1-30-2013

Evaluating Simulated Military Training Exercises

George Boese

George Boese

Candidate

Economics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

Dr. Margaret Blume-Kohout, Chairperson

Dr. Catherine Krause

Dr. Robert Valdez

EVALUATING SIMULATED MILITARY TRAINING EXERCISES

by

GEORGE BOESE

B.S., ECONOMICS, UNITED STATES AIR FORCE ACADEMY, 2009

THESIS

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

Economics

The University of New Mexico

Albuquerque, New Mexico

December, 2012

EVALUATING SIMULATED MILITARY TRAINING EXERCISES

by

George Boese

B.S., Economics, United States Air Force Academy, 2009

M.A., Economics, University of New Mexico, 2012

ABSTRACT

The Joint Kill Chain Event (JKCE), designed by the Air Force's Distributed Mission Operations Center (DMOC), provides an interesting case study for evaluation of simulated military training. JKCE places Army Air Defense Artillery Fire Control Officers (ADAFCOs) into an Air Force Control and Reporting Center (CRC) simulator. From the CRC, the ADAFCO leads an air missile defense scenario, protecting friendly assets (airfields, bases, missile launchers, etc.) and coordinating surface-to-air missile strikes. This scenario serves as the operational capstone event in the Army's ADAFCO certification course, which qualifies officers to perform a vital missile authorization function in major combat operations.

The current course configuration consists of two weeks of classroom-based academics followed by one week of simulator practice and testing. The didactic portion of the course is designed to teach the students those skills required of an operational ADAFCO. The simulator-based portion of the course, JKCE, tests those new skills in a

simulated operational environment. However, this analysis reveals major testing inconsistencies and statistically significant practice effects during JKCE. Additionally, the didactic portion of the course proved incapable of generating the desired operational result, successful completion of JKCE.

Changes to the structure and implementation of JKCE could improve this program, by shifting focus toward desired outcomes. In this thesis, I develop relevant performance metrics for JKCE, capture and analyze those metrics, and finally provide recommendations for course improvement, based on empirical evidence and program evaluation best practices.

TABLE OF CONTENTS

LIST OF FIGURES vii

LIST OF TABLES viii

INTRODUCTION 1

BACKGROUND 3

Cost Considerations 3

Simulator Fidelity 6

Pilot Utility and Welfare Maximization 8

Operator Acceptance 9

Improved DMO Evaluation 11

Experimental Design 11

CASE STUDY MOTIVATION 13

ADAFCO Responsibilities and Training 16

ADAFCO Course 17

The DMOC and JKCE 18

JKCE Schedule 19

Scenario Overview 20

Simulated vs. Live JKCE 22

METHODOLOGY 26

Performance Metrics 26

Four Level Analysis 28

Data Collection 29

Preliminary Experimental Design (JKCE 12-1) 31

RESULTS 33

Preliminary Results (JKCE 12-1) 33

Preliminary Observations (JKCE 12-1) 34

Secondary Analysis (JKCE 12-2) 36

Course Configuration 41

CONCLUSION 45

ACRONYMS 47

REFERENCES 48

APPENDICES 50

Appendix A: JKCE 12-1 Collected Data 51

Appendix B: JKCE 12-2 Collected Data 52

Appendix C: Reaction Times 53

Appendix D: Equations and Test Statistics 54

LIST OF FIGURES

Figure 1: ADAFCO Organizational Structure 16

Figure 2: JKCE Scenario 21

Equation 1: Regression Model 37

Equation 2: Hypothesis Test 40

LIST OF TABLES

Table 1: JKCE 12-1 Experimental Design 32

Table 2: JKCE 12-1 Reaction Times 34

Table 3: JKCE 12-2 Experimental Design 36

Table 4: JKCE 12-2 Reaction Times 37

Table 5: Regression Analysis 38

Table 6: JKCE 12-1 and 12-2 Record Run Reaction Times 38

Table 7: All Practice and Record Run Reaction Times 39

Table 8: Test Statistics 40

Table 9: Course Reconfiguration 2x2 Factorial Design 43

INTRODUCTION

In upcoming years the Department of Defense is expected to face a series of deep budget cuts, like many other federal agencies. As a result of the 2011 Budget Control Act, the Department of Defense will reduce spending by "approximately $487 billion over the next decade or $259 billion over the next five years" (Department of Defense, 2012, p. 1). The cause is a combination of the end of the War in Iraq and drawdown of troops from the War in Afghanistan, coupled with slow economic growth. A stagnant economy produces relatively less tax revenues, which provide the funding base for federal programs like those in the Department of Defense. The military's reaction has been a scramble to prioritize funding and preserve war-fighting capabilities.

Training expenditures are an area where the military could potentially save money by leveraging technological advancements. Military operators require a significant amount of specialized training on sophisticated equipment, which can be expensive. Deciding that training is too costly and then producing lower-skilled operators is not an acceptable alternative. The military's goal is to provide the same product, highly skilled operators, for less money by exploiting previously unrealized efficiencies. One way the Department of Defense seeks to achieve this goal is through increased use of modeling and simulation. As fuel, maintenance, and weapons system procurement costs rise, simulated training becomes more appealing compared to its live counterpart. From a cost effectiveness standpoint, modeling and simulation appears favorable to live forms of training.

Simulators provide an artificial environment for training, testing, and experimentation. The term simulator, as used in the military today, appeared shortly after the world's first aircraft about a hundred years ago. At that time a simulator was nothing more than a stationary aircraft cockpit used to practice basic procedures, which provided zero user feedback. While simulator-based military training started with aircraft pilots, it later spread throughout the Department of Defense.

Today, computer-based simulators are used to train nearly every form of military operator, i.e., those members of the armed services who are directly engaged in the act of warfare. Though the technology and complexity of military simulation has evolved, the premise remains unchanged: simulators exist to reduce the cost of training. These costs take many forms; including land, fuel, maintenance, ordnance, time, negative externalities, and even human lives. The military uses simulated training exercises to lower the cost of training and maintain a force of combat-capable operators.

Aside from high costs, a problem is prevalent in both live and simulated military training. Both forms, live and virtual, typically do not link training metrics to desired operational outcomes. In many cases, operationally relevant training metrics are not captured whatsoever. This is one reason why the question of simulated versus live training is commonly approached with a cost-effectiveness analysis (CEA).

Military leaders prefer less-costly training, assuming constant effects across the potential alternatives. However, there is a more fundamental question than which alternative is cheapest. That question is, does the training program, live or simulated, actually create or improve operational war-fighting skills? To execute a CEA, one must first establish some measure of the training effect. Military training programs should be justified by quantifiable results, in the form of operationally relevant skills.

BACKGROUND

Returning to the original objective, the military wants to deliver the same product, highly trained operators, at lower cost. One way to achieve this is by shifting training away from the live environment toward simulators in Distributed Mission Operations (DMO). However, this approach relies on the assumption that simulator training can effectively replace at least some live training requirements. This section provides an overview of the simulated military training literature with an emphasis on effects-based analyses.

Cost Considerations

As fuel, maintenance, and procurement costs rise, simulated training through DMO becomes relatively more appealing compared to its live counterpart. While simulators still have procurement and maintenance costs, in general these costs are orders of magnitude cheaper than their live counterparts.

Military equipment, for example planes and tanks, depreciates with use. Therefore an opportunity cost is associated with using this equipment for training, because theoretically this use causes equipment depreciation. Usage can shorten the equipment's lifecycle and increase maintenance costs. Again this is a benefit of simulation; military leaders are able to save equipment for its intended purpose, which is actual combat.

Aside from these strictly financial concerns, there are additional social costs to consider. Training with live military equipment may create significant negative externalities. Some examples include: noise pollution from aircraft and other heavy equipment, environmental pollution from vehicle exhaust or weapon tests, civilian property damage and or loss of life from training accidents. Simulation enables the military to avoid these social costs typically attributed to training programs.

Costs aside, sometimes live training is simply not feasible due to safety regulations and or test range limitations. Military test ranges are slowly disappearing as a result of population growth and increased environmental concerns. For instance, if the military wanted to test a new nuclear weapon it is unlikely that they would be able to use the ranges of Nevada, which was once commonplace. Again, simulation is unaffected by these constraints.

Basic Simulator Effectiveness Studies

When analyzing military training, distinguishing between live and simulated training can be difficult. Schank et al. (2002) distinguish between the two, saying "live training implies the exercise of the operational platform while simulated training suggests the operational platform is not exercised but rather is replaced by another device" (p. 47). Put simply, "live" implies that the operator is using the actual weapon system, for example an airplane or a tank. Simulators on the other hand are typically computer-based systems that emulate those live assets.

Jacobs et al. (1990) performed a meta-analysis of 26 experiments on simulated training, finding that simulated training combined with live training improved performance of basic flight procedures when compared to live aircraft training alone. This report establishes that simulator training can improve operators' skills. The finding is important; however, it should not be surprising. One expects that additional training, whether it occurs in the aircraft or simulator, would improve performance. However, the report is unable to offer any help in comparing live and virtual training effectiveness. Additionally, the authors promote the use of simulators as a supplement to live training rather than a replacement.

Bell and Waag (1998) were able to increase the scope of the virtual training analysis beyond basic flight procedures. They concluded that technological improvements have expanded the actual and potential usages of simulators in training (Bell & Waag, 1998, p. 232). The study demonstrated that military operators and policymakers were becoming more open-minded about the applications of simulated training. Finally, although they do not provide empirical evidence, they believe that simulation technology has reached the point necessary to effectively train combat-oriented skills (Bell & Waag, 1998, p. 232). It is important to remember that creating and developing combat skills is the fundamental goal of military training programs.

