the latent traits (Vickers, 2014:398–400).

When evaluating a human behavior model, alternative statistical models need to be considered together with the best fit. The best model by statistical criteria may not always be the most plausible model when considering human behavior. Models that are close to the best fit also need to be considered along with their plausibility. The model evaluation should also include steps to explain the relationship between statistical parameters and behavior. A human behavior model cannot explain every single human behavior but it needs to be able to explain enough to provide a reasonable estimate (Vickers, 2014).

Simulation models, whose goal is to capture multiple dimensions of human behavior—such as personality, job performance, and environmental factors that reflect real world observations, need to consider past studies and literature from behavioral scientists. Singh, Mayuri, Duggirala, Hayatnagarkar, Patel, and Balaraman (2016) explain how they were able to model human behavior in seven steps. Step zero selects the variables to use in the model. When deciding what variables to include, it is critical for past research to have already found a mathematical linkage between the variables. The first step appropriately defines the variables and ensuring that they are not highly correlated. The second step creates a measurement model that develops selection criteria for behavioral relationships from past research. Currently, most simulations use a theory-driven approach for the selection of variables. The third step produces behavior fragment sections that address the theoretically linked relationships between variables based on common definitions. Currently, a common method for linking the variables is linear regression. The fourth step uses common definitions to identify and theoretically link the fragments. The simulation model then uses these linkages to study their interactions and steady state. The fifth step ensures that the relationships between the variables discussed in the previous steps have numerical thresholds. The last step converts the variable relationships into a simulation model, which consists of: developing logic to represent human behavior for agents, configuring a process model and assigning agents to it, and finally defining the environment model and inserting the process model into it (Singh and others, 2016).

One of the main considerations when modeling human behavior is finding a mathematical model that includes only the variables of interest as well as understanding what each variable is describing. Singh and others (2016) were able to achieve a more accurate model by integrating human behavior into the model. Since their model also had to integrate multiple behaviors, these had to be researched and behavioral models had to be combined to fit into the simulation.

There is not a lot of existing literature that addresses the interpretation and translation of reported findings regarding human behavior for the use and implementation in simulation. That is why it is important to conduct an in-depth survey of the literature that shows the relationships between variables of interest in order to develop cognitive models for synthetic agents. While behavioral models do exist, their integration into simulation models is typically missing due to poor mathematical definition. Integrating human behaviors into models improves the realism of the entity or agent behavior (Silverman, 2004).

Understanding why human beings do things requires knowledge of how individuals interact with each other and how these individual interactions influence group behavior. Agent-Based Modeling (ABM) can help to capture this behavior, which is unique for different individuals but contributes to overall group behavior. ABM can enable the study of systems that have multiple interacting entities and systems that exhibit emergent properties. Simply averaging or summing the properties of the entities does not lead to emergent properties. The goals of ABM can take on six forms: empirical description, empirical prediction, normative analysis, behavioral understanding, heuristic understanding, and methodological advancement (Zacharias, MacMillian, and Hemel, 2008). For empirical description, this involves ensuring that the model regenerates the observed macro level regularities. Empirical prediction refers to seeing how the agents react to injections introduced over time. Normative analysis tries to evaluate whether the polices and institutional agreements for various types of social systems result in desirable system performance over time. Behavioral understanding using ABM has the goal of illuminating the accumulation of effects from diverse behavioral rules as well as the extent to which slight variations in behavior rules have substantial effects. The fifth form is heuristic understanding, summed up as the study of how fundamental causal mechanisms in a social system lead to greater insight. It can be hard to anticipate the full consequences of simple forms of interaction. Methodological advancement refers to the capability of experimentally generated theories to model real world data. Methodological advancement considers methodological principles in terms of the practical development of programming, visualization, and validation tools (Zacharias, MacMillan, and Hemel, 2008).

Table 3. ABM Structures (Zacharias, MacMillan, and Hemel, 2008:240)

Model	Number of Agents	Agent Representation	Cognitive Sophistication	Social Sophistication	Grid-based
Cognitive	Few	Rules	High	Low	No
Dynamic-network	Many	Equations + rules	Moderate	High	No
Cellular automata	Few to Many	Equations or Rules	Low	Low	Yes
Rule-based	Few to Many	Rules	Low	Low	Often

ABMs have different structures, as Table 3 shows. The actual level of realism depends on the degree to which agent attributes are based on actual data and the degree to which agent behavioral rules faithfully represent real world processes. Also, some of the limiting properties of ABMs are computational powers, because each agent has different properties that interact with other agents (Zacharias, MacMillan, and Hemel:2008).

Previous Research

The study presented here takes into consideration the previous research that analyzed such as the Air Force Reserve Officer Training Corps (ROTC), Turkish F-16 pilot training, human behavior influences on an Army recruiting station, and Air Force personnel retention. Multiple methods, used in previous research, all modeled systems that rely on human behavior.

Captain Marisha Kinkle (2012) created a multi-stage Optimization Model for Air Force Reserve Officer Training Corps (AFROTC) Officer Candidate Selection. Captain Kinkle (2012) examined how to select personnel to receive AFROTC scholarships and how to fill Air Force Specialty Code jobs from AFROTC. A Quiz problem, a class of stochastic scheduling problems, represents an alternative method of scholarship allocation that allows the decision maker to consider an applicant's probability of accepting a scholarship and commissioning given his or her individual composite or SAT equivalent score when offering a scholarship. Insights gained from factors contributing toward field training completion and commissioning are a take away from her research. A limit of this research is the separate consideration of every candidate's entry to the AFROTC program (Kinkle, 2012).

First Lieutenant Adem Okal (2015) of the Turkish Air Force modeled Turkish F-16 pilot training. He took into consideration some human factors, such as annual leave for the instructors. First Lieutenant Okal took an experimental design approach when setting up his simulation. It is important to understand which factors can be changed and which affect the simulation but cannot be changed. It is also important to understand how to verify the model.

The limitation of this research lies in the number and variety of human factors considered (Okal, 2015).

Captain Edward L. McLarney (1999) used simulation to model an Army recruiting station. His work shows that it is very important to first clean the data in Excel and then move onto other computer-based tools in order to gain theoretical distributions that describe behavior. When trying to fit behavior to a theoretical distribution, it is important to understand correlations and to try different models in order to reduce the unexplained variance. Modeling behavior using theoretical models is possible, although the method of data collection affects the accuracy of these models. Primary and secondary categories both influence the model and should both be considered (McLarney, 1999).

First Lieutenant Zabrina Y. Hoggard (2008) studied how to identify enlisted stay and leave population characteristics using discriminant analysis. When an internal human resource system generates a database, it is important to identify which variables are solely for internal purposes, which are extraneous, and which variables are pertinent to the research questions. Understanding how the variables interact is also important when trying to build a theoretical model. Once the correlations are found, then the variables can be screened again to find the ones that are going to create a more accurate model. Not every variable needs to be used as this creates noise that reduces the model accuracy (Hoggard, 2008).

Summary

This background and literature review examined the ROTC policies regarding cadet flow, some of the current research on using simulation to model human behavior, and past research that modeled retention. Multiple approaches were discussed for modeling human behavior, with many using some type of computer simulation as done in this research.

