IMPROVING THE ENTERPRISE REQUIREMENTS AND ACQUISITION MODEL'S DEVELOPMENTAL TEST AND EVALUATION PROCESS FIDELITY

THESIS

Jason W. Sutherlin, Captain, USAF

AFIT-ENV-14-M-60

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-14-M-60

IMPROVING THE ENTERPRISE REQUIREMENTS AND ACQUISISITON MODEL'S DEVELOPMENTAL TEST AND EVALUATION PROCESS FIDELITY

THESIS

Presented to the Faculty

Department of Systems Engineering and Management

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 Systems Engineering

Jason W. Sutherlin, BS

Captain, USAF

March 2014

AFIT-ENV-14-M-60

IMPROVING THE ENTERPRISE REQUIREMENTS AND ACQUISISITON MODEL'S DEVELOPMENTAL TEST AND EVALUATION PROCESS FIDELITY

Jason W. Sutherlin, B.S.

Captain, USAF

Approved:

//signed//	14 March 2014
Dr. John M. Colombi, Ph.D. (Chairman)	Date
//signed//	14 March 2014
Lt. Col. Joseph R. Wirthlin, Ph.D. (Member)	Date
//signed//	14 March 2014
Mr. Jerry J. Vandewiele (Member)	Date
//signed//	14 March 2014
Lt. Col. Kyle F. Oyama, Ph.D. (Member)	Date

AFIT-ENV-14-M-60

Abstract

Shrinking budgets and dynamic military conflicts have driven Department of Defense (DoD) leadership to reform how the military acquires weapon systems with the goal of decreasing program schedules and costs, while maximizing performance. Yet with more than fifty years of acquisition reform, the DoD has been unable to adequately control program schedule objectives. Previous research attempted to support acquisition reform through computer modeling and simulation. One model, called the Enterprise Requirements and Acquisition Model (ERAM), captures a program's progression through the Defense Acquisition Management System (DAMS) to gain insight into significant delays that impact a program's schedule and probability of completion. A past unexpected result included the insignificant impact that Developmental Test and Evaluation (DT&E) activities had to a program's overall schedule. This ERAM research improves the fidelity of the Air Force (AF) DT&E activities through data collection, subject matter expert (SME) feedback, computer modeling and simulation, and Monte Carlo analysis. Interventions included modifying the probability of passing the Test Readiness Review, System Verification Review, decreasing the maximum delay to a program's first test mission, improvements in Responsible Test Organization resource availability, test item quality, and test item quantity. Several interventions significantly reduced major program schedule by 15% or 21 months. The research demonstrates a methodology for quantitatively supporting acquisition reform interventions by characterizing key DT&E activities and delay factors.

Acknowledgments

When I first came to the Air Force Institute of Technology I heard someone in the hallway say, "graduate school is a team effort." I did not appreciate the wisdom in those words until the last few months of my research. Everything I accomplished was because I had a team who supported me every step of the way. Dr. John Colombi provided limitless enthusiasm, advice, and direction to ensure I was properly focused. Dr. John Miller established my modeling and simulation capabilities and instilled a joy for simulation making the long nights executing simulations more enjoyable. Even after leaving AFIT, Lt. Col. Rob Wirthlin found time to answer my emails, answer my phone calls, and exhibit a genuine interest in my research. Mr. Jerry Vandewiele connected me with almost all of my SMEs which were critical to completing this research. No journal article or policy document could have replaced the wisdom garnered from my discussions with the SMEs in DAU, SAF/AQXC, AFLCMC/AQTB, OUSD/AT&L, and the test community. I will miss our conversations and enthusiastic collaboration. To all my committee members, SMEs, faculty, and friends, thank you. But most of all, thank you to my family whose smiles, encouragement, and love gave me the strength to wake up early every morning and make it to school on time.

Jason W. Sutherlin

Table of Contents

List of Figures

Page

List of Tables

Page

I. Introduction

The political and economic environment the United States (US) found itself in during the beginning of the twenty first century forced the Department of Defense (DoD) to research new methods to improve the processes by which the US military acquires its weapons systems. The ability to observe system level impacts of acquisition reforms could assist leadership in making reforms which not only have local process benefits, but positively impact the entire system. This thesis focuses on refining previous research efforts in an attempt to provide Air Force (AF) senior leadership a different capability to assist in addressing acquisition reform.

General Issue

As technology advanced during the 21st century, it was integrated into military weapon systems increasing the time required to produce them. Efforts to evolve an effective acquisition process culminated into the 2008 DoD Instruction (DoDI) 5000.02 which was the official instruction on conducting DoD acquisitions and is summarized in the Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management System Chart in Figure 1. Pending modification to DoDI 5000.02 have been proposed at the time of writing this thesis (USD, 2013).

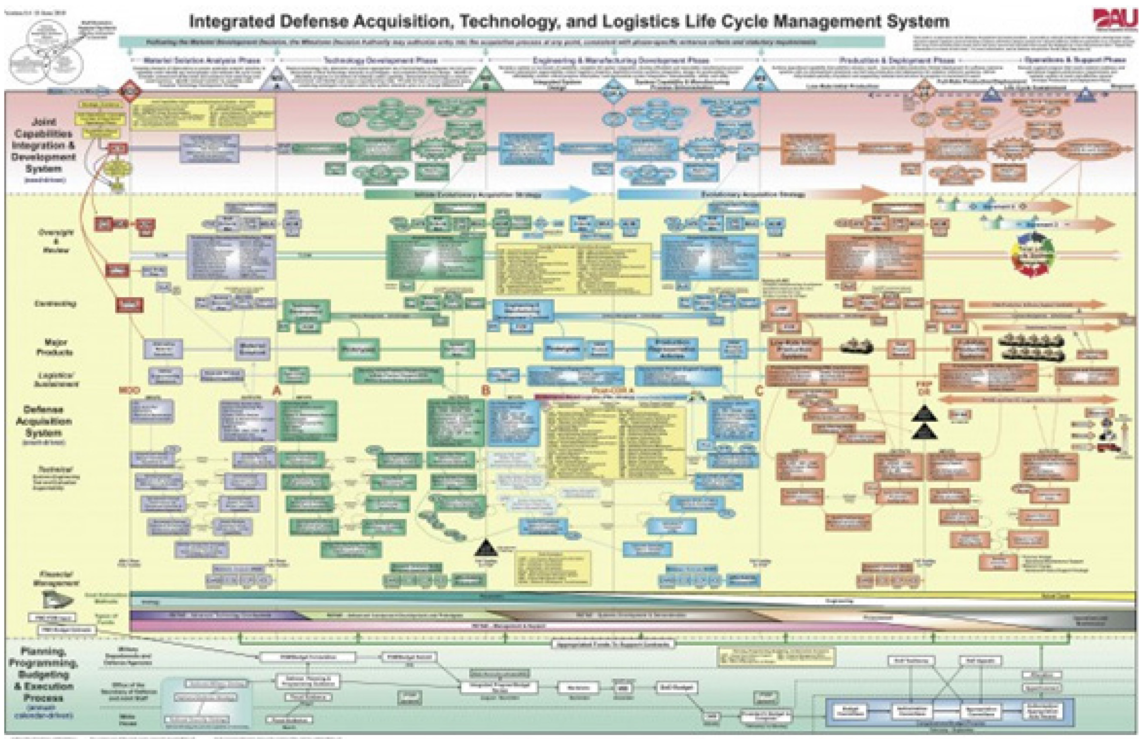

Figure 1: DoD Acquisitions (DAU, 2013)

As complicated as the weapon systems being obtained, the process to acquire these systems can be viewed as equally complex. DoD leadership formally acknowledged problems with how the military produces weapon systems in the Hoover Study conducted in 1949 and since then over 128 acquisition studies have been conducted (Kadish, 2005). Even with over 60 years of evolution, the acquisition process has not consistently performed at an acceptable level (Eide, 2012). From a recent Government Accountability Office (GAO) report, programs in the 2012 acquisition portfolio were on average more than two years behind schedule (2013). The consistent findings in many similar reports combined with the economic crisis the US found itself in during the beginning of the twenty-first century has captured senior DoD leadership's interest in acquisition reform. Both President Obama and Secretary of Defense (SECDEF) Hagel have addressed the issue. President Obama expressed his concern, that the US could no longer afford acquisition's poor performance and must become more efficient in delivering weapon systems to the warfighter (Obama, 2009), by signing into law the Weapons System Acquisition Reform Act in 2009. The President's viewpoint was supported by SECDEF Hagel's who expressed his concern during a speech at the National Defense University in 2013.

> "We need to continually move forward with designing an acquisition system that responds more efficiently, effectively and quickly to the needs of troops and commanders in the field. One that rewards cost-effectiveness and efficiency, so that our programs do not continue to take longer, cost more, and delivers less than initially planned and promised."

Problem Statement

Joseph R. Wirthlin investigated the Defense Acquisition Management System (DAMS) for complex relationships which could be causing emergent behaviors within the system. Wirthlin postulated that acquisition reform may have been implemented without the ability to accurately predict system impacts due to these complex relationships. Wirthlin (2009) recognized the opportunity for research in this area and became the focus of his dissertation. His research created the Enterprise Requirements Acquisition Model (ERAM), an extensive simulation model of the DAMS. The purpose of ERAM was to investigate the DAMS process relationships in order to characterize how the system worked, why it behaved the way it did, and if there were ways to improve it. ERAM provided the capability to simulate policy reforms in the simulation model and observe the system impacts. Due to the complexity of the DAMS, many low level processes were purposefully abstracted. Wirthlin discovered several unexpected results in his dissertation and suggested them as areas for future research including: the Test Readiness Review (TRR), Developmental Test and Evaluation (DT&E), and System Verification Review (SVR) activities (Wirthlin, 2009: 189-190). These three areas will be the focus of this thesis.

Investigative Questions

The following questions were identified for this thesis:

1. How can the fidelity of ERAM 1.0 DT&E activities be improved?
2. What insight can be gained from the improved fidelity ERAM with regard to supporting previous research conclusions regarding the TRR, DT&E, and SVR activities?
3. What DT&E process interventions can significantly reduce program schedule?

Impacts

The ability to simulate acquisition policy reform in a simulation model and observe system level impacts, before implementing the policy in reality, could be useful to DoD leadership. ERAM is not viewed by the author as a tool for DoD leadership to directly use to make reforms. Rather, ERAM is viewed as a demonstration of how computer modeling and simulation could be utilized to support acquisition reform. With adequate resources, an advanced model, similar to ERAM, could be developed to as a tool to support quantitative-based acquisition reform through computer modeling and simulation. Lastly, ERAM could also be an educational tool for the Defense Acquisition University to assist in teaching future acquisition professionals about the complex relationships between process, technology, people, and the resulting emergent behaviors.

II. Literature Review

Chapter Overview

Chapter II is divided into several sections: Modeling and Simulation Overview, ERAM, ERAM Evolution (2010-2013), DT&E's Role in Program Schedule Delays, and Literature Synthesis. The first section provides a brief introduction of modeling and simulation focusing on its advantages, disadvantages, and limitations. The chapter will continue with an in-depth review of the previous ERAM research. The Other Acquisition Modeling Efforts section will be review other similar research projects. Chapter II concludes by synthesizing key concepts from the extant literature.

Modeling and Simulation Overview

From automobile factory production lines to shipping distribution centers, these collections of processes can be viewed as systems. Often there is desire to improve some aspect of system performance such as decreasing production line down time or cycle time. However, for complex systems, it may be difficult to estimate the impact changing local variables will have on the entire system. The most direct way to observe system impacts would be to implement the change in the actual system. This method is generally not used because of the feasibility and potential financial loss should such a change result in unintended negative consequences. Another method is to utilize modeling and simulation. "A simulation is an abstraction of an operation in a real-world process or system over time" (Banks, 2005: 3). Coupled with the computational capabilities of computers, modeling and simulation enables system analysis difficult, if not impossible, to attain from any other method. For example, in a simulation model, performance results are directly traceable to changes the experimenter executed in the system. If acquisition reform was instead executed in the DAMS in reality, it would be difficult to correlate system improvements to the implemented reform because of the multitude of policy reforms consistently enacted on a monthly basis. Figure 2-3 each presents a three year timeline of DAMS reform implementation.

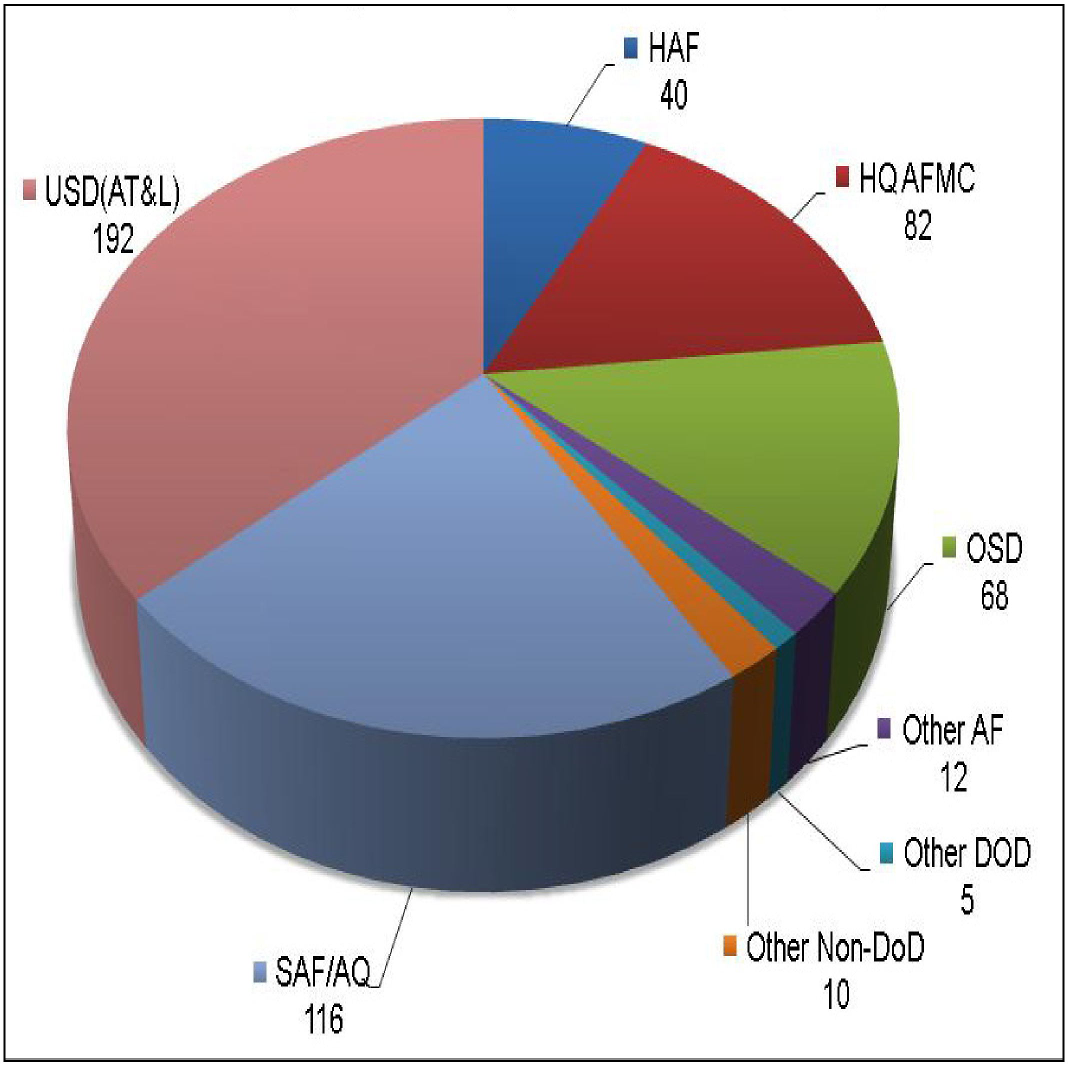

Figure 2: New DAMS Policy Reforms by Organization 2008-2011 (Milam, 2012)

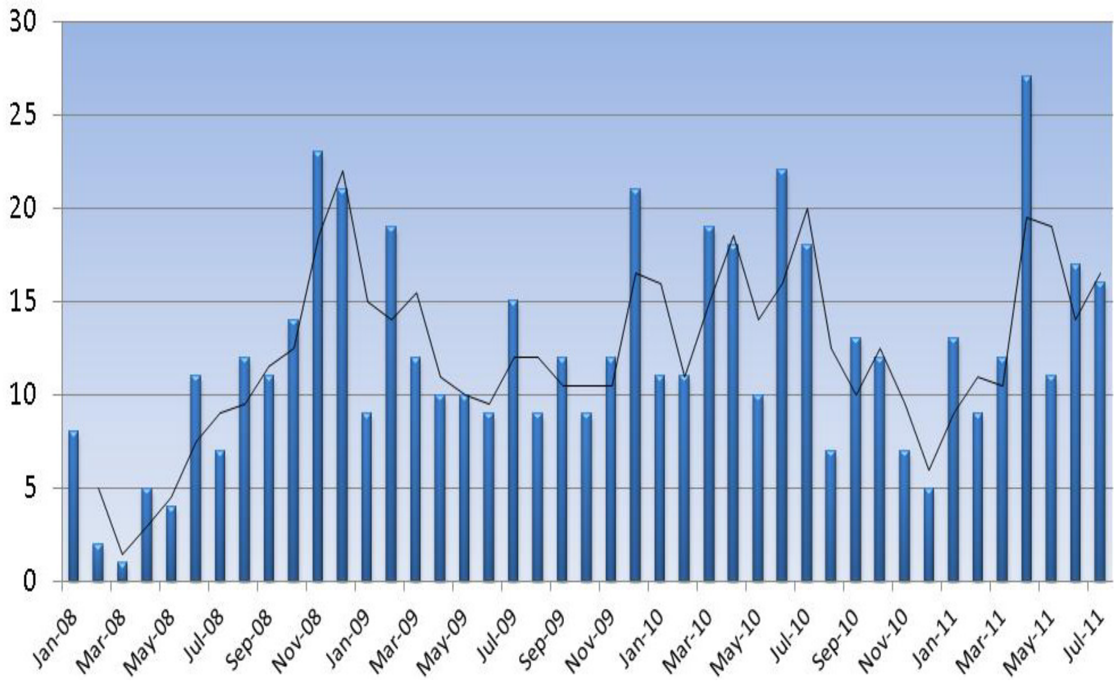

Figure 3: Number of DAMS Policy Reforms from 2008-2011 (Milam, 2012)

The constant process change may result in a state of causal ambiguity where policy reforms and system improvements are difficult to correlate with one another. In addition, long DAMS program cycles would require years of observation before adequate sample sizes are collected. With computer modeling and simulation, time is less of a limiting factor as it can be manipulated. Data representing hundreds of years can be collected in a few hours. Collecting the same information by observing the reality is impossible.

As powerful as modeling and simulation can be, it should only be used in certain situations. Four situations, directly relatable to this research project, are discussed below (Banks, 2005: 4):

1. "The goal of the study or experimentation is the interactions of a complex system or of a subsystem within a complex system."

- ERAM may be viewed as an investigation to characterize a complex system (DAMS) and understand what impacts the subsystems (DT&E activities) had on the overall system (Wirthlin, 2009).

2. "The knowledge gained from a simulation model could be used to suggest improvements in the real system."
    - ERAM investigated interventions to identify process improvements which could result in improved system performance in reality.
3. "Changing simulation inputs and observing the resulting outputs could produce valuable insight into which variables are the most important and how those variables interact."
    - One of the main goals of ERAM and this research project was to observe what changes in the system would result in system schedule performance benefits (Wirthlin, 2009).
4. "Many modern systems are so complex (automobile factory, wafer fabrication plant) that the internal interactions cannot be understood without the use of computer simulation."
    - Over sixty years of acquisition reform has failed to create a system which adequately controls program schedule. This result may be, in part, due to the system's complexity hinting at the requirement to utilize computer simulation to better characterize the system and its emergent behaviors.

ERAM

ERAM is a discrete event simulation model of acquisition category (ACAT) I, II, and III which attempts to capture the "idea" of a program in pre-Milestone A all the way to Milestone (MS) C. Included are the functional areas of the Joint Capabilities Integration Development System (JCIDS), Acquisitions, the Planning Programming Budgeting & Execution (PPBE), and Contractors. Created through investigation of official policy and refined by SMEs, a single program progresses through the model. Through Monte-Carlo analysis and the stochastic nature of the model, thousands of potential outcomes are characterized to create a distribution of program schedule and probability of successfully navigating DAMS up to MS C.

Validation of Arena Model

Model validation was conducted by comparing ERAM results to historical data. Data were collected primarily from the System Metrics and Reporting Tool (SMART). Student t-Tests compared ERAM results to the historical data. Specifically, the program time from MSB to MSC, for different ACAT groups (all ACATS, ACAT I, ACAT II, and ACAT III), was analyzed. Hypothesis testing indicated that ERAM was a valid representation of the DAMS for all ACAT categories at a 95% confidence level (Wirthlin, 2005: 138-146). The validation results of ERAM are important because they will enable validity of this research project discussed in Chapter III.

ERAM Evolution (2010-2013)

Since 2009, other researchers have realized the potential benefits of utilizing ERAM to investigate DAMS. These include work by Leach and Searle (2010), Montomery (2011), and Baldus and others (2013). Below is a summary of their research efforts.

Table 1: Overview of ERAM Research Projects

Author	Year	Version Number	Simulation Program	Changes
Wirthlin	2009	ERAM 1.0	Arena	Baseline translation from Arena to ExtendSim
Leach and Searle	2010	ERAM 1.1	ExtendSim	Updates by the Aerospace Design Team and served as new baseline model
		ERAM 1.2	ExtendSim	Implemented new DoD 5000.02 policies
		ERAM 2.0	ExtendSim	Incorporated the global variables that modify acquisition capabilities
		ERAM 2.1	ExtendSim	Incorporated JCIDS review process
Montgomery	2011	ERAM 2.2	ExtendSim	Added more capabilities for ACAT II/III and Rapid Acquisition Process
Baldus and others	2013	2.4	ExtendSim	Integrated space launch process delays

ERAM Research Vectors

Two research vectors were identified when reviewing the research in Table 1. The original purpose of the ERAM was to improve understanding of how the DAMS operated in order to conduct system level improvements. This research vector can be categorized as improving system schedule performance. Since 2009, the focus shifted from system schedule performance to prediction of a single program's schedule. Figure 4 provides a summary of previous ERAM research and their respective vectors.

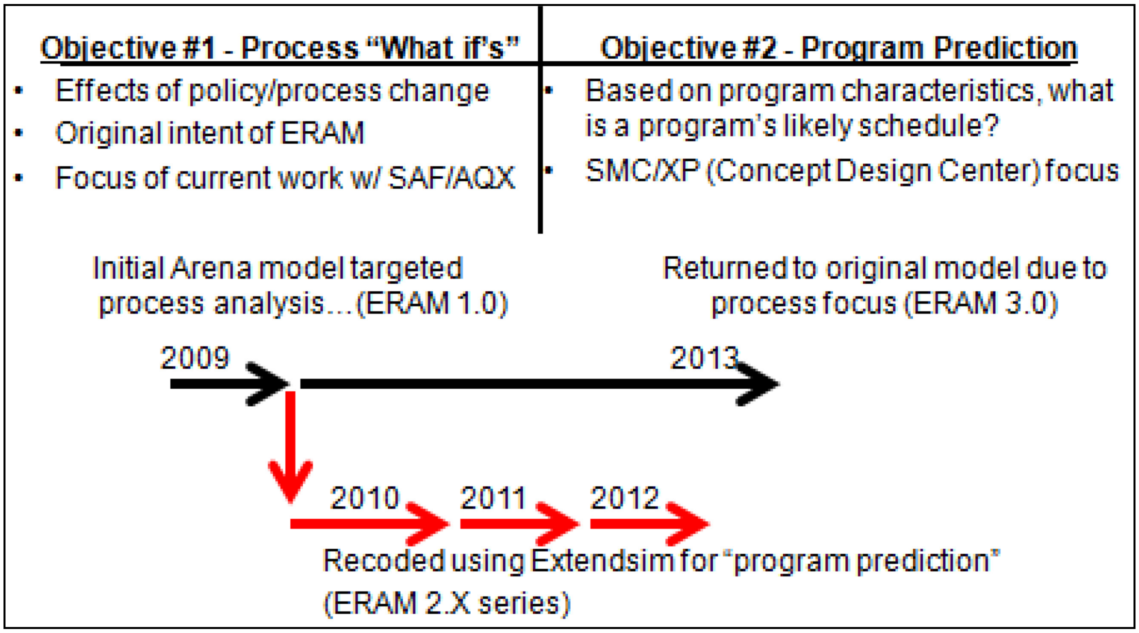

Figure 4: ERAM Research Vectors

None of the projects listed in Table 1 addressed the DT&E areas of concern identified by Wirthlin which presented the opportunity to proceed with either research vector. The author was advised to "Go where the research interest is" by the research committee and the author queried the acquisition community for input. Discussions with SAF/AQXC, OUSD/ AT&L, and DAU indicated that the system schedule improvement vector would be more relevant and directed this research project to investigation of the DT&E activities in the original ERAM (ERAM 1.0) for the purpose of acquisition system reform.

Other Acquisition Modeling Efforts

Acquisition Document Development Model (ADDM)

Senior DoD leadership identified one problem in government acquisitions was the lack of document control and listed seven document control issues (ASC/RCC, 2010: 2):

1. Milestone dates delayed due to non-timely document preparation.
2. Creating documents consumes a large amount of time and resources.
3. The rationale in tailoring program documents is not captured in a formal way.
4. There is no strong linkage between program documentation.
5. The quality and content is inconsistent across a program's documents.
6. There is no capability to support cross-cutting changes to acquisition documents with minimal effort.
7. A lack of insight into Milestone readiness

These issues were found to be especially prevalent when a program changed Program Manager's and were approaching a MS review. The AF created an interactive model, called the Acquisition Document Development Model, capable of tracing program documents and processes to address this issue. The four ADDM objectives were (ASC/RCC, 2010:4):

1. Provide a roadmap that identifies what documents are required and when based on a program's ACAT level and MS.
2. Provide the ability for a PM to modify the program's document roadmap to meet specific program requirements.

3. Provide a set of validated document templates which are linked to current guidance, references, and Program Executive Officer (PEO) specific instructions
4. Provide a quick, visual indicator on program review and Milestone document status.

The intent of ADDM is to provide PM's with document situational awareness through several key model features. One such feature was that ADDM created a unique, customized document roadmap based on the acquisition program's ACAT level and the next MS. Another feature was ADDM automatically created a set of standardized and validated document templates according to the program's roadmap. Each of these documents was linked to the current policy and instruction. As a program progressed through acquisitions, ADDM captured decisions and updated the program's roadmap and documents as required. In addition, ADDM was continuously updated to ensure the most relevant information was accessible to PMs.

ADDM, as shown in Figure 5, provides PMs a tool, to assist in moving the program from the current situation to the next MS review, listed the documents required at the next MS review, provided standardized templates for documents, updated document status, and provided current document guidance and instruction. Future plans for ADDM include the addition of DoD space and business systems roadmaps.

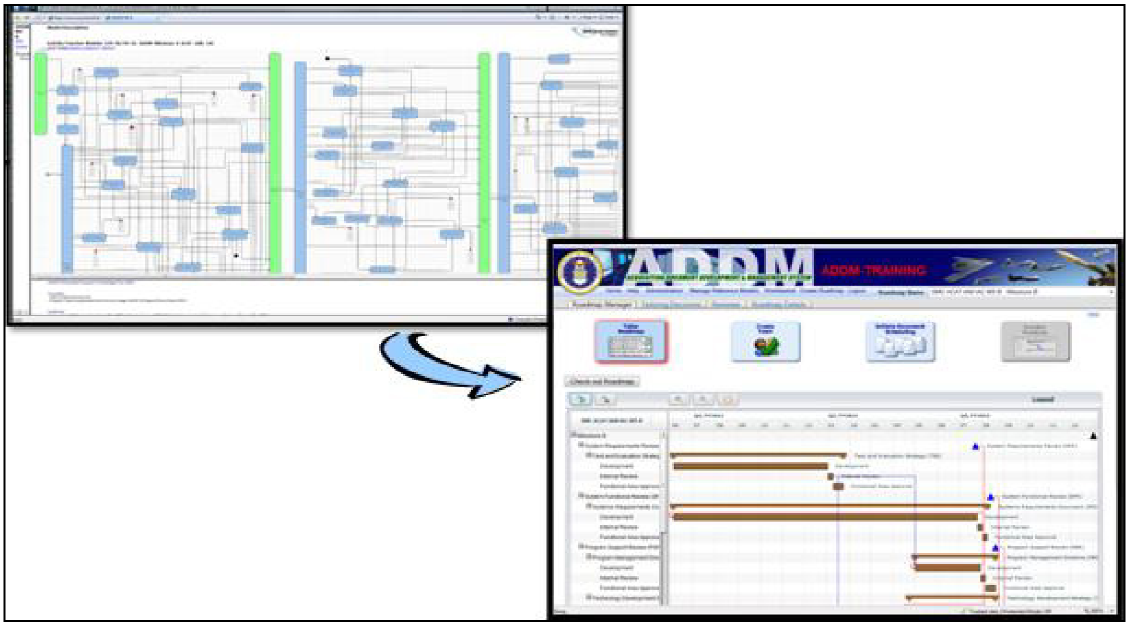

Figure 5: ADDM (ASC/RCC, 2010: 8)

Acquisition Process Model (APM)

The DAMS can be viewed as a complex system of processes and the ability to guide a program through the processes is critical to success. In 2009, the AF Acquisition Chief Process Office initiated a project to create an official, authoritative process model of the DAMS. The Acquisition Process Model was the culmination of their efforts. APM provides an interactive process model for ACAT I programs from the point of view of the Program Executive Officer (PEO) covering the DoD 5000 instruction, the JCIDS, and the PPBE activities. APM's goal is to provide a standardized, authoritative acquisition process model with six objectives (ACPO, 2011):

1. Establish standard definition and activities associated with AF acquisition.
2. Provide process decomposition from Defense Acquisition Executive/Service Acquisition Executive through PEO level actions.
3. Provide an integration context for other external/related process models.

4. Provide the process input to Acquisition Enterprise Architecture and other Enterprise Architectures.
5. Provide a standard reference model for all stakeholders.
6. Provide a common context for process improvement initiatives.

APM utilizes an interactive model to capture document and process relationships placing additional emphasis on the requirements generation (JCIDS), acquisitions (DoDI 5000.02 series), and funding (PPBE) activities. Key information (including process definition, owner, reference document, performer, and links to current documentation) is available for each process as shown in the APM preview in Figure 6.

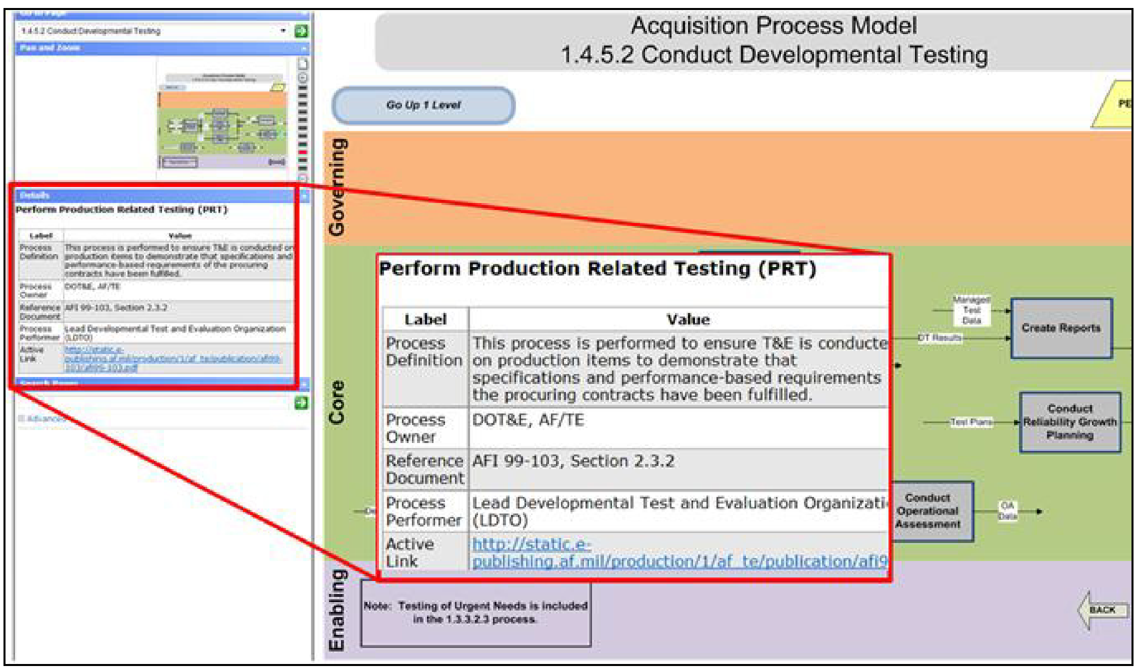

Figure 6: APM Developmental Test (ACPO, 2011)

APM is updated on a routine basis to incorporate the current policies. Future plans include improving model fidelity to the PM level processes.

Requirements and Acquisition Management Plan (RAMP)

Space system acquisition is different from conventional acquisitions in many aspects as discussed in Air Force Instruction 99-103. However, space acquisition also suffers from similar schedule problems observed in conventional acquisitions. The Air Force Space Command's Directorate of Requirements (AFSPC/A5) investigated the space acquisitions and identified that quality and speed of requirements generation are critical areas of concern (Gilchrist, 2011:24). AFSPC/A5 chose to use modeling as a method of investigating these problems and created the Requirements and Acquisition Management Plan.

The goal of RAMP is improve the requirements generation and acquisition processes through a "standard, consistent, and transparent" requirements and acquisition management process (Gilchrist, 2011:3). RAMP is a work breakdown structure tailored for acquisitions which provide users the ability to schedule activities, assign responsibilities, and access activity relationships. Figure 7 shows an example the RAMP model.

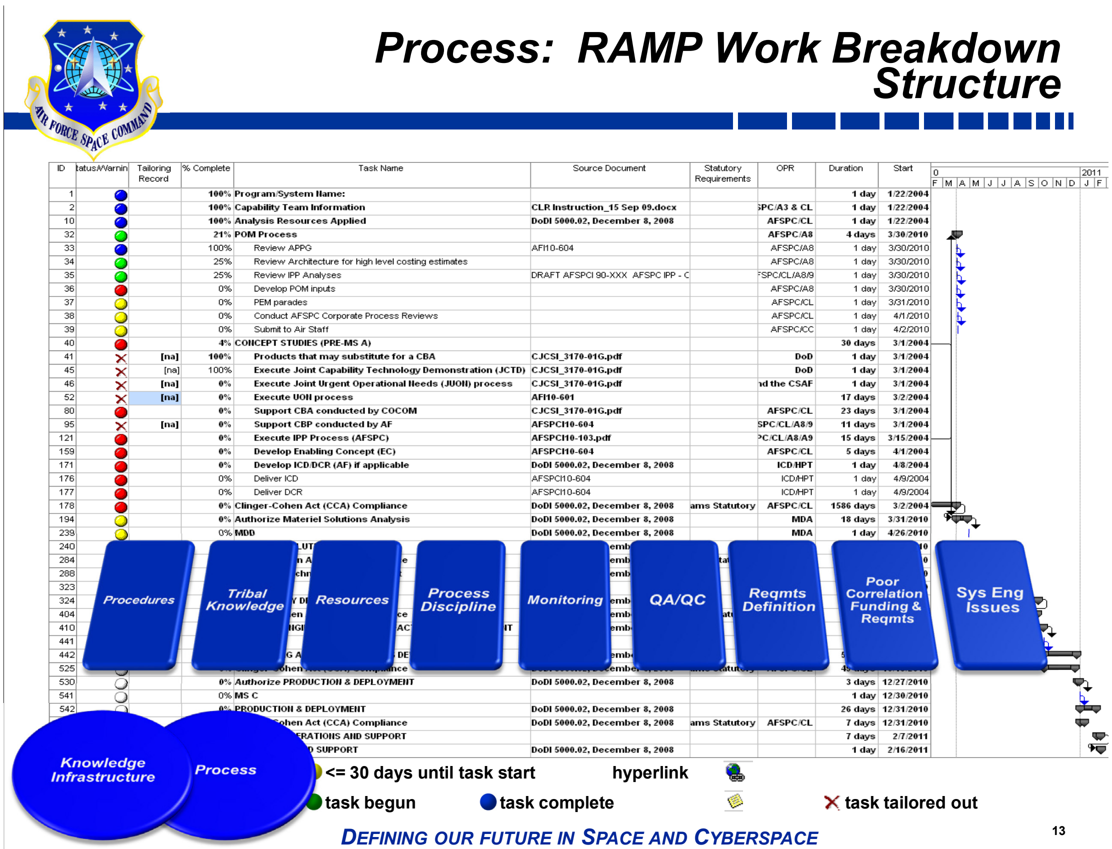

Figure 7: RAMP (Gilchrist, 2011:13)

Literature Synthesis

Chapter II provided several insights. The first insight was that modeling and simulation are capable of providing insight into understanding complex process (DAMS) when utilized appropriately. However, there are disadvantages inherent in all models and the corresponding results must be analyzed by modeling and simulation experts who understand these limitations. Another insight was that there are different methods (system dynamics, agent based modeling, interactive charts, and other options) to model a system and each method can provide a different viewpoint. All the research projects discussed in Chapter II were modeling the same system but from different viewpoints. The different methodologies were driven by the type of problem each model was addressing. The

ADDM approached acquisition reform from a documents control perspective in order to bring PMs situational awareness and control over the multitude of documents required during procurement. APM provided PEOs a standardized, validated, and traceable model of the DAMS processes. Requirements issues were addressed by RAMP through an integrated, work schedule structure in order to decrease the requirements generation schedule and increase quality. The DAMS suffers from diverse problems of which only a very small sample were discussed here. However, as diverse as the problems encountered were, a commonality among these research projects is that they used modeling as a method for investigating a complex system.

III. Methodology

Methodology

A simulation study methodology was utilized for this research. See the Appendix for a figure of the methodology. The first step was reviewed in Chapters I and II. Chapter III will address data collection and the iterative process of building a simulation model and verifying the model. The remaining steps are addressed in Chapters IV and V.

Data Collection

There are two fundamental modeling constructs utilized in ERAM: processes and decisions. Processes are tasks which take a stochastic amount of time to accomplish and are modeled using triangular distributions. Decisions represent reviews where an entity may progress through different model paths. Figure 8 contains graphical representations of the two constructs encountered in ERAM. This area of ERAM was constructed with three process blocks (represented by the rectangles) and a decision block (represented by the diamond). The lines connecting the blocks represent the possible paths from one process or decision to another and identify how an entity could progress through the model. The model logic (in Figure 8) is from read left to right. A program (the entity) enters the “Developmental Test and Evaluation” process block and a random number from a triangular distribution will be randomly selected representing the time required to perform the process. Next, the program progresses to the “Trades Needed” decision block which will direct the entity to either the “Dev test rework and delay” or “Early Operational Assessment” block based on the random value compared to a percent true criteria. A program not requiring any rework will take the path around the “Dev test rework and delay" block (not incurring a delay due to accomplishing the blocks process) and proceed to the "Early Operational Assessment" block. In this block, the program will incur another process delay as specified by random number chosen from the block's distribution.

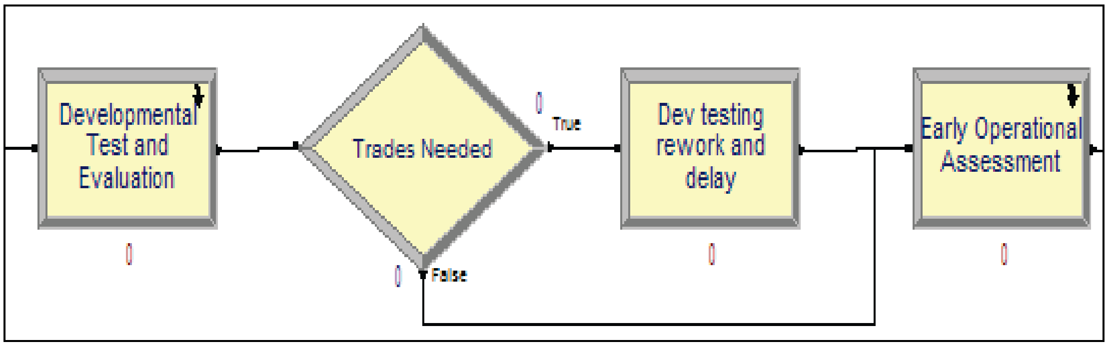

Figure 8: ERAM Systems Engineering Activities (Wirthlin, 2009: 318)

Triangular Distributions

All ERAM processes, relevant to this research, are populated with triangular distributions as shown in Figure 9.

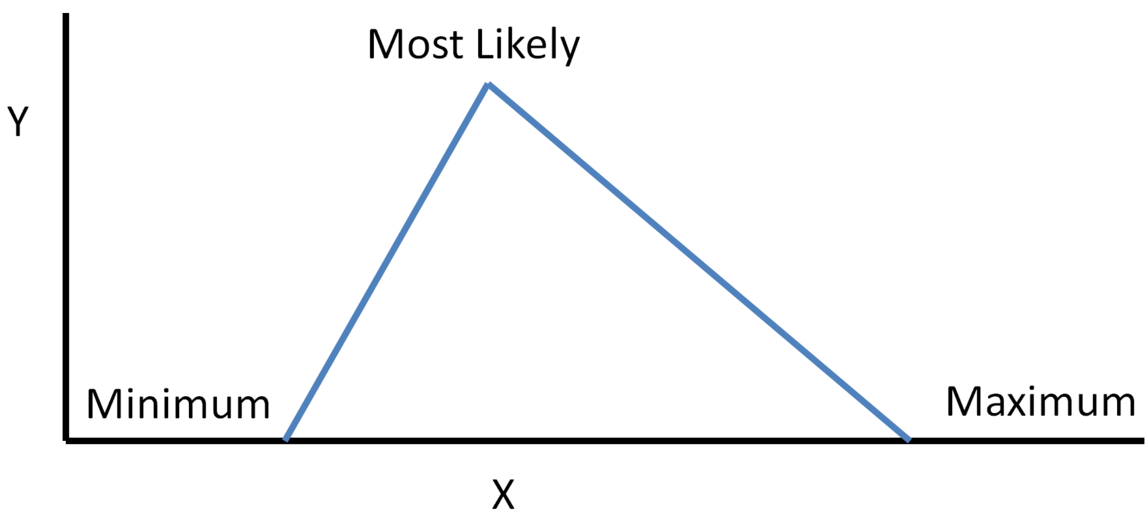

Figure 9: Task Block Triangular Distribution

Defining a triangular distribution requires definition of a minimum, mean, and maximum value. This type of distribution was utilized because SMEs were easily able to estimate the minimum, mean, and maximum time required to complete a process based on their personal experience. All the processes and activities relevant to this research project were constructed from SME opinion. If real data could be collected, it could increase the validity of ERAM.

Historical Data

Historical schedule data for ACAT I, II, and III programs were desired for validation purposes. The Defense Acquisition Management Information Retrieval (DAMIR) and SMART databases were identified by SMEs as possible sources of the desired data. The author's request for DAMIR access was denied but SMART access was granted. Unfortunately, the author experienced technology problems with the SMART application and no data was collected. Data from Wirthlin's research on program schedule times from MS B-C were available and utilized. Test mission data and factors which resulted in cancelations, aborts, test mission effectiveness, and other metrics were received from a Major Range and Test Facility Base (MRTFB). The test mission data were imported into Microsoft Excel in order to construct model probability inputs of categorical factors identified during the SME discussions. The test mission data are labeled "For Official Use Only" (FOUO) and are not included in this research paper. If the reader would like a copy of the data, please contact a member of the research team whose contact information is provided in the Appendix.

SME Discussions

Purposeful sampling was utilized to select possible SMEs who were familiar with the DT&E activities. These individuals were contacted through phone calls and email. Semi-formal discussions were conducted between the author and SMEs who were available to participate in the research. A list of general discussion topics can be found in the Appendix. If the SME was in the local area, the author conducted the discussion in person at the SMEs office. If the SME was not local, the discussion was conducted over the phone with supporting documents provided through email. Although a predetermined set of topics were utilized to initiate and direct the discussions, conversations were allowed to deviate. SME answers were transcribed on paper by the researcher. At the end of the discussion, SMEs was asked to provide contact information for any additional references that may be able to provide additional information or have interest in this research. This proved to be a very useful technique and how a majority of the SMEs were identified.

The discussions with SMEs were the most enlightening source of knowledge for this research. Issues only hinted at in literature were discussed frankly without the need for political correctness allowing a different perspective of DT&E. The next few pages will present quotes from SMEs which were particularly enlightening and relevant to this research.

Almost all of the SMEs identified unrealistic schedule expectations as the most common and significant source of DT&E program delay. Poor quality estimates were mentioned to originate from both the SPO and DT&E communities. Several DAU SMEs discussed how this trend may be linked to official policy that is not always as cohesive or direct as required in regard to dealing with planning for problems. "Policy does not direct planning for a problem. We plan for success with minimum schedule" commented one SME. Interestingly, the Defense Acquisition Guidebook (DAG) directly addresses this observation (2012: 52):

> "Experience indicates that most programs use a success based timeline when planning the integrated program schedule, meaning that each event or activity is based on positive results and moving to the next activity or phase of the acquisition effort. Experience also indicates that this concept is a major fault in most program planning."

Another issue is the limited number of test resources available. The following quotes reveal how a limited number of test resources (test personnel and dedicated test aircraft) can cause interdependencies between test programs and outside organizations (not a DT&E organization) negatively impacting schedule.

> "We are constantly trying to find qualified test personnel. It is forcing me to borrow people [from other test organizations] in order to execute my own tests. Now my test is dependent on whether or not someone outside my organization is available. Luckily, we have a pretty good relationship with those organizations and they are in the same boat as us. They help us when they can and we do the same. However, I've been here long enough to know it is not always like that." (RTO SME)

> "We don't own our test aircraft. When we want to execute a test we have to coordinate with the ops guys to get one of their birds. Sometime our tests last a few weeks and we need the aircraft the entire time. But they have a mission to do as well. They don't want to give up a bird for that long and they own it so if they don't want to or need it for something else, we don't test unless the test is important enough that we start climbing the chain [of command] and get one of them to set the priority. It's a constant struggle. And when we do get one, they don't give us their best aircraft, they give us the one that is having maintenance issues. So now I'm fighting maintenance issues while trying to execute test." (RTO SME)

> "The RTO gave us a schedule estimate about a year out. Problem was we ended up a low priority program and we were constantly fighting for range

> time which we never got. It took longer to get done than we first thought." (SPO SME)

The next quote reveals how acquisition leadership can encourage negative behavior. Although not directly a process issue, it reveals how acquisition reform will have to address cultural issues in addition to process reform.

> "We try to plan for bad things to happen, but when I take that padded schedule to senior leadership, I get punched in the face for planning for failure so I take it out or try to hide it in other places. Funny thing is, when bad things do happen, I get punched in the face by the same leadership because my schedule slipped." (SPO SME)

Another common theme discussed with several of the DT&E SMEs was poor test item quality resulting in unplanned, additional work to fix and test the configuration changes. There were two aspects identified: initial test item problems which occurred before test execution and test item deficiencies discovered during test execution. The next quote discusses how many test items are brought to the RTO in less than optimal conditions and how schedule may be impacted.

> "The reality is when a customer comes to us with a poor test item and we find problems, we don't just give it back to them and tell them to fix it on their own and bring it back. That doesn't help the customer or the warfighter. So we find the problem, then fix it, test it, then we find another problem, we fix that problem, then test that problem, and this goes on until we finish. It is not the most efficient way to execute test because we end up spending a lot of time fixing and testing the problems we find. Is it our [DT&E community] fault that the test item was of poor quality? No. The customer brought us a bad test item to begin with. We are merely the messenger of bad news and they [the SPO] are trying to shoot the messenger." (DT&E SME)

> "If we are consistently finding problems with the test item, 90% of the time the program is behind schedule, over budget, or both." (DT&E SME)

The DAG states "Although T&E is best managed as event-driven, in most cases it is not practical in practice" (DAG, 2012: 52). Several SMEs supported the DAGs observation.

> "Test should be event driven, but in reality we do not always follow that. Especially with larger programs, the SPO comes to us and tell us how long we have to test. Right now I am fighting with the **** program because they gave us *** months to test. I really need *** months to adequately test this system. The truth is we will test what we can test in that time, find and fix as many of the deficiencies as we can, and the SPO will hope that we don't find any big issues that delay the program. As for the less important deficiencies, most will just get fixed along the way. Sometimes though, one will get buried or carried to the next phase of testing. The program has the support and need to push its way through. But when they give me half the time I need to properly test the system and its gets pushed through to OT, is it really our fault that problems are discovered in OT that we would have found had I been given the time I requested?" (DT&E SME).

When the SME indicated that deficiencies were buried, clarification was requested.

> "We only find deficiencies and report them. It is up to the SPO to decide whether or not to fix them. For really big problems, the SPO will fix these because they can be show stoppers. But for smaller problems that don't seem to have a large impact on the system, sometimes they are played down as unimportant, do not get fixed, and are swept under the rug as unimportant. However, later in OT the problem surfaces, only this time the OT guys think it's a big problem and are upset with us [RTO] because we didn't find the problem. Well, truth is we actually did find it and reported it, but the SPO downplayed or hid it because they didn't want to spend the time or money to fix it." (DT&E SME)

One SME postulated that the location of DT&E in the acquisitions cycle may be a source for delay. Located at the end of a program's life cycle, by the time the programs arrives at DT&E, any delay potentially planned for and built into the schedule has been utilized in other phases of acquisition and must have unrealistic performance results in order to finish on time. In the AF, generally the RTO will direct what testing needs to be accomplished and is supportive of more thorough testing (which takes time and costs money) while the SPO must balance cost, schedule, and performance objectives which rarely allow for thorough system testing. The SPO and RTO objectives are contentious and can result in a hostile relationship between the SPO and DT&E community.

> "Test falls in a poor location on a programs schedule. Program schedules are planned and approved sometimes years in advance and they [SPO] try to take into account schedule delays. They lay it out and it all looks nice with plenty of time for everything. But then there is a problem with manufacturing, the software is late, and then something else takes longer to fix than anticipated. All these eat up schedule and if it takes too long it eats schedule from somewhere else. By the time the program gets to test all that padding is gone, the money is tight, and they hope nothing goes wrong. I see hope as a risk management strategy way too often." (DT&E SME)

SME Demographics

The SMEs that participated in this research were required to have experience in either the System Program Office (SPO) or DT&E community. Unfortunately, individuals in the SPO community with PM experience and interest in participating proved to be challenging as they tended to be higher ranking individuals, busy, and a majority politely declined to participate. It is worth noting that at this particular point in time the government shutdown of 2013 had just concluded. Had this event not occurred, interest from the SPO community may have been greater. Regardless, two SPO SMEs with PM experience participated in this research. The response from the DT&E community was more positive and consisted of the majority of SME demographics. These individuals were GS-15/ Lieutenant Colonel and below personnel. In total, eleven SMEs participated in this research which included active duty/ retired Army and AF officers, DAU professors, RTO test conductors, SPO managers, and RTO test directors. The experience ranged from over twenty-five years of acquisition experience (one of the

PMs) to four years (for a RTO test conductor). All SMEs had spent time executing test, one had written AF test policy, three were DAU T&E professors, and three had worked in a program office. Many of the individuals were identified through snowball sampling.

SME Discussion Summary

The SME discussions brought to light several issues which may contribute to DT&E program delays. Initial schedule estimates are created and approved long before test execution occurs which may not be representative of the program once it reaches DT&E. SPO estimates appear to be overly optimistic possibly due to senior leadership cultural issues. For those programs that do plan for delays, the estimated delay is still optimistic due to problems encountered early in the acquisition lifecycle which supports overly optimistic DT&E schedule. RTO estimates are based on current organizational manning and resource conditions which may not be representative of the future state at the time the program arrives at DT&E. In addition, substandard test item quality may be forcing the RTO to execute a suboptimal test management methodology (fly, fix, fly) with limited resources in order to provide the warfighter a system of limited capability sooner rather than a perfect solution later. Unfortunately, differing opinions on what deficiencies require additional schedule to address can be motivated by the good intention of getting a weapon system to the warfighter as soon as possible but at the risk of overlooking or missing a critical deficiency. These observations originate from a small sample population of the acquisition community, with a majority originating from the DT&E perspective, and do not necessarily represent the general consensus of the acquisition community. Even if the opinions presented here represent a sound basis of the acquisition community, they are purely subjective in nature. A quantitative method of analyzing DT&E activities and delays was required.

DT&E Conceptual Model

The knowledge obtained from the Chapter II was utilized as a foundation to identify critical DT&E activities and significant delays for creation of an initial conceptual model. The author's initial intent was to create a one-one process model, modeled at the PEO level of abstraction, with simulation capabilities. A breakthrough in creating the conceptual model came during review of the APM. As mentioned in Chapter II, there were several similarities between the APM and ERAM including that both modeled the AF DAMS processes. The major difference between the two models was APM was a process model which did not have simulation capabilities. Realizing the potential to transform APM into a simulation based model, the initial plan was to utilize the APM as an initial starting point to build a simulation model of the DT&E processes.

The author chose to model from a top down approach supported by Banks (2005:14). Based on the APM, the DT&E processes and their relationships to other acquisition processes were identified and assembled into a conceptual model. Several challenges were encountered during the course of creating a conceptual model due to the software ERAM was created in and ERAMs graphical size. None of the SME had access to Arena software and due to ERAM's size and complexity, it was impractical to transfer the design onto common software found on government furnished computers. This supported creation of a conceptual model in Microsoft Visio which all the SMEs had access to. After the first conceptual model was created, discussions with SMEs were accomplished to check the model's validity. During the discussions, it became apparent that a majority of the conversations were focused on processes and interactions not displayed in the conceptual model of this research but contained in other areas of ERAM. After several iterations of refining and verification of the model with SMEs, the size and complexity of the conceptual model increased to the point there was concern if the research project would finish on schedule. The research scope was narrowed to only focus on DT&E execution activities. In addition, the majority of the research project delay was attributed to verifying the process model. However, a process model was not required to answer the investigative questions and seen as replicating Wirthlin's original work, the author chose not to structure the simulation model as a process model. This approach was supported by the idea that "It is not necessary to have a one-to-one mapping between the model and the real system. Only the essence of the real system is needed" (Banks, 2005: 14). This decision simplified the model considerably. A figure of the conceptual model is available in the Appendix. Several iterations of SME discussions and modifications to the model were required in order to arrive at a general consensus that the model reasonably represented the system in reality. At this point, the conceptual model was considered to have face validity.

DT&E Simulation Model

The conceptual model was translated into a simulation model in the Arena software separate from ERAM 1.0 in order to decrease verification and simulation run time. The separate model is referred to as the DT&E Model (DTEM). Several iterations of model building, inputs from SMEs, and refinement were conducted resulting in the final DTEM version as presented in Figure 10. A detailed description of the model is available in the Appendix. Chapter IV will assess model validity and investigate DT&E activities and delays through interventions to assess activity/ delay significance at the system level.

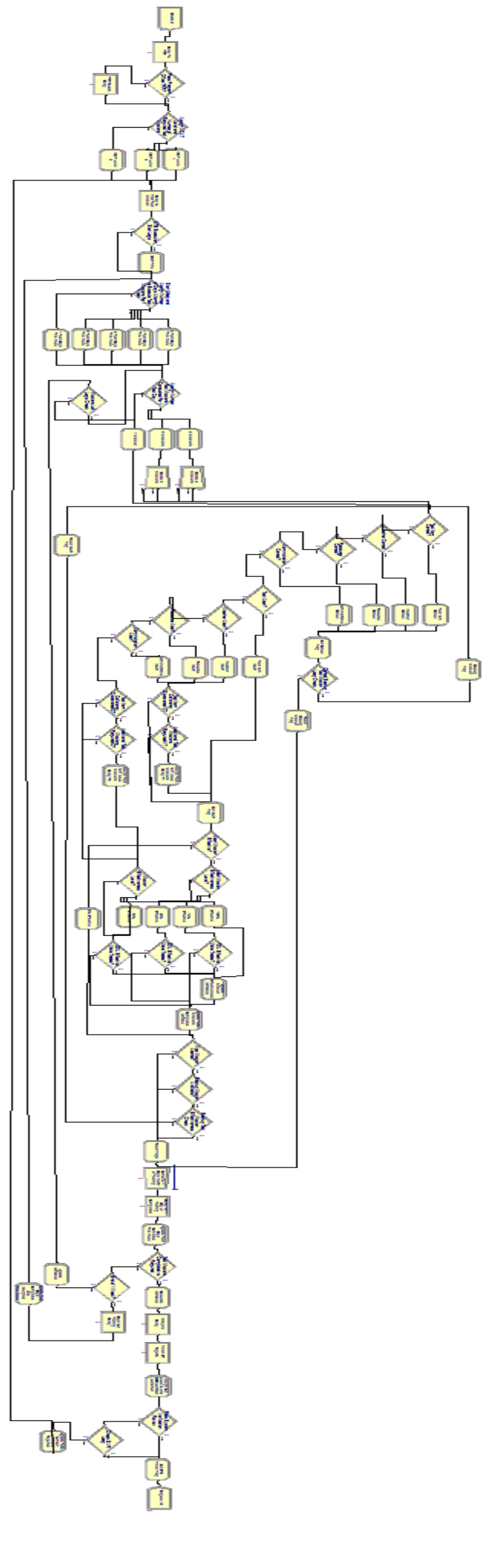

Figure 10: Final DTEM

IV. Analysis and Results

DTEM

Assumptions

Several model assumptions were required for model abstraction and simplification of several concepts and activities. ERAM 1.0 assumptions were incorporated into DTEM for consistency and model integration purposes. The most relevant ERAM 1.0 assumptions were: the entity passing through the model is a program which can be represented by an ACAT dependent number of required test missions and there are no memory effects in the model (Wirthlin, 2009: 148-149). DTEM assumptions were constructed with input from SMEs and are listed below:

- Backup missions are executed the same day as the primary mission. In reality, backup test missions are not necessarily executed on the same day as the primary as a risk management technique.
- A single backup mission is planned for every primary mission. In practice, depending on the criticality of a test mission, several backup missions could be scheduled. However, this assumption simplified the model while still capturing the intent of backup missions.
- Each test mission is independent of any test missions.
- If a test mission is more than 60% effective, it will not execute a backup mission.
- At least one and no more than five days of testing will occur each week.

- The historical test mission data utilized in constructing the DTEM are valid representation of the reality. Three SMEs (collocated at the MRTFB where the test mission data were collected) discussed how the test mission data are purely representative of the people who report the data and the process may incentivize reporting optimistic values or conducting unorthodox behavior to improve organizational performance statistics.

> “If we have an aircraft problem early in the day, maintenance will push the test mission right hoping that there will be bad weather in the afternoon. If weather occurs, then maintenance will cancel the test mission which now gets labeled as a weather cancel when it was really a maintenance cancel” (DT&E SME)

This behavior could potentially skew the data and model results. However, the average data values utilized in the model were discussed with DT&E SMEs, from the same MRTFB where the test mission data was collected, who did not observe any gross abnormalities.

DTEM Verification

Verification of the DTEM model was accomplished through several of Arena’s built in verification capabilities. When executing a simulation, Arena will ensure all blocks in the model are appropriately connected, populated with parameters, and defined. If any of these conditions are not met, Arena will display an error window identifying the category of error and location. The model cannot be executed until all issues are corrected. Arena also has the capability to display variables, processes, statistics and other categories as a group in a single spreadsheet. This allows for easy verification that all items have the correct units, logical expression, or parameters. In addition, the user may display values for variables/ statistics while stepping through the simulation. By utilizing animation and displaying the current value variables at each step in the simulation, the user can observe the simulation progress, verify the models mathematical logic, and ensure the reports generated at the end of the simulation were displaying the correct values. If an anomaly occurs during the simulation run (such as division by zero) Arena will terminate the run and display a warning window identifying the type, time, and location of the issue. Several iterations of model refinement and calibration utilizing the techniques discussed were required before the model would run error free with no unusual results or observed behavior. At this point the model was verified.

DTEM Validation

No historical data for ACAT DT&E schedules were available for this research. However, historical data from Wirthlin's research regarding program schedule from MS B-C were available and utilized once DTEM and ERAM 1.0 were integrated. Two aspects of validation are presented: face validity of DT&E execution time, ERAM 3.0 MS B-C time.

DTEM Face Validity

DTEM exports data files of user specified system performance parameters and ACAT time spent in test execution. These data were imported into Microsoft Excel 2007 and analyzed using Excel's Analysis Tool-pack. Histograms and descriptive statistics were compiled (Figures 11-15) and presented to SMEs who reviewed the results providing confidence in model face validity. SME feedback on the DTEM results was positive. The comments ranged from “They look good” and “The histograms look realistic considering all of the variables that come into play” to “The histograms do seem to tell a story.” Based on the comments from SMEs, DTEM results were considered to be a representation of reality accrediting DTEM with face validity. The next step was to integrate DTEM into ERAM 1.0 and statistically compare ERAM 3.0 (the integrated DTEM and ERAM 1.0 model) to the historical data gathered from Wirthlin’s research.

ACAT I Statistics	
Mean	958
Standard Error	2.04
Median	933
Standard Deviation	316
Sample Variance	100221
Range	1863
Minimum	64
Maximum	1927
Count	24158
Confidence Level(95.0%)	4

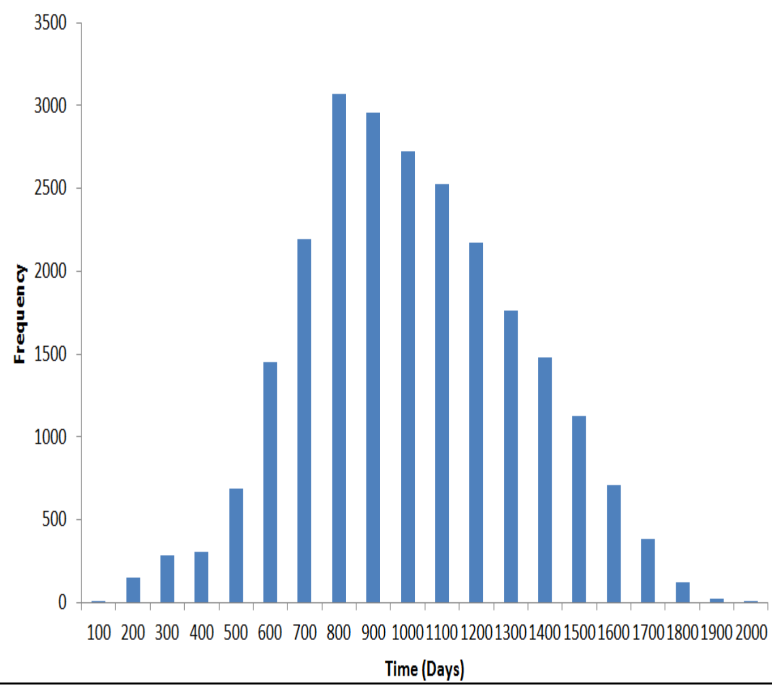

Figure 11: DTEM ACAT I Schedule

ACAT II Statistics	
Mean	310
Standard Error	0.80
Median	303
Standard Deviation	94
Sample Variance	8899
Range	644
Minimum	15
Maximum	659
Count	13994
Confidence Level(95.0%)	1.56

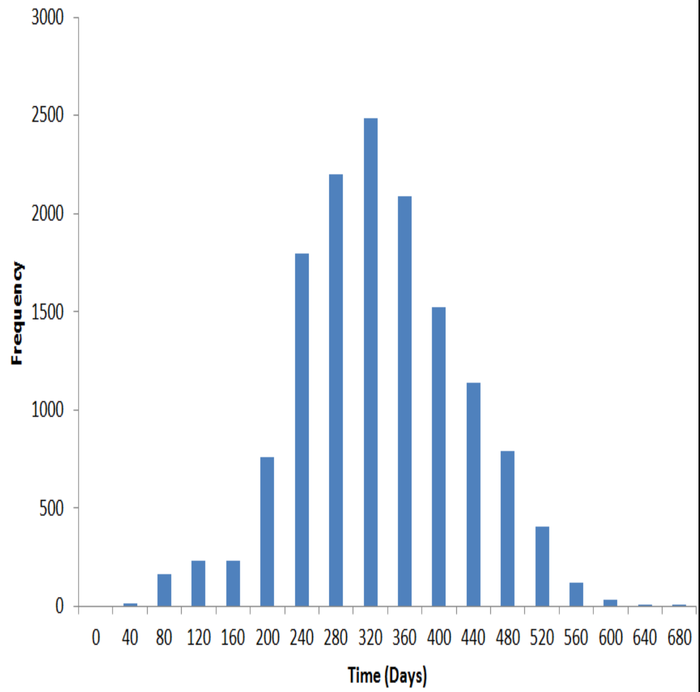

Figure 12: DTEM ACAT II Schedule

ACAT III Statistics	
Mean	233
Standard Error	0.32
Median	226
Standard Deviation	79
Sample Variance	6252
Range	546
Minimum	8
Maximum	554
Count	61848
Confidence Level(95.0%)	0.62

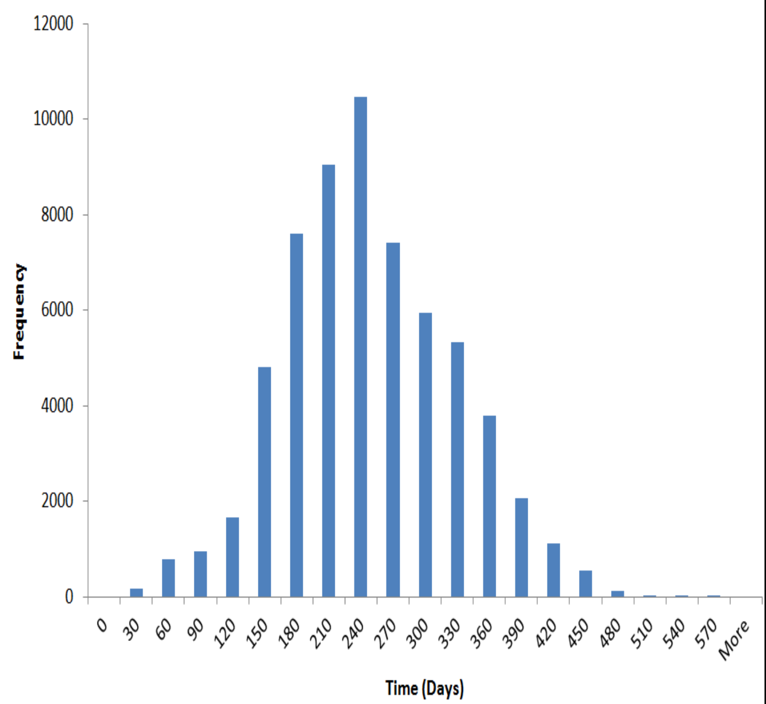

Figure 13: DTEM ACAT III Schedule

Test Mission Growth Statistics	
Mean	37
Standard Error	0.05
Median	34
Standard Deviation	17
Sample Variance	290
Range	234
Minimum	0
Maximum	234
Count	100000
Confidence Level(95.0%)	0.11

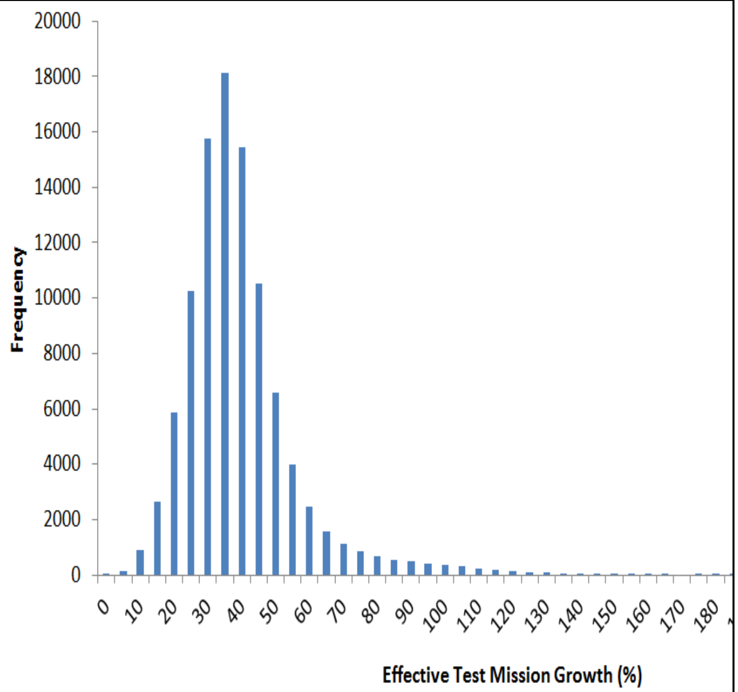

Figure 14: DTEM Effective Test Mission Growth

Test Mission Completed Per Day	
Mean	0.34
Standard Error	0.0007
Median	0.29
Standard Deviation	0.23
Sample Variance	0.052
Range	6.87
Minimum	0.23
Maximum	7.14
Count	100000
Confidence Level(95.0%)	0.001

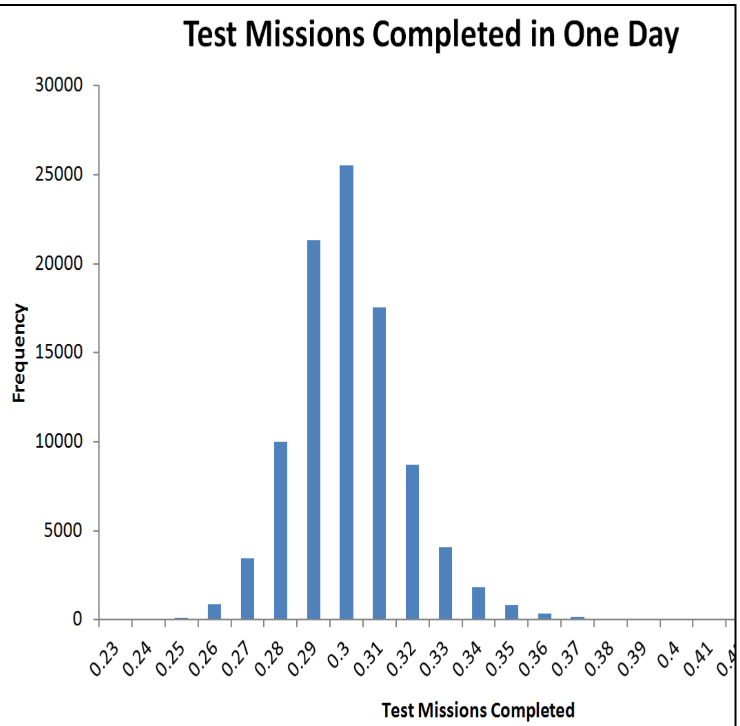

Figure 15: DTEM Average Number of Effective Test Missions Executed in One Day

ERAM 1.0 and DTEM Integration

Integration efforts provided confidence that no unintended configuration changes occurred in merging ERAM 1.0 and DTEM. While creating DTEM, ERAM 1.0 interface boundaries were investigated starting with identification of ERAM DT&E activities and their operations within the model. The DT&E activities of interest are contained in the red system boundary in Figure 16.

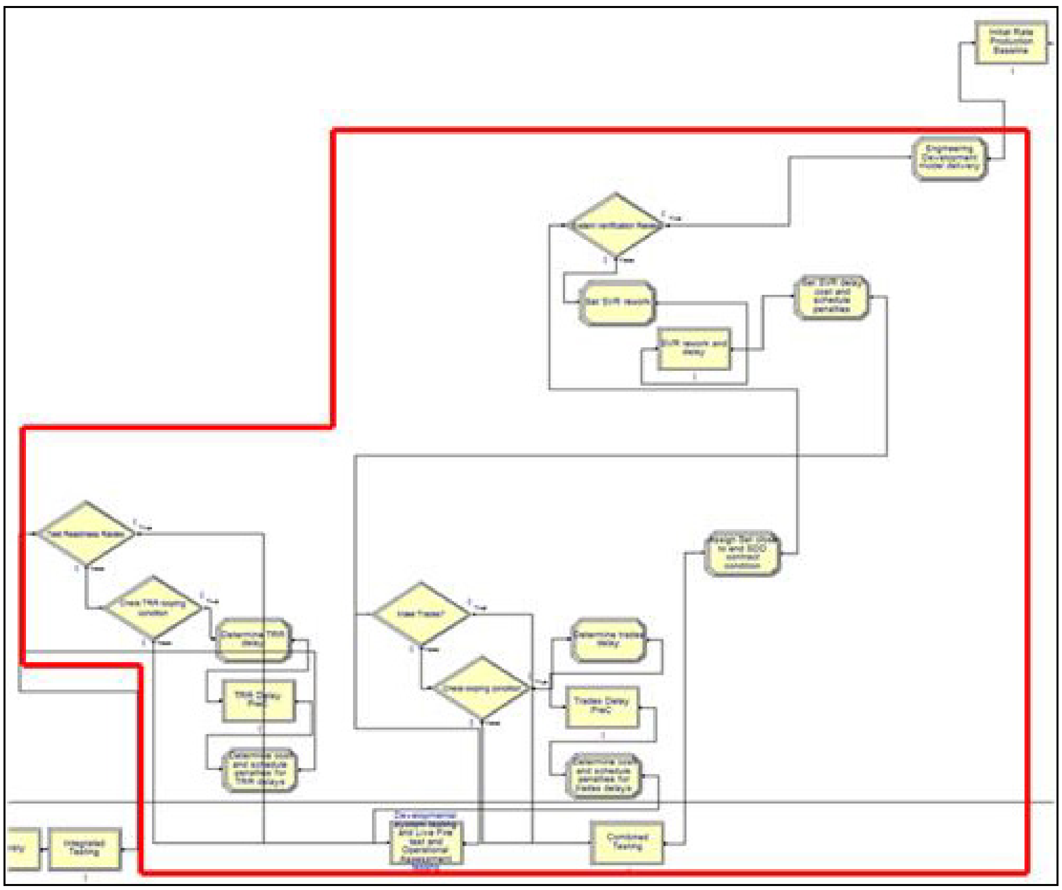

Figure 16: ERAM 1.0 DT&E Activities

Discussions with Wirthlin were conducted to verify the block and variable operations and how they could potentially impact integration efforts. Exploratory runs with adjusted activity distributions were conducted on ERAM 1.0 in order to increase confidence in their relationships and operations. The final integrated model, ERAM 3.0, replaces the blocks contained in the red system boundary in Figure 16 with a single, hierarchical block labeled “DTEM.”

ERAM 3.0 and ERAM 1.0 Analysis

The program time spent in DT&E activities for ERAM 3.0 and ERAM 1.0 were analyzed. Details of the analysis (including histograms/ cumulative distributions of the data, KS Test Results, and a table of percent differences between the models for each ACAT category) are available in the Appendix. KS tests concluded that the two models were statistically different for each ACAT category.

ERAM 3.0 Validation

Hypothesis testing with the unequal variance student t-Test were utilized to calculate ERAM 3.0 validity with respect to the historical data. The student t-Test requires the assumption that sample data are assumed to be approximately normally distributed. The test calculates a t-statistic and compares it to a critical value obtained from a t-Test table which indicates if there is enough information to support rejection of the null hypothesis. The null hypothesis, H_o, is that the difference between the ERAM 3.0 sample mean and the historical data sample mean is zero. The calculated t-statistic utilizes the means of each sample ($\overline{X}$), an estimate of each sample’s standard deviation (S), and the number of observations for each data set (n). The t-Test equation is shown in (1). The subscripts delineate between the two sample sets and were assumed to have unequal variances. Equation (2) is calculates degrees of freedom (df).

$$\text{t}\,statistic = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (1)$$

$$df = \frac{[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}]^2}{\frac{[\frac{S_1^2}{n_1}]^2}{n_1 - 1} + \frac{[\frac{S_2^2}{n_2}]^2}{n_2 - 1}} \quad (2)$$

(Banks, 2005: 438)

If the calculated t-statistic is greater than the critical t-statistic (or p-value less than 0.05), there is strong evidence to support rejection of the null hypothesis. Otherwise there is not enough information to support a statistical difference between the means of the two data sets. Wirthlin collected a limited data sample of historical program schedule from MS B-C (2009: 132-133) which were compared to the equivalent time frame in ERAM 3.0. A total of 10, 000 replications were utilized to construct the model data samples (Wirthlin, 2009: 137). Histograms and t-Test results of the historical and ERAM 3.0 data are presented in the following pages for each ACAT grouping. Figures 17-18 and Table 2 are the results for the All ACAT category.

Under the null hypothesis, H_o, there was a significant difference between the ERAM 3.0 and the historical data for the All ACAT category based on the results in Table 2. The analysis was repeated for the individual ACAT categories. The ACAT I results are presented in Figures 19-20 and Table 3.

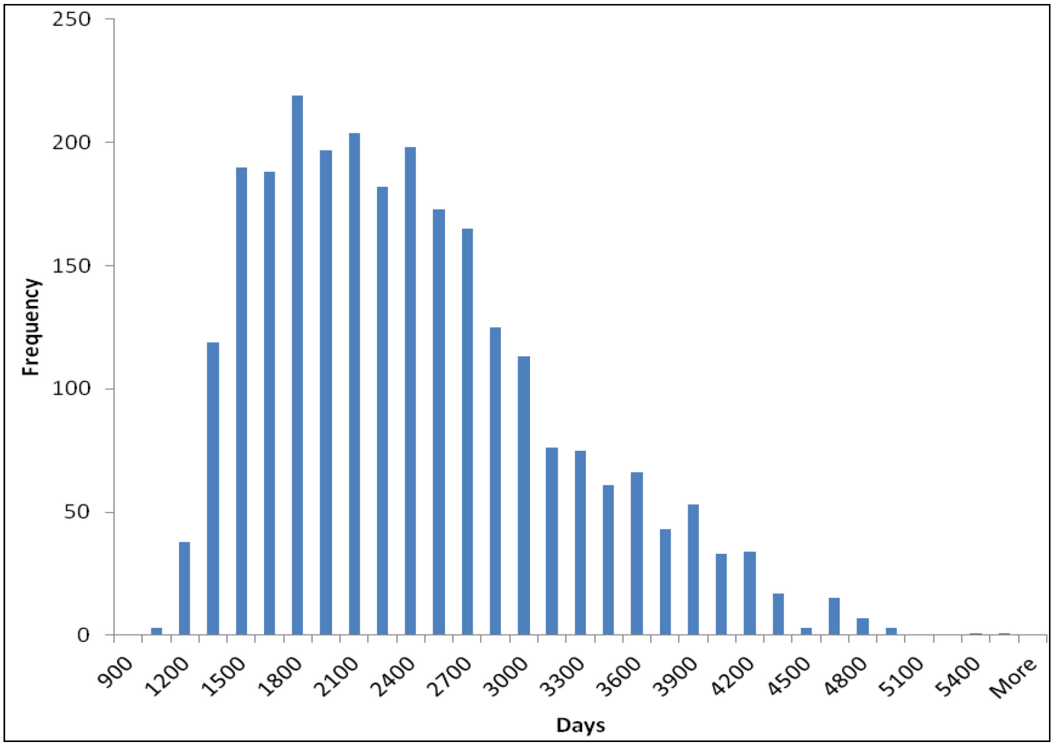

Figure 17: Histogram of ERAM 3.0 All ACAT MS B-C Schedule

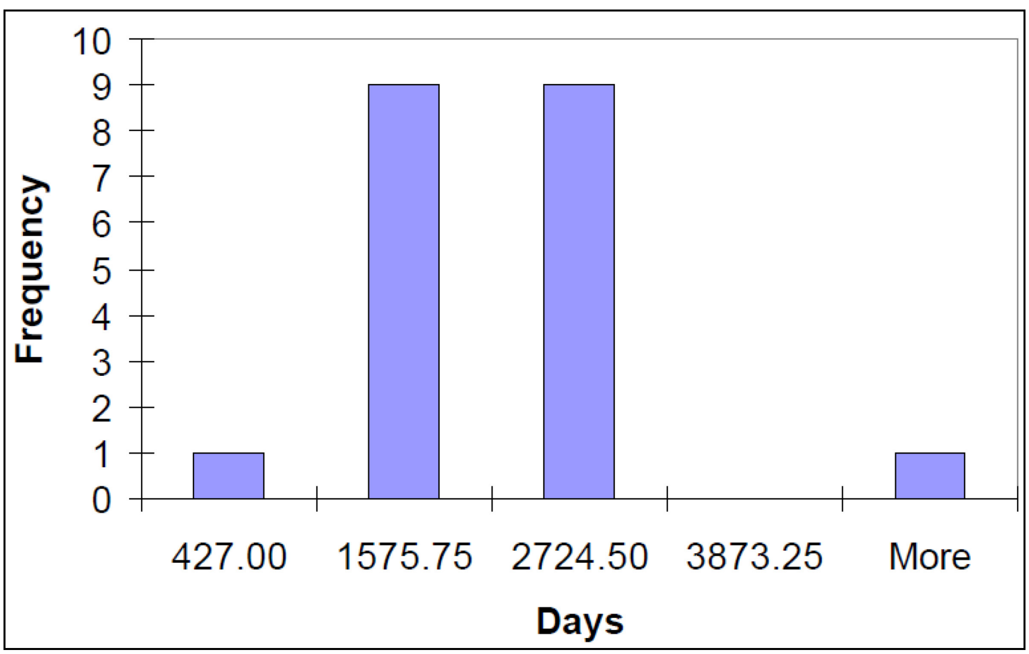

Figure 18: Historical Data All ACAT MS B-C Schedule (Wirthlin, 2009: 139)

Table 2: All ACAT MS B-C t-Test Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Historical Data**	**Model Data**
Mean	1620	2334
Variance	991072	601220
Observations	20	2602
df	19	
T Critical	2.09	
T Calculated	3.20	
P -Value	0.00	

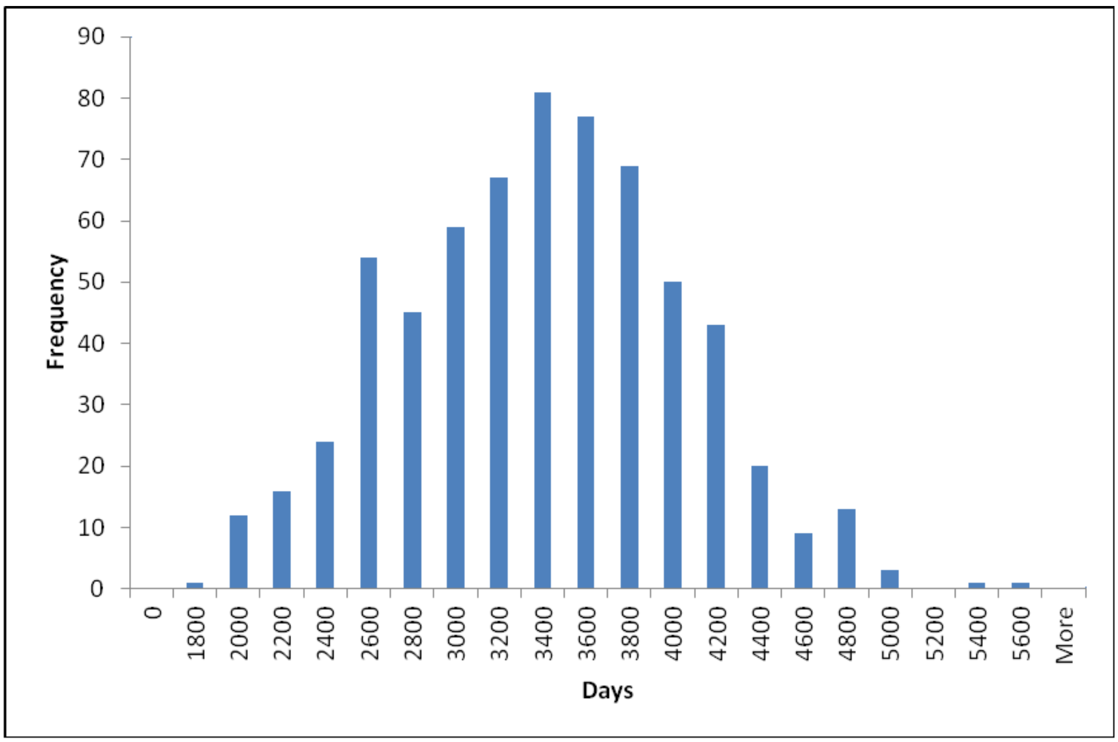

Figure 19: Histogram of ERAM 3.0 ACAT I MS B-C Schedule

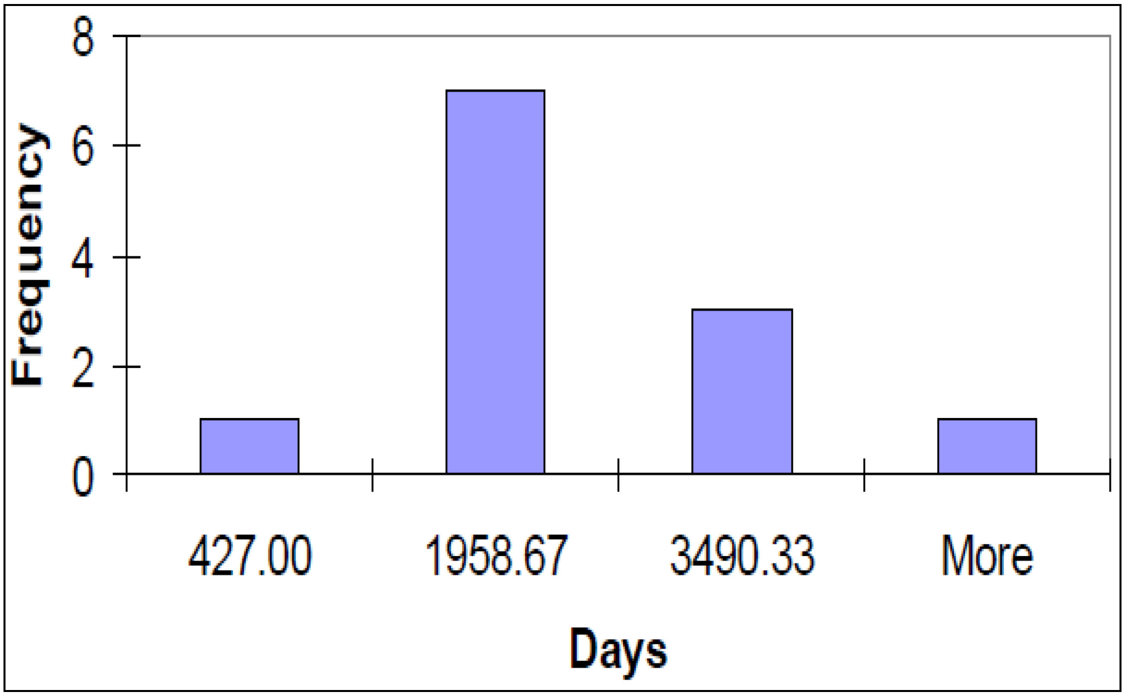

Figure 20: Historical Data ACAT I MS B-C Schedule (Wirthlin, 2009: 141)

Table 3: ACAT I MS B-C t-Test Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Historical Data**	**Model Data**
Mean	1801	3297
Variance	1435250	417191
Observations	12	645
df	11	
T Critical	2.20	
T Calculated	4.32	
P -Value	0.00	

Results in Table 3 show a significant difference between the means of ERAM 3.0 and historical data for the ACAT I category and the null hypothesis was rejected. The ACAT II analysis is presented in Figures 21-22 and Table 4.

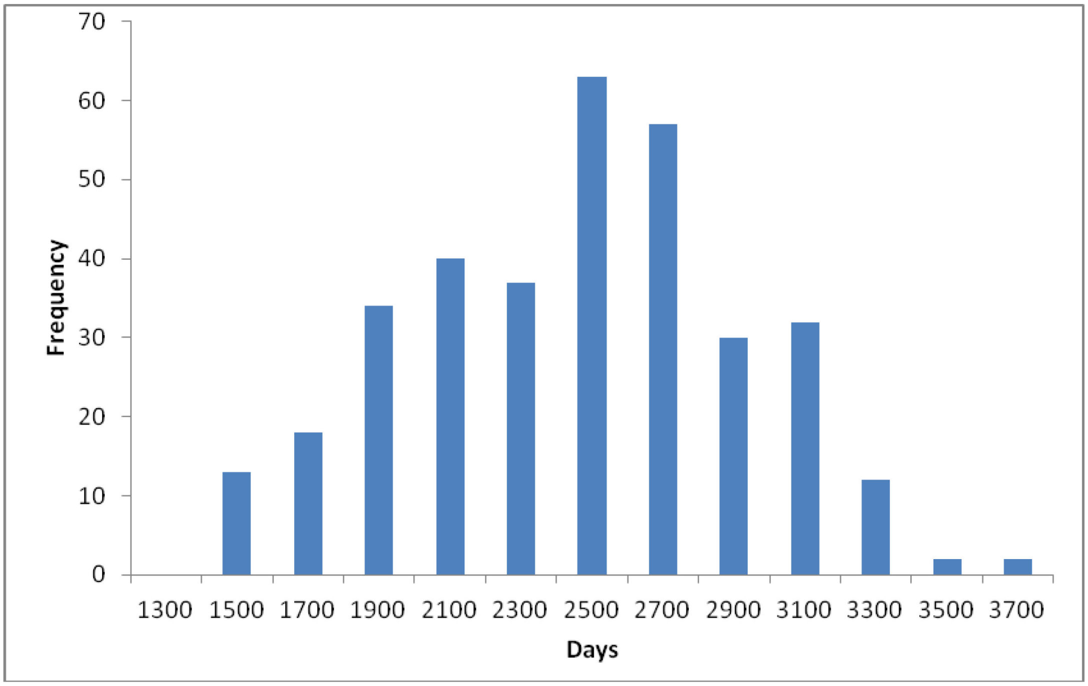

Figure 21: Histogram of ERAM 3.0 ACAT II MS B-C Schedule

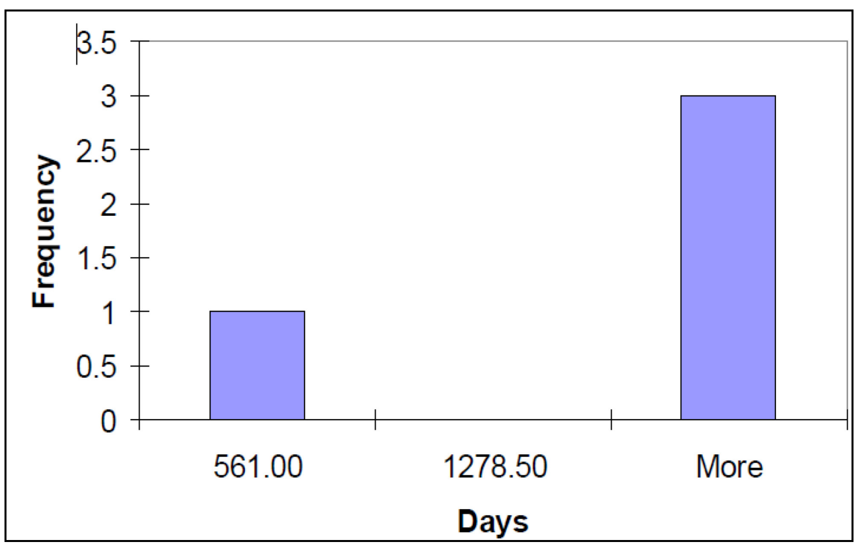

Figure 22: Historical Data ACAT II MS B-C Schedule (Wirthlin, 2009: 143)

Table 4: ACAT II MS B-C t-Test Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Historical Data**	**Model Data**
Mean	1476	2363
Variance	422276	217613
Observations	4	340
df	3	
T Critical	3.18	
T Calculated	2.72	
P -Value	0.07	

A p-value of 0.07 dictated that the null hypothesis was not rejected for the ACAT II category. Figures 23-24 and Table 5 display the ACAT III results.

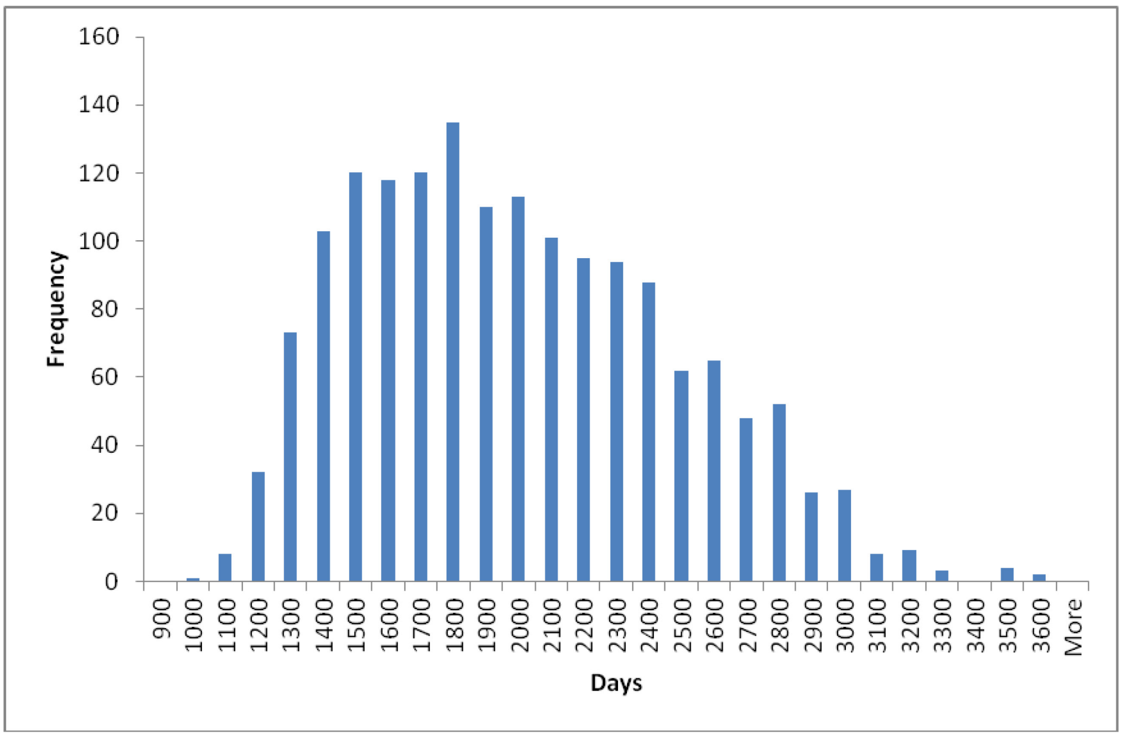

Figure 23: Histogram of ERAM 3.0 ACAT III MS B-C Schedule

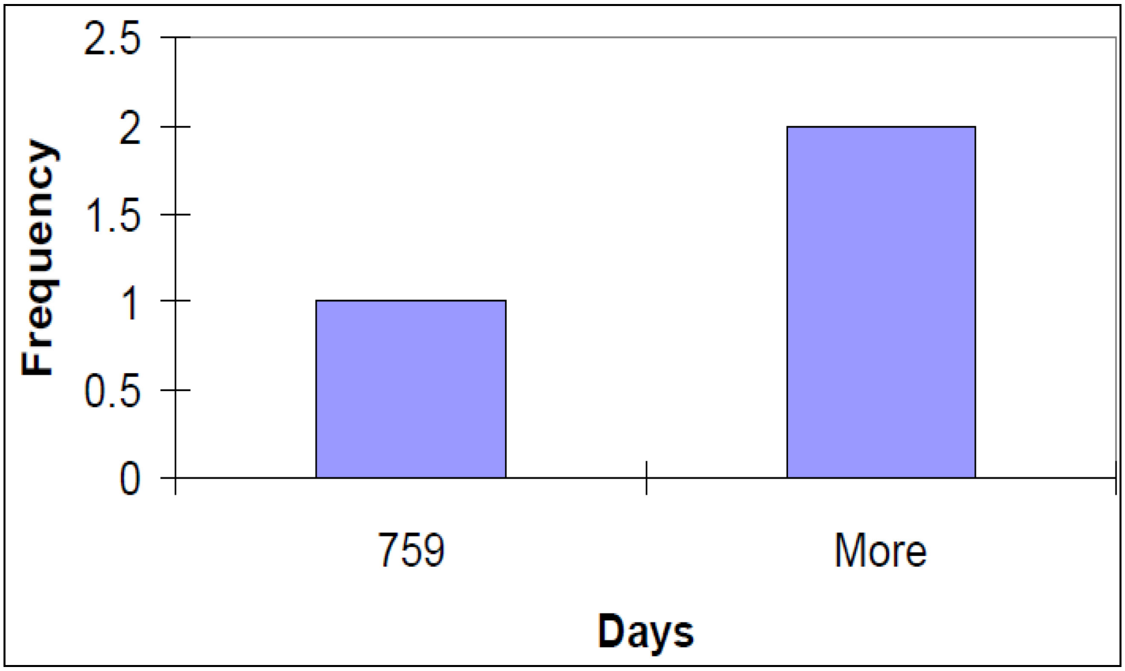

Figure 24: Historical Data ACAT III MS B-C Schedule (Wirthlin, 2009: 145)

Table 5: ACAT III MS B-C t-Test Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Historical Data**	**Model Data**
Mean	1224	1945
Variance	224564	234093
Observations	4	1617
df	3	
T Critical	3.18	
T Calculated	3.04	
P -Value	0.06	

The null hypothesis was not rejected for the ACAT III category. The analysis results for comparing the program schedule from MS B-C for ERAM 3.0 and the historical data is summarized in Table 6.

Table 6: Summary of Results

t-Test Results		
ACAT Group	**p-value**	**Result**
ALL	0	Reject H_o
I	0	Reject H_o
II	0.07	Fail to reject H_o
III	0.06	Fail to reject H_o

A second iteration of model refinement and data collection would have been beneficial in addressing the ERAM 3.0 validity for ACAT I and All ACAT categories, but there was not adequate project schedule to accomplish this. However, based on the available sample data, ERAM 3.0 was valid for ACAT II and III programs. For academic purposes, this level of model validity was adequate to continue the research. Next, the author will demonstrate how acquisition reform policies may be simulated in ERAM 3.0 to quantitatively support policy implementation in reality and further characterize DT&E's role in acquisitions.

ERAM 3.0 Interventions

This section demonstrates how potential acquisition reform policy may be executed in ERAM 3.0 and the resulting impacts analyzed to support reform implementation. Referred to as interventions, ERAM 3.0 was modified in an explicit method with results compared to the baseline ERAM 3.0 data through hypothesis testing. A one tailed, unequal variance t-test was utilized. The null hypothesis for all interventions was: "The difference between the intervention mean and baseline mean is

0.” The interventions were chosen based on discussions with SMEs and a select few identified by Wirthlin as having unexpected results in ERAM 1.0. The interventions are based on concepts of improved program quality, test item quality, test item quantity, and RTO resource availability. Table 7 provides a list of the different types of excursion that are investigated. The t-Test analysis was only conducted for the All ACAT category. The results are presented in tabular format with additional information regarding the differences between the model's descriptive statistics available in the Appendix.

Table 7: All ACAT Interventions Summary

Intervention	Program Quality	Test Item Quality	Test Item Quantity	RTO Resource Availability
TRR	X			
SVR	X	X		
RTO Test Resource Availability				X
Test Item Quantity			X	
Additional Test Missions	X	X		
Decrease Maximum Delay to First Test Mission		X		X
Decrease Test Item Deficiencies		X		
Aggregate	X	X	X	X

TRR Intervention

ERAM 1.0 concluded that the TRR activities did not significantly impact program schedule. This result was surprising because SMEs indicated scheduling of test ranges was a significant source of program delay (Wirthlin, 2009: 189). For this intervention in ERAM 1.0, Wirthlin adjusted the probability of passing the TRR from 70% to 100%

which represented an increase in the quality of a program. The same intervention strategy was executed in ERAM 3.0 where the baseline value of 90% successes was adjusted to 100%. The results of the t-test (p-value of 0.41), shown in Table 8, indicate that there is not enough evidence to support rejection of the null hypothesis at the 95% confidence level. This supports Wirthlin's original conclusion that the TRR is not a critical activity for acquisition programs in regards to program schedule. The increase in the intervention mean in Table 8 is attributed to the insignificance of the activity combined with the stochastic nature of the model because the value remains within the standard error.

Table 8: TRR Intervention Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4243
Variance	2867719	2885263
Observations	6582	6592
Hypothesized Mean Difference	0	
df	13172	
t Stat	-0.21	
P(T<=t) one-tail	0.41	
t Critical one-tail	1.64	

SVR Intervention

The SVR ensures that programs have adequately conducted DT&E and addressed major test item deficiencies with a baseline probability of 85% passing the review. ERAM 1.0 implemented the intervention with the acquisition reform concept of programs adequately addressing all test item issues before the SVR resulting in a 100% probability of passing the review. The same intervention strategy was implemented in ERAM 3.0 where the baseline value of 95% was increased to 100%. The results are in Table 9.

Table 9: SVR Intervention Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4233
Variance	2867719	2874276
Observations	6582	6594
Hypothesized Mean Difference	0	
df	13174	
t Stat	0.12	
P(T<=t) one-tail	0.45	
t Critical one-tail	1.64	

The two tailed t-Test resulted in a p-value=0.45 meaning that there was not enough evidence to support rejection of the null hypothesis and the difference between the baseline and intervention data are insignificant. This result supports SMEs observations that indicated that by the time a program arrives to the SVR there is a very high probability that it will pass regardless of whether there are still deficiencies. The DAMS supports pushing a less capable product to the warfighter in less time than providing the 100% solution in a longer time frame. This concept is sometimes referred to as the "%80 solution" in the acquisition community.

RTO Test Resource Availability Intervention

SPO SMEs indicated that many programs experienced significant program schedule delays because of a lack of RTO test resources while executing tests. This delay factor included priority conflicts over test ranges (the factor most commonly mentioned), RTO test personnel, test range personnel, maintenance, test support aircraft, and other RTO test infrastructure. The acquisition reform this intervention represents in reality would be the procurement of more test ranges, test personnel, maintenance personnel, test support aircraft, and other RTO test infrastructure to decrease the probability of delays due to this factor. For this intervention, the probability that a test mission cancelation or abort occurs is reduced from the baseline value (FOUO) to 0%. The results are summarized in Table 10.

Table 10: RTO Test Resource Intervention Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4247
Variance	2867719	2881313
Observations	6582	6574
Hypothesized Mean Difference	0	
df	13154	
t Stat	-0.32	
P(T<=t) one-tail	0.37	
t Critical one-tail	1.64	

The p-value is 0.37 and the null hypothesis was not rejected. The availability of RTO test resources during test execution does not significantly impact program schedule to MS C. This result is surprising considering the number of SMEs who indicated that there was an availability issue with RTO test infrastructure resources significantly impacting programs. This result warrants further investigation and is discussed in Chapter V.

Additional Test Missions Intervention

Several DT&E SMEs addressed how additional test schedule would be of value to address test item deficiencies. How much more time could be spent in DT&E without significantly impacting the programs schedule to MS C? This time could be utilized to execute additional test missions and potentially find more test item deficiencies resulting in a higher quality weapon system delivered to the warfighter in statistically the same amount of time. This intervention was executed by increasing the initial required number of test missions required to progress through DT&E by 10%. The intervention results are in Table 11.

Table 11: 110% Additional Test Missions Intervention Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4284
Variance	2867719	2941144
Observations	6582	6596
Hypothesized Mean Difference	0	
df	13175	
t Stat	-1.57	
P(T<=t) one-tail	0.06	
t Critical one-tail	1.64	

This intervention did not have a statistically significant impact on the program schedule (p-value = 0.06) indicating that a program could execute 10% add test missions without significantly impacting schedule. The intervention was repeated at 115% (results in Table 12) which had a significant impact to on schedule and the null was rejected. This set of interventions indicated that a program could be required to execute between 10%-15% addition test missions without significantly impacting schedule to MS C.

Table 12: 115% Test Missions Required Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4306
Variance	2867719	2980340
Observations	6582	6604
Hypothesized Mean Difference	0	
df	13181	
t Stat	-2.32	
P(T<=t) one-tail	0.01	
t Critical one-tail	1.64	

Decrease Maximum Delay to Execution of First Test Mission Intervention

After passing the TRR, SMEs identified a delay before test execution begins and was attributed to several factors including poor test item quality, delay due to test range availability, and the RTO technical reviews. These delays are represented by a single abstract process block, "Delay to First Test Mission," with a triangular distribution of (1, 30, 365). SMEs indicated that the maximum value in the distribution was representative of poor test item quality and RTO test range unavailability. If better quality test items were produced through use of technology with higher technology readiness levels, increased systems engineering efforts earlier in acquisitions, better-trained personnel, and other engineering practices then the maximum observed value in the triangular distribution could be decreased. In this intervention, the maximum delay is decreased to 45 days. This value was suggested by SMEs as the maximum delay to complete the RTO technical reviews without any RTO test resource or test item issue delays. This decrease was acknowledged to be unrealistic but was a practical starting point because if this value was found insignificant, then no values between 45 and 365 would be either. Intervention results are summarized in Table 13.

Table 13: 45 Days Maximum Delay to Execution of First Test Mission Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4148
Variance	2867719	2854192
Observations	6582	6574
Hypothesized Mean Difference	0	
df	13154	
t Stat	3.01	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.64	

The intervention results reject the null hypothesis and the two models are statistically different. A program will save 2% of schedule time (see analysis results in Table 33 in the Appendix) to MS C if the program can decrease the maximum amount of time to the execution of the first test mission to 45 days. However, decreasing the maximum delay to 45 may be unrealistic. Another intervention was simulated with the max delay adjusted to 182.5 or 50% of the baseline. Table 14 contains the results.

Table 14: 182.5 Days Maximum Delay to Execution of First Test Mission Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4185
Variance	2867719	2850115
Observations	6582	6601
Hypothesized Mean Difference	0	
df	13181	
t Stat	1.77	
P(T<=t) one-tail	0.04	
t Critical one-tail	1.64	

This iteration had a significant impact on program schedule (p-value= 0.04) with a 1% decrease in the mean (see Table 34 in Appendix I). The intervention was repeated at a maximum delay of 228.125 (or a 37.5% decrease in the baseline). The p-value was calculated at 0.08 and the t-Test failed to reject the null hypothesis (refer to Table 15).

Table 15: 228.125 Days Maximum Delay to Execution of First Test Mission Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4196
Variance	2867719	2857758
Observations	6582	6570
Hypothesized Mean Difference	0	
df	13150	
t Stat	1.40	
P(T<=t) one-tail	0.08	
t Critical one-tail	1.64	

This set of interventions revealed that a decrease in the maximum delay of the execution of the first test mission to greater than approximately 200 days will result in a significant impact to program schedule to MS C.

Test Item Deficiencies Intervention

The most commonly mentioned DT&E program delay factor was overly optimistic DT&E schedule based on optimal weapon system performance. The historical test mission data collected tracked test mission cancelation and aborts due to test item issues. Test deficiencies may also be discovered but not result in a test mission cancelation or abort. For this intervention, the probability of a cancelation, abort, or discovery of a test item deficiency was decreased from the baseline values (FOUO) to 0%. Although this value may be unrealistic, it was an efficient analysis technique.

Decreasing test item deficiencies could be executed in reality through increasing the quality of test items through more emphasis on early systems engineering activities, utilization of more mature technologies, early prototyping, and other engineering efforts. Table 16 summarizes the intervention results.

Table 16: 100% Decrease in Test Item Deficiencies Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4134
Variance	2867719	2737611
Observations	6582	6574
Hypothesized Mean Difference	0	
df	13148	
t Stat	3.54	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.64	

The null hypothesis was rejected at a p-value=0 with a mean decrease of 2% (see Table 35 in Appendix I). A second iteration was simulated with a value of 50% fewer test deficiencies.

Table 17: 50% Decrease in Test Item Deficiencies Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4183
Variance	2867719	2781795
Observations	6582	6587
Hypothesized Mean Difference	0	
df	13164	
t Stat	1.85	
P(T<=t) one-tail	0.03	
t Critical one-tail	1.64	

The models were significantly different (as indicated in Table 17) with a p-value of 0.03 and the null hypothesis was rejected. The intervention was repeated at a 37.5% reduction in test item deficiencies and the results are presented in Table 18. A calculated p-value of 0.08 resulted in failure to reject the null hypothesis. This set of interventions revealed that a decrease in test item deficiencies between 50%-37.5% would be required to have a significant impact on program schedule.

Table 18: 37.5% Decrease in Test Item Deficiencies Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	4195
Variance	2867719	2800061
Observations	6582	6608
Hypothesized Mean Difference	0	
df	13185	
t Stat	1.44	
P(T<=t) one-tail	0.08	
t Critical one-tail	1.64	

Test Item Quantity Intervention

For large programs, often there is only a single test article available for testing. How would having two test articles impact program schedule? This intervention investigated the idea that if the RTO had sufficient, qualified test personnel and test infrastructure to effectively execute test missions for two test articles, the number of potential test missions executed per day would increase by a factor of two. The intervention results are presented in Table 19. This intervention resulted in significant difference between models (p-value=0). The mean decreased by 14% (see Table 37 in Appendix I for analysis).

Table 19: Test Item Quantity Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	3655
Variance	2867719	2310744
Observations	6582	6605
Hypothesized Mean Difference	0	
df	13024	
t Stat	20.77	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.64	

Aggregate Intervention

This intervention investigated a system approach to acquisition reform. The following combination of factors was utilized to provide a realistic combination of reforms: Maximum Delay to First Test Mission (228.125), Maximum Delay to TRR (135), TRR (100%), SVR (100%), Test Item Quantity (2), and Test Item Deficiency (-25%). The results are shown in Table 20.

Table 20: Aggregate Intervention Results

t-Test: Two-Sample Assuming Unequal Variances		
Statistic	**Baseline**	**Intervention**
Mean	4237	3613
Variance	2867719	2300524
Observations	6582	6603
Hypothesized Mean Difference	0	
df	13017	
t Stat	22.30	
P(T<=t) one-tail	0.00	
t Critical one-tail	1.64	

The null hypothesis was rejected at a p-value= 0. This intervention resulted in a mean schedule decrease of approximately 15% (see Table 38 for analysis).

Intervention Analysis and Results Summary

DTEM increased the fidelity of the ERAM 1.0 DT&E activities further characterizing DT&E's role in the DAMS. The higher fidelity DT&E activities enabled investigation of several interventions attainable by no other practical method. A summary of the intervention results is available in Table 21.

Table 21: Intervention Results Summary

Intervention	Results
TRR	Not significant
SVR	Not significant
RTO Test Resource Availability	Not significant
Test Item Quantity	Significant at 2 test items, 14% mean schedule decrease
Additional Test Missions	Significant at greater than 10% additional test missions
Maximum Delay to First Test Mission	Significant at greater than 37.5% maximum delay decrease, 2% mean schedule decrease
Test Item Deficiencies	Significant at greater than 37.5% decrease in the number of deficiencies, 2% mean schedule decrease
Aggregate	Significant, 15% mean schedule decrease

The Null Program

Previous research by Baldus and others (2013) presented the concept of executing a null program that "did nothing" which effectively investigated how much time a program spent in system was due to process. A similar methodology was utilized for this investigation. DTEM was adjusted to execute a single test mission in order to observe how much time a program would spend in DT&E executing the process. For this intervention, the DT&E time was defined as the time from passing the TRR to passing the SVR. The results are displayed in Figure 25 and Table 22.

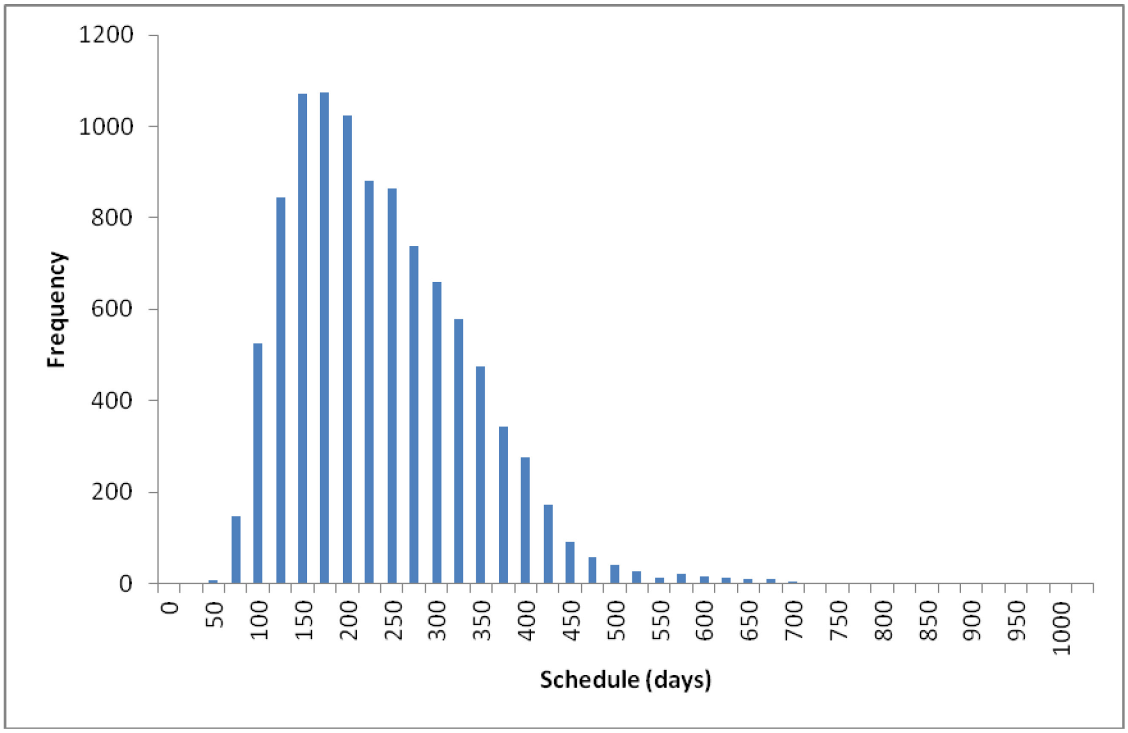

Figure 25: Histogram of DT&E Time to Execute One Test Mission

Table 22: DT&E Time to Execute One Test Mission

Results	
Mean	225
Standard Error	1
Median	209
Standard Deviation	100
Sample Variance	10008
Range	951
Minimum	33
Maximum	984
Count	10000
Confidence Level (95%)	2

The 95% confidence interval for executing a single test mission (with 10,000 replications) was calculated to be 224 +/- 2 days. This was a surprisingly high value with two implications. First, if the results are valid representations of reality, they suggest that a large amount of program schedule delay is due to the process itself. The same conclusion was reached by Wirthlin (2008: 211). In addition, this could hint at a possible acquisition “bottleneck” located at DT&E where programs are waiting in the “DT&E queue” to conduct testing. However, if the results are not valid, then the large amount of time required to execute one test missions suggests DTEM requires additional validation efforts and refinement for executing small numbers of test missions. Future work is necessary to investigate the validity of this result and is discussed in Chapter V.

Chapter Summary

This research did not exhaust all means by which ERAM may prove beneