



**CREATING DATA-DRIVEN DECISIONS IN ALLOCATING C-17 AIRCREW
TRAINING**
GRADUATE RESEARCH PAPER

Christopher M. Stevens, BA
Major, USAF

AFIT-ENS-MS-22-J-053

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

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

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the United States Government and is not subject to copyright protection in the United States. Additionally, this report discusses performance in terms of capabilities. All scores or results referring to capability are a measure created by the author in order to quantify performance. These values are in no way tied to any form of real-world readiness capabilities or ratings.

AFIT-ENS-MS-22-J-053

CREATING DATA-DRIVEN DECISIONS IN ALLOCATING C-17 AIRCREW
TRAINING

GRADUATE RESEARCH PAPER

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 Management

Christopher M. Stevens, BA

Major, USAF

May 2022

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

AFIT-ENS-MS-22-J-053

CREATING DATA-DRIVEN DECISIONS IN ALLOCATING C-17 AIRCREW
TRAINING

Christopher M. Stevens, BA
Major, USAF

Committee Membership:

Lt Col Jesse Wales, PhD
Chair

Abstract

The United States Air Force is currently in a largescale paradigm shift concerning how capability requirements are viewed and assessed. Increased emphasis on accurately measuring aircrew performance and forecasting training requirements has forced senior leaders to rethink whether the training programs currently in place genuinely achieve their desired effects. Specifically, the process of training resource allotment lacks the ability to quantify the amount of resources required to achieve a necessary level of performance or the flexibility needed to be altered with changes in resources. This study proposes a methodology to identify factors that have a significant impact on aircrew performance. Variables are identified in C-17 simulator data utilizing linear regression to develop a prediction model. Using this model, different courses of action are created where resources can be modified in order to achieve a desired capability. While the final model utilized within this study produces mixed results, the concepts, methodology, and proposed application provide insights into the value of using models like those described in this study in creating data-driven decisions in training resource allocation.

Acknowledgements

I want to express my personal gratitude to all those who helped make this research possible. First, to my academic advisor for all the countless phone calls, draft versions, and spreadsheets you helped me through. Your guidance and mentorship were crucial in getting me through this project. Additionally, my sincerest thanks to my sponsor for his support and encouragement in sponsoring this research.

This paper would not be possible without the expertise offered by the Air Mobility Command Staff. Furthermore, assistance from the members out pushing the mission was instrumental in pointing this research in the right direction. I would be remiss if I left out my Advanced Studies of Air Mobility class. Thank you to every one of you for shaping not only my paper, but also the future leader I hope to become thanks to your influence.

Last, and most importantly, to my family. Thank you for your endless support, incredible patience, and countless sacrifices you both make. I am unbelievably appreciative to have you in my corner.

Major Christopher M. Stevens

Table of Contents

	Page
Abstract	iv
Acknowledgements	v
List of Figures	viii
List of Tables	ix
List of Equations	x
 I. Background and Research Problem	 1
1.1 General Issue	1
1.2 Problem Statement	3
1.3 Research Questions	3
1.4 Overall Research Methodology	3
1.5 Summary of Findings	4
 II. Literature Review	 5
2.1 Overview	5
2.2 Defining Commonly Used Terms	5
2.3 Current C-17 Aircrew Training Program	6
2.3.1 <i>Overall Structure</i>	6
2.3.2 <i>How Training Requirements are Made</i>	7
2.3.3 <i>Flight Training Level (FTL)</i>	8
2.3.4 <i>Currency and RTM</i>	8
2.4 C-17 Aircrew Flight Training Resources	9
2.5 AF Force Generation (AFFORGEN)	11
2.6 Previous Training Reform Initiatives	13
2.7 Current C-17A Training Initiatives	15
2.8 Regression Analysis	17
2.8.1 <i>What Data Does Regression Provide</i>	17
2.8.2 <i>Uses of Regression Analysis</i>	19
 III. Methodology	 21
3.1 Overview	21
3.2 Assumptions and Limitations	21
3.3 Establishing the Sample	23
3.4 Calculating a Dependent Variable - ATS	24
3.5 Cleaning the Independent Variables	25

3.6 Descriptive Statistics	28
3.7 Simple Linear Regression	28
3.8 Multiple Regression	29
3.9 Assessing Model Performance	29
IV. Results and Analysis	32
4.1 Overview	32
4.2 Descriptive Statistics	32
4.3 Simple Linear Regression	36
4.3.1 Overall Findings	36
4.3.2 Selection of Significant Variables	37
4.3.3 Interpreting Significant Indicator Variables	40
4.3.4 Interpreting Insignificant Indicator Variables	42
4.3.5 Limitations of Simple Linear Regression Applications	43
4.3.6 Implications and Takeaways	44
4.4 Multiple Variable Regression	45
4.4.1 Overall Findings	45
4.4.2 Proposed Model	46
4.4.3 Application of Model	49
4.5 Model Effectiveness	53
4.5.1 Findings	54
4.5.1 Implications	55
4.6 Summary	55
V. Conclusion	56
5.1 Research Conclusions and Significance	56
5.2 Recommendations for Future Actions and Research	58
Bibliography	59

List of Figures

Figure 1. Air Force Flying Hour Model (AFI11-102, 2020, 6).	11
Figure 2. AFFORGEN Model (Secretary of the Air Force Public Affairs, 2022).	12
Figure 3. Study's Overall Methodology Flow.	21
Figure 4. Confusion Chart Example	30
Figure 5. ATS Box and Whisker Plot	33
Figure 6. Distribution of Aircrew Tactics Scores.	34
Figure 7. Model Confusion Matrix Results	54

List of Tables

Table 1. Independent Variables and Descriptions	26
Table 2. Mean ATS for Given Independent Variable Subcategories	36
Table 3. Simple Linear Regression Independent Variables' Significance Levels.....	38
Table 4. Simple Linear Regression Coefficient Values.....	41
Table 5. Multiple Variable Regression Coefficients.....	47
Table 6. Possible Aircrew Tactics Score Based off Combination of Independent Variables	48
Table 7. Example Decision Matrix Using Final Model Values.....	50
Table 8. Sample ATS Decision Matrix with no <i>Tactics Planning</i> Available.....	51
Table 9. Sample Calculation of Amount of Time to Complete a Given Combination of Training.....	52
Table 10. Comparison of ATS and Time Required to Achieve.....	53

List of Equations

Equation 1: Simple Linear Regression Model	17
Equation 2: Multiple Variable Regression Model	17
Equation 3: Aircrew Tactical Score Formula	25
Equation 4: Estimated Simple Linear Regression Model	29
Equation 5: Estimated Multiple Variable Regression Model	29
Equation 6: Confusion Matrix Accuracy Formula.....	30
Equation 7: Confusion Matrix Sensitivity Formula.....	31
Equation 8: Confusion Matrix Specificity Formula.....	31
Equation 9: Final ATS Prediction Model	47

CREATING DATA-DRIVEN DECISIONS IN ALLOCATING C-17 AIRCREW TRAINING

I. Background and Research Problem

1.1 General Issue

The United States Air Force (USAF) is currently in a largescale paradigm shift concerning the way units are deployed and their overall mission capability requirements. The recent transition of focus from decades of operations in the Middle East towards competition with near-peer adversaries (Everstine, 2021) by the 2018 National Defense Strategy has compelled Airmen to challenge the status quo as they look to find new and innovative ways to project US military might (Mattis, 2018). Likewise, USAF Chief of Staff General Brown charged all Airmen in accelerating this change with the ominous understanding of wide national security impacts without a corresponding increase in USAF capabilities (Brown, 2020).

The need for Rapid Global Mobility in future conflicts has specifically been identified as a crucial capability as “without a credible, interoperable mobility force capable of projecting and sustaining Joint Force combat power at strategically relevant speed, scale, and distance, the United States cannot compete in the modern world” (Van Ovost, 2021, 1). Air Mobility Command (AMC) is responsible with providing the USAF core competency of Rapid Global Mobility (AFDP 3-36, 2019) but must now adapt to operate in environments with multiple-domain threats in order to provide “resilient and agile logistics” (Mattis, 2018, 7).

While AMC has made strides in the right direction over the past two years, the 2021 Government Accountability Office (GAO) report on the Department of Defense's abilities to operate in a contested environment assessed that there is a large gap between AMC's current versus desired capabilities and recommended 59 areas for improvement (GAO, 2021). One common theme from the report is a continued need for intentional training aimed at cultivating individual aircrew competency in hopes of improving overall flying unit capabilities (GAO, 2021).

Additionally, the USAF currently lacks the ability to gauge an individual aircrew member's capability. In a 2020 GAO report on training effectiveness, Air Combat Command's aircrew training program was unable to evaluate capability accurately due to a lack of ability to quantify proficiency measures. The effect of this deficiency is that the USAF is unable to project training or quantify resourcing requirements, which produces a "system [that] does not lend itself to analysis that could be used to gain efficiencies" (GAO, 2020)

The problem is further exacerbated for future AF leaders with the recent push to the AF Force Generation (AFFORGEN) model. In this construct, commanders are required to make decisions concerning the resources and training required for their unit to achieve a set level of readiness capability by a given time. Units and their corresponding training programs are then evaluated through a certification event that will determine their ability to accomplish their assigned mission (Secretary of the Air Force Public Affairs, 2022). Due to a lack of relevant data, these resourcing decisions are largely based off the opinions of others and "gut instinct."

1.2 Problem Statement

Unit commanders and their respective training programs lack the required information to make an accurate, data-driven decision concerning allocation of resources. Specifically, there is a need for an understanding of the relationship between allocating a given resource and its effect on performance in order to ensure members receive access to the training required for success. Through this research, the author proposes a process for unit commanders to enable real-time, data-driven decisions based on their unit's current capabilities in order to drive resource allocation to achieve a desired end state.

1.3 Research Questions

It is the ultimate goal of this research to answer the following questions:

- 1) Given a method to quantify aircrew proficiency and performance, can the magnitude, direction, and effects of training resources and crew demographics be measured to identify statistically significant indicator variables of performance?
- 2) Can performance be predicted by measured indicator variables?
- 3) Can a model be constructed that provides alternative avenues for allocation of crew resources while maintaining a minimum performance requirement?

1.4 Overall Research Methodology

This research proposes a methodology to identify factors that have a significant impact on aircrew performance. Variables are identified in C-17 simulator data utilizing various techniques of linear regression to develop a prediction model. Using this model, different courses of action are created where independent variables can be modified in order to achieve a desired capability. The final model's validity in predicting passing and

failure scores is assessed using a confusion matrix to compare predicted scores to actual scores.

1.5 Summary of Findings

Overall, single and multiple variable linear regression produce only a few statistically significant models. However, a great deal of the information and interpretation gathered from both the significant and insignificant models is useful for the given problem set. The final selected model, while it only explains a small amount of the variance in scores, does provide a great deal of practical significance to a decision maker.

II. Literature Review

2.1 Overview

This literature review provides an outline of the current structure, processes, and expectations of the C-17 aircrew training program. An understanding of program weaknesses and problems frame areas where improvements should be made. Additionally, the exploration of civilian and military aviation training reform efforts provides different methodologies that can be used to affect change. Lastly, discussion of linear regression provides relevant background on the terminology, usage, and efficacy of the studies' selected methodology.

2.2 Defining Commonly Used Terms

A resource is any asset given or needed by a unit or individual in order to accomplish a set task or mission. There are four categories of resources: personnel, equipment availability, equipment condition, and training (AFI10-201, 2018). For the purposes of this study, a resource is considered any academic, simulator, or flight training event along with its obligatory personnel and equipment that can be allocated to an individual or crew.

Capability is an assessment of ability. Generally, this is assigned by a commander on a unit's ability to accomplish their assigned mission (AFI10-201, 2018). Individual capability is the ability to effectively complete the mission of their assigned unit. Discussion on crew capability will be in terms of ability to accomplish the desired learning objectives of the training event and does not relate to a readiness capability.

Currency is defined in accordance with AFMAN11-2C-17V1 as either the volume or frequency an event must be accomplished through “Continuation Training” in order to be considered mission ready within a given period. If a person is unable to accomplish the prescribed amount of training events within the required time period, then that individual is considered non-current (AFMAN11-2C-17v1, 2020).

Proficiency for events refers to the ability to accomplish a task correctly and within the given parameters (AFMAN11-202v1, 219). In order to be considered “proficient,” an individual must be able to perform the task to their required skill-level.

2.3 Current C-17 Aircrew Training Program

2.3.1 Overall Structure

The C-17 aircrew training program is directed by three main documents: AFI 11-202V1 *Aircrew training*, AFMAN11 2C-17V1 *C-17 Aircrew Training* (commonly referred to as the “Vol 1”), and the *C-17 Ready Aircrew Program (RAP) Tasking Memo (RTM)*. Both Vol 1 documents set the framework for how a unit’s training program is to be structured, key definitions of processes, and waiver authorities with the goal of creating a training program that creates and maintains proficient aircrew members (AFMAN11-202v1, 2019). The RTM, on the other hand, is much more specific on what an individual is required to accomplish as far as currency requirements (AFMAN11-2C-17v1, 2020). The combination of these documents drives the focus areas for each unit’s training program.

2.3.2 How Training Requirements are Made

The creation of aircrew training requirements begins with the identification of some need or capability. Once identified, AMC's aircrew training division creates a plan to accomplish that goal based off which skills they believe to be required. This draft plan is provided to the subordinate wings for review and comments. The wings' comments are collected and adjudicated into one document that becomes the new working version. That working version is then boarded in front of a group of training subject matter experts where it is discussed to ensure that everyone agrees with the method (Becker, 2022; AFMAN11-2C-17v1, 2020).

The main issue with this process is the lack of data within the planning process. As previously mentioned, there is a verified need for a skill, however, the current process relies heavily on opinion in order to fix new problems. Quantities and frequencies are often determined from "gut decisions" that can vary based off who is there to discuss them (Becker, 2022). This leads to other issues later because these currencies become not only the basis of aircrew training programs, but also one of the main ways aircrew training resources are requested and allocated. After a training requirement is validated, training resources are allotted based off a unit's makeup. Contributing factors include the number of members (crew ratio), the experience level of those within the unit, and what the command's pilot seasoning requirements are (AFI11-102, 2020).

2.3.3 Flight Training Level (FTL)

FTL is used as an indicator of a member's overall experience, which determines the minimum number of training events they must accomplish. Overall, a member's FTL is determined by the following process:

- 1) Members are first divided into two categories: are you an aircraft commander (MP) or a copilot (FP)
- 2) Within the two categories there are subgroups of A-C with A being considered the most senior and C being the most junior. (AFMAN 11-2C-17v1, 2020)

Determining FTL is very subjective with only a small amount of details on the subject within the Vol 1. In general, the process is considered more of an “art” than a science (Becker, 2022). As noted in Chapter 1, this is an area with very little literature and garners further explanation into how the designation is determined.

2.3.4 Currency and RTM

Depending on which overall FTL an aircrew member is assigned will determine the quantity or frequency of an event that crew member will have to accomplish (AFMAN11-2C-17v1, 2020). Generally, those designated as less experienced are required to accomplish more training items and vice versa. The combination of the number of FTLs within a unit and the training requirements of each FLT determines the total amount of training events required. A unit's overall training item totals are then used to determine the number of allocated training resources (AFI11-102, 2020).

2.4 C-17 Aircrew Flight Training Resources

Aircrew training resources generally fit into one of three categories: ground (academic), simulator, and flight training. Each category has a corresponding currency requirement with it. Ground training is commonly thought of as baseline or refresher training with focus primarily on the academic learning of the individual (AFMAN11-2C-17v1, 2020). In general, ground training requires fewer resources such as funds, equipment, or instructor staffing to accomplish versus those required for simulator and flight training.

Since 2018, Boeing has been the contracted provider for C-17 simulator training for both pilots and loadmasters (McCullough, 2018). The contract includes everything from the equipment of simulators, parts/software, and the instruction/instructor, and boasts of robust training within a realistic environment that either cannot be reproduced in the actual aircraft or is available at a much cheaper price (McCullough, 2018). From a financial aspect, the simulator is a force multiplier. In FY20, it is projected that the C-17 simulator contract saved the USAF a total of \$462.4M and over 4,000 training sorties (A3TR, 2021).

While the simulator can be used to accomplish various training events, this study specifically looked at the Mobility Air Forces (MAF) Mission Profile (MMP) simulator. This training event is concerned with preparing crews for the high-end fight through complex integration tasks and overall emphasis on the employment of tactics, techniques, and procedures required for the given threat. A third party ensures that training profile

and desired learning objective requirements are accomplished administers the MMP for all USAF C-17 locations (AFMAN11-2C-17v1 2020).

The Flight Hour Program (FHP) is directed by AFI 11-102 with the goal of appropriately distributing resources in the form of flight training hours in order to “train aircrews to safely operate aircraft while sustaining them in numbers sufficient to execute the core tasked mission as well as experiencing or aging requirements for aircrew management” (AFI11-102, 2020, 3). It is, on average, a \$9 to \$10 billion program that drives not only training, but also maintenance programs, aircraft upgrade timelines, and the overall parts supply system (Mills, 2018).

The number of hours a particular unit is allotted is calculated through an assessment of Major Weapon System (MWS) crew ratio, unit special mission sets, specific aircrew training requirements, and current seasoning requirement (AFI11-102, 2020). This is ultimately a balancing of maintaining today’s readiness through aircrew currency with future force capability requirements in the form of aircrew seasoning and development (Wilcox, 2021). Crew ratio (also known as force structure) is established in accordance with AFI 65-503, *Financial Management*, and is a standardized value based off the number of crew members per aircraft assigned. This is held constant across all units (AFI65-503, 2018).

Aircrew seasoning is concerned with developing new aircrew members into the next experienced qualification (AFI11-102, 2020). For pilots, it is primarily focused on taking a brand new copilot and upgrading them into an aircraft commander. As a standard, calculations are based off an agreed upon ratio that is held constant for all bases

(A3TR, 2021). An MWS average seasoning requirement is established in the form of hours per month that a member must fly.

The last component used to determine flying hour allocation is FTL and Vol 1 currency requirements. As previously discussed, FTL is used to indicate the overall experience of an individual and, in turn, determines how often or how many training events they need to accomplish (AFI11-102, 2020). For flight hour computations, unit FTL breakdowns (aircrew data) are calculated to determine how many total training events (requirements) are necessary which are combined with all the previous variables in order to get a total requirement (AFI11-102, 2020). Figure 1 provides a visual of the overall flow of the process, with each previous step informing the requirements of the subsequent steps.

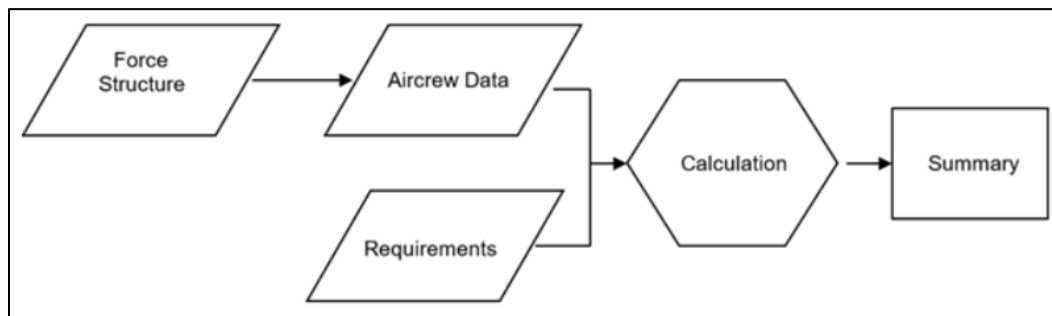


Figure 1. Air Force Flying Hour Model (AFI11-102, 2020, 6).

2.5 AF Force Generation (AFFORGEN)

AFFORGEN is the USAF's proposed model for force development and deployment. In this construct, units are assigned to one of four buckets that correspond to an overall requirement for that phase (Everstine, 2021). Figure 2 provides a visual representation of each phase and the tasks accomplished in that period. In the past,

members were trained and deployed to execute their mission as an individual assigned to a deployed unit (Secretary of the Air Force Public Affairs, 2022). This led to inefficiencies in training and allocation of resources (Secretary of the Air Force Public Affairs, 2022; Everstine, 2021) as units were often trying to fill the tasking versus looking at the holistic view of readiness of who to send (Parker, 2012). The new AFFORGEN model instead deploys whole units together in order to execute the mission that they are trained and equipped for to accomplish (Everstine, 2021) which offers greater utilization of individual skillsets (Parker, 2012) and personal job satisfaction while deployed (Kirby, 2000).

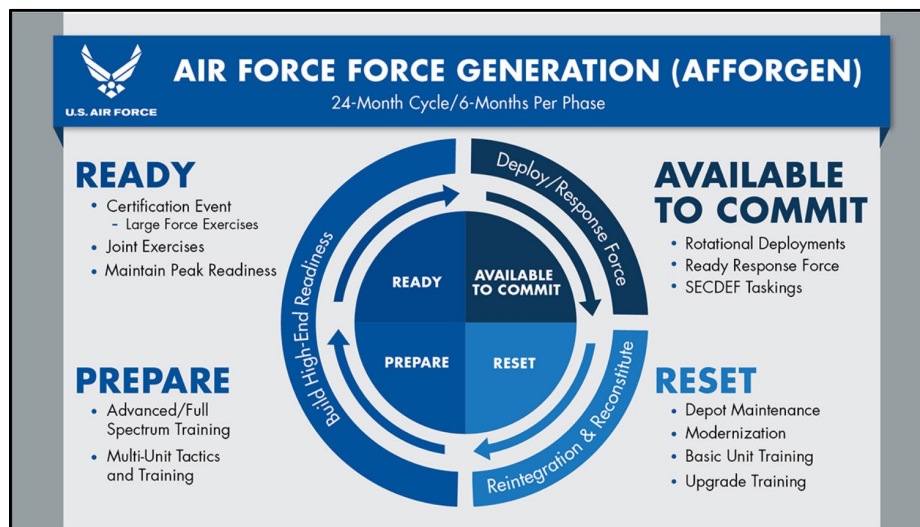


Figure 2. AFFORGEN Model (Secretary of the Air Force Public Affairs, 2022).

Specifically, there are two new added requirements that this study is primarily concerned with. The first is the addition of the unit commander assessed training plan. This plan serves as a roadmap from the commander for the unit on how they intend to get their unit ready. This includes requests for training resources, classes, and exercises for

the prepare and ready bins (Secretary of the Air Force Public Affairs, 2022). The issue is that these commanders are required to accomplish this assessment with very little prior knowledge of their unit's background or needs. As it currently stands, there is a need for more data to make a more educated decision.