III. Methodology

Chapter Overview

This chapter provides a detailed description of the methodology developed to clean the data, exploring the data, and finding mathematical models at different points of entry and exit for individual cadets. In order to simulate this, the Army ROTC pipeline becomes a multi-stage model with three inputs: freshman year, sophomore year, and junior year. These three inputs include the National Scholarship Process, campus-based recruiting during freshmen year, and campus-based recruiting up to junior year. This chapter considers the disenrollment of applicants and discusses simulation development, verification, and validation. Results and analysis follow in Chapter 4.

Data Cleaning and Conceptual Model Validation

The Cadet Command provided three databases and one workbook that included: a onetime entry of every cadet who was in ROTC with 197 columns of information per cadet, a list of cadets who disenrolled together with 13 columns of information per cadet, a yearly snapshot of all cadets in the ROTC program with 43 columns of information per cadet, and eight worksheets with definitions for the codes used. The data in these spreadsheets covered the time period from 1981 to the present. Multiple database changes occurred during this time, making it difficult to ensure a comparison of the same variables from year to year. The data chosen from this database for the purposes of the research presented in this paper dates from 2011 onwards, when the data began to stabalize due to U.S. recovery from the recession and the drawdown in manpower requirements resulting from policy changes in Iraq and Afghanistan.

After deciding on the time-frame for data consideration, the selection of variables for the simulation took place based on the input from the thesis sponsor and following an evaluation of the dataset. The variables chosen for examination were: gender, race, scholarship type, and scholarship type, as Table 4 shows.

Table 4. Data Variables

Gender	Male
	Female
Race	African American
	Hispanic
	Native American
	White
	Asian
	Other
Scholarship type	non-Stem Scholarship
	STEM Scholarship

The database grouped the cadets' majors into five different categories: Generalist, Tech Management, Physical Sciences, Engineering, and Nursing. In order to simplify the model, the study grouped the variables into either STEM or non-STEM. The majors included in the STEM category were Physical Sciences and Engineering, while the non-STEM category contained the other three majors. A search for the characteristics of the cadets who chose to disenroll, performed using a VBA code, matched their employee identification numbers (EMPID) from the disenrollment spreadsheet to the yearly cadet rollup sheet. The VBA code further separated the cadets' first year from the yearly rollup in order to obtain the characteristics of the changing cadet body by year. Then R code was used to determine how many cadets in each year had the variable of interest. This information became an input into the simulation model. Some of the records were missing data under the race variable

REDCAT, so the author used the race_cd variable to translate the race into a REDCAT variable. The military science (MS) year determined the education level (year of education) for each cadet upon entering ROTC. Data was not consistent for year of education and resulted in the following assumptions: If a cadet was in MS1 they could only be a 4-year scholarship or non-scholarship cadet. A cadet who entered at MS2 could be a 4-year, 3-year, or non-scholarship cadet. A cadet who entered at MS3 could either be a 3- or 2-year scholarship or non-scholarship cadet. Since the number of cadets who entered at MS4 was small, the study's author decided to regard them as late enrollees in MS3 and to add them to that count.

In order to discover what year a cadet disenrolled, the study sorted the yearly rollup and referenced only the last entry for each cadet ID to the disenrollment excel sheet. The MS year was used to classify the number of years a cadet was in ROTC. The date cadets started ROTC was not consistent in the database; thus, using their MS year was the best way to categorize what year of college the cadets were in.

The study then employed the summary statistics gathered using the R code to calculate the percentages of the variables of interest for seven years, subsequently obtaining an average of these percentages (as Appendix A, B, and C shows) to use as input for the simulation model.

Conceptual Model Validation

Performing an examination of the conceptual model determined that the assumptions underlying the model were correct and that the model's logic and mathematical relationships are "reasonable" for the intended purpose of the model. The conceptual model for the simulation came from averaging the seven years of data provided by the CCIMM. An examination of this data found that there was an issue with the enrollment and disenrollment data, as Table 5 shows.

Table 5. Entry Data

year	enrollment	disenrollment	graduated	Enrollment – disenrollment
2011	16,936	15,631	5,467	1,305
2012	16,819	17,024	5,798	-205
2013	15,005	13,799	5,524	1,206
2014	16,551	15,398	5,406	1,153
2015	16,218	15,156	5,590	1,062
2016	11,876	7,260	5,538	4,616
2017	13,140	15,631	4,440	-2,491
average	15,220.71	14,271.29	5,394.714	949.4286
stand dev	1,844.838	2,993.051	406.1761	

From Table 5 it is apparent that there was an issue in how the CCIMM accounted for enrollments and disenrollments. The number of cadets enrolling did not equal the number of cadets disenrolling and graduating. The Cadet Command provided a chart that showed historic trends of enrollments and graduations. As the chart in Figure 3 shows, modifying the disenrollment numbers used in the model to reflect the enrollment numbers minus the graduated numbers for each year provided the percentage of cadets that disenrolled each year, which was necessary in order for the model to have a reasonable graduation number. The yearly enrollments were input into the model as a triangular distribution with a percentage of cadets entering in freshman, sophomore, and junior years. This approach provided a valid conceptual model of the ROTC process.

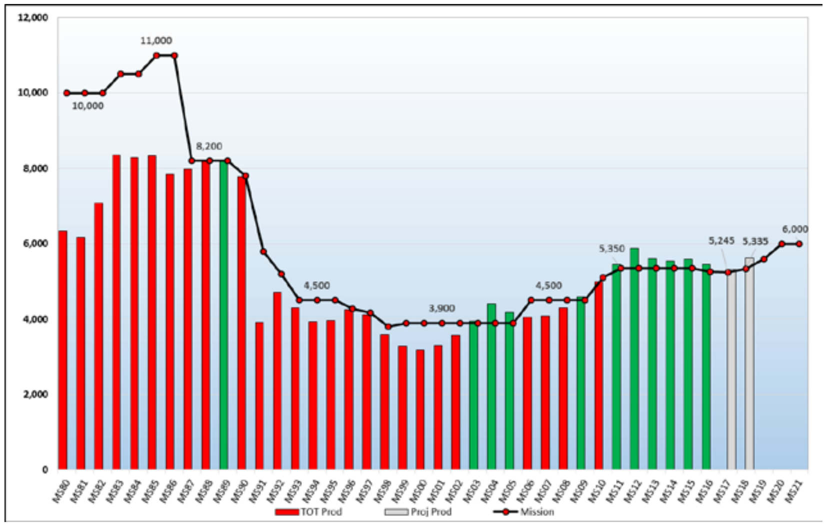

Figure 3. Historical missions and productions (USACC Intro to NPS, 11OCT18).