Distributed Mission Operations (DMO) takes the idea of standalone simulators to the next level by integrating multiple systems over a network. DMO requires linking multiple simulators in a common virtual environment, which allows multiple operators from different weapon systems to train together. The appeal is clear; these once stand-alone simulators can now be integrated into more complex training events. These simulated training events allow operators to not only practice individual skills, but also learn how their actions fit into the larger war-fighting system. Trainees can practice the communication and coordination skills that are also crucial in combat operations.

The Air Force conducted an all-virtual training exercise called RoadRunner 98 with the goal of assessing the training potential of DMO, previously known as

Distributed Mission Training. Crane et al. (2000) provided the first-ever DMO effectiveness analysis and concluded "technologies are capable of providing effective, warfighter training in simulation that complement aircraft training" (p. 45). However, this assessment, like the others discussed up to this point, was based entirely on subjective analysis. There was no detailed investigation of operationally relevant performance metrics. Essentially, the contribution is that the basic training effectiveness results attributed to simulators generally carry over to combat-oriented tasks exercised in DMO. An interesting take away from this study is that the writers still regarded DMO and simulation as a complement rather than a replacement for live training.

Simulator Fidelity

The literature discussed up until this point has advocated simulation and DMO as supplements to a live training regimen for high-risk and infeasible training events (Schank et al., 2006, p. 53). The intent of the first flight simulators was to facilitate task familiarization through repetition. For instance, new pilots use the simulator to practice landing procedures and radio calls.

The argument for DMO as a live training replacement comes down to the issue of fidelity. In this context, fidelity refers to the degree to which the simulation replicates the live environment. For an aircraft simulator, fidelity can mean the cockpit configuration, equipment behavior, and environmental characteristics of the virtual environment, to name a few. Military policymakers and simulator designers assume that operators want a simulator that looks, feels, and acts like the asset it is emulating.

The logic of the simulation fidelity argument is simple. Live training more closely resembles live war-fighting; therefore live training must be more effective than simulated training. When comparing identical training events, one virtual and one live, the prevailing assumption is that the live event better prepares operators for combat. Likewise, a higher-fidelity simulator is preferred to a lower-fidelity simulator. This argument is intuitively appealing, but to this writer's knowledge there has never been a pure simulator versus live training effectiveness study. Additionally, both simulated and live training share a common dilemma, they are not always supported by studies that link training to tangible improvements in combat skills. The burden of demonstrating training effectiveness falls on live and simulated programs alike.

The difficulty of providing a high-fidelity simulator varies greatly across weapons systems. For example, consider the Airborne Warning and Control System (AWACS). AWACS is essentially a radar control tower that exists on an aircraft. The cabin of the aircraft is filled with rows of computer workstations, rather than seating as in a traditional passenger aircraft. Typically, a four-person crew consisting of two pilots, a navigator and a flight engineer flies the plane. The other twenty or so personnel on the plane work at computer stations tracking and directing other aircraft. For these workstation operators, simulator fidelity is relatively easy to achieve, because one only has to replicate voice, data links, and radar traffic. However, accurately simulating the in-flight behavior of the plane is a much more challenging problem. Rather than train the entire crew together, separating the aircrew from the radar operators simplifies the fidelity problem.

Schank et al. (2002) analyzed simulator effectiveness across a few different weapon systems. The researchers examined three United States Navy weapon systems including a fighter aircraft called the F/A-18, an anti-submarine and radar aircraft called the P-3C, and a ship called DDG-51 destroyer (2002, p. iii). One of the major finding is that "simulators are used more often for events that involve the analysis of input data" and "used least often for events that attempt to replicate situational or environmental conditions" (2002, p. 46). This can be demonstrated in the context of the AWACS system operator and the fighter pilot. It is much easier to simulate the AWACS operator's computer-based equipment than the physical behavior of a fighter jet.

Unlike the AWACS, transitioning fighter aircraft toward increased use of simulation is difficult, because pilots' performance is so dependent on physical effects, such as high-speed maneuvering and G-force management. These effects are very difficult if not impossible to accurately simulate, thus in that sense simulator fidelity must increase until it can benefit fighter pilot training (Schank, 2002, p. 49).

Pilot Utility and Welfare Maximization

The fighter pilot is an interesting example to discuss within the context of a welfare maximization problem. There is a possibility that fighter pilots dislike simulator-based training, due to inadequate simulator fidelity, and that a shift from live to virtual training could cause a reduction in operator utility. The theoretical benevolent social planner would be interested in this utility reduction when calculating the cost portion of a cost-benefit analysis.

However, the fact that these individuals are military members bound to the lawful orders of their superiors simplifies the problem. While service-members' utility is an important concern, it should be considered lower in weight than military readiness.

Command structure aside, this issue should be argued on the grounds of training effectiveness and not operator utility.

From an economist's perspective, military training programs are evaluated on their ability to create or improve operational capabilities and cost. If the simulator-based training is cheaper and creates the same or better operational outcomes than live training, then the pilots would have no justification for complaint. Program effectiveness trumps operator utility when evaluating training alternatives.

However, recent literature suggests that simulator training is achieving higher levels of operator acceptance, even among fighter pilots. To clarify, the term operator acceptance refers to a subjective estimation of training effectiveness from the operator's perspective.

Operator Acceptance

The Air Force Research Laboratory (AFRL) hosted a DMO-based training program in 2006 with the goal of quantifying training effectiveness. As part of their research, the writers carried out an extensive survey process, aimed at the event's participants, F-16 fighter pilots and AWACS radar controllers. These operators participated in a weeklong training program and post-course survey. In their post-course survey, 49 of 50 AWACS controllers and 311 of 327 fighter pilots revealed that they would recommend this training event to fellow operators (Schreiber, Rowe, & Bennett, 2006, p. 18).

The overwhelming support demonstrated by operators for this DMO event is a milestone for simulated training. As Schrieber, Rowe, and Bennett conclude, "user opinion does not bestow an adequate measure of training success or effectiveness, but it can establish user acceptance, which is often necessary condition for continued system use" (2006, p. 1). This support suggests that, in the eyes of the operators, simulation fidelity has reached an adequate level of fidelity.

Fidelity is important to operator acceptance of simulated training, with the assumption that more is better. Considering most military decision makers are or were operators, fidelity will likely remain a simulation focal point. However, improving fidelity is not the underlying purpose of virtual training.

The goal of military training is improving operational performance, which does not necessarily require high-fidelity simulations. Salas, Bowers, and Milham state "fidelity, however, is not the only component of effective training" but rather "the ability of trainees to perform trained behaviors in operational settings" (2003, p. 4-5, 13). They maintain that increasing operational performance may or may not necessitate higher-fidelity simulators. Schreiber and Bennett reinforce this point, saying the "DMO training environments exist first and foremost to improve warfighter competence, not necessarily to create the most realistic environment as an end unto itself" (2006, p. 4). Military decision makers must concentrate on training effectiveness, which may or may not require improved simulator fidelity.

The literature in this section, particularly AFRL's survey results, suggests that simulation fidelity is improving, and as a result so is operator approval. Unfortunately, increased simulation fidelity and even operator acceptance is not the underlying goal. Improving operational performance is the purpose of any legitimate military training program, regardless of whether or not it occurs in a DMO environment.

Improved DMO Evaluation

The AFRL study from 2006 was the first to gather task-oriented performance data from a DMO event. As previously stated, they conducted a weeklong training course for F-16 pilots and AWACS operators. The course structure was as follows. The pilots:

> Received some simulator familiarization training and then were immediately benchmarked, or tested on their pre-training point defense scenario performance. Post-training reassessment with those same pilots using mirror-image point defense scenario benchmarks occurred at the completion of five-day DMO training. (Schreiber & Bennett, 2006, p. 5)

The researchers demonstrate that DMO training can have a significant positive effect on simulated performance. They provide statistically significant evidence that the DMO training program improved operator competency. This is an important step toward quantifying operational improvement as the result of simulated training.

This study represents an exception to the standard military training evaluation, because it was able to demonstrate tangible training benefits. However, the overarching problem with this study is that the analysis cannot reach into the operational realm. The results are relevant, but do not answer the question, does this training improve war-fighting performance? Once again, this is not uniquely a DMO problem, but rather a problem in evaluating any military training program.

Experimental Design

The literature discussed above suggests that simulated training in addition to live training improves live performance. Additionally, the AFRL DMO study demonstrated that simulator training can significantly increase performance in the simulator. However, "transfer of skill from simulation to the operational environments has not been established" (Salas, Bowers, & Milham, 2003, p. 13). Schreiber and Bennett (2006) stress that "application-oriented studies must also be undertaken, such as what is the degree of transfer to a live-fly training event" (p. 21). Schank (2006) suggests the following setup to investigate the transfer of skills from the simulator to the live environment:

> Such experimental studies would involve setting up two groups to accomplish specific training objectives. One group would use simulators; the other would use live training. The performance of the two groups along several dimensions would then be measured and compared to understand the impact of the two methods of training. (p. 44).

No one has ever conducted this type of experiment for military training, because of its difficulty and cost. However, even this explanation assumes a considerable amount of difference between the simulated and live environment. When evaluating JKCE, the high level of simulator fidelity makes such an experiment unnecessary. Additionally, replicating JKCE in a live environment is unrealistic for multiple reasons discussed later in this paper.