The second addition of concern is the certification event. The intent of the certification event is to serve as the validation of the unit commander's training plan and to ensure that the unit can accomplish their assigned mission and solidify lessons learned (Air Combat Command Public Affairs, 2022). There are currently no ways to accurately predict a unit's ability to pass the event or the requirements needed to get there. This provides two main issues: (i) commanders may be expending too few resources or too many on different individuals with the hopes of achieving the same overall "grade," and (ii) due to a lack of performance forecasting abilities, there are no opportunities for commanders to intervene until it is too late.

2.6 Previous Training Reform Initiatives

The idea of modifying training requirements due to outside inputs is not new as there are a plethora of historical examples of aircrew training being tailored due to changes in the environment. A prime example can be seen in the Federal Aviation Administration's adoption of the Advanced Qualification Program (AQP) in 1991. The aim of the program was to remove previous barriers to training improvement by allowing companies to modify their programs based off their specific requirements and changes in technology. The AQP emphasizes proficiency over numbers and uses data to validate a program's effectiveness (U.S. Department of Transportation, 2017).

Similarly, the success of AMC's Flight Safety Program through the implementation of the Military Flight Operations Quality Assurance program, Airman Safety Action Program, and Line Operation Safety Action program, serve as a good example of changing the way performance is tracked. Through the rigorous evaluation of the data provided by these mechanisms, the root cause of pilot monitoring skills was identified as accounting for almost 35% of all errors (Altenburg, 2015). Based off that analysis, a push to add Crew Resource Management in both ground and flight training was emphasized (Wagener, 2014).

Another idea for optimizing training resources includes removing training that is irrelevant and refocusing those resources to other requirements. The 2021 GAO report lauded AMC for the identification of skills that were viewed as "unrealistic, outdated and unnecessary" in favor of more "high end training" (GAO, 2021, 20). While not implemented, Clinton's 2015 study demonstrates that between \$70-\$100 million a year can be save if air refueling training was accomplished primarily in the simulator versus the aircraft. Additionally, he argues that the USAF would experience little loss in overall capability due to a lack of use operational air refueling (Clinton, 2015).

The Ukrainian Air Force also noticed a need to revamp their training program in order to keep up with a rapidly changing environment. A study on the efficacy of a competence-based education model for evaluation of their training process supports the notion of teaching towards individual competency. This style of instruction resulted in a noted increase in overall aircrew performance (Roman, 2018).

During conversion from the F-16 to the F-35, the Norwegian Air Force attempted to balance mission effectiveness with training resources. Using an integer linear program, they were able to maximize the timeline of conversion and also make appropriate resource decisions for training to minimize lack of fighter capability now and in the future (Fauske & Hoff, 2016). Likewise, the Royal Australian Air Force identified a need to improve their aircrew training for mission command and concluded that a change in training mentality, equipment, and overall training content was required as funds and resources, namely instructors, became more limited (Slocombe, 2012).

The civilian sector has also looked at ways to streamline training proficiency in order to save on resources. The commercial airline Continental optimized their aircrew training with the implementation of the Crew Resource Solver decision-support system and saw an annual savings of \$10 million just by looking at how their resources were being used (Gang et al., 2004). Similarly, Delta was forced to relook at their training scheduling following the downturn in revenue after 9/11. What they found was that by streamlining their scheduling processes, they could decrease both pilot downtime for training and overall training costs without impacting the quality of training required (Sohoni et al., 2003).

2.7 Current C-17A Training Initiatives

Currently, there are two main training reform efforts within the C-17. The first is based primarily off the recommendations from McConville's 2019 study of C-17 aircrew training program's effectiveness and is referred to as the Sortie Based Training (SBT) program. In his study, McConville advocates for the departure from the traditional

volume and quantity tracking of currency to a program more focused on quality of training (McConville, 2019). Expanding on these ideas, SBT identified that the current RTM construct of training was prone to large swings in mission ready rates due to calendar-based timelines for currency versus the ability of the member to accomplish the mission (Vidt, 2021). SBT's solution was a rolling timeline with less total tracked events that emphasized proficiency over quantity. The initial trial run showed that SBT is effective at reducing mission-ready rate drop offs over the calendar year, however, they had no way to prove if there was any change in aircrew proficiency due to a lack of a standardized grading scale (McCormick, 2022). This, in turn, makes SBT unable to modify training requirements based off individual proficiency requirements.

The other effort being tested is the Tactics Advancement Course (TAC) model. The three-week course utilizes academics, planning events, and flight training events with the desire of providing aircrew members the tools that are required to be able to operate in a complex environment (Morales, 2022). Like SBT, TAC also lacks the ability to communicate the effectiveness that can be attributed to the program due to a lack of a standardized proficiency model (Leeming, 2022). This leads to a similar issue of the program not being able to be modified based off individual performance or background.

2.8 Regression Analysis

Regression is a form of supervised learning that explains the relationship of the data input, known as the independent variables, with an identified output, known as the dependent variable (Gutman & Goldmeier, 2021). The product of regression analysis is referred to as a regression model. This model can be used to help explain the relationship between the two variables and can be used to make dependent variable predictions based off changes in the independent variable (Gutman & Goldmeier, 2021). Linear regression is a specific type of regression for data that is believed to have a linear relationship between variables (Camm et al., 2018). The resulting model of simple linear regression is an equation that takes on the form in equation (1), where y is the dependent variable, x_1 is an independent variable, and ϵ is the error.

$$y = \beta_0 + \beta_1 x_1 + \epsilon \quad (1)$$

Multiple variable regression examines the relationship of two or more independent variables on one dependent variable. This model is represented by equation (2), where k is the number of independent variables:

$$y = \beta_0 + \beta_1 x_1 \dots + \beta_k x_k + \epsilon \quad (2)$$

2.8.1 What Data Does Regression Provide

The probability of significance of the relationship between the independent and dependent variable, known as the p-value, is the first area of interest in regression analysis. The p-value is used to test the likelihood that the relationship between the data occurs under the null hypothesis of there being no relationship (Gutman & Goldmeier,

2021). If the p-value is greater than the selected significance level, then we fail to reject the null hypothesis and assume the current data does not provide enough information in order to support a relationship. If a p-value less than the significance level is calculated, then the null hypothesis of there being no relationship is rejected. This supports the hypothesis that there is a linear relationship between the two variables (Camm et al., 2018).

The magnitude and direction of this relationship is provided in the form of coefficients (Gutman & Goldmeier, 2021). The coefficient shows the amount of change in the dependent variable given a change in the independent variable. For example, a coefficient of 2 would indicate that for every one unit change of the independent variable yields a change in the dependent variable by 2, keeping everything else constant. Referring back to our model, the coefficient is represented by the value of β_i for independent variable i . For linear regression, this is also referred to as the slope of the line (Camm et al., 2018).

Lastly, regression analysis provides a measurement of the coefficient of determination, which is represented as R^2 . R^2 offers insight into how much of the variation in the data can be explained by the proposed model (Camm et al., 2018). When calculated, R^2 is a value between zero and one with results closer to one representing a model that has little difference between the model predicted and actual value while models with a R^2 closer to zero providing little explanation of the variation in the data. This is commonly referred to as the “goodness of fit” of the model (Camm et al., 2018). In multiple variable linear regression, an adjusted R^2 is used to compensate for the

addition of more independent variables to the model. This ensures that additional variables are providing value and are not artificially increasing the R^2 (Ragsdale, 2017).

2.8.2 Uses of Regression Analysis

Regression analysis is a common technique used to predict values or explain relationships between variables (Gutman & Goldmeier, 2021) and can be seen in multiple applications. An example from the airlines is seen in Choi's 2019 study for American Airlines that utilizes linear regression to help determine which factor is the best indicator for predicting the airline's profitability. The results offer insight into what areas American Airlines should invest their money to see higher profits (Choi et al, 2019). Klapper and Ruff-Stahl also use linear regression to explain the relationship of decreasing pilot numbers for regional airlines with overall profitability of the regional airlines. What is found validates that pilot manning numbers are a significant indicator for fleet size, which in turn have a significant relationship to the number of passenger enplanements. The results show there is a linear relationship between the number of pilots and the overall profitability of that airline (Klapper & Ruff-Stahl, 2019).

Regression has also been used for military applications as well. Regression was used to determine the effects of different environmental factors such as weather or type of defensive positions on operation success rates for the US Army (Jaisloal, 1995). A 2020 Rand study examined if there were any significant predictor variables for aircraft mishap rates to help the Air Force's flight safety program. The age of an aircraft, when it was purchased, number of engines, and the type of mission it is assigned to were identified as significant indicators of aircraft mishap rates (Light & Pfeifer, 2020). Additionally,

Wells' 2020 study on C-17 taxi procedures employed a regression model to identify significant predictors for fuel usage on the ground. The results of his regression analysis show that fuel costs are most impacted by the number of operational engines. This narrows the focus of the study on what independent variable would produce the most change in fuel usage during aircraft taxi operations (Wells et al., 2020). Lastly, regression has also been used to predict USAF retention rates by looking at the relationship between Air Force Specialty Codes (Pujats, T., 2020) and macroeconomic metrics (McGee, 2021).

III. Methodology

3.1 Overview

C-17A Mobility Air Forces (MAF) Mission Profile (MMP) simulator data is utilized to develop an aircrew performance score as a way to evaluate aircrew proficiency. Using this score, descriptive statistics are evaluated in order to establish any baseline trends that exist in the data. Significant relationships between independent variables and performance are then identified by taking 80% of the available score data as training data for both simple linear and multiple regression analysis. The values attained from regression analysis are compiled into a model where given resources are modified and a predicted performance score is calculated. Finally, the model's effectiveness is evaluated through a confusion matrix that compares the prediction capability of the model created utilizing the 80% training data to the actual values of the remaining 20% testing data set (Camm et al., 2018). Figure 3 provides a general flow of the study's process.

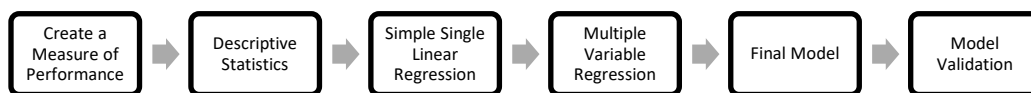


Figure 3. Study's Overall Methodology Flow.

3.2 Assumptions and Limitations

As highlighted in the 2020 GAO report, the Air Force lacks an assessment tool that can objectively quantify a crew member's proficiency or overall performance in a mission set (US GAO, 2020). In order to accomplish analysis on the effects of resources

on performance, a measure of performance would have to be created by the author for this study. In order to complete this requirement, the average completion rate of desired learning objectives (DLOs) achieved by a crew in the MMP is calculated in order to create what the author is calling the Aircrew Tactics Scorecard (ATS) as a method to quantify performance for this study. Validity of the ATS in truly measuring aircrew performance is outside of the scope of this study. However, the ATS is used as a construct to demonstrate the level of insight that may be gained if given an accurate methodology to measure performance. Additionally, ATS only reflects the ability to complete DLOs from the author's point of view associated with the MMP simulator as outlined by the C-17 Vol 1. ATS does not indicate a crew or unit's actual capability and, therefore, cannot be translated to any other training events, mission sets, or assessments outside of this study.

While it is the desire of the author for this research to be applicable to all of the Mobility Air Forces, this research is limited in scope to the C-17A community. Specifically, the primary focus is on the baseline airland and airdrop pilot qualification in hopes of creating a universal toolset that can be applicable to any unit. Special mission qualification such as Special Operations Low Level II (SOLLII) and Prime Nuclear Airlift Force (PNAF) are not considered when determining an individual's overall experience.

This study is limited based off the quantity and type of information gathered in the MMP. In general, most information is recorded in broad categorical format and lacked specificity. For example, crew qualification only looks at the highest qualified

individual and does not consider overall experience. The sortie may have had an instructor on the aircraft, but the information gathered did not differentiate between a new instructor versus one who with a lot of experience as an instructor.

Lastly, the information gathered and overall performance is attributed to the crew versus an individual. It is not clear on who was making the specific decisions or flying the aircraft at a given time, but instead is a measure of the cumulative effects of the crew's decision-making. While inconvenient for making individual forecasts, this is viewed as acceptable for this study as it is the collective crew efforts and dynamics that result in successful mission accomplishment in the C-17, and is sufficient to demonstrate the proposed methodology.

3.3 Establishing the Sample

The MMP simulator is selected as the training event to be evaluated for a couple of reasons. First, the simulator is administered by a third-party organization not associated with any individual unit (AFMAN11-2C-17v1, 2020). This compensates for any biases that may exist from a training program measuring its own effectiveness. Additionally, the MMP managers use simulator data and recorded information to analyze and score a crew's ability to accomplish an event. This is a level of fidelity that usually cannot be accomplished in the aircraft as accurately due to human factors in measuring performance (Goldsmith & Johnson 2002). Lastly, the MMP data collected offers insight into the cumulative effects of multiple skillsets versus just observing one component. In general, other training events' effectiveness are limited to only the accomplishment of

one event and is unable to provide a holistic view of how well the overall mission can be completed.

The initial sample was comprised of 1186 individual sorties. Incomplete data or missing demographic data was analyzed to see if it could be rectified based off comments or other information within the sortie information. If it could not, it was removed and not considered for further analysis. This resulted a sample size of 1075 sorties with viable information. Of those 1075, a randomly selected 80% of sorties are identified as the training data for the model with the remaining 20% as the test data sample.

For timeline calculations, the cleaned sample of 1075 is reduced to a sample of sorties only accomplished by one selected base. This produces a sample size of 276. Aviation Records Management System data is then matched to MMP sortie data in order to establish the previously accomplished training events of the individuals in the simulator. This process yields a total of 49 crews and sorties.

3.4 Calculating a Dependent Variable - ATS

As previously mentioned, this study does not have an approved metric to measure aircrew performance. ATS is the average of 13 data points recorded in the MMP simulator identified by the simulator administrators to assess a crew's ability to accomplish C-17 Vol 1 desired learning objects for the MMP simulator. Each area is initially recorded in a "yes/no" format that requires conversion into dummy variables to allow for further calculation. The technique of using dummy variables involves converting categorical data into numerical values, generally 1's or 0's (Camm et al., 2018). ATS is calculated for each of the 1075 sorties using the following formula:

$$\text{Aircrew Tactics Score} = \frac{x_1 + x_2 \dots + x_{13}}{13} \quad (3)$$

3.5 Cleaning the Independent Variables

Independent variables are broken down into two categories: demographic and timeline-based variables. Demographic variables are given in both categorical and continuous formats. Categorical data is converted into usable data using the same dummy variable process described Section 3.4. In total, there are 12 variables (see table 1, x1-x9 and x10-13) that used this method. Previous MMP experience is the only demographic independent variable that is recorded in continuous format (ex: an individual could have between 0-k previous MMPs). Analysis of previous MMP experience is made using the continuous and dummy variables in order to examine both the effect of overall previous simulator experience (categorical) and the difference between additional simulators (continuous).

Timeline calculations are primarily concerned with looking at the relationship between aircrew performance and time since last MMP accomplishment, time since last tactical event of any format, and type of last event. Timeline information is first created through plotting all the previous tactical training events an individual had accomplished previously. These timeline values are matched to the 49 crews previously created. Analysis is conducted using the crew's shortest, largest, and average time between events. Event type analysis is conducted by matching the individual with the smallest and largest time between events with the type of training they had previously

accomplished. The type of event is considered a categorical variable of either ground, simulator, or flying training event and is converted using the dummy variable format for further calculations.

Table 1. Independent Variables and Descriptions

Independent Variable Number	Description	Explanation
1	Crew Qualification	Crew qualification is given based off the highest crew qualification (or experience level) in the simulator. Values are given as either copilot (0), aircraft commander (1), or instructor pilot (2).
2	Number of Crewmembers	The simulator profile requires a minimum of two pilots; however, a third crewmember can be added to increase the crew's situation awareness through accomplishing delegated tasks to alleviate workload on the pilot flying and monitoring. Values are given as either a two-member crew (0) or three-member crew (1).
3	Airdrop Experience	Airdrop experience is used to differentiate between crews with no airdrop experience (0) and those who have any member that have previous airdrop experience (1).
4	Weapons Officer	Weapons officers are those who hold a special qualification after attending an additional tactical training course. Weapons officers are considered the subject matter experts in their aircraft. Values are given as either a crew without a weapons officer on board (0) or crews with a weapons officer (1).
5	Mission Planning	Mission planning is offered by the simulator administrators as a call-in academic course the day prior to execution. Values are given as either crews who had not accomplished this training (0) or those who had (1).
6	Tactics Planning	Tactics planning is accomplished through additional crew studies over approved tactics, techniques, and procedures. In general, this is meant to be instructed by a tactics officer, however, can be accomplished through self-study. Values are given as either crews

		who had not accomplished this training (0) or those who had (1).
7	Intelligence Planning	Intelligence planning is accomplished through additional crew studies with an intelligence officer. Values are given as either crews who had not accomplished this training (0) or those who had (1).
8	Total Mission Planning Events	Total mission planning events looks at the total number of prior training events a crew had accomplished between mission planning, tactics planning, and intelligence planning. Values are given as either no events accomplished (0), one event (1), two events (2), or three events (3).
9	Previous MMP Experience	Previous MMP experience is given as a measurement of whether the crew had accomplished this training event before. Values are given as either a yes (1) or no (0) depending on if they had accomplished a MMP.
10	MMP Quantity (Continuous)	MMP quantity (continuous) is the number of simulators that had previously been accomplished on a continuous scale between 0-6.
11	MMP Quantity (Categorical)	MMP quantity (categorical) looks at the number of simulators that had previously been accomplished using a categorical method. Values are provided in 7 categories (MMPs accomplished equal to 0-6) and a corresponding 0 (those who had not accomplished that number of simulators) or 1 (the crew had accomplished that indicated number of simulators) is assigned to that category. A value of 1 is assigned only once based off the total number of previous simulators accomplished.
12	Base of Assignment	The base that the crew is assigned to is converted into a categorical measurement and calculated similar to MMP quantity (categorical). There are a total of 12 bases reported, which were converted into either a 1 (base of assignment) or 0 (not base of assignment).
13	Profile Type Flown	Different profiles modify the type of threats are to be encountered. Each profile, however, still evaluates the same types and levels of desired learning objective.
14	Time Since Last MMP	Time since last MMP evaluates the number of days since the crewmember had accomplished a MMP simulator.

15	Time Since Last Tactical Event	Time since last tactical event evaluates the number of days since the crewmember had accomplished any form of tactical training to include academic classes, simulator training, or flight training.
16	Last Training Event Type	Last event type looks to evaluate which type of training (academic, simulator, or flight training) was last accomplished prior to the MMP and uses categorical values.

3.6 Descriptive Statistics

A baseline measurement of the effects of the independent variables on the dependent variables is accomplished through a review of the descriptive statistics of the sample. These statistics are commonly used for analysis in training programs due to their ease of calculation, however, they only describe the data and do not provide complete insight into the interactions. Because aircrew training allocations are currently made using demographic information (crew qualification, airdrop qualification, etc.), an analysis of the descriptive statistics of demographic variables is accomplished in order to compare the predicting capability of descriptive statistics versus the regression model.

3.7 Simple Linear Regression

Linear regression analysis of the 860 training data is conducted utilizing Microsoft Excel's regression tool in order to explore the relationship between an individual independent variables and dependent variable (ATS). To be considered a valid indicator for statistical inference, the calculated regression coefficient first had to produce a p-value of 0.05 or less. Additionally, the residuals of the model must demonstrate qualities of homoscedasticity, or the ability to produce residuals of equal variance around a mean of zero (Camm et al., 2018). Those variables that meet both requirements are

deemed significant indicators and their corresponding results are recorded. This results in an estimated regression equation seen in equation (4), where b_o is the y-intercept, b_1 is the slope of x_1 , and $\hat{y}$ is the estimated average for y when x_1 is changed.

$$\hat{y} = b_o + b_1x_1 \quad (4)$$

3.8 Multiple Regression

Multiple variable regression is used to examine the relationship between two or more independent variables and ATS. A stepwise selection method is used for model creation using the significant findings from the singular regression models. First, significant indicator variables are used as the baseline for model creation. Next, both significant and insignificant variables are added one at a time to the regression analysis to see if a significant model could be created. Analysis of model significance is performed using the same methodology previously described in the singular regression model. This process is repeated using previously identified insignificant indicator variables as the baseline to examine if adding additional variables would create a significant model. This results in an estimated regression equation seen in equation (5) using k independent variables.

$$\hat{y} = b_o + b_1x_1 + b_kx_k \quad (5)$$

3.9 Assessing Model Performance

Assessment of the final model's prediction capability is conducted with a confusion matrix. Independent variable values from the test sample are input in to the final model resulting in a predicted ATS. Comparison of the model's ability to predict

either a passing or failing ATS and the actual results from the test sample then reveals trends in the prediction capability of the model (Gutman & Goldmeier, 2021). In this analysis, a cutoff score of 69.2% is used to delineate a passing or failing score. Figure 4 provides an example of a confusion matrix where z is determined by the count of scores that fall in to each category. For example, z_{00} is the number of ATSs that are predicted to fail by the model and actually do fail.

		Model Prediction	
		Fail	Pass
Actual Results	Fail	z_{00}	z_{10}
	Pass	z_{01}	z_{11}

Figure 4. Confusion Chart Example

Measures of interest include the model's accuracy, sensitivity, and specificity. All three metrics are percentages with values closer to 100% indicating a stronger model. Model accuracy is a proportion of model predictions that are correct versus the total number of observations. Accuracy is calculated using equation (6):

$$Accuracy = \frac{z_{00} + z_{11}}{z_{00} + z_{01} + z_{10} + z_{11}} \quad (6)$$

Sensitivity is the ability of the model to identify actual passing scores correctly and can be calculated with equation (7):

$$Sensitivity = \frac{z_{11}}{z_{01} + z_{11}} \quad (7)$$

Conversely, specificity describes the model's ability to predict actual failures and is described by equation (8):

$$Specificity = \frac{z_{00}}{z_{00} + z_{10}} \quad (8)$$

IV. Results and Analysis

4.1 Overview

Results from the single linear and multiple variable regression analysis are of mixed results. Due to limitations in the independent variables available, the models within this study provide very limited ability in aircrew performance prediction. The application of the final model, however, did exhibit a moderate ability in identifying failures when compared to actual MMP scores. These findings provide practical significance to a decision maker as it provides a proactive means to recognize individuals that require attention and offers different courses of actions to attain a required capability.

4.2 Descriptive Statistics

Descriptive statistics from the training data are examined first to establish a baseline for prediction capability within the data and to delineate if any indicators of performance might be discerned from the measured independent variables. The first significant feature is the sample's range of scores. With 14 possible score values ranging from 0/13 to 13/13, an absence of scores of 0% is something of note (see Figure 5).

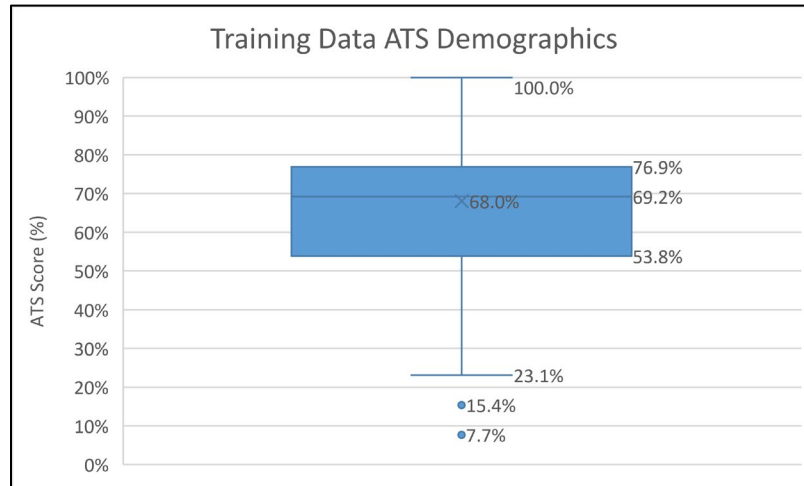


Figure 5. ATS Box and Whisker Plot

Furthermore, the median (69.2%) being higher than the average (68%) indicates that slightly more than half of the data points are above the average score. The examination of the distribution of scores provides additional insight into this issue (see Figure 6). The overall distribution of the scores can be categorized as having a negative skew (or left skew) with the preponderance of scores grouped in the higher region and very few scores within the lower score region. A negative skew is typical for exam scores due to the fact that scores cannot be higher than 100% and it is rare to see scores in the extreme lower regions (Camm et al., 2018). In conjunction with there being no scores of 0% within the range of data, the fact that 82.4% of the distribution lies between scores of 46.2% to 84.6% and 99% of the distribution lies between 23.1% to 100% supports the notion that crews are coming in with some form of baseline knowledge or skillset. This supports that no crew is starting from zero when it comes to their tactical skills as even a baseline crew (indicated by all zeros in the demographic independent variables) is not scoring a 0%. The box and whisker plot in Figure 5 shows that a crew's baseline score is

above 23.1% as scores of 7.7% (1/13) and 15.4% (2/13) are unlikely as they are labeled as outliers by the box and whisker plot.

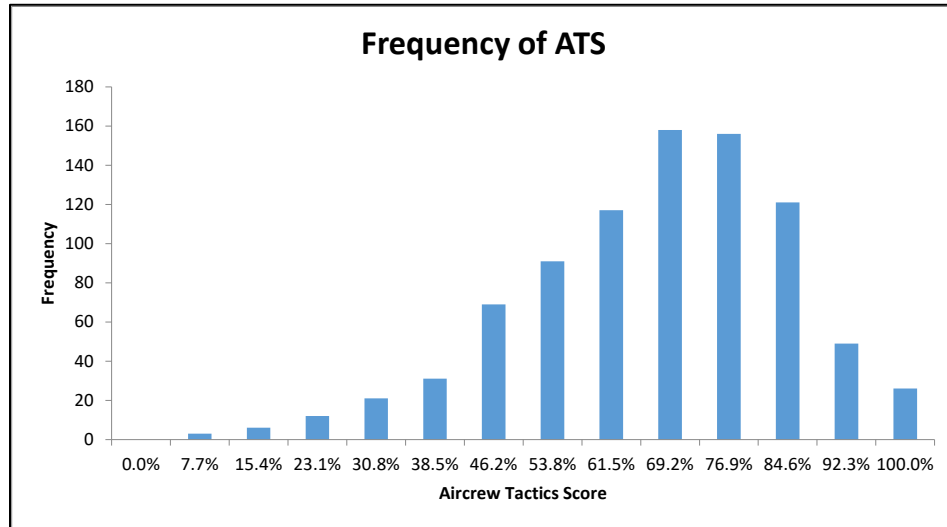


Figure 6. Distribution of Aircrew Tactics Scores.

The distribution of the data also provides a baseline for performance prediction capability. 41% of the sorties result in a passing grade using a cutoff score of 10/13 (76.9%) or a 59.3% pass rate utilizing a score of 9/13 (69.2%). Employing these percentages, if a commander assumes all of their members are going to pass, they would, on average, be correct 41% or 59.3% of the time depending on which cutoff score they selected. Based off those numbers, a commander could then assume the majority of their unit would either pass or fail and try to draw conclusions for resources founded off that assumption. Both of these “pass rates” offer a better success rate than making a random 50/50 decision, but requires generalizing performance of their unit. What is lacking from these metrics is specificity of who is going to pass or fail, the magnitude by which they

would pass or fail, and what is needed to take a crew from a failing score to a passing one.

The mean score for each independent variable, displayed in Table 2, provides some insight into which variables might contribute to the scores. By filtering the ATSs into categories based off the classification of a given independent variable, some evidence of viable indicator variables are identified. In general, it could be assumed that an independent variable might be a good predictor of performance if the ATS of a category has a large difference between when the crew possessed a resource or demographic than when they did not (and vice versa). Expanding on that concept, a couple of intriguing patterns begin to emerge. First, areas that have minor differences in averages or do not appear to reflect practical significance of the real world highlight potentially insignificant variables. For example, crews in which the highest qualified individual was a copilot (69.2%), aircraft commander (67.2%), or instructor pilot (67.2%) show that increasing crew qualification actually has a negative effect on a crew's performance, which is counter intuitive. Additionally, there are some independent variables where there are larger differences in average scores, which could serve to identify potential indicator variables. Three prime examples can be seen in the tactics planning (an increase of 5.8 percentage points), intelligence planning (an increase of 5.3 percentage points), and weapons officer (an increase of 4.5 percentage points) averages.

Table 2. Mean ATS for Given Independent Variable Subcategories

Independent Variable	Mean ATS	Independent Variable	Mean ATS	Independent Variable	Mean ATS
<i>Crew Qualification</i>		<i>Weapons Officer</i>		<i>Prior Experience</i>	
CP	69.2%	No	67.0%	No	60.2%
AC	67.2%	Yes	71.5%	Yes	68.5%
IP	67.2%	<i>Tactics Planning</i>		<i>MMP Quantity</i>	
<i>Number of Crew Members</i>		No	66.7%	0	60.2%
2 Members	67.1%	Yes	72.5%	1	66.1%
3 Members	67.3%	<i>Mission Planning</i>		2	69.8%
<i>Airdrop</i>		No	65.5%	3	73.0%
No	66.5%	Yes	68.0%	4+	74.7%
Yes	68.2%	<i>Intelligence Planning</i>			
		No	67.1%		
		Yes	72.4%		

4.3 Simple Linear Regression

4.3.1 Overall Findings

Scrutiny of the simple linear regression results answers a couple of necessary questions required for further analysis. First, can tracking measured resource allotment and demographic information serve as an indicator to predict future performance? The majority of the items that are tracked and/or the extent that they are currently observed are not significant statistical indicators for predicting future performance. Of the 16 independent variables examined, five variables have a statistically significant relationship within the 860 training data sample that meet or exceed the predetermined significance level (p-value) of 0.05.

The second question is to what extent can said indicators predict future performance now that significant indicator variables are identified? In general, the recognized indicators only exhibit a limited ability to predict performance as represented by their very low goodness of fit values (R^2) and large standard errors. This inability to

consistently explain the variance in data, even with a statistically significant relationship, exhibits the complexity of predicting aircrew performance in a dynamic environment.

The implications of these findings, while small, are not inconsequential. The ability to categorize variables as either statistically insignificant or significant helps to identify areas of focus when it comes to predicting performance. This affords commanders the capability to understand which resources produce what results as opposed to relying on misleading descriptive statistics or anecdotal evidence. Additionally, the results from simple regression assist in identifying which variables should be evaluated when creating a multiple variable model in an attempt to expand the performance predicting capability of the research's final proposed model.

4.3.2 Selection of Significant Variables

Initial analysis of the simple linear regression results provides a holistic look at any trends or relationships between an independent variable and ATS within the given training sample set. These results aid in focusing areas of interest for further examination and eliminate others that might not have any relationship at all. By analyzing the regression's p-value output, the null hypothesis of there being no relationship between the independent and dependent variable is tested. In order to support the hypothesis that there is a relationship and to be considered a significant indicator variable, the p-value of the model must be below 0.05 and exhibit qualities of homoscedasticity of the residuals as described in Chapter 3. Variables that do not meet the 0.05 threshold, but are below 0.1, are reviewed to determine if they demonstrate a practical significance and should be considered for future use.

Within the 860 training data sample, only 5 out of the 13 measured demographic independent variables are statistically significant (see Table 3 for a comprehensive list of all variables). While not considered statistically significant, mission planning (p-value of 0.053) and weapons officer (p-value of 0.09) are variables of interest that may be evaluated for use in the multiple regression model. Additionally, analysis of the three time-related independent variables fail to yield any significant relationships with ATS and are not considered for future models, based on evaluating 49 crews.

Table 3. Simple Linear Regression Independent Variables' Significance Levels

Independent Variable	
<i>Crew Qualification</i>	
<i>Number of Crewmembers</i>	
<i>Airdrop Experience</i>	
<i>Weapons Officer</i>	*
<i>Mission Planning</i>	*
<i>Tactics Planning</i>	***
<i>Intelligence Planning</i>	
<i>Total Mission Planning Events</i>	***
<i>Previous MMP Experience</i>	***
<i>MMP Quantity (Continuous)</i>	***
<i>MMP Quantity (Categorical)</i>	***
<i>Base of Assignment</i>	
<i>Profile Type Flown</i>	
<i>Time Since Last MMP</i>	
<i>Time Since Last Tactical Event</i>	
<i>Last Training Event Type</i>	
* achieved 0.1 level of significance	
** achieved 0.05 level of significance	
*** achieved 0.01 level of significance	
No stars indicate variable did not achieve 0.1 level of significance	

The evaluation of significant variables emphasizes a few important points. First, exclusion of certain variables, such as weapons officer, can indicate a lack of representative data to make an accurate assessment. Crews with a weapons officer only

account for 5.2% of all the sorties within the sample. The incongruity in group sizes between crews with or without weapons officers may be skewing the analysis and not giving an accurate depiction of the true effects of having a weapons officer.

Furthermore, the analysis of significance may be limited based off how the variable is measured. Since the majority of the independent variables utilize dummy variables representing categorical yes or no values, the independent variables can only assume a value of one or zero based off whether the independent variable is present (1) or not (0). This allows for almost no variation in the data and do not allow for much interpretation. For example, there is no differentiation between a crew who has current versus previous airdrop experience or which level of qualification they held. A crew with someone who was a copilot airdrop member three years ago is counted the same as a crew who has a current lead airdrop member.