Assumptions

While developing the simulation to represent the real world Army ROTC cadet flow, this research made several assumptions. First, the military science year determined which year of school the cadet was in. There were cases of some cadets who might have entered their senior year if they attended a military school in which military science classes were mandatory, but these numbers would be statistically insignificant. Thus, the junior year percentages included all cadets shown entering their senior year in the CCIMM data. Of the five different categories of academic majors, only those classified as Physical Sciences and Engineering counted as STEM scholarships. The simulation tracked and processed STEM and non-STEM cadets separately throughout. It is assumed that race and gender demographics are independent of scholarship types and the simulation only tracked them individually at enrollment and graduation stages. The numbers of cadets entering ROTC included those that have a scholarship and those who only took an ROTC class. Because the simulation included both non-scholarship and non-contracted cadets, who may have never planned to commission, the disenrollment numbers are very large. The simulation based the incoming and disenrollment percentages on STEM and non-STEM scholarships. The percentage of cadets with STEM and non-STEM major and no scholarship, were within three percentage points of those for cadets with a scholarship. Therefore, using the STEM and non-STEM scholarship percentages is reasonable. The model did not take into account all the different scholarship codes but did include cadet input at all standard entry points. The research captured those cadets who changed their scholarship type from STEM to non-STEM at the end of the senior year.

Simulation Methodology

The objective of this research study was to create a simulation for the number and demographics of cadets that commission from ROTC, with the ability to modify those parameters that are of interest to Cadet Command. The primary focus was on changing the numbers and the demographic mix at various cadet entry points, shown in Figure 1, in order to produce the desired number and demographics of graduating cadets. For the demographic mix, this study only captured the STEM and non-STEM percentages explicitly in its disenrollment logic.

Creating the Simulation

The simulation modeled each cadet as an entity in SIMIO, a simulation production planning and scheduling software. The author used variables from the provided data to assign appropriate attributes to each cadet. The study then assigned the percentages for race, gender, and STEM/non-STEM scholarships as Table 6 shows; as entity properties immediately following entity creation. For the placement of individual cadets into the ROTC process, the study used all the demographic properties, while it only handled the STEM/non-STEM scholarship cadets separately for disenrollments. Table 7 shows all the demographic disenrollment percentages pulled from the CCIMM. Statistics are still collected on all other demographic factors for graduating and disenrolling cadets.

Table 6. Demographic Enrollment Data

Year	Gender		Race: Redcat						Scholarship Type				Discipline	
	F	M	B-African American	H-Hispanic	O-Other	R-Native American	W-White	Y-Asian	0-no Scholarship	2-2yr	3-3yr	4-4yr	non-STEM	STEM
Freshman	0.2901	0.7099	0.1928	0.122	0.0599	0.0025	0.5833	0.0395	0.9109	0	0.0002	0.0889	0.703	0.297
Sophomore	0.2589	0.7411	0.1728	0.12	0.0589	0.0017	0.5991	0.0475	0.9171	0.0014	0.0183	0.0632	0.751	0.249
Junior	0.2243	0.7757	0.1785	0.1007	0.0459	0.0014	0.6342	0.0393	0.6095	0.0839	0.0905	0.2161	0.8332	0.1668

Table 7. Demographic Disenrollment Data

Year	Gender		Race: Redcat						Scholarship Type				Discipline	
	F	M	B-African American	H-Hispanic	O-Other	R-Native American	W-White	Y-Asian	0-no Scholarship	2-2yr	3-3yr	4-4yr	non-STEM	STEM
Freshman	0.2901	0.7099	0.1928	0.1220	0.0599	0.0025	0.5834	0.0395	0.9109	0.0000	0.0002	0.0889	0.7455	0.2545
Sophomore	0.2589	0.7411	0.1728	0.1200	0.0589	0.0017	0.5991	0.0475	0.9171	0.0014	0.0183	0.0632	0.7405	0.2595
Junior	0.2243	0.7757	0.1785	0.1007	0.0459	0.0014	0.6341	0.0393	0.6094	0.0839	0.0905	0.2161	0.7969	0.2031
Senior	0.2039	0.7961	0.1228	0.0991	0.0397	0.0028	0.6847	0.0509	0.3937	0.1490	0.1281	0.3292	0.8404	0.1596

Simulation Flow

The study examined the enrollment data and selected a random triangular distribution to model the number of cadets entering the simulation. It set the mode at 15,220, roughly representing the mean from 2011–2017. The minimum (11,800) and maximum (17,000) represented the smallest and largest enrollment numbers from 2011–2017, rounded up to the nearest hundred. The data for each year determined the percentages of the incoming cadets, along with their STEM/non-STEM, male/female, and racial demographics. The study tracked disenrollments separately for STEM and non-STEM cadets after each year, segregating an appropriate percentage of the disenrolled cadets from those continuing to the next year, as Figure 4 shows.

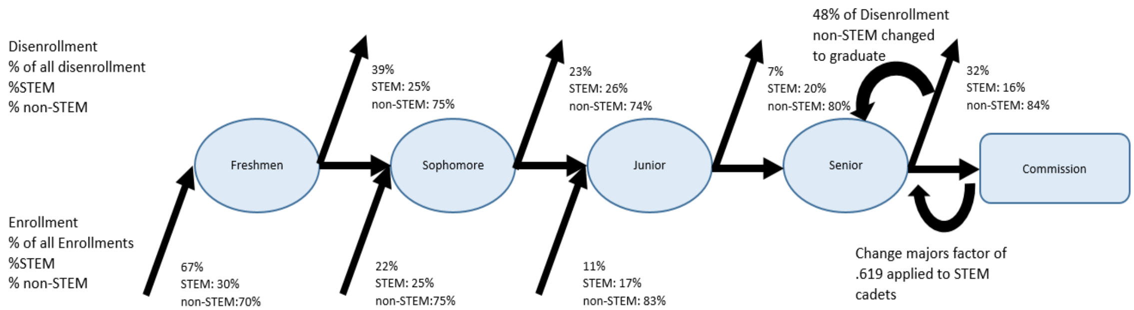

Figure 4. Cadet flow in simulation with percentages.

The data made it difficult to identify the number of cadets that changed from a STEM to a non-STEM major in specific years of their schooling. Consequently, the study applied a factor of .619 at the end of the senior year to account for cadets changing their major from STEM to non-STEM. To derive this factor that handled the STEM graduates well, the study used data found in the CCIMM. However, an additional factor was needed to similarly adjust the number of non-STEM graduates. To accomplish this, a factor was added to direct 48% of non-STEM cadets identified as disenrollments to continue on to graduation and commission.

The number of replications for the study were selected to achieve a 95% half-width that was the same level of magnitude as the variation in the CCIMM data. The initial replication number started at three, then increased to five, incremented by fives, and ended up with a total of fifteen replications producing the desired half width. Each random number draw within the simulation was assigned a unique random number stream in order to use common random numbers for variance reduction between replications.

Model Verification

The study also examined model behavior by decreasing the number of entries to 10 and conducting a step function with a stop at disenrollment and graduation. Thus, the study verified the logical flow of cadets by following the entries, as Figure 4 shows.

The model ran for 11 years, including a four-year warm up period. The length of the warm up period resulted from the fact that it takes at least three years for cadets to appear in each year of the process. Thus, the four-year warm up length allowed for a full set of cadets at each year in the system. The study collected a full seven years of data to match the seven years of historical data used.