CASE STUDY MOTIVATION

This section begins with background on the Phased Array TRacking Intercept of Target (PATRIOT) weapon system and its performance during Operation Iraqi Freedom (OIF). Next, the discussion shifts toward the purpose and training of the U.S. Army's Air Defense Artillery Fire Control Officers (ADAFCOs). This leads into justification for the development and implementation of JKCE as an operational capstone event in the Army's formal ADAFCO training course. The section concludes by explaining why a simulation-based JKCE is preferred to theoretical live training alternative.

The PATRIOT is a surface-to-air missile system designed to defend military assets (personnel, bases, equipment, etc.) against enemy aircraft, cruise missiles, and tactical ballistic missiles (United States Army). Surface-to-air simply means that the PATRIOT missile launches from the ground and targets airborne assets, i.e. planes or other missiles. Basically, the system protects friendly assets from airborne threats.

Despite its longstanding and successful relationship with the Army, the PATRIOT missile defense system received significant negative attention in the early months of the War in Iraq. The PATRIOT weapons system was involved in two fratricide incidents, which killed three friendly aircrew members (Hawley, 2007, p.1). Two different friendly aircraft were misclassified as enemy missiles, then subsequently shot and destroyed. The PATRIOT system was employed eleven times during OIF, and every engagement resulted in successful destruction of the target. Unfortunately, two of those targets were misclassified friendly aircraft. Within the first few months of the Iraq War the PATRIOT missile system committed fratricide in eighteen percent of its engagements, an unacceptable statistic by American military standards.

The Defense Science Board, by direction of the Under Secretary of Defense for Acquisition, Technology, and Logistics, formed a Task Force to investigate PATRIOT system performance. Between August 2003 and June 2004, the Task Force investigated the two incidents, as well as the training and employment procedures related to the PATRIOT system (Williams & Delaney, 2005, p. 1).

The Task Force attributed the fratricide incidents to three major shortfalls: poor performance of the identification friend or foe (IFF) system, insufficient situational awareness across the members of the joint air defense system, and overly automated PATRIOT operating procedures. Subsequently, the Task Force provided recommendations addressing each shortfall.

The first shortfall was that the "combat identification capability embodied in the Mode IV IFF system performed very poorly" (Williams & Delaney, 2005, p. 2). Avoiding unnecessary technical details, the Mode IV IFF system on an aircraft sends an electronic signal that identifies itself as a friendly, i.e. "don't shoot me". This system obviously did not work properly during the fratricide incidents. However, given the sheer number of PATRIOT to friendly aircraft interactions, outside of defensive engagements, "even very-low probability failures could result in regrettable fratricide incidents" (Williams & Delaney, 2005, p. 2). The PATRIOT's radar will pick up anything airborne within its range, so if the system is located near an airfield, then there is a significant amount of activity. Moreover, there is a large amount of friendly air traffic in the beginning stages of a war, when there are still a high number of supply shipments and offensive strikes. Not surprisingly, the Task Force's recommendation was to "find and fix Mode IV IFF problems" (Williams & Delaney, 2005, p. 3).

The next shortfall was the lack of situational awareness, directed at the air missile defense system and command and control infrastructure in general. Command and control, as defined by Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms (2012), is "the exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission" (p. 59). This term describes the centralized control of decentralized assets within a particular area of responsibility.

William and Delaney (2005) summarize the problem by saying "we tend to assume that data are routinely communicated from one system to the other, that targets are correlated, and target information is shared and assimilated by all" (p. 2). Even without IFF, the command and control structure should be robust enough to correctly identify and categorize friendly aircraft. The problem is that this structure is not failsafe, because the military has a significant amount of aircraft tracking equipment that is not necessarily perfectly integrated. The fact that a friendly aircraft properly displays on one radar or data link system does not mean that it properly appears on every system.

The third and final shortfall was that the PATRIOT crews' protocol became too automatic and dependent on the system. This is what Hawley (2007) defines as automation bias or an "unwarranted over-reliance on automation" (p. 4). The Task Force recommended that the Army develop "a protocol that allows more operator oversight and control of major system actions" (Williams & Delaney, 2005, p. 3). In essence, the Task Force wanted to inject human reasoning into the launch decision, rather than placing one hundred percent trust in the equipment.

Two of the aforementioned shortfalls directly apply to the employment of the PATRIOT weapon system. First, communication across the aircraft command and control infrastructure needs improvement. Second, PATRIOT operators need to move away from their "trust the system" mentality. PATRIOT operational protocol needs increased human interaction and verification. One way the Army sought to implement these recommendations was with a reinvigoration of the ADAFCO position.

ADAFCO Responsibilities and Training

The ADAFCO acts as the upper-level approval authority for PATRIOT missile launches in an operational environment. Each ADAFCO can direct multiple PATRIOT battalions, which in turn direct multiple PATRIOT batteries. Figure 1 below provides an example of the basic organizational structure. Put simply, the ADAFCO is responsible for giving the "engage" order, which results in an interceptive PATRIOT launch against Anti-Radiation Missiles (ARMs), Theater Ballistic Missiles (TBMs), cruise missiles, or enemy aircraft.

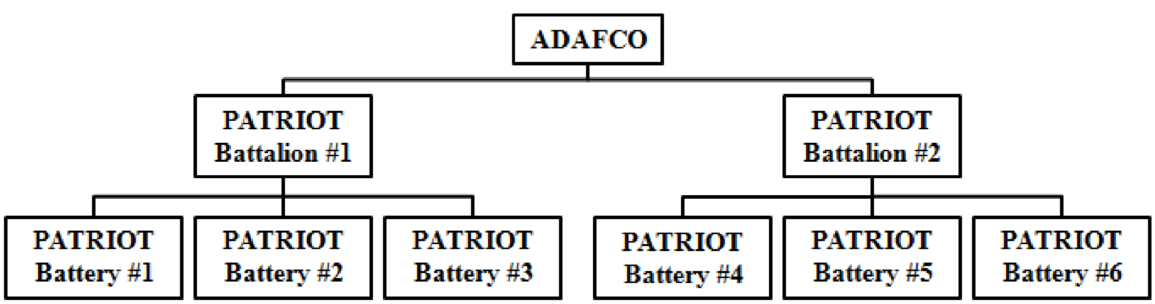

Figure 1: ADAFCO Organizational Structure

The ADAFCO is somewhat unique in that he or she operates within a joint command and control platform, that is, with the Navy, Marine Corps, Air Force, or any combination thereof. Although the Department of Defense is responsible for all branches of the military, the integration of their personnel and equipment is by no means a trivial task. Each branch operates somewhat differently, and in wartime these differences can reduce operational effectiveness or worse, lead to friendly or civilian casualties. The Army designated the ADAFCO position to bridge the intra-service gaps with the Air Force, Navy, and Marine Corps.

Some examples of the joint platforms that the ADAFCOs can operate within include; the Marine Corps's Tactical Air Operation Center, and the Air Force's AWACS, Control and Reporting Center (CRC), and Battlespace Command and Control Center. The ADAFCO integrates into the crew of one of the above command and control stations and acts as their PATRIOT liaison. Not only is the ADAFCO a system expert, but also he or she is the missile firing authority for the PATRIOT battalions. This setup ensures coordination of the individual PATRIOT batteries and provides a direct line of communication to the overarching joint command and control authority.

ADAFCO Course

In response to the Defense Science Board's recommendations, the Army "launched the development of a structured Army Air Defense Artillery Control Officer (ADAFCO) training school resulting in major changes to the USA Air Defense Artillery (ADA's) Training and Standard Operating Procedures (TSOP)" (Cavanagh, 2011, p. 1). The resultant course is three weeks in length and based on nine training objectives. For assessments, the course contains two written examinations and a DMO-based operational test known as JKCE. JKCE stems from the course objectives to familiarize the

ADAFCOs with the command and control agencies from which they will operate and execute joint kill chain operations (Lybarger, 2008).

The DMOC and JKCE

A tri-service working group congregated to implement the recommendations of the Defense Science Board Task Force. The working group concluded the Air Force's CRC would be the ideal location for the ADAFCO to prevent future PATRIOT-related fratricide incidents (Cavanagh, 2011, p. 1). Additionally, the CRC is the most-likely command and control system to be deployed in a wartime environment (Maule, 2010). In short, the creators of JKCE focused their efforts on the CRC, because it was the most frequently used platform in a deployed environment and best potential counter-fratricide system.

In addition to being a suitable counter-fratricide node and highly utilized system during deployments, the physical structure of the CRC van promotes a high-level of scenario fidelity. The simulator possesses the same hardware, software, and functionality of its live counterpart. Because the CRC is essentially a hardened windowless trailer, there is negligible difference between the live and simulated environment. To the CRC crew and ADAFCO, the scenario appears as it would in the real world. The system behavior and observable features are consistent with live operations. Because the CRC van does not have to move or even interact with its physical surroundings, the scenario designers avoid the typical fidelity issues associated with many other tasks, such as flight simulation. This relatively high level of fidelity blurs the line between live and simulated training, which is one reason why the JKCE presents such an interesting case study.

With the CRC in mind, the ADA school approached the DMOC to develop and execute the first-ever JKCE in June of 2006. Since the event's inception, the DMOC has hosted two JKCEs per year, typically every January and August. The ADA school chose the DMOC not only for its CRC simulator, but because the DMOC is an Air Force DMO center of excellence. The DMOC designs, executes, and supports a variety of DMO test, training, and experimentation events.

JKCE consists of much more than just the CRC; constructive simulators are necessary to render the friendly aircraft, enemy aircraft, enemy missiles and enemy missile sites. Simulated PATRIOT battalions and radio capabilities are also necessary. The DMOC is able to provide these capabilities, as well as the expertise to design virtual combat scenarios while adhering to customer requirements.