Lastly, just because an independent variable is identified as significant does not necessarily mean that the data it provides is practical for future use. For example, two of the identified significant indicators are related to how many previous MMP simulators a crew had previously accomplished. What is different is whether the independent variable is a categorical or continuous variable. When using prior experience as a continuous variable, adding more experience produces a constant slope. This indicates that as the number of previous MMP simulators increases, the ATS will increase at the same proportionate rate (in this case 3.3 percentage points for each simulator added). However, when examining prior experience as a categorical variable, the benefit from additional MMP simulators is not constant. This produces a model that shows that each simulator

added has different levels of effects with the biggest increase occurring going from no prior MMPs to 1 (5.9 percentage point increase) and a diminishing increase rate when going from 3 to 4 or more MMPs (1.7 percentage point increase). In some instances, the continuous model under or overestimates the crew's ATS as compared to the categorical model.

Examining the difference in intercept values also reveals a difference between the models as the continuous model estimates the intercept at 62.4% and the categorical model produces an intercept of 60.2%. It can be argued that the intercept of 60.2% is more accurate as it matches the average score of all crews within the sample that have never accomplished an MMP before. For both of these reasons, prior MMP experience is analyzed using the categorical model as opposed to the continuous, even though both were considered statistically significant.

4.3.3 Interpreting Significant Indicator Variables

As identified in Table 4, there are five statistically significant variables and two independent variables of interest when using the 860 training sample set. Each of these variables creates a model that can be used to predict performance based off whether or not a single independent variable is changed. Each model then produces three factors that this study is concerned with: (i) the baseline score without the independent variable (using the intercept); (ii) the amount of change that is expected, on average, when the independent variable is changed (the coefficient of the variable); and (iii) the amount of variation in the data that is explained by the model (the R^2). Table 4 summarizes these values from the 860 training sample.

Table 4. Simple Linear Regression Coefficient Values.

Parameter	R ²	Intercept	Coefficient	p-value
<i>Tactics Training</i>	0.01	66.7%	5.8%	< 0.01
<i>2 or More Training Events</i>	0.01	66.7%	5.3%	< 0.01
<i>Previous MMP Experience</i>	0.03	60.2%	8.3%	< 0.01
<i>MMP Quantity (0 MMP)</i>	0.05	60.2%		
<i>1 MMP</i>			5.9%	< 0.01
<i>2 MMPs</i>			9.7%	< 0.01
<i>3 MMPs</i>			12.8%	< 0.01
<i>4+ MMPs</i>			14.5%	< 0.01
<i>Mission Planning</i>	0.004	65.5%	2.5%	0.053
<i>Weapons Officer</i>	0.003	67.0%	4.5%	0.09

These models are limited in their ability to predict ATS. First, examining the R² value of the model provides insight into its ability to predict ATS. The highest R² value of any of the independent variables is 0.05. This means that the best model created using simple linear regression can only explain approximately 5% of the variation in the data. While R² values are generally lower when performing a study involving human factors versus those in the engineering realm, statistically and practically speaking, simple linear regression does not produce strong models for prediction.

Additionally, these models have a large standard error which makes the confidence interval of predictions very wide. All of the models created using simple linear regression have a standard error of approximately 17.5 percentage points. This gives each ATS a prediction interval of plus or minus 35 percentage points from the predicted score (a total range of 70 percentage points). Based off the results of simple linear regression, multiple linear regression is used in order to examine if the prediction capability of a model can be improved by considering multiple independent variables.

4.3.4 Interpreting Insignificant Indicator Variables

Some substantial findings are possible from the statistically insignificant variables. For starters, filtering variables into either significant or insignificant indicators of performance helps to determine which programs a commander should pay attention to using the descriptive data. Intelligence planning, for instance, has one of the larger differences in means between those who had accomplished the training or not. Looking at only the descriptive data, commanders would assume that they should increase intelligence training with the expectation of raising their units overall performance. While this could potentially be beneficial, the current data does not support that expectation. Identifying intelligence training as statistically insignificant when predicting performance can provide a decision maker an assessment on the effectiveness of the training and drive them to either request more data on the subject, temper their expectations on the potential outcome, or to revamp their program completely in order to achieve the desired outcome.

Identifying a specific variable as statistically insignificant can be viewed as constructive as well. For example, base of assignment not being a significant indicator could be considered a positive as that indicates that no one base is doing better than the others. While the negative side of that is that means there is not an individual base whose training program should be emulated, it also means that no crew is at a disadvantage due to where they are stationed. Since some of the bases that are in the computations are purely Guard or Reserve, this could also indicate that component within the total force does not have any significant influence on a crew's ATS. While more data would be

required to confirm that conclusion, it could be a good indication of expected results. Furthermore, base of assignment not being a significant indication of performance could be helpful from an operations planning perspective, as there is no advantage to tasking one base versus the other.

Lastly, these insignificant variables can be used to identify what data is being tracked that we may be tracking either improperly or not to the level that is required for proper analysis when it comes to ATS prediction. While this study categorized variables as either significant or insignificant indicators of performance, it does not mean that there is not a relationship between the independent and dependent variables. What it could actually indicate is that there is not enough data to support the hypothesis that there is a relationship. If it is still believed that there is a relationship between a given independent variable and ATS, then the data that is collected regarding a variable should be re-evaluated. A prime example of this would be crew qualification. Anecdotally, the consensus thought process is that there should be a significant difference in performance between an aircraft commander and an instructor pilot. The current data, however, cannot differentiate between a brand new C-17 instructor pilot and a seasoned aircraft commander who might have prior instructor experience in other airframes. Increasing data granularity may support these premises or reveal new trends that were not thought of before.

4.3.5 Limitations of Simple Linear Regression Applications

There are some major restrictions of simple linear regression that limit its ability to accurately and practically predict performance. The first issue is that of extrapolation

or prediction capability of the model being bound by the values of the given experimental range (Camm et al., 2018). For example, the model predicting ATS based off whether or not a crew had tactics planning prior to the MMP was limited to a “yes” (1) or “no” (0) answer. By going from a zero to a one, on average, the crew’s ATS increased by 5.8 percentage points. This does not mean, however, if a commander had a crew who went through two tactics planning sessions prior to the simulator that the commander should expect their score to increase by 11.6 percentage points. While the model comes in the form of an equation, that equation is bounded by the limits of the data that was used to create it. This could cause some confusion if not interpreted correctly.

Additionally, simple linear regression is limited by its ability to only change one independent variable while maintaining all other things constant. This means that an individual cannot look at the results of all the simple linear regression and just add all the coefficients of the models that were identified as significant. For example, it would be incorrect that adding one MMP and tactics planning session would result in an increase of 11.7 percentage points (the sum of their coefficients). In order to get that kind of fidelity, multiple variable regression would have to be performed, shown in Section 4.4.

4.3.6 Implications and Takeaways

In general, there are three major takeaways from simple linear regression analysis. The first is that there is no one “silver bullet” to make a crew better. This is, in general, a good thing as if there was only one thing that needed to be fixed then the answer should have been obvious prior to this analysis and would have already been corrected.

Furthermore, simple linear regression shows the current limitations of tracking performance as it cannot be attributed to the individual level. For example, an “instructor crew” is analyzed as if the instructor is the one flying the simulator and making all the appropriate decisions. This may or may not be true as the less qualified individual may be the one flying the simulator as part of an upgrade training program. Any “mistakes” made by this less qualified individual are, however, attributed to the instructor category, which may be providing an incorrect picture. In order to eliminate this error, greater emphasis on tracking who is making what decision or flying which portion is needed.

Finally, given a way to measure performance (ATS), a process can indeed be formed that identifies statistically significant indicator variables of performance. The efficacy of the models created as a performance predictor, however, leave a good deal to be desired as the variables tested in this study could at best only describe 5% of the variation in the data. While not ideal, this does support the idea that performance can be predicted, but better predictor variables would be required.

4.4 Multiple Variable Regression

4.4.1 Overall Findings

The intent of using multiple variable regression analysis is to create a model that could utilize different combinations of independent variables together in order to predict a crew’s projected performance. The final model applies three different independent variables (tactics planning, mission planning, and MMP quantity) utilizing the 860 training sample set. The different combinations of independent variables provide different avenues for a commander to employ in order to achieve or maintain a required

level of performance. An example color-coded decision matrix is created to serve as an example of how the results of regression analysis could be utilized to inform a commander when making resourcing decisions.

4.4.2 Proposed Model

The use of the stepwise method for selecting independent variables for use within the final model tests the addition or subtraction of different independent variables on the model's capability of predicting ATS. As previously established in the simple linear regression, in order for a variable to be considered for further use within the model it has to demonstrate indicators of statistical significance. This process examines both variables identified as significant and insignificant indicator variables by simple linear regression in order to determine if the combination of different variables made selected indicator variables significant. The final model includes three independent variables, all of which are significant below a 0.05 p-value: tactics planning, mission planning, and MMP quantity (categorical). The use of these variables produces a model with an intercept and coefficient values that can be seen in Table 5.

Table 5. Multiple Variable Regression Coefficients

Parameter	Point Estimator	Coefficient	p-value
<i>Intercept</i>	b_0	57.1%	< .01
<i>Tactics Planning</i>	b_1	4.8%	0.02
<i>Mission Planning</i>	b_2	3.5%	< .01
<i>1 Prior MMP</i>	b_3	6.0%	< .01
<i>2 Prior MMPs</i>	b_4	10.1%	< .01
<i>3 Prior MMPs</i>	b_5	13.6%	< .01
<i>4+ Prior MMPs</i>	b_6	15.6%	< .01

Applying the above values for the selected independent variables, a model utilizing the formula in equation (9) is constructed.

$$ATS = 57.1 + 4.8x_1 + 3.5x_2 + 6x_3 + 10.1x_4 + 13.6x_5 + 15.6x_6 \quad (9)$$

In this model, variables are still restricted to values of 1 or 0 depending on if that independent variable is present or not. Additionally, variables $x_3 - x_6$ (number of MMPs previously accomplished) are mutually exclusive variables as only one or none of the variables can be present. Using these three selected indicator variables, 20 different combinations of independent variables are possible. Applying coefficient values to these combinations then yields a projected ATS for a given combination as seen in Table 6.

Table 6. Possible Aircrew Tactics Score Based off Combination of Independent Variables

<i>Parameter</i>	<i>Intercept</i>	<i>Tactics Planning</i>	<i>Mission Planning</i>	<i>1 MMP</i>	<i>2 MMPs</i>	<i>3 MMPs</i>	<i>4+ MMPs</i>	
<i>Point Est</i>	b_0	x_1	x_2	x_4	x_5	x_6	x_7	ATS
<i>Coefficient</i>	0.571	0.048	0.035	0.06	0.101	0.136	0.156	
1	1	0	0	0	0	0	0	57.1%
2	1	1	0	0	0	0	0	61.9%
3	1	0	1	0	0	0	0	60.6%
4	1	1	1	0	0	0	0	65.4%
5	1	0	0	1	0	0	0	63.1%
6	1	1	0	1	0	0	0	67.9%
7	1	0	1	1	0	0	0	66.6%
8	1	1	1	1	0	0	0	71.4%
9	1	0	0	0	1	0	0	67.2%
10	1	1	0	0	1	0	0	72.0%
11	1	0	1	0	1	0	0	70.7%
12	1	1	1	0	1	0	0	75.5%
13	1	0	0	0	0	1	0	70.7%
14	1	1	0	0	0	1	0	75.5%
15	1	0	1	0	0	1	0	74.2%
16	1	1	1	0	0	1	0	79.0%
17	1	0	0	0	0	0	1	72.7%
18	1	1	0	0	0	0	1	77.5%
19	1	0	1	0	0	0	1	76.2%
20	1	1	1	0	0	0	1	81.0%

This model is selected as the best model for two reasons. First, this model offers the ability to manipulate the most independent variables. Other models created utilizing multiple regression only had two significant indicator variables. While simpler models are usually preferred, this quality is important in order for the model to represent some form of practical significance as commanders have a large amount of resources that they are able to choose from. A model with only two options would not truly represent the complexity of the decision with regard to this problem set.