Additional verification of the model took place in steps for each demographic variable. The first one verified was STEM and non-STEM, followed by gender and race. A visual examination of the model outputs determined whether they followed the coded logic. Since the study tracked disenrollments only by the STEM and non-STEM variables, the gender and race demographics for graduation and disenrollments did not match the historical data as it was not expected to be equal. Additional logic could be added to the simulation to track disenrollments by gender and race as well.

Model Validity

In order to ensure the simulation model's graduation numbers for STEM and non-STEM scholarship cadets reasonably compare to the real ROTC process, the study compared them to the data found in the CCIMM. To do so, it used tally statistics and a confidence interval from the simulation, comparing it to the CCIMM data.

Table 8. SIMIO Output Compared to CCIMM Data

Metric	SIMIO Output for 7 Years		CCIMM Data 2011-2017
Graduation	**Mean**	**95% Confidence Interval**	**Value**
Total	36,000	(35,403 36,597)	37,763
STEM	8,557	(8,427 8,687)	7,302
STEM Pct	23.77%	(23.40%, 24.13%)	19.34%
non-STEM	27,443	(26,971 27,915)	30,461
non-STEM Pct	76.23%	(74.92%, 77.54%)	80.66%

The data from the CCIMM for 2011-2017 contained a significant variation from year to year in both enrollment numbers and graduation numbers. For validation, the study modeled the input distribution for annual enrollments as a triangular distribution, using the mean (15,220) instead of the median (16,218). The minimum (11,800) and maximum (17,000) represented the smallest and largest enrollment numbers, rounded up to the nearest hundred. This distribution had a significantly larger lower tail, resulting in overall lower enrollment numbers within the simulation than those shown in the CCIMM data. Hence, the statistically significant (at the 95% level) lower overall graduation numbers from the simulation were not surprising. For total graduates, the upper limit of the simulation confidence intervals are just over 1000 less than the CCIM value (~3%), a practically insignificant difference. The simulation was not as close in capturing the expected mix of STEM and non-STEM scholarship cadets graduating. The results for percentages were still within 4% of the CCIMM data (STEM percentages higher and non-STEM percentages lower). These results were close enough to provided positive validation results for using the simulation to examine the mix of STEM and non-STEM cadets through the ROTC system. For analysis, the next chapter uses a desired enrollment number (picked from the historical data) as the mode in a triangular distribution, with the maximum and minimum set at plus or minus 10%.

Summary

This methodology chapter introduced the data used, explained the data cleaning process, and outlined the assumptions made in cleaning the data. It also discussed the making of the simulation as well as model verification and validation. Furthermore, the chapter addressed the overall model assumptions and logic, along with their impact on the analysis. The next chapter uses this simulation to vary the numbers of cadets entering at various points, along with the changes in the demographic mix, and an analysis of the results.

IV. Analysis and Results

Chapter Overview

This chapter describes the analysis conducted with the simulation by changing the input parameters for the percentage of scholarship cadets per year and the overall number of cadets receiving a scholarship. The baseline model used in this chapter is based on the model validated in Chapter Three. The author assumes that the corresponding percentages of disenrollment attributes remain consistent with Chapter Three. This chapter discusses the expected change in output of cadets graduating and commissioning from changes in the inputs.

Designing the Parameters for Analysis with the Simulation Model

As mentioned in the previous chapters, the goal of this project was to create a model that could be used to determine how many cadets need to be recruited at each year in order to obtain the required number and demographic mix (just STEM vs non-STEM currently) of graduates. To do this, this study conducts two separate analyses. For each analysis, the study employs all the original simulation input parameters developed from the CCIMM, except those modified for the particular experiment. The first analysis increases the incoming cadets to discern how this impacts the number of cadets who graduate. This analysis should provide insight regarding the level of increase that is necessary to reach an overall goal in terms of the number of graduates, as well as the desired mix of STEM and non-STEM majors, using the historical demographics. The second analysis changes the percentage of the cadets entering each year, which also changes the mixture of STEM and non-STEM majors since there are separate percentages for scholarship types for each entry year. The study maintains these at the values derived from the CCIMM, but they could also be varied for additional analysis if desired. Each analysis uses a length of eight years, with a four-year warm up period for each replication, running a total of 15 replications.

Changing Input Quantities of Cadets

The mission and the number of required graduates change from year to year. The purpose of this first analysis is to explore how changes in the admission numbers change the expected numbers of graduates. This research does this by creating a baseline simulation and then creating another scenario that increases the number of cadets entering, thus exploring how such an increase impacts the ultimate number of graduates. All paired-t confidence intervals use a 95% confidence.

The baseline model selected for both analyses uses a triangular distribution with a mode of 16,551 cadets, minus or plus 10%, for the minimum (14,896) and maximum (18,206). The mode chosen was how many cadets enrolled in ROTC in 2014. Setting the minimum and maximum to plus or minus 10% of the mode provides a reasonable range from a desired yearly enrollment. For the experiment, the mode increases by 2,000 (18,551), setting the minimum (16,696) and maximum (20,406) respectively, to minus or plus10% from the mode. A comparison between the baseline and the experiment takes place, using a paired-t test, and Table 9 displays the results.

Table 9. SIMIO Results for Baseline and 2000 Enrollment Increase

Metric	Baseline	Increase	Difference	95% Paired-t
Disenrollment	Mean	Mean	Mean (Increase-Baseline)	Confidence Interval (Increase – Baseline)
Total	47,972	54,030	6058	(5,488 6,628)
STEM	6,682	7,482	800	(721 878)
STEM Pct	14%	14%	-	-
non-STEM	41,290	46,548	5258	(4,744 5,773)
non-STEM Pct	86%	86%	-	-
Graduation				
Total	18,042	20,255	2,213	(2,084 2,342)
STEM	4,316	4,842	526	(463 588)
STEM Pct	24%	24%	-	-
non-STEM	13,726	15,413	1687	(1,600 1,773)
non-STEM Pct	76%	76%	-	-

None of the confidence intervals contain zero, so there is a significant statistical impact of recruiting 2,000 more cadets a year. With a mode of a 2,000 cadet increase per year (8,000 over four years) entering ROTC, the graduating numbers rose by an average of 553 cadets each year (2,213 over four years). This follows the CCIMM data where it was found that approximately 27% of cadets that enter ROTC continue on to graduate and commission. Note also that the percentage of STEM and non-STEM disenrollments and graduates do not change, as expected. The next experiment looks at indirectly changing these percentages for enrollments.

Changing Input Percentages of Cadets

The year in which a cadet enters ROTC and receives a scholarship affects how much the Cadet Command spends. Based on this fact, it is important to explore the percentages of cadets entering each year in order to understand how this affects the percentages of STEM and non-STEM cadets.

The distribution for the number of incoming cadets is the same as the baseline in the first analysis. Baseline percentages for entering scholarship year are the same as those used in Chapter 3 for the original model: freshmen: 67%, sophomores: 22%, and juniors: 11%. Since the freshmen scholarships are over four years, they are the most expensive. In addition, a scholarship student entering as a freshmen also has a longer time over which they could change from a STEM to a non-STEM major. For the experiment, freshman scholarships were set at 50% of yearly enrollments, with the remaining 50% spread in a 20%-30% mix between sophomore and junior scholarships.