From 2006 through 2009, an Army base in Texas called Fort Bliss provided the PATRIOT capability for the exercise. Because of this, JKCE was a two-site DMO event with the CRC simulator and constructive simulators located at the DMOC on Kirtland Air Force Base, and the PATRIOT simulator at Fort Bliss. In 2010, the DMOC expanded its support with the addition of a PATRIOT emulator. The entire ADAFCO course moved to the DMOC, which created savings by eliminating unnecessary travel costs and reducing the amount of simulator integration.

JKCE Schedule

As previously stated, the entire ADAFCO course lasts three weeks, with two weeks of classroom learning and the week of simulator practice and testing, otherwise known as the JKCE. The JKCE itself is also broken into three parts: simulator familiarization, practice runs, and record runs.

On the simulator familiarization day, the CRC crews and ADAFCOs observe the simulators and learn about the scenario and their performance objectives. During the next two days, each ADAFCO gets a one-hour practice run in the CRC. The course instructors do not grade these practice runs, and in past JKCEs the engineers at the DMOC had not recorded these runs either (until JKCE 12-2).

Finally, during the last two days each ADAFCO gets a single record run. The instructors grade the record runs on a pass or fail basis, which determines whether or not the student will receive their ADAFCO certification. Starting in January of 2012, the DMOC began recording the graded runs, so that the information would be available to course instructors, scenario developers, and system engineers for analysis.

Scenario Overview

JKCE is an integrated air missile defense scenario, which requires the ADAFCO students to respond to a myriad of enemy threats; Figure 2 below provides a simplified visual overview. The friendly assets are green, enemy assets are red, and the blue represents a body of water. There are more PATRIOT battalions, aircraft, and missiles in the actual scenario; however the figure below provides a basic outline. Additionally, JKCE uses real-world terrain and locations; the figure below suffices while avoiding possible security classification issues.

The CRC and PATRIOT battalions are located on one side of a border, with enemy TBM sites and SCUDs on the other. Constructive simulators generate the TBM and SCUD sites and launch missiles at pre-programmed intervals throughout the scenario. A pair of PATRIOT emulators represents the two battalions, each equipped with a radio feed to the CRC. This setup enables just one operator to simulate an entire PATRIOT battalion, which would normally require significantly more personnel.

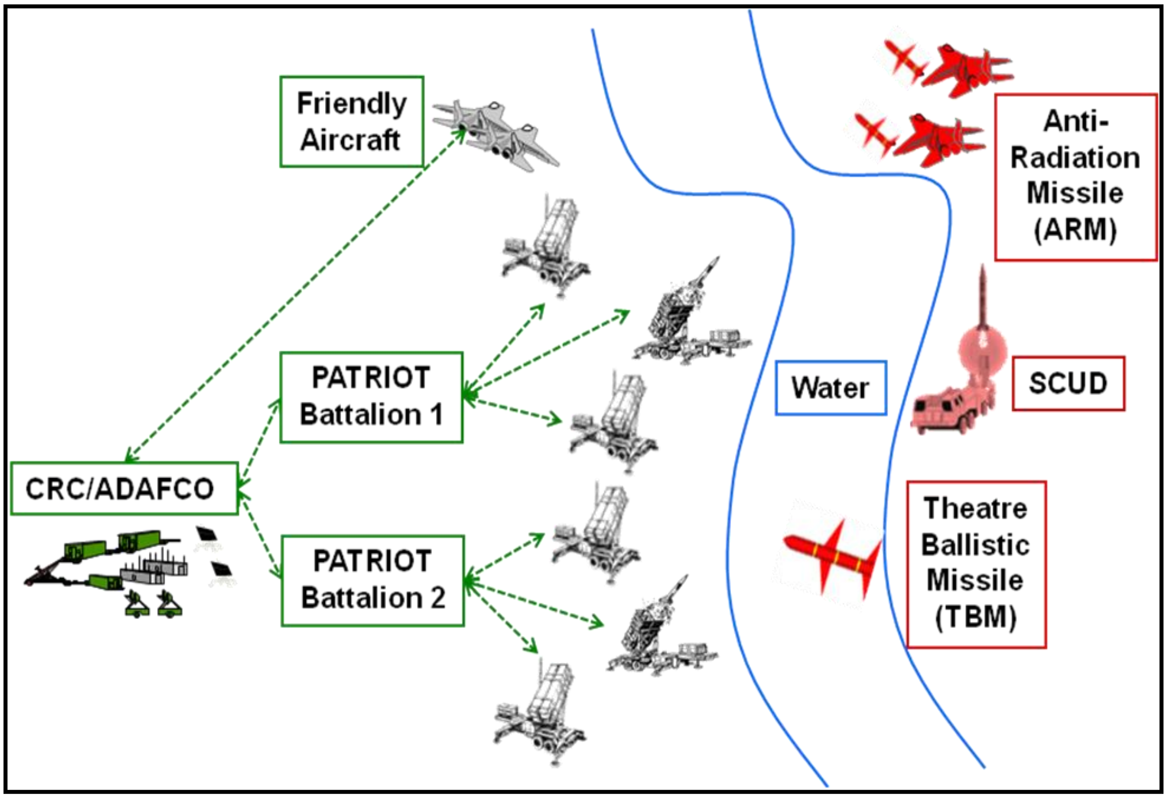

Figure 2: JKCE Scenario

From the start of the scenario, there are both friendly and enemy fighter aircraft flying near the border areas. The DMOC's red air constructive simulator generates the enemy aircraft, which run autonomously on pre-programmed routes. Another constructive simulator generates friendly aircraft, which are controlled by other active duty personnel or government contractors. The scenario developers chose to require that personnel fly the friendly aircraft, so that the CRC crew would be able to perform their normal aircraft control function. The CRC crew has no role when not communicating with pilots, tracking, and directing aircraft. This feature improves the event's realism and enables the CRC crew to get practice during the event. While the CRC crew does accomplish some of their own continuous training requirements in JKCE, the DMOC designed the scenario to train the ADAFCOs first and foremost.

In the context of the scenario, the Air Force CRC crew performs aircraft command and control functions. That is, they direct friendly aircraft to defend military bases against an onslaught of attacking enemy aircraft. This provides the ADAFCO with the situational awareness necessary to prevent potential fratricide. If the PATRIOT incorrectly classifies an aircraft, the ADAFCO is able to crosscheck with the CRC crew before giving an engagement order. This engagement de-confliction is the anti-fratricide function alluded to in the previous section. The CRC crew with the ADAFCO identifies aircraft and missiles as hostile or friendly, and then coordinates PATRIOT missile strikes as applicable. The span between threat detection and engagement is called the kill chain, hence the event's title.

At its core, JKCE is an exercise in task saturation. The ADAFCO must direct two PATRIOT battalions that are almost constantly engaging enemy aircraft and missiles throughout the scenario. At the same time, they are de-conflicting targets with the CRC to ensure that the PATRIOT battalions don't accidently engage friendly aircraft. The radio chatter is almost nonstop, and the ADAFCO must quickly recognize and react to missile attacks, in order to succeed.

Simulated vs. Live JKCE

The political and operational motives for the ADAFCO course and JKCE are clear. The military wants to avoid future PATRIOT-related fratricide because it is counterproductive to successful military operations and most importantly costs friendly lives. Assuming that the Army now must provide formal ADAFCO training, the DMO environment offers many benefits over and above a live alternative.

The first benefit of a simulated JKCE is the high level of fidelity from the ADAFCO's perspective. As previously stated, in JKCE the ADAFCO operates from within a CRC simulator, which is identical to the live system. The ADAFCO only interacts with radar systems, data link equipment, and radios all of which are relatively simple to emulate. From their perspective, the scenario appears as it would in the real world, except that the CRC is located indoors rather than in the field. This high-fidelity setup is expected to improve the operators' acceptance of the training. Additionally, this makes the difference between the simulated scenario and a theoretically identical live scenario negligible, except that the students know the simulation is not real. The analyst can confidently assume that the ADAFCO's performance in the simulation would mimic their performance in the same live situation.

Second, due to significant safety concerns JKCE is an infeasible scenario to conduct in a live environment. As elaborated in the scenario overview section, the ADAFCO conducts multiple PATRIOT strikes against incoming enemy missiles and aircraft. To recreate these effects in a live environment would require pilots to fly the enemy aircraft that the PATRIOT battalions are shooting down, as well firing missiles at the PATRIOTs and other friendly assets. While an unsuccessful engagement could mean a failed run in the simulated environment, the same situation live would result in destroying a PATRIOT battery and killing its crewmembers. The military has not nor will ever engage in a training scenario with such significant risks and consequences.

Third, JKCE in a live environment is not feasible due the test range limitations. The scenario occurs in an area of about sixty thousand square miles. As a comparison, the area of White Sands Missile Range in New Mexico is only about three thousand, two hundred square miles. There are not any ranges in the United States large-enough to cover that area.

Fourth, executing JKCE in a live-fire environment would be incredibly expensive from a monetary standpoint. According to a Department of Defense's *Selected Acquisition Report* (2011) the unit cost of a single PATRIOT missile is approximately six million dollars. Considering each ADAFCO student orders approximately twenty missile strikes during the course, in both practice and record runs, this drives the live cost of JKCE up to approximately two hundred million dollars for missiles alone. That figure towers over the approximately forty thousand dollars in contractor support that it requires the DMOC to host a virtual JKCE.