The second reason is that it yields the best adjusted- R^2 value (0.063) of any combination tested (the adjusted R^2 is used to account for the number of independent variables in the model). This indicates that the most variation in the data can be

explained by using this model. While 0.063 is still extremely low, it is a slight improvement from the 0.05 from the simple linear regression.

4.4.3 Application of Model

While ultimately not a great statistical model, the application of a multiple variable regression model does have various real world applications. In order to demonstrate the potential applications of utilizing a methodology and model such as the one created in this study, the model for predicting ATS in this section will be referred to as if it had a much larger R^2 and a low standard error. This allows for further discussions to be in terms of actual predicted scores with little variation. The coefficient values and calculated scores are, however, still from the final model using multiple linear regression.

For starters, a decision matrix for commander's use can be made by taking the values calculated in Table 6 in order to delineate which combinations of variables yield a desired threshold of performance. Table 7 provides an example where ATS values are conditionally formatted to where scores of 0% - 69.9% are red, 70% - 74.9% are yellow, and values 75% or higher are green. This matrix provides a quick reference for a

commander to be able to predict if, on average, a given combination of variables could produce a desired ATS.

Table 7. Example Decision Matrix Using Final Model Values

		MMP Only	Mission Planning	Tactics Planning	Mission + Tactics
MMP Quantity	0	57.1%	60.6%	61.9%	65.4%
	1	63.0%	66.5%	67.8%	71.3%
	2	67.2%	70.7%	72.0%	75.5%
	3	70.7%	74.2%	75.5%	79.0%
	4	72.7%	76.2%	77.5%	81.0%

Additionally, this matrix output provides a quick visual on the effects that adding or subtracting a resource would potentially have on a crew's performance. This informs the decision making process by directing a commander where more or less resources might be required. For example, if there is a crew that has previously accomplished 2 MMP simulators, a commander would be able to say that, on average, adding either mission planning or tactics planning will bring that crew's score above the 9/13 (approximately 70%) threshold, but they would need to accomplish both forms of training if they wanted to give them the best chance at achieving a 10/13 score. Of course, prediction confidence intervals should also be considered when making declarations like these. This process also helps to provide a proactive mitigation plan if a resource was to be taken away. For example, if a member that was scheduled to accomplish both tactics and mission planning prior to their fourth MMP sim (meaning three had been previously accomplished) was suddenly unable to accomplish one of the planning events due to a personal matter, a commander could direct that tactics planning should be prioritized over mission planning. This is based off the fact that tactics

planning keeps the member in the green range while mission planning only keeps them in the yellow.

This model could also be used to inform impacts of constrained or altered resources in order to construct a viable plan of action. For example, if the person who instructs the tactics planning portion of the simulator was sent on a last minute deployment, a commander could now see the effects of not having that training resource available. Table 8 provides what the updated decision matrix would look like with all tactics planning options now being equal to zero. This figure shows that without tactics training, a commander would need to increase their number of simulators sorties available as the only way to get a “green crew” would be to accomplish four MMPs and mission planning. Based off availability of simulators, this may or may not be a viable option. This may then drive the discussion to what kind of resources would be required to reestablish the unit’s tactics planning program such as moving a member from a different unit into theirs in order to compensate for the loss.

Table 8. Sample ATS Decision Matrix with no *Tactics Planning* Available

		MMP Only	Mission Planning
MMP Quantity	0	57.1%	60.6%
	1	63.0%	66.5%
	2	67.2%	70.7%
	3	70.7%	74.2%
	4	72.7%	76.2%

Lastly, the use of this model could help to optimize the use of other forms of resources other than just the training resources themselves. For example, a commander might need to minimize the amount of time that it would take to generate a crew that

should on average score above a 75% in order to fill a last minute tasking. By assigning time coefficients (the red values in Table 9) to each independent variable based off how many hours it would take to accomplish said training event, a total time to complete required training metric can be calculated (see Table 9 for an example computation).

Table 9. Sample Calculation of Amount of Time to Complete a Given Combination of Training

Parameter	Intercept	Tactics Planning	Mission Planning	1 MMP	2 MMPs	3 MMPs	4+ MMPs		
Point Est	b_0	x_1	x_2	x_4	x_5	x_6	x_7		
Coefficient	0.571	0.048	0.035	0.06	0.101	0.136	0.156	ATS	Cost: Time (in hrs)
Time in Hours	0	1.5	2	8	16	24	32		
1	1	0	0	0	0	0	0	57.1%	0
2	1	1	0	0	0	0	0	61.9%	1.5
3	1	0	1	0	0	0	0	60.6%	2
4	1	1	1	0	0	0	0	65.4%	3.5
5	1	0	0	1	0	0	0	63.1%	8
6	1	1	0	1	0	0	0	67.9%	9.5
7	1	0	1	1	0	0	0	66.6%	10
8	1	1	1	1	0	0	0	71.4%	11.5
9	1	0	0	0	1	0	0	67.2%	16
10	1	1	0	0	1	0	0	72.0%	17.5
11	1	0	1	0	1	0	0	70.7%	18
12	1	1	1	0	1	0	0	75.5%	19.5
13	1	0	0	0	0	1	0	70.7%	24
14	1	1	0	0	0	1	0	75.5%	25.5
15	1	0	1	0	0	1	0	74.2%	26
16	1	1	1	0	0	1	0	79.0%	27.5
17	1	0	0	0	0	0	1	72.7%	32
18	1	1	0	0	0	0	1	77.5%	33.5
19	1	0	1	0	0	0	1	76.2%	34
20	1	1	1	0	0	0	1	81.0%	35.5

By adding this time aspect to the model, a commander can now not only look at a score projection but also the time that it would take to accomplish that training combination. This could then aid a commander in making a decision on selecting the appropriate course of action. Using the time coefficients from Table 9, Table 10 ranks a training combination by the projected ATS and the time it would take to accomplish this training. If using a 70% threshold, combination 11 would be their fastest option (18 hours) while combination 12 (19.5 hours) would be required to meet a 75% requirement.

This also helps to differentiate between combinations that yield the same ATS. For example, combination 14 and 12 both produce an ATS of 75.5%, however, combination 14 takes 4 hours longer to accomplish. This would make option 12 the preferential option if a commander was looking to optimize training time. This process is replicated using other areas of interest such as financial costs.

Table 10. Comparison of ATS and Time Required to Achieve

<i>Independent Variable Combination</i>	<i>Model Projected ATS</i>	<i>Model Cost: Time (in hrs)</i>
20	81.0%	35.5
16	79.0%	27.5
18	77.5%	33.5
19	76.2%	34
14	75.5%	25.5
12	75.5%	19.5
15	74.2%	26
17	72.7%	32
10	72.0%	17.5
8	71.4%	11.5
13	70.7%	24
11	70.7%	18
6	67.9%	9.5
9	67.2%	16
7	66.6%	10
4	65.4%	3.5
5	63.1%	8
2	61.9%	1.5
3	60.6%	2
1	57.1%	0

4.5 Model Effectiveness

As previously discussed in Section 4.4, the ability of the final model to explain the variation in the data is very low ($R^2 = 0.063$). Additionally, the standard error, while lower than the 17.5 percentage points calculated by the simple linear regression model, is equal to 17 percentage points. This indicates that the prediction interval of ATS could actually be as much as 34 percentage points higher or lower than the model's projection

(a range of 68 percentage points). With such a small R^2 value and large standard error, the question becomes whether or not this model is actually effective at projecting performance. In order to test this question, a confusion matrix is constructed.

4.5.1 Findings

The constructed confusion matrix compares the ability to predict a passing or failing score of the multiple variable regression model produced using the 860 randomly selected training data sample to the actual results of the remaining 215 testing dataset. For this analysis, a score of 69.2% is used as the cutoff score for passing, as it is the closest to the desired score of 75% without going over. Figure 7 provides the results of the confusion matrix analysis.

		Model Prediction			
		Fail	Pass	Accuracy %	55.8%
Actual Results	Fail	56	17	Sensitivity	45.1%
	Pass	78	64	Specificity	76.7%

Figure 7. Model Confusion Matrix Results

The model's accuracy rate of 55.8% summarizes the ability of the model to predict whether a crew is going to pass or fail as 120 out of 215 times the model predicts the correct result. While it might not seem to be a positive, the fact that the model produces more false negatives (54.9%) than false positives (23.3%) can represent a strength of the model. This means that a commander is more likely to give extra resources to a crew that does not need them versus depriving crews of resources that they actually do need. One

of the biggest highlight from this analysis, however, is the overall specificity of the model. 76.7% of actual failures are correctly predicted by the model. The ability to predict failures accurately is vital in the discussion of resource allocation as the decision maker generally spends the preponderance of their time trying to elevate “failing” members into passing ones.

4.5.1 Implications

While a 55.8% accuracy rate does not have much of a statistical significance, it does show that the model is producing better success rates than making a 50/50 generalization. This demonstrates that, while not perfect, a model can be constructed that can provide a data-driven decision with results better than a random guess. Additionally, the model’s prediction of 76.7% of the actual failures aids in directing where resources should be focused. This provides a practical significance for a decision maker as it helps to scope where resources are needed most with the hopes to mitigate any future issues.

4.6 Summary

Overall, the use of both simple linear and multiple variable regression produces only a few statistically significant models. Furthermore, the models that are considered significant produce low correlation values, low levels of ability to explain the variation in the data, and a large possible score range for a given prediction. However, a great deal of the information and interpretation gathered from both the significant and insignificant models is useful for the given problem set. The final selected model, while not largely statistically significant, does provide a great deal of practical significance to a decision maker. It demonstrates that given appropriate data, the methodology from this study can

be used to identify the relationships between resources and aircrew performance and ultimately optimize the way training resources are then allocated.

V. Conclusion

5.1 Research Conclusions and Significance

From a statistical prospective, the models created within this study possess marginal capability to predict aircrew performance, however, there are a good deal of practical applications of the process and paradigm described. First, the ability to identify which training resources have a significant relationship with performance assists in informing a data-driven decision when it comes to resource management. A commonly held belief within the aircrew training community is that not all training is created equal when it comes to efficacy. The value of a given training event, however, is largely based off opinion and limited by personal experience or exposure. The process utilized in this study provides a baseline for empirically evaluating resources off not only their costs, but also their prospective effects. This could theoretically save time, money, and effort on training programs that are not able to accomplish their desired goals either through a lack of appropriate resources or areas of focus.

Moreover, scrutiny of common factors amongst those factors identified as significant can reveal trends that can be translated to other areas. Examining those variables identified as significant in both the simple linear and multiple variable regressions suggests that exposure to concepts (as measured by prior experience) and quality of preparation prior to the event (as measured by mission and tactics planning).