Table 10 shows the resulting data from changing the percentages for the scholarship years. The overall number of disenrollment's and graduates can be seen from the table. The difference between the two experimental mixes, illustrates a larger percentage of cadets disenrolling during their sophomore year than junior year as shown in Table 7. Table 7 also shows that freshman and sophomore year have almost identical disenrollment percentages, but sophomores have fewer STEM majors entering which can also be seen in Table 10. However, moving the freshmen entries to junior year, produces about a 5% increase in disenrollments for non-STEM scholarships in the CCIMM data. This matches the increase seen with both mixes for increased disenrollments from the baseline. Clearly the increase in disenrollments for the experimental mixes result in a subsequent decrease in the number of graduates for the mixes. Table 11 provides a summary of results for three 95% paired-t confidence intervals between the baseline and two experimental mixes.

Table 10. SIMIO Results for Changing in Scholarship Year

Metric	Baseline	30% Sophomore 20% Junior	20% Sophomore 30% Junior
Disenrollment	Mean	Mean	Mean
Total	47,972.00	48,566.00	49,033.00
STEM	6,682.00	6,260.00	5,995.00
STEM Pct	14%	13%	12%
non-STEM	41,290.00	42,305.00	43,038.00
non-STEM Pct	86%	87%	88%
Graduation			
Total	18,042.00	17,423.00	17,125.00
STEM	4,316.00	4,065.00	3,998.00
STEM Pct	24%	23%	23%
non-STEM	13,726.00	13,358.00	13,127.00
non-STEM Pct	76%	77%	77%

Table 11. Paired-t Results: Mean Difference/95% C.I.

Metric	Baseline – 20%/30% Mix	Baseline – 30%/20% Mix	20%/30% Mix-30%/20% Mix
Disenrollment			
Total Diff	1,061	594	467
Total C.I.	(646 1,475)	(29 1,158)	(-1,000 65)
STEM Diff	687	422	265
STEM C.I.	(611 763)	(345 498)	(181 349)
non-STEM Diff	1,748	1,015	733
non-STEM C.I.	(1,400 2,095)	(517 1,513)	(269 1,197)
Graduation			
Total Diff	917	619	298
Total C.I.	(770 1,065)	(448 791)	(142 453)
STEM Diff	318	251	67
STEM C.I.	(279 356)	(200 302)	(24 109)
non-STEM Diff	599	368	231
non-STEM C.I.	(477 722)	(241 496)	(98 365)

Table 11 shows that the disenrollments between the 30%/20% and 20%/30% mixes contain zero in there confidence intervals. All other mixes do not contain zero, indicating statistically significant changes. The total disenrollments between 30%/ 20% and 20%/30% are not statistically different; however, the disenrollments are significantly different when looking at the STEM/non-STEM majors individually. Almost 700 of the commissioned graduating cadets would no longer be STEM. The Cadet Command would then have to consider how much should be spent to receive the desired commissioned demographics. By changing the percentages of cadets entering in sophomore and junior years, the percentages of STEM disenrollments are significantly different from each other. There is a larger impact between the baseline and the mix due to the 17% reduction in freshmen enrollments for each mix.

Summary

This chapter contained two different analyses, both of which used the paired-t test to evaluate whether the changes to the base model make a statistically significant impact on the number of graduating cadets. The insight achieved through increasing the numbers or mix of entering cadets and the percentages of cadets given a scholarship depending on the year in which they enter ROTC is going to benefit the Army Cadet Command by improving its understanding of how many cadets and in what year should be given a scholarship. The results stated in this chapter support the fact that the year in which the scholarships are given and the number of recruited cadets change the overall numbers of cadets in the STEM and non-STEM majors. Chapter five provides recommendations for future research.

V. Conclusions and Recommendations

General

This chapter reviews the key points of this simulation study followed by recommendations for further research topics. This study examined the element that Cadet Command can easiest control, scholarships. It explored cadet entry and exit points, based on whether they belonged to a STEM or non-STEM major. The simulation model provided an initial tool for modeling how cadet flow changes based on changing mission numbers and the school year which STEM and non-STEM scholarship cadets enter.

Conclusion

This research study shows that simulation can effectively model the cadet flow though ROTC. This research began by exploring the CCIMM data and determining the variables of interest. The CCIMM data from Cadet Command used in this study contained a lot of information that needed to be mined for the sake of relevance. One of the largest research obstacles for this study lay in understanding the data that the CCIMM provided and how to clean and interpret it in order to create the simulation input variables. The data proved to be unclean and the recruitment of cadets had additional outside influences, such as the economy, that the simulation does not model. This simulation considered the entry and exit points of ROTC cadets during their undergraduate education. The data from the CCIMM validated the simulation.

Once the study created and validated the simulation, the analysis phase began. During this phase, the author explored changes in initial cadet entry points as well as the percentages of cadets receiving a scholarship at which year of their college experience. Changing the number of cadets entering ROTC with a 2,000-cadet annual increase produced a significant change in the number of graduates, but the overall demographic makeup of the cadets did not change. Changing the percentage of cadets entering during the sophomore and junior years resulted in a statistically significant change in the number of graduates as well as the mix. If cadets enter in their junior year, they are less likely to disenroll from ROTC, however, cadets are less likely to have a STEM major.

Recommendations for Future Research

To improve the simulation model, future research should add non-scholarship cadets and logic that changes non-scholarship to scholarship cadets based on the CCIMM data. Thus, research could examine various combinations of contracted scholarships over time versus non-scholarship cadets in order to provide insight on the demographics, costs, and risks associated with incentives given to students to join ROTC. A more vigorous model could add race and gender demographics to the disenrollment data. To create a more robust simulation, extracting the race and gender demographics from the data, along with the type of scholarship, would allow for a model to show how changing specific scholarships affects the demographic make-up of the graduating cadets. Not only do scholarships need increased fidelity into types for the model to be more representative, but also adding logic to the disenrollment demographics is necessary in order for the model to give an accurate prediction of those variables. Focusing just on STEM/non-STEM scholarship cadets, data (or the means to pull this data from CCIMM) needs to track when a particular cadet changes from a STEM to a non-STEM major during a particular school year. Not just the total number of STEM or non-STEM cadets enrolled in each year, but specific individual changes of major. This would significantly improve the simulation by being able to more accurately model disenrollments each year.

Summary

The recruitment of people is a complex job with many factors that affect the students' willingness to join ROTC. This model provides a basic tool to model how changing the cadet entry year affects the graduating class and how increasing the initial recruitment of cadets results in an expected change in graduation numbers.

Appendix A. Graduation Data Rollup

Graduation																	
sclr_award			gender		Race: Red cat						scholarship type: sclr_award_cd					Academic discipline: acad_disc_mix_cd	
Year	None	Scholar ship	F	M	B-African American	H-Hispanic	O-Other	R-Native American	W-White	Y-Asian	none	2yr	3yr	4yr	5yr	non-STEM	STEM
2011	1571	3896	1097	4370	569	490	205	8	3840	313	1585	798	847	2195	42	4576	891
	0.29	0.71	0.20	0.80	0.10	0.09	0.04	0.00	0.70	0.06	0.29	0.15	0.15	0.40	0.01	0.84	0.16
2012	1775	4023	1238	4560	610	488	240	11	4126	274	1806	784	457	2695	56	4828	970
	0.31	0.69	0.21	0.79	0.11	0.08	0.04	0.00	0.71	0.05	0.31	0.14	0.08	0.46	0.01	0.83	0.17
2013	1969	3555	1141	4383	574	465	195	10	3892	289	2110	415	742	2313	44	4525	999
	0.36	0.64	0.21	0.79	0.10	0.08	0.04	0.00	0.70	0.05	0.38	0.08	0.13	0.42	0.01	0.82	0.18
2014	2064	3342	1164	4242	549	492	215	10	3839	292	2091	534	523	2215	42	4405	1001
	0.38	0.62	0.22	0.78	0.10	0.09	0.04	0.00	0.71	0.05	0.39	0.10	0.10	0.41	0.01	0.81	0.19
2015	1914	3676	1181	4409	633	556	208	15	3851	323	1638	905	1194	1468	86	4511	1079
	0.34	0.66	0.21	0.79	0.11	0.10	0.04	0.00	0.69	0.06	0.29	0.16	0.21	0.26	0.02	0.81	0.19
2016	1743	3795	1172	4366	629	542	201	10	3795	361	43	738	1278	1687	83	4232	1306
	0.31	0.69	0.21	0.79	0.11	0.10	0.04	0.00	0.69	0.07	0.01	0.13	0.23	0.30	0.01	0.76	0.24
2017	1151	3289	994	3446	443	448	47	39	3116	347	36	624	900	1708	56	3384	1056
	0.26	0.74	0.22	0.78	0.10	0.10	0.01	0.01	0.70	0.08	0.01	0.14	0.20	0.38	0.01	0.76	0.24
mean	0.32	0.68	0.21	0.79	0.11	0.09	0.03	0.00	0.70	0.06	0.24	0.13	0.16	0.38	0.01	0.81	0.19
stand dev	0.04	0.04	0.01	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.16	0.03	0.06	0.07	0.00	0.03	0.03

Appendix B. Disenrollment Data Rollup

Disenrollment																
year	ms_cls _enrol	gender		Race: Red cat						scholarship type: sclr_award_cd					Academic discipline: acad_disc_mix_cd	
		F	M	B- African America n	H- Hispa nic	O- Other	R-Native American	W- White	Y- Asian	0-non Scholarship	2-2yr	3-3yr	4-4yr	5-Advanced degree scholarship	Non- STEM	STEM
2011	1	1632	4220	1057	762	340	16	3457	219	5061	0	0	791	0	4441	1411
		0.28	0.72	0.18	0.13	0.06	0	0.59	0.04	0.86	0	0	0.14	0	0.76	0.24
2011	2	814	2622	535	418	230	9	2073	141	3040	7	75	313	0	2555	881
		0.24	0.76	0.16	0.12	0.07	0	0.6	0.04	0.88	0	0.02	0.09	0	0.74	0.26
2011	3	278	916	202	128	51	4	754	51	618	106	83	379	0	958	236
		0.23	0.77	0.17	0.11	0.04	0	0.63	0.04	0.52	0.09	0.07	0.32	0	0.8	0.2
2011	4	996	4153	506	425	184	9	3700	287	1591	798	740	1978	0	4416	733
		0.19	0.81	0.1	0.08	0.04	0	0.72	0.06	0.31	0.15	0.14	0.38	0	0.86	0.14
2012	1	1756	4451	1160	722	335	11	3724	255	5824	0	1	382	0	4788	1419
		0.28	0.72	0.19	0.12	0.05	0	0.6	0.04	0.94	0	0	0.06	0	0.77	0.23
2012	2	1008	2869	690	452	243	3	2297	189	3539	1	35	302	0	3304	873
		0.24	0.69	0.17	0.11	0.06	0	0.55	0.05	0.85	0	0.01	0.07	0	0.79	0.21
2012	3	266	791	208	97	66	1	621	51	604	84	88	275	0	859	198
		0.25	0.75	0.2	0.09	0.06	0	0.59	0.05	0.57	0.08	0.08	0.26	0	0.81	0.19
2012	4	1135	4448	529	469	229	8	4033	271	1901	829	374	2426	0	4746	837
		0.2	0.8	0.09	0.08	0.04	0	0.72	0.05	0.34	0.15	0.07	0.43	0	0.85	0.15
2013	1	1448	3614	993	540	324	16	3013	175	4756	0	0	306	0	3846	1216
		0.29	0.71	0.2	0.11	0.06	0	0.6	0.03	0.94	0	0	0.06	0	0.76	0.24
2013	2	710	2197	515	316	164	4	1731	176	2764	0	38	105	0	2198	709
		0.24	0.76	0.18	0.11	0.06	0	0.6	0.06	0.95	0	0.01	0.04	0	0.76	0.24
2013	3	162	628	144	76	44	1	496	29	492	48	46	203	0	642	148
		0.21	0.79	0.18	0.1	0.06	0	0.63	0.04	0.62	0.06	0.06	0.26	0	0.81	0.19
2013	4	1011	4032	495	442	180	12	3559	255	1817	469	669	2046	0	4226	814
		0.2	0.8	0.1	0.09	0.04	0	0.71	0.05	0.36	0.09	0.13	0.41	0	0.84	0.16
2014	1	1651	4273	1280	682	366	12	3357	226	5600	0	0	324	0	4393	1531
		0.28	0.72	0.22	0.12	0.06	0	0.57	0.04	0.95	0	0	0.05	0	0.74	0.26
2014	2	813	2372	629	358	174	2	1857	164	3011	0	48	126	0	2373	812
		0.26	0.74	0.2	0.11	0.05	0	0.58	0.05	0.95	0	0.02	0.04	0	0.75	0.25
2014	3	177	809	182	80	45	1	652	26	682	97	94	102	0	770	216
		0.18	0.82	0.18	0.08	0.05	0	0.66	0.03	0.69	0.1	0.1	0.1	0	0.78	0.22
2014	4	1041	4262	495	466	215	9	3844	270	2278	511	436	2037	0	4396	907
		0.2	0.8	0.09	0.09	0.04	0	0.72	0.05	0.43	0.1	0.08	0.38	0	0.83	0.17
2015	1	1725	3792	1092	769	342	15	3043	256	5015	0	0	502	0	3966	1551
		0.31	0.69	0.2	0.14	0.06	0	0.55	0.05	0.91	0	0	0.09	0	0.72	0.28
2015	2	903	2606	665	477	185	6	2024	152	3252	1	89	167	0	2502	1007
		0.26	0.74	0.19	0.14	0.05	0	0.58	0.04	0.93	0	0.03	0.05	0	0.71	0.29
2015	3	201	602	157	97	22	0	489	38	476	86	121	115	0	617	186
		0.25	0.75	0.2	0.12	0.03	0	0.61	0.05	0.59	0.11	0.15	0.14	0	0.77	0.23
2015	4	1080	4247	588	515	208	17	3682	313	2029	841	1091	1277	0	4421	906
		0.2	0.8	0.11	0.1	0.04	0	0.69	0.06	0.38	0.16	0.2	0.24	0	0.83	0.17
2016	1	1324	2908	810	490	260	9	2361	174	3430	0	4	357	0	3000	1232
		0.31	0.69	0.19	0.12	0.06	0	0.56	0.04	0.81	0	0	0.08	0	0.71	0.29
2016	2	621	1691	390	303	131	5	1336	115	1960	12	52	150	0	1597	714
		0.27	0.73	0.17	0.13	0.06	0	0.58	0.05	0.85	0.01	0.02	0.06	0	0.69	0.31
2016	3	126	452	88	58	26	0	382	18	406	37	62	66	0	462	116
		0.22	0.78	0.15	0.1	0.04	0	0.66	0.03	0.7	0.06	0.11	0.11	0	0.8	0.2
2016	4	33	106	37	24	7	1	65	5	77	33	17	10	0	114	25
		0.24	0.76	0.27	0.17	0.05	0.01	0.47	0.04	0.55	0.24	0.12	0.07	0	0.82	0.18
2017	1	1632	4220	1057	762	340	16	3457	219	5061	0	0	791	0	4441	1411
		0.28	0.72	0.18	0.13	0.06	0	0.59	0.04	0.86	0	0	0.14	0	0.76	0.24
2017	2	814	2622	535	418	230	9	2073	141	3040	7	75	313	0	2555	881
		0.24	0.76	0.16	0.12	0.07	0	0.6	0.04	0.88	0	0.02	0.09	0	0.74	0.26
2017	3	278	916	202	128	51	4	754	51	618	106	83	379	0	958	236
		0.23	0.77	0.17	0.11	0.04	0	0.63	0.04	0.52	0.09	0.07	0.32	0	0.8	0.2
2017	4	996	4153	506	425	184	9	3700	287	1591	798	740	1978	0	4416	733
		0.19	0.81	0.1	0.08	0.04	0	0.72	0.06	0.31	0.15	0.14	0.38	0	0.86	0.14

Appendix C: Incoming Cadet Data Rollup

Incoming Cadets																
		gender		Race: Red cat						scholarship type: sclr_award_cd					Academic discipline: acad_disc_mix_cd	
year	ms_cls _enrol	F	M	B-African American	H-Hispanic	O-Other	R-Native American	W-White	Y-Asian	0-non Scholarship	2-2yr	3-3yr	4-4yr	5-Advanced degree scholarship	Non-STEM	STEM
2011	1	2758	8379	1589	1162	542	25	7368	448	7868	0	0	3267	0	8277	2858
		0.25	0.75	0.14	0.10	0.05	0.00	0.66	0.04	0.71	0.00	0.00	0.29	0.00	0.74	0.26
2011	2	852	2704	531	427	178	11	2220	181	3025	20	391	118	2	2772	784
		0.24	0.76	0.15	0.12	0.05	0.00	0.62	0.05	0.85	0.01	0.11	0.03	0.00	0.78	0.22
2011	3	449	1650	311	242	102	8	1319	116	1294	626	87	45	47	1795	304
		0.21	0.79	0.15	0.12	0.05	0.00	0.63	0.06	0.62	0.30	0.04	0.02	0.02	0.86	0.14
2011	4	19	127	16	9	3	0	106	12	104	20	8	10	4	121	25
		0.13	0.87	0.11	0.06	0.02	0.00	0.73	0.08	0.71	0.14	0.05	0.07	0.03	0.83	0.17
2012	1	2677	8336	1577	1020	542	19	7341	514	9285	1728	0	0	0	8229	2784
		0.24	0.76	0.14	0.09	0.05	0.00	0.67	0.05	0.84	0.16	0.00	0.00	0.00	0.75	0.25
2012	2	930	2920	606	464	256	4	2316	202	3533	0	157	160	0	3032	818
		0.24	0.76	0.16	0.12	0.07	0.00	0.60	0.05	0.92	0.00	0.04	0.04	0.00	0.79	0.21
2012	3	387	1396	318	206	86	5	1063	105	1220	409	73	40	41	1513	270
		0.22	0.78	0.18	0.12	0.05	0.00	0.60	0.06	0.68	0.23	0.04	0.02	0.02	0.85	0.15
2012	4	31	142	18	15	6	0	128	6	119	41	1	10	1	150	23
		0.18	0.82	0.10	0.09	0.03	0.00	0.74	0.03	0.69	0.24	0.01	0.06	0.01	0.87	0.13
2013	1	2429	7901	1549	987	498	22	6827	445	8716	0	0	1610	0	7513	2817
		0.24	0.76	0.15	0.10	0.05	0.00	0.66	0.04	0.84	0.00	0.00	0.16	0.00	0.73	0.27
2013	2	756	2363	590	373	161	5	1812	176	2771	0	271	70	0	2391	728
		0.24	0.76	0.19	0.12	0.05	0.00	0.58	0.06	0.89	0.00	0.09	0.02	0.00	0.77	0.23
2013	3	314	1146	280	179	68	1	852	79	1043	306	50	21	29	1239	221
		0.22	0.78	0.19	0.12	0.05	0.00	0.58	0.05	0.71	0.21	0.03	0.01	0.02	0.85	0.15
2013	4	21	75	21	4	2	0	65	4	78	2	5	3	0	77	19
		0.22	0.78	0.22	0.04	0.02	0.00	0.68	0.04	0.81	0.02	0.05	0.03	0.00	0.80	0.20
2014	1	2711	8348	1782	1137	611	19	7045	464	9156	0	1	1902	0	7795	3264
		0.24	0.75	0.16	0.10	0.05	0.00	0.64	0.04	0.83	0	.00	0.17	0	0.70	0.29
2014	2	949	2637	683	401	172	7	2099	223	3130	0	283	167	0	2658	928
		0.26	0.74	0.19	0.11	0.05	0.00	0.59	0.06	0.87	0.00	0.08	0.05	0.00	0.74	0.26
2014	3	401	1440	385	252	83	6	1014	101	997	562	183	24	70	1554	287
		0.22	0.78	0.21	0.14	0.05	0.00	0.55	0.05	0.54	0.31	0.10	0.01	0.04	0.84	0.16
2014	4	18	47	11	12	1	0	37	4	43	12	3	6	0	56	9
		0.28	0.72	0.17	0.18	0.02	0.00	0.57	0.06	0.66	0.18	0.05	0.09	0.00	0.86	0.14
2015	1	2905	7822	1619	1257	606	20	6721	504	8443	0	0	2283	0	7336	3391
		0.27	0.73	0.15	0.12	0.06	0.00	0.63	0.05	0.79	0.00	0.00	0.21	0.00	0.68	0.32
2015	2	1018	2621	647	515	180	3	2114	180	3115	2	354	167	1	2656	983
		0.28	0.72	0.18	0.14	0.05	0.00	0.58	0.05	0.86	0.00	0.10	0.05	0.00	0.73	0.27
2015	3	450	1315	376	263	64	2	932	128	901	565	209	14	75	1438	327
		0.25	0.75	0.21	0.15	0.04	0.00	0.53	0.07	0.51	0.32	0.12	0.01	0.04	0.81	0.19
2015	4	19	68	22	11	4	0	41	9	62	18	2	2	2	73	14
		0.22	0.78	0.25	0.13	0.05	0.00	0.47	0.10	0.71	0.21	0.02	0.02	0.02	0.84	0.16
2016	1	2319	6167	1142	870	430	14	5356	418	5523	6	15	2051	0	5502	2984
		0.27	0.73	0.13	0.10	0.05	0.00	0.63	0.05	0.65	0.00	0.00	0.24	0.00	0.65	0.35
2016	2	714	1766	445	328	122	5	1325	152	1784	29	338	109	0	1776	703
		0.29	0.71	0.18	0.13	0.05	0.00	0.53	0.06	0.72	0.01	0.14	0.04	0.00	0.72	0.28
2016	3	242	631	158	128	42	0	440	75	432	268	11	1	0	681	192
		0.28	0.72	0.18	0.15	0.05	0.00	0.50	0.09	0.49	0.31	0.01	0.00	0.00	0.78	0.22
2016	4	7	31	10	5	2	0	20	1	23	8	2	0	0	34	4
		0.18	0.82	0.26	0.13	0.05	0.00	0.53	0.03	0.61	0.21	0.05	0.00	0.00	0.89	0.11
2017	1	2456	6279	1130	945	516	11	5512	423	6097	5	25	2608	0	5605	3130
		0.28	0.72	0.13	0.11	0.06	0.00	0.63	0.05	0.70	0.00	0.00	0.30	0.00	0.64	0.36
2017	2	905	2187	533	391	162	5	1678	196	2443	37	465	145	0	2228	864
		0.29	0.71	0.17	0.13	0.05	0.00	0.54	0.06	0.79	0.01	0.15	0.05	0.00	0.72	0.28
2017	3	342	908	245	171	71	4	619	100	472	480	194	18	0	1001	249
		0.27	0.73	0.20	0.14	0.06	0.00	0.50	0.08	0.38	0.38	0.16	0.01	0.00	0.80	0.20
2017	4	16	47	18	4	2	0	33	5	40	16	5	0	0	49	14
		0.25	0.75	0.29	0.06	0.03	0.00	0.52	0.08	0.63	0.25	0.08	0.00	0.00	0.78	0.22