Fifth, an all-live version of JKCE would create significant negative externalities in the form of noise and environmental pollution. Both missile launches and fighter aircraft, which comprise a majority of the JKCE scenario, generate a significant amount of noise. This exercise could impose negative externalities on citizens if the test range was not sufficiently isolated from local populations. Additionally, live aircraft and missile strikes generate fire, smoke, and exhaust that can cause negative environmental impacts. The military avoids these potential social costs by utilizing a virtual training environment.

Assuming the execution of JKCE is a binding requirement for the military, DMO is strictly preferred to a live environment. The simulated version offers near-live fidelity at significantly lower monetary and social costs. Additionally, a live JKCE is infeasible due to test range limitations and basic safety considerations.

Simulation is clearly the preferred method for conducting JKCE from the cost standpoint; however one crucial question remains unaddressed. Does JKCE or the ADAFCO course as a whole actually improve the operational performance of ADAFCOs? Though not possible to analyze their wartime operational performance, it is assumed that JKCE replicates the challenges of actual warfare. Each ADAFCO directs approximately twenty strikes in just an hour and a half of total simulator time during JKCE. Remember that there were only eleven PATRIOT engagements during the first three months of OIF for the entire United States Army. This high level of task saturation serves a stress test for the ADAFCOs prior to awarding their certification. Therefore, improvements in JKCE performance should translate into improved real-world operational ability.

METHODOLOGY

This section includes a justification for the performance metrics and description of the data gathering process. This study required the collection of an original dataset, using only the JKCE pre-recorded scenarios. Previously, the DMOC only captured the number of students who completed the course.

Performance Metrics

The first important metric is successful event completion, i.e. did the student pass JKCE? This pass or fail criteria is based on whether or not the ADAFCO successfully defended all of the friendly assets. To elaborate, the ADAFCO must successfully order defensive PATRIOT strikes against all of the oncoming enemy missiles.

In the January 2012 event JKCE 12-1, every student accomplished the overarching air defense goal during record runs. This pass rate may suggest that the scenario is too easy; however the scenario only needs to challenge students to the extent that it establishes the baseline performance necessary to serve as an ADAFCO. From an institutional perspective, course completion equates to an ADAFCO certification and authority to hold that position. Based on the expert opinions of the instructors and system experts that designed the event, JKCE successfully tests that baseline.

It is possible that the high pass rate observed in JKCE 12-1 reflects selection bias or significant prior experience. As former PATRIOT operators, the students could already be familiar with ADAFCO procedures and thus relatively unchallenged by the scenario. Similarly, JKCE 12-1 may have just contained and abnormally large concentration of talented students. These potential issues are discussed in more detail later on.

The more descriptive performance metric, from an analytical perspective, is response time. For the purpose of this analysis, response time is the time from the emergence of the enemy threat to the ADAFCO's engagement order. In other words, the time between the enemy's missile launch and the ADAFCO's "engage" command. As previously stated, this is the time when the ADAFCO students are coordinating between the CRC crew, friendly aircraft, and PATRIOT battalions.

In the military, the ability to rapidly respond is essential, and this is especially true for ADAFCOs. From a purely defensive perspective, a quicker response means that enemy aircraft and missiles are shot down further away from friendly assets. In the case of a miss, this might mean that the PATRIOT battalions have enough time to fire a second shot at the target. On the other hand, more time may also improve the ADAFCO's anti-fratricide function. PATRIOT operating procedures dictate mandatory system defense once an unidentified flyer enters within a certain distance. The ADAFCO must positively identify aircraft before they cross the PATRIOT's self-defense threshold, at which point the battalions will fire. The PATRIOT operators will be forced to assume the potential target is hostile, if the ADAFCO cannot properly identify the target in time. Therefore, a quicker ADAFCO has relatively more time to investigate unknown and potentially misclassified threats. For both general defense and anti-fratricide coordination a quicker response is a desirable trait.

The operational outcome of interest at the core of the ADAFCO course is reduced response time. Therefore, one could evaluate the effectiveness of the ADAFCO course based on its ability to improve that response time. Unlike many other military training programs, this course could justify its existence by demonstrating tangible improvement of a desired operational skill, quicker ADAFCO response times.

Four Level Analysis

Kirkpatrick (1998) developed a now widely utilized model for training program analysis. This model evaluates training programs at the following four levels: reaction, learning, behavior, and results. Reaction is a measure of operator acceptance or trainee satisfaction with the program. Learning is an increase in the students' knowledge or skill levels. Behavior refers to changes in participants' future actions as a result of training. Lastly, results are the final outcomes caused by the training.

This study does not directly measure the students' reaction to training during JKCE or the ADAFCO course in general. An end of course survey would be an easy way to incorporate this element into future analyses. However, it is assumed that the students positively react to the ADAFCO course, given the alternative is no training at all. The course provides students with at least some level of confidence and experience, before they serve as ADAFCOs in a wartime environment.

The next level for consideration is learning, which this analysis measures in two ways. The first and more general measurement of operational learning is whether or not the student passed JKCE. The second and more direct measurement is reaction time and more importantly improvements in reaction time.

The third level of evaluation concentrates on behavioral changes. In the case of the ADAFCO course this is an easy criterion to meet. After the course, the students will serve as ADAFCOs, which is a duty that they have not performed before. The fact that the students will assume a new position, with its own unique requirements, guarantees that they will display a behavioral change. However, it would be difficult to determine whether the ADAFCO course or the job itself caused those behavioral changes.

The final level of training program analysis is results. Currently, the course does not capture any final results. However, one might measure course results with a follow-up JKCE record run post-training. For instance, former students could come back to the DMOC sometime after completing the ADAFCO course to retest their skills in JKCE. Then analysts could determine whether or not the course is creating lasting operational results, rather than temporary skill level improvements. Ideally, the analyst would investigate results in an operational wartime environment. However, gathering reaction time data in a deployed environment could be troublesome given the relatively low number of PATRIOT engagements, day-to-day unpredictability, and overall danger.

Due to data limitations and practicality issues, this analysis focuses on the learning level of evaluation. More specifically, skill level improvement as demonstrated by reduced ADAFCO reaction time is the primary interest. The following subsections describe the data collection process and experimental design used to evaluate the level of learning in the ADAFCO course.

Data Collection

An application called the DMOC replay tool enables detailed analysis of the ADAFCO's response time. This tool allows an analyst to record all of the simulation traffic for a given event or test; in this case JKCE 12-1 in January of 2012 and JKCE 12-2 in August 2012. Aside from the simulated assets, i.e. missiles, planes, and PATRIOT batteries, the replay tool also captures the radio conversations. This allows the analyst to observe the virtual battle from a bird's eye view, while listening to the radio chatter.

With the virtual events on the record, the next step is scenario familiarization. The students have basic intelligence on the enemy's capabilities and know that they are entering a hostile environment, however they do not know the number or mix of threats they will face. A full thirty-minute record run requires the students to identify, de-conflict, and "engage" three questionable missile attacks. Questionable means that these missiles are not easily classified and could be friendly aircraft, therefore PATRIOT engagement requires ADAFCO direction. After identifying these three areas of interest the data collection began.

As discussed in the previous section, response time is the time between the enemy's missile launch and the ADAFCOs "engage" command. Using the DMOC replay tool, the analyst can visually observe and document enemy missile launches. Then he or she must listen for the "engage" command and record the time accordingly. The difference between these times is the response time performance metric.

It may seem that the missile interception or detonation would serve as a better end point for the response time. However, the ADAFCO students are the targeted training audience and not the PATRIOT operators. The "engage" command more accurately captures the completion of the ADAFCO's task, because the missile launch is then in the hands of the PATRIOT operators. This approach avoids potential measurement errors that the PATRIOT battalions could cause with varying reaction times. If the goal were to test and analyze the entire kill chain, then it would make sense to use missile launch or detonation as a stopping point. However, it would then be necessary to utilize fully manned PATRIOT battalions and batteries, rather than the simplified JKCE set-up. While this theoretical training scenario may have merit, it is not the most efficient way to accomplish the current goals of the ADAFCO course. The "engage" command best captures ADAFCO response time, which is the performance metric of interest.

This manual recording process likely introduces some human measurement error into the data. For instance, it is possible that the data became more precise over time as measurement technique improved. However, that error is probably less than one or two seconds. As demonstrated in the following sections, errors of such small magnitude would not change the recommendations presented in this analysis.

Preliminary Experimental Design (JKCE 12-1)

Due to the aforementioned data constraints, the preliminary analysis of the ADAFCO course follows a basic one-group posttest-only quasi-experimental design, represented in Table 1 below. That is, there is one group of test subjects who all receive treatment (ADAFCO course), and all subjects receive a single test after that treatment (the JKCE). The absence of random assignment makes this a quasi-experiment, because there is nothing random about ADAFCO students' assignment to the course. In addition, as Shadish, Cook, and Campbell (2002) write "the absence of a pretest makes it difficult to know if a change has occurred, and the absence of a no-treatment control group makes it difficult to know what would have happened without treatment" (p. 106). This statement is a significant strike against evaluation of the current course's relevancy and effectiveness.

Treatment	Posttest
X	O1

Table 1: JKCE 12-1 Experimental Design

It is arguable that the students could not successfully complete JKCE without training, which is the didactic portion of the ADAFCO course. Therefore, the fact that all thirteen students passed JKCE 12-1 might constitute a skill-level improvement. The problem with this argument is that most of the students are familiar with the ADAFCO duties and have likely worked with one in past assignments. There is a chance that some of the more-talented students could make it through the scenario without the course. For this reason, successful event completion alone is not a sufficient argument for the course's effectiveness. JKCE sets a baseline for the students' operational skills, but cannot really validate any improvement of said skills as a result of the ADAFCO course. Some sort of pretest is necessary to measure improvement in the students' skill level or responsiveness.

RESULTS

This section presents the results from the overall analysis, based on data from JKCE 12-1 and 12-2. The preliminary analysis, JKCE 12-1, faced significant limitations, but laid groundwork for a more in-depth follow on study. Despite the constraints, the preliminary analysis served as a proof of concept for gathering operationally relevant performance metrics in DMO training events. The secondary analysis focused on JKCE 12-2 and benefited from previous lessons learned. Merging the data from both exercises enabled a more robust analysis with recommendations for future course configuration.

Preliminary Results (JKCE 12-1)

The results of the JKCE 12-1 analysis are available in Appendix A, while Table 2 below contains some basic summary statistics. The table lists average response time across students and the corresponding standard deviations for each of the three enemy missile threats. The table in Appendix A displays the complete dataset, and contains each student's response time by threat. The grey sections of that table indicate that a particular student did not face the corresponding threat. This happened because students either started late after certain threats already launched or finished the scenario prematurely. Testing inconsistency is a topic of discussion in the following subsection, because all students should have encountered the same mix of threats.

As seen in the table below, the data is compiled by individual threats. This method is preferred to lumping all three reaction times together into a single composite score for each student. As stated above, testing inconsistency was a problem during JKCE 12-1. Therefore, only response times with respect to a single threat are comparable. Assigning composite scores to students that only faced one or two threats would introduce bias. In order to avoid this issue, the threat-based sample calculations simply exclude the empty values.

	Threat 1	Threat 2	Threat 3
Average	220.18	106.75	203.33
Std Dev	61.85	45.05	100.42

Table 2: JKCE 12-1 Reaction Times

As previously stated, this was the first iteration of JKCE that utilized the replay tool, but unfortunately the DMOC only recorded the graded runs. Therefore, it is not possible to examine the statistical relationship between the students performance in their practice versus record runs. Despite the data limitations, this analysis retains importance as a proof of concept. It demonstrated that, given an event recording exists, measuring the students' response times is relatively straightforward. Furthermore, this analysis established a foundation for the follow-up effectiveness study discussed later.

Preliminary Observations (JKCE 12-1)

The preliminary JKCE analysis revealed two obvious areas for improvement. First, the instructors inconsistently tested students, despite the use of a standardized scenario. Second, course directors administer the practice scenario too closely to the record run, given the scenarios are identical.

Testing consistency is the first obvious area for improvement. As demonstrated by the grey cells in Appendix A, not all of the students took the "full" test. In the extreme example, a student didn't have to react to a single questionable threat, while others responded to all three. This inconsistency damages the credibility of the course, particularly when graduating means that the student is qualified to serve as an ADAFCO.

According to Shadish, Cook, and Campbell (2002) "casual inference from any quasi-experiment must meet the basic requirements for all causal relationships: that cause precede effect, that cause covary with effect, and that alternative explanations for the causal relationship are implausible" (p. 105). For the ADAFCO course, only the first requirement is easy to achieve. The ADAFCO course or "cause" does precede the implied effect of improved performance during JKCE.

The second criterion is somewhat more troublesome, because at this time it is unknown whether or not the cause and effect co-vary. The cause or didactic instruction does co-vary with the effect of passing JKCE. However, as previously stated it is undetermined whether or not the students could succeed in JKCE without any instruction, i.e. the ADAFCO course. If the ADAFCO course improves JKCE performance in a statistically significant way, then this condition will hold.

The third requirement is also difficult to satisfy, because there are plausible threats to internal validity. The primary threat here is the practice effect, particularly given the shortness of the event. The time difference between the practice and record runs is only about two days, and the scenario is exactly the same. Letting students practice the test only a day or two before they take it for a grade is questionable. Practice effects are expected given this course configuration. The didactic portion of the ADAFCO course cannot claim any performance difference between the two runs, because the treatment precedes both tests.

During the initial review of the JKCE 12-1 recordings there was only data available on the actual test or record run. Therefore a detailed investigation of potential practice effects was not possible. However, in the follow-up analysis of JKCE 12-2, the DMOC recorded both the practice and graded runs. This enables identification of performance differences between the two runs.

Secondary Analysis (JKCE 12-2)

In order to investigate practice effects and provide further course improvement recommendations, a follow-up analysis was conducted after JKCE 12-2 in August 2012. The performance metrics and data gathering mirror the initial analysis, except that the experimental design now includes two posttests as shown in Table 3 below. Unfortunately, it was not possible to pair an individual student's practice and record run results, because the recording files do not contain the student names.

Treatment	Posttest 1 (Practice Run)	Posttest 2 (Record Run)
X	O1	O2

Table 3: JKCE 12-2 Experimental Design

The new experimental design and accompanying dataset, Appendix B, offers the first detailed look at practice effects in JKCE. As shown in Appendix B, all but one of the students has at least one "HIT" during their practice run. "HIT" means that the student failed to give the engagement command before the enemy missile impacted its target. That means that all but one student would have failed JKCE if they were tested for a grade directly after the classroom training.

In order to create a more robust dataset, a reaction time equal to the time between launch and impact is assigned to the hits. This methodology actually causes downward bias, because only a few students were on the brink of an engagement command when the missiles impacted. Therefore, a majority of the students would have even longer reaction time then the maximums that they were assigned. Table 4 below presents summary statistics from the JKCE 12-2 analysis.

	Threat 1		Threat 2		Threat 3	
	Record	Practice	Record	Practice	Record	Practice
Average	84.18	267.67	91.17	230.42	103.82	325.55
Std Dev	63.66	71.17	36.20	52.45	44.00	29.55

Table 4: JKCE 12-2 Reaction Times

The practice run itself appears to spur an improvement in student performance from a reaction time standpoint. A simple regression model sheds light on the magnitude of these improvements. As shown, the dummy variable is set to zero for a practice run and one for a record run.

$$\text{ReactionTime} = \beta_0 + \beta_1(\text{DummyVariable})$$
$$\text{RecordRun} : DV = 1$$
$$\text{PracticeRun} : DV = 0$$

Equation 1: Regression Model

Because the dependent variable only assumes two values, one or zero, the best-fit line runs through the reaction time means for the practice and record runs. Hence, the intercepts shown in Table 5 below correspond to sample means for the JKCE 12-2 practice runs. The slopes are the differences between the record and practice run sample means. Consider Threat 1 for instance, the model predicts that a student's record run reaction time will be just over three minutes faster than their practice run reaction time. However, data restrictions limit the power of this model, because other presumably important performance factors like individual skill level or prior experience are not included.

	B0	B1
Threat 1	267.67	-183.48
Threat 2	230.42	-139.25
Threat 3	325.55	-221.73

Table 5: Regression Analysis

Approximately forty-seven percent of the assessed enemy threats leaked through during the JKCE 12-2 practice runs. This appears to be evidence to the contrary of the statement that some high-skilled students could pass JKCE even without didactic portion of the ADAFCO course. However, it is possible that the poor performance is the result of selection bias. For instance, the students in JKCE 12-2 might be an unusually untalented group, but this theory is unlikely given the selective nature of the ADAFCO career field. Officers are chosen for the ADAFCO course because they have shown proficiency in previous positions. Seemingly, this reduces the likelihood of sampling errors.

Table 6 presents the averages and standard deviations for the JKCE 12-1 and 12-2 record runs side by side. If anything, this data suggests that the JKCE 12-2 students were a relatively talented group, at least when compared to the JKCE 12-1. The fact that the entire JKCE 12-2 class failed their practice runs cannot be attributed to a difference in skill level between the JKCE 12-1 and 12-2 samples. Based on the data available, it is likely that the JKCE 12-1 class also struggled with the practice scenario. Unfortunately, there is no data available to confirm that assertion.

	Threat 1		Threat 2		Threat 3	
	12-1	12-2	12-1	12-2	12-1	12-2
Average	220.18	84.18	106.75	91.17	203.33	103.82
Std Dev	61.85	63.66	45.05	36.20	100.42	44.00

Table 6: JKCE 12-1 and 12-2 Record Run Reaction Times

The more likely explanation for the unfavorable practice performance is that the classroom-based portion of the ADAFCO course does little to increase the operational skills of the students. If the goal of course is to improve the operational capacity of the ADAFCOs and JKCE is a legitimate assessment of that capacity, then the didactic portion of the course should foster favorable performance in JKCE. However, the classroom learning alone appears inadequate to garner the baseline of JKCE success.

Table 7 below shows the sample averages and standard deviations for all of the available practice and record runs. To be clear the practice sample only consists of JKCE 12-2 data, while the record run sample captures the JKCE 12-1 and 12-2 results. Clearly the record run sample demonstrates improved performance when compared to the practice run sample. Nevertheless, additional hypothesis testing is used to expand the evidence of practice effects.