This guides the decision maker to find opportunities to increase training opportunities but also to maximize that training event by creating a culture valuing preparation versus just “checking the box.”

Additionally, aircrew performance can be potentially explained using specific indicator variables and is not a completely random event. While the indicator variables used within this study are limited in their ability to explain the variation in the data, there could be other variables that might be better indicators of performance that are not currently tracked either to the appropriate level or at all. This suggests that effort should be made in tracking not only performance, but also in the process that led to that crew’s capability. This has two benefits: (i) it creates a repeatable and predictable training program and (ii) it transforms a reactive unit training program into a proactive one.

Lastly, the ability to modify resources while maintaining a desired level of capability demonstrates that aircrew training can be accomplished through multiple avenues. The current RTM style of training allotment provides only one approved methodology for resource allocation based off the assumption that completion of the assigned events will produce a proficient pilot and failure to complete results in a non-proficient pilot. What RTM fails to do is to differentiate the risk of not getting enough of a given resource or the actual quantity required to achieve a desired goal. By creating different courses of actions, the decision maker is given not only flexibility but also the knowledge of the amount of risk assumed by deviating from the baseline.

5.2 Recommendations for Future Actions and Research

Much of this research was limited based off the inability to either measure a desired variable or the availability of the data. This study demonstrated that given a way to measure capability, impactful decisions regarding aircrew training and resourcing could be made with a wide span of applications. In order to gain these benefits, AMC must generate an approved way to evaluate capability accurately. Having an approved definition of what a proficient crewmember is and the tools by which to measure their performance will be necessary for future applications. Additionally, further refinement of indicator variables may yield better results.

Furthermore, this study also was concerned with crew versus individual performance. Supplementary research into individual factors would aid in creating a unique training program that could optimize the training resources requirements of the individual while maximizing their performance.

Finally, linear regression as a technique for analysis is limited in its ability to evaluate only linear relationships and smaller data sets. Different types of evaluation, such as logistic regression, should be considered in order to establish if a different model could be constructed. The USAF as a whole would also benefit from expanding research into different aircraft training programs and training forums. Lessons learned evaluating other programs may reveal new applications or trends that could be applied on a larger scale.

Bibliography

- Air Combat Command Public Affairs. (2022). *ACC Completes Strategic Validation Exercise 4 on ACE, CAFFORGEN*. Accessed 2 April 2022.
<https://www.af.mil/News/Article-Display/Article/2794036/acc-completes-strategic-validation-exercise-4-on-ace-cafforgen/>
- Air Force Doctrine Publication (AFDP) 3-36. *Air Mobility Operations*, 28 June 2019.
- Air Force Instruction (AFI) 10-201. *Force Readiness Reporting*, 22 December 2020.
- Air Force Instruction (AFI) 11-102. *Flying Hour Program Management*, 08 December 2020.
- Air Force Instruction (AFI) 65-503. *Air Force Cost and Planning Factors*, 13 July 2018.
- Air Force Manual (AFMAN) 11-202v1, *Aircrew Training*, 27 September 2019.
- Air Force Manual (AFMAN) 11-2C-17v1, *C-17 Aircrew Training*, 12 February 2020.
- Altenburg, M. (2015). *An Effective Pilot Monitoring Mitigates Threats and Keeps Errors at Bay*. Mobility Forum: The Journal of the Air Mobility Command's Magazine, 24(1), 7–8.
- AMC/A3TR Briefing. (2021). *AMC Flying Hour Program*.
- Becker, P. (2022). Personal Correspondence: *Training Program Processes*. To the author. Email, 16 February 2022.
- Brown, C. (2020). *CSAF Action Orders to Accelerate Change Across the Air Force*. From AF.Mil. Accessed 14 August 2021.
https://www.af.mil/Portals/1/documents/2020SAF/ACOL_booklet_FINAL_13_Nov_1006_WEB.pdf
- Camm, J., Cochran, J., Fry, M., Ohlmann, J., Anderson, D., Sweeney, D., & Williams, T. (2018). *Business Analytics* (3rd ed.). Boston, MA: Cengage Learning Inc.
- Choi, W. J., O'Connor, M. B., & Truong, D. (2019). *Predicting the U.S. airline operating profitability using machine learning algorithms*. International Journal of Aviation, Aeronautics, and Aerospace, 6(5).
- Clinton, K. (2015). *Rethinking C-17A Training Requirements: Air Refueling*. Graduate Research Paper, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.

- Everstine, B. (2021). *CSAF Outlines the Air Force's New Deployment Model*. Air Force Magazine. Accessed 30 April 2022. <https://www.airforcemag.com/csaf-outlines-air-force-new-overhauled-deployment-model/>
- Fauske, M., & Hoff, E. (2016). From F-16 to F-35: *Optimizing the Training of Pilots in the Royal Norwegian Air Force*. *Interfaces*, 46(4), 326–333.
- Gang Yu, Pachon, J., Thengvall, B., Chandler, D., & Wilson, A. (2004). *Optimizing Pilot Planning and Training for Continental Airlines*. *Interfaces*, 34(4), 253–264.
- Goldsmith, T., Johnson P. (2002). *Assessing and Improving Evaluation of Aircrew Performance*. *International Journal of Aviation Psychology*, 12(3), 223–240.
- Gutman, A., & Goldmeier, J. (2021). *Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning*. Indianapolis, IN: Wiley.
- Jaiswal, N. K. (1997). *Military Operations Research: Qualitative Decision Making*. Norwell, MA: Kluwer Academic.
- Klapper, E. S., & Ruff-Stahl, H. J. K. (2019). *Effects of the pilot shortage on the regional airline industry: A 2023 forecast*. *International Journal of Aviation, Aeronautics, and Aerospace*, 6(3).
- Kirby, S. N., & Naftel, S. (2000). *The Impact of Deployment on the Retention of Military Reservists*. *Armed Forces & Society*, 259–12.
- Light, T., Hamilton, T. Pfeifer, S. (2020). *Trends in U.S. Air Force Aircraft Mishap Rates (1950-2018)*. RAND Studies.
- Leeming, T. (2022). Personal Correspondence: *Tactics Advancement Course*. To the author. Email, 31 March 2022.
- Mattis, J. (2018). *2018 National Defense Strategy of the United States*. From the US Department of Defense. Accessed 14 August 2021. <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>
- McConville, S. (2019). *Scheduling for success: tailoring C-17 training and apportionment for the evolving mission environment*. Graduate Research Paper, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- McCormick, J. (2022). Personal Correspondence: *Aircrew Training Efforts*. To the author. Email, 14 February 2022.

- McCullough, A. (2018). *Boeing Recaptures C-17 Trainer Contract*. Air Force Magazine. <https://www.airforcemag.com/Boeing-Recaptures-C-17-Trainer-Contract/>
- McGee, M. (2021). Modeling Air Force Retention with Macroeconomic Indicators. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- Morales, J. (2022). *Developing Mobility Air Force Tacticians through Tactics Advancement*. Accessed 2 March 2022. <https://www.jbmdl.jb.mil/News/Article/2905623/developing-mobility-air-force-tacticians-through-tactics-advancement/>
- Parker, K. (2012). *Support the Combatant Commander, Develop the Force, or Roll the Dice?* Air & Space Power Journal, 26(1), 68–77.
- Pujats, T. S. (2020). Forecasting Attrition by AFSC for the United States Air Force. Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio.
- Ragsdale, C. (2017). *Spreadsheet Modeling and Decision Analysis: A Practical Introduction to Business Analytics*. Cengage Learning.
- Roman Nevzorov. (2018). *A Competence Approach Implemented in Training of Future Pilots at Higher Education Military Institutes to Apply Professional Knowledge into Practice*. Вісник Національного Авіаційного Університету, 76(3), 117–125.
- Secretary of the Air Force Public Affairs. (2022). *New Force Generation model builds high-end readiness, sustainability for Joint Force*. Accessed 2 April 2022. <https://www.af.mil/News/Article-Display/Article/2720333/new-force-generation-model-builds-high-end-readiness-sustainability-for-joint-f/>
- Slocombe, G. (2012). *Training Aircrew Officers for Mission Command*. Asia-Pacific Defence Reporter, 38(2), 14–17.
- Sohoni, M. G., Glenn Bailey, T., Martin, K. G., Carter, H., & Johnson, E. L. (2003). *Delta Optimizes Continuing-Qualification-Training Schedules for Pilots*. Interfaces (Vol. 33, Issue 5).
- U.S. Department of Transportation, Federal Aviation Administration (2017). *Advisory Circular AC 120-54A Chg 1: Advanced Qualification Program*. Accessed 15 January 2022. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-54A_CHG_1.pdf

- U.S. Government Accountability Office (GAO) (2020). *Ready Aircrew Program: Air Force Actions to Address Congressionally Mandated Study on Combat Aircrew Proficiency*. Report to Congressional Committees.
- U.S. Government Accountability Office (GAO) (2021). *TRANSPORTATION DOD Can Better Leverage Existing Contested Mobility Studies and Improve Training*. Report to Congressional Committees.
- Van Ovost, J. (2021). *Accelerating Change for Rapid Global Mobility Delivering Joint Force Success in the High-End Fight*. The Mitchell Forum. Mitchell Institute for Aerospace Studies., 39.
- Vidt, C. (2021). Personal Correspondence: *Sortie Based Training*. To the author. Email, 28 August 2021.
- Wagener, F. (2014). Crew Resource Management Application in Commercial Aviation. *Journal of Aviation Technology & Engineering*, 3(2), 2–13.
- Wells, M., Kretser, M., Hazen, B., & Weir, J. (2020). Modified C-17 taxi procedures: a fuel cost savings exploration. *Journal of Defense Analytics and Logistics*, 4(2).
- Wilcox, A. (2021). Personal Correspondence: *Flight Hour Program*. To the author. Email, 9 September 2021.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 074-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 03/06/2022		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From - To) July 2021 - June 2022	
4. TITLE AND SUBTITLE Creating Data-Driven Decisions in Allocating C-17 Aircrew Training				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Stevens, Christopher M, Major				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENS) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865				8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENS-MS-22-J-053	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Mobility Command A38 Colonel Nicholas Leonelli 510 POW/MIA Dr Scott Air Force Base, IL 62225-5011 Email: nicholas.leonelli@us.af.mil				10. SPONSOR/MONITOR'S ACRONYM(S) AMC/A38	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.					
13. SUPPLEMENTARY NOTES This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.					
14. ABSTRACT The United States Air Force is currently in a largescale paradigm shift concerning how capability requirements are viewed and assessed. Increased emphasis on accurately measuring aircrew performance and forecasting training requirements has forced senior leaders to rethink whether the training programs currently in place genuinely achieve their desired effects. Specifically, the process of training resource allotment lacks the ability to quantify the amount of resources required to achieve a necessary level of performance or the flexibility needed to be altered with changes in resources. This study proposes a methodology to identify factors that have a significant impact on aircrew performance. Variables are identified in C-17 simulator data utilizing linear regression to develop a prediction model. Using this model, different courses of action are created where resources can be modified in order to achieve a desired capability. While the final model utilized within this study produces mixed results, the concepts, methodology, and proposed application provide insights into the value of using models like those described in this study in creating data-driven decisions in training resource allocation.					
15. SUBJECT TERMS Aircrew Training, C-17A, Resource Allocation, Linear Regression					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 74	19a. NAME OF RESPONSIBLE PERSON Wales, Jesse G, Lt Col
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (937) 255-3636 x4740

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