Bibliography

Applebaum, Sandra, DesRoches, Catherine, Donelan, Karen, Hinshsw, Ada Sue, Romano, Carol, Schoneboom, Bruce A., Ward, Johanna. "National Surveys of Military Personnel, Nursing students, and the Public: Drivers of Military Nursing Careers," *Military Medicine*, Vol 179, 565-571 (May 2014).

Army Cadet Command. *Army Reserve Officers' Training Corps Retention Guide*. CC Pam 145-37 Rev: 03 December 2008.

Cadet Command Information Management Module (CCIMM), "CDT Disenroll TRAC", pulled 10 May 2018.

Castleman, Benjamin L., Long, Bridget T., Mabel, Zachary "Can Financial Aid Help to Address the Growing Need for STEM Education? The Effects of Need-Based Grants on the Completion of Science, Technology, Engineering, and Math Courses and Degrees" *Journal of Policy Analysis and Management*, Vol. 37, No. 1: 136-166 (Winter 2018).

Clerkin, Richard M., Ngaruiya, Katherine M., Taylor, Jami Kathleen, Velez, Anne-Lise Knox. "Public Service Motivation and Institutional-Occupational Motivations Among Undergraduate Students and ROTC Cadets," *Public Personnel Management*, Vol 43(4), 442-458. (2014)

Congress of the United States Congressional Budget Office, *Social Representation in the U.S. Military*. Washington: Government Printing Office, 1989.

Giglio, Kate, Haddad, Abigail, Keller, Kirsten M., Lim, Nelson, McInnis, Brian, Sims, Carra S., Wong, Anny. *The Use of Standardized Scores in Officer Career Management and Selection*. Santa Monica: RAND National Defense Research Institute. 2012.

Gilbert, Arthur C. F., Wilson, Lucy B. *Career Attitudes of ROTC Cadets and College Students*. U.S. Army Research Institute for the Behavioral and Social Sciences, 413-418 (25 Mar 1983).

Hall, Kimberly Curry, Lim, Nelson, Marquis, Jefferson P., Schulker, David, Zhuo, Xiaohui. *Officer Classification and the Future of Diversity Among Senior Military Leaders: A case Study of the Army ROTC*. Santa Monica: RAND National Defense Research Institute. 2009.

Hemel, Susan B. , MacMillan, Jean, Zacharias, Greg L. (Eds). (2008). Behavioral Modeling and Simulation From Individuals to Societies. Washington, D.C.: The National Academies Press

Hoggard, Zabrina Y. *Identifying Enlisted Stay and Leave Population Characteristics with Discrimnate Analysis*. MS thesis, AFIT/GOR/ENS/02-08. School of Engineering and

Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2002.

Kinkle, Marisha T. *A Multi-Stage Optimization Model for Air Force Reserve Officer Training Corps Officer Candidate Selection*. MS thesis, AFIT/OR/ENS/16-16. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2012

Law, Averill M. *Simulation Modeling and Analysis, Fourth Edition*. McGraw-Hill. 2007.

McLarney, Edward L. Using Simulation to Model the Army Recruiting Station with Multi-Quality Prospects. MS thesis, AFIT/GOR/ENS/99M-10. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 1999.

Silverman, Barry G., Toward Realism in Human Performance Simulation. J. Ness, V. Tepe, D. Ritzer (Eds), *The Science and Simulation of Human Performance, Advances in Human Performance and Cognitive Engineering Research, Volume 5* (pp 469-489). Oxford, England: Elselvier Ltd.

Singh, Meghendra, Duggirala, Mayuri, Hashal Hayatnagarker, Sachin Patel, and Balarman, Vick, Towards Fine Grained Human Behavior Simulation Models. Roeder T., Frazer P., Szechtman R., Zhou, E., Huschka T., and Chick S. (Eds), *Proceedings of the 2016 Winter Simulation Conference. (3452-3463).*

Okal, Adem. Simulation Modeling and Analysis of F-16 Pilot Training Squadron. MS thesis, AFIT/GOR/ENS/08-15. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2008.

U.S. Army. "ROTC Basic Camp Bonus, Earn Cash for Completing ROTC Basic Camp." Goarmy.com. Retrieved on 2 May 2018.

United States Army Cadet Command. *Army Reserve Officers' Training Corps Army ROTC Incentives Policy*. USACC Regulation 145-1. Rev. 2 August 2016

Vickers, Ross R. Jr., Statistics and the Art of Model Constuction. J. Ness, V. Tepe, D. Ritzer (Eds), *The Science and Simulation of Human Performance, Advances in Human Performance and Cognitive Engineering Research, Volume 5* (pp 393-467). Oxford, England: Elselvier Ltd.

Wang, Xueli "Why Students Choose STEM Majors: Motivation, High School Learning, and Postsecondary Context of Support" *American Educational Research Journal*, Vol. 50, No. 5: 1081-1121 (October 2013).