	Threat 1		Threat 2		Threat 3	
	Record	Practice	Record	Practice	Record	Practice
Average	152.18	267.67	98.96	230.42	125.14	325.55
Std Dev	92.71	71.17	40.75	52.45	69.54	29.55

Table 7: All Practice and Record Run Reaction Times

A hypothesis test is constructed to test differences between two population means, practice versus record runs, with unknown population standard deviations (Anderson, Sweeney, & Williams, 2009, p. 403). The null hypothesis is the practice run population mean for a single threat minus the record run population mean for the same threat is less than or equal to zero. A failure to reject the null hypothesis means there is not a statistically significant difference between the practice and record run population means, or worse that the practice scores might actually be better. The alternative hypothesis is the practice run population mean for a single threat minus the record run population mean for the same threat is greater than zero. Thus rejecting the null hypothesis, suggests there is a statistically significant and positive difference between the practice and record runs, and the record run reaction times are better.

$$H_0 : \mu_{\mathrm{Pr}\,actice} - \mu_{\mathrm{Re}\,cord} \leq 0$$
$$H_1 : \mu_{\mathrm{Pr}\,actice} - \mu_{\mathrm{Re}\,cord} > 0$$

Equation 2: Hypothesis Test

With the test hypotheses in place, the next step is generation of test statistics and corresponding degrees of freedom. The general equations and results are included in Appendix D, while just the individual degrees of freedom, test statistics, and p-values are presented in Table 8 below. All three tests, one for each threat, are upper-tailed tests with positive rejection regions. As shown below, the outcome is to reject the null hypothesis in all three cases with over ninety-nine percent confidence.

	DOF	Test Stat	P-Value
Threat 1	28.26	4.07	1.75E-04
Threat 2	18.07	7.62	2.42E-07
Threat 3	18.48	9.67	7.43E-09

Table 8: Test Statistics

The results of the secondary JKCE analysis suggest that there is a statistically significant and positive difference between the ADAFCO students' practice and record run performance. To simplify, the students display quicker reaction times in the graded scenario runs than the practice runs. This supports the claim that the current ADAFCO course configuration elicits significant practice effects.

The classroom-based portion of the course cannot logically cause this performance improvement, because both events occur afterwards. The claim that the didactic portion of the course trains students to operate as ADAFCOs and further that successful completion of JKCE certifies that ability is false. The more likely explanation, is that the students pass JKCE because they perform a practice test, identical to their graded test, only a day or two beforehand. Based on the data collected, one expects that most students would fail JKCE if not for the practice runs. While this discredits the ADAFCO course's current configuration, the results of this analysis are actually quite positive.

As previously stated, the goal of the course is to improve the operational capability of soon to be ADAFCOs. The JKCE scenario establishes a baseline for that operational capability, under the assumption that students who successfully complete the scenario are good enough to serve as ADAFCOs. While the didactic portion of the course appears incapable of achieving this baseline, simulated operational training in the form of practice runs rendered statistically significant performance increases. An increase in simulated training could spur further improvements to the students' operational skills.

Course Configuration

Restructuring the ADAFCO course is a worthwhile endeavor given the didactic element's inability to generate the desired operational result that is, passing JKCE. This is important, because tax revenue funds the ADAFCO course like all other military endeavors. The military must wisely and responsibly use those funds to create and improve war-fighting capabilities. The ADAFCO position and accompanying skill set is indeed a legitimate military requirement. However, ADAFCO training must be designed with the development and improvement of operational capabilities as the top priorities.

The first step in reconfiguration is evaluating the didactic portion of the course. The course instructors have some purely knowledge-based objectives not directly tested by JKCE, which are still highlighted as ADAFCO course learning objectives. They must concentrate on teaching and testing that information within the classroom-based portion of the course. Additionally, if the academic test scores were available to researchers it would enable a more robust course analysis.

With all the relevant learning objectives identified, both operational skills and general knowledge, one can start experimenting with different course configurations. Table 9 below presents a two-by-two factorial design created to examine alternatives. This experimental design varies two factors, which are the amount of didactic and simulator training. Additionally, each factor contains two levels measured by duration in weeks. For simplification, the two levels of measurement are one and two weeks. However, these levels could be adjusted to different lengths of time, given this design is just an example. Similarly, additional levels of measurement could be added, which would provide for a more detailed analysis.

		Classroom Academics	
		1Week	2 Weeks
Simulator Training	1 Week	Condensed Course	Current Configuration
	2 Weeks	Reduced Academics with Additional Simulation	Current Configuration with Additional Simulation

Table 9: Course Reconfiguration 2x2 Factorial Design

The ADAFCO course directors could implement this experimental design in two different ways. First, they could randomize an entire class into different treatments. Second, they could assign an entire class to one of the new types of treatment. The latter is the more practical option, given the small class size and students' potential fairness concerns. If a single class of only twelve students was divided into four treatments that would only provide three-student sample sizes. There is a high risk that sampling variance might exaggerate the differences between treatments. Additionally, some students might think that the course is unfair given they are not receiving the same treatment as their classmates, but they all take the same tests. While the economist is not necessarily concerned with fairness, the students would have a legitimate complaint if the treatments produced differing levels of operational readiness.

Regardless of how the course directors implement the experiment, they must restructure the classroom academics, so that all of the general knowledge objectives can be addressed in just one week. Reducing the didactic portion requires the shortening or even deletion of certain learning blocks. This task is necessary in order to explore the condensed course and reduced academics with additional simulation alternatives.

Expanding the simulation portion of the course is a much simpler task, assuming simulator availability. At twelve students per course and the current rate of approximately twelve thirty-minute runs per day, each student receives an additional five practice runs in a week. This is necessary for the current configuration with additional simulation and reduced academics with additional simulation alternatives. With these configurations each practice run serves as both a treatment and measurement point. One major benefit of this design is the increased number of operational performance measurement points.

Decreasing marginal returns to scale are expected to appear with the additional simulator training. That is, reaction times are expected to improve by narrowing margins as the students approach a theoretical performance maximum. For instance, it is possible that the students' performance improves throughout the first four simulator runs but then levels off for the final three runs. This may be reason to eventually scale back the simulator practice and either shorten the course or add back some didactic training.

Introducing additional data analysts might also be beneficial, as that one or two seconds of human measurement error may become relevant on the margin. To alleviate this problem, course planners could add one or two data analysts and then use the average reaction time measurement across analysts. This should help reduce the measurement error and provide more precise datasets.

The overarching goal of the course remains improvement of the ADAFCOs' operational performance, and this experiment is designed with that in mind. To achieve this goal, the experiment introduces three new ADAFCO course treatments. The recommendation is that the instructors test the three alternative course designs during the next three iterations of the course. Finally, they can use the new data to further improve the course.

CONCLUSION

During the beginning stages of the War in Iraq, the PATRIOT weapon system committed two incidents of fratricide resulting in three friendly deaths and two lost aircraft. In response to these incidents, the military formed a working group to investigate the underlying causes and recommend future courses of action. One policy that stemmed from this was the creation of a formal ADAFCO training course.

The goal of the course is to train and certify ADAFCOs for their role in major combat operations. The capstone event in the ADAFCO course is an operationally focused DMO exercise called JKCE. During this event, the ADAFCO operates from within an Air Force CRC and directs two simulated PATRIOT battalions to fire against a myriad of enemy threats.

The DMO-based scenario is strongly preferred to a theoretical live version of the exercise for various reasons. The first benefit is that the CRC simulator is identical to the live equipment it emulates. From the training audience's perspective the scenario appears as it would in the real world. Next, a live version of JKCE is simply infeasible due to test range limitations and safety concerns. Finally, the simulated event carries significantly lower monetary and social costs.

The Department of Defense faces considerable budget cuts in years to come. The goal is saving money while preserving war-fighting capability. One possible avenue for savings is shifting from live to simulated training through DMO. However, it is not acceptable to simply reduce costs by substituting simulation for live training while ignoring program effectiveness. Both live and simulated training events must be objectively analyzed on the basis of effectiveness. The goal of military training is to improve the operational capability of soldiers, marines, airmen, and sailors.

JKCE was used as a case study to explore training effectiveness analysis and ADAFCO reaction times for three engagements common to JKCE 12-1 and 12-2 comprise the study's dataset. The didactic portion of the ADAFCO course proved inadequate to achieve the baseline level of operational performance necessary to pass JKCE. In fact, the only element of the course that demonstrated a measurable and statistically significant improvement in ADAFCO response time was the practice run in the simulator.

These findings are crucial to improving future iterations of ADAFCO training. This analysis suggests that the course might be more effective if more time was spent in the simulator practicing operational skills. With this in mind, I presented a recommendation to explore course reconfiguration.

The recommended experimental model provides three alternatives for course reconfiguration. This set up enables the ADAFCO course directors to test different training designs, while still achieving their basic training objectives. After using this experimental design to gather additional data, the course can be readjusted for the sake of optimization.

The ultimate objective of improving operational skills in wartime must be kept in mind when designing training programs, whether they occur in a live or DMO environment. Additionally, in an era of fiscal constraint training programs will face more pressure to justify their existence by demonstrating improvements to operational capabilities. This thesis provides a blueprint for how military decision-makers might approach future evaluations of training program.

ACRONYMS

ADA - Air Defense Artillery

ADAFCO - Air Defense Artillery Fire Control Officers

ARM - Anti-Radiation Missile

AWACS - Airborne Warning and Control System

CEA - Cost-Effectiveness Analysis

CRC - Control and Reporting Center

C2 - Command and Control

DMO - Distributed Mission Operations

DMOC - Distributed Mission Operations Center

JKCE - Joint Kill Chain Event

OIF - Operation Iraqi Freedom

PATRIOT - Phased Array Tracking Intercept of Target

TBM - Theatre Ballistic Missile

REFERENCES

Bell, H. H. & Waag, W. (1998). Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A review. *The International Journal of Aviation Psychology, 8(3), 223-242.*

Cavanagh, J. L. (2011). *Bullet Background Paper: Control and Reporting Center Involvement with Integrated Air and Missile Defense Initiative*. Unpublished Manuscript.

Crane, P., Schiflett, S., & Oser, R. (2000). *RoadRunner 98: Training Effectiveness in a Distributed Mission Training Exercise* (AFRL-HE-AZ-TR-2000-0026, AD A391423). Mesa AZ: Air Force Research Laboratory, Warfighter Training Research Division.

Hawley, J. K. (2007). *Looking Back at 20 Year of MANPRINT on Patriot: Observation and Lessons*. Maryland: Army Research Laboratory.

Jacobs, J. W., Prince, C., Hays, R. T., & Salas, E. (1990). *A Meta-Analysis of the Flight Simulator Training Research*. Florida: Naval Training Systems Center Human Factors Division.

Kirkpatrick, D. L. (1998) *Evaluating Training Programs: The Four Levels*. California: Berrett-Koehler Publishers.

Maule, E. D. (2010). *Memorandum for Headquarters Air Combat Command, Langley AFB: Joint Kill Chain Exercise*. Unpublished Manuscript.

Salas, E., Milham, L. M., & Bowers, C. A. (2003). Training Evaluation in the Military: Misconceptions, Opportunities, and Challenges. *Military Psychology, 15(1), 3-16.*

Schank, J. F., Thie, H. J., Graf, C. M., Beel J., & Sollinger J. (2002). *Finding the Right Balance: Simulator and Live Training* for Navy Units. California: RAND Corporation.

Schreiber, B. T. & Bennett, W. Jr. (2006). *Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary report.* (AFRL-HE-AZ-TR-2006-0015-Vol I). Arizona: Air Force Research Laboratory, Warfighter Readiness Research Division.

Schreiber, B. T., Rowe, L. J., & Bennett, W. Jr. (2006). *Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Participant Utility and Effectiveness Opinions and Ratings.* (AFRL-HE-AZ-TR-2006-0015-Vol IV). Arizona: Air Force Research Laboratory, Warfighter Readiness Research Division.

United States Army. (2012). *PATRIOT*, <http://www.army.mil/factfiles/equipment/airdefense/PATRIOT.html>.

United States Department of Defense. (2012). *Defense Budget Priorities and Choices: January 2012*, 1.

United States Department of Defense. (2012). *Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms*.

United States Department of Defense. (2011). *Selected Acquisitions Report: PATRIOT/MEADS CAP*, <http://www.dod.gov/pubs/foi/logistics_material_readiness/acq_bud_fin/SARs/DEC%202011%20SAR/PATRIOT%20MEADS%20CAP%20-%20SAR%20-%2031%20DEC%202011.pdf>.

APPENDICES

Appendix A: JKCE 12-1 Collected Data

Appendix B: JKCE 12-2 Collected Data

Appendix C: Reaction Times

Appendix D: Equations and Test Statistics

Appendix A: JKCE 12-1 Collected Data

Threat 1			Threat 2			Threat 3		
Detect	Engage	React	Detect	Engage	React	Detect	Engage	React
18:25:35	18:26:50	0:01:15	18:26:13	18:28:10	0:01:57	18:36:49	18:41:44	0:04:55
20:10:35	20:15:49	0:05:14	20:11:13	20:13:10	0:01:57	20:21:51	20:25:30	0:03:39
20:55:35	20:59:06	0:03:31	20:56:13	20:58:16	0:02:03			
21:40:36	21:44:05	0:03:29	21:41:15	21:42:36	0:01:21			
			22:11:12	22:11:54	0:00:42			
15:10:34	15:14:11	0:03:37	15:11:12	15:12:06	0:00:54			
15:55:34	15:58:55	0:03:21	15:56:14	15:57:43	0:01:29			
16:51:52	16:56:59	0:05:07	16:52:05	16:54:39	0:02:34			
17:26:20	17:29:55	0:03:35	17:26:58	17:28:40	0:01:42			
18:10:35	18:14:09	0:03:34	18:11:10	18:13:12	0:02:02			
19:40:34	19:44:26	0:03:52	19:41:12	19:42:12	0:01:00			
20:25:49	20:29:36	0:03:47	20:26:28	20:28:28	0:02:00	20:37:06	20:38:42	0:01:36

* This table contains all recorded data for JKCE 12-1 presented in five-digit times. The differences between detection times and engagement times (reaction times) are also included.

Appendix B: JKCE 12-2 Collected Data

	Threat 1			Threat 2			Threat 3		
Run Type	**Detect**	**Engage**	**React**	**Detect**	**Engage**	**React**	**Detect**	**Engage**	**React**
Practice	14:12:33	14:17:27	0:04:54	14:12:52	14:16:42	0:03:50	14:19:36	**HIT**	**0:05:40**
Practice	15:27:47	15:32:57	0:05:10	15:27:58	15:31:51	0:03:53	15:35:41	**HIT**	**0:05:40**
Practice	16:42:41	**HIT**	**0:05:40**	16:43:19	**HIT**	**0:04:49**	16:49:43	**HIT**	**0:05:40**
Practice	18:43:14	18:45:37	0:02:23	18:43:23	18:47:10	0:03:47	18:49:40	**HIT**	**0:05:40**
Practice	19:57:57	20:02:24	0:04:27	19:58:30	20:00:40	0:02:10	20:04:51	**HIT**	**0:05:40**
Practice	21:12:36	21:16:18	0:03:42	21:12:47	21:16:33	0:03:46			
Practice	14:12:34	**HIT**	**0:05:40**	14:12:48	14:15:58	0:03:10	14:19:31	14:24:04	0:04:33
Practice	15:27:36	**HIT**	**0:05:40**	15:27:49	**HIT**	**0:04:49**	15:34:34	15:38:54	0:04:20
Practice	16:42:34	16:46:05	0:03:31	17:04:46	**HIT**	**0:04:49**	16:49:33	**HIT**	**0:05:40**
Practice	18:42:36	**HIT**	**0:05:40**	18:43:13	18:45:58	0:02:45	18:49:34	**HIT**	**0:05:40**
Practice	19:57:35	20:01:36	0:04:01	19:57:48	**HIT**	**0:04:49**	20:04:34	20:10:02	0:05:28
Practice	21:12:46	21:15:30	0:02:44	21:13:24	21:16:52	0:03:28	21:19:43	**HIT**	**0:05:40**
Record	14:12:34	14:13:29	0:00:55	14:13:13	14:14:59	0:01:46	14:19:34	14:21:11	0:01:37
Record	14:57:35	14:59:00	0:01:25	14:58:11	14:59:00	0:00:49	15:04:31	15:07:05	0:02:34
Record	15:42:35	15:44:32	0:01:57	15:43:13	15:44:59	0:01:46	15:49:32	15:51:50	0:02:18
Record	16:27:34	16:28:24	0:00:50	16:28:12	16:30:13	0:02:01	16:34:33	16:36:16	0:01:43
Record	18:42:34	18:44:30	0:01:56	18:43:12	18:44:30	0:01:18	18:49:32	18:50:19	0:00:47
Record	19:27:35	19:28:15	0:00:40	19:28:14	19:28:55	0:00:41	19:34:31	19:35:55	0:01:24
Record	20:57:33	21:00:02	0:02:29	20:58:12	21:00:36	0:02:24	21:04:33	21:06:58	0:02:25
Record	21:42:35	21:43:10	0:00:35	21:43:14	21:43:49	0:00:35	21:49:32	21:50:24	0:00:52
Record	14:12:36	14:13:20	0:00:44	14:13:12	14:14:17	0:01:05	14:19:32	14:21:23	0:01:51
Record	14:57:35	14:58:10	0:00:35	14:58:16	15:00:18	0:02:02	15:04:33	15:05:17	0:00:44
Record	15:42:48	15:44:55	0:02:07	15:43:25	15:45:23	0:01:58	15:49:45	15:52:32	0:02:47
Record	16:27:37	16:29:06	0:01:29	16:28:13	16:30:02	0:01:49			

* This table contains all recorded data for JKCE 12-2 presented in five-digit times. The differences between detection times and engagement times (reaction times) are also included.

Appendix C: Reaction Times

	Record Runs		
	Threat 1	Threat 2	Threat 3
	75	117	295
	314	117	219
	211	123	
	209	81	
		42	
	217	54	
	201	89	
	307	154	
	215	202	
	214	122	
	232	60	
	227	120	96
	55	106	97
	85	49	154
	117	106	138
	50	121	103
		78	47
	40	41	84
	249	144	145
	35	35	52
	44	65	111
	35	122	44
	127	118	167
	89	109	
Average	152	99	125
Std Dev	93	41	70

	Practice Runs		
	Threat 1	Threat 2	Threat 3
	294	230	340
	310	233	340
	340	289	340
	143	227	340
	267	130	340
	222	226	
	340	190	273
	340	289	260
	211	289	340
	340	165	340
	241	289	328
	164	208	340
Average	268	230	326
Std Dev	71	52	30

* This table contains all recorded reaction times, JKCE 12-1 and 12-2, converted from six digit times to seconds. The times are separated by run type, practice versus record, as well as threat. Sample means and sample standard deviations are also included.

Appendix D: Equations and Test Statistics

Hypothesis Tests

$$H_0 : \mu_{Practice} - \mu_{Record} \leq 0$$
$$H_1 : \mu_{Practice} - \mu_{Record} > 0$$

Degrees of Freedom

$$DOF = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{\left(s_1^2 / n_1 \right)^2}{(n_1 - 1)} + \frac{\left(s_2^2 / n_2 \right)^2}{(n_2 - 1)}}$$

Test Statistics

$$TestStat = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Results

	DOF	Stat	P-Value
Threat 1	28.26	4.07	1.75E-04
Threat 2	18.07	7.62	2.42E-07
Threat 3	18.48	9.67	7.43E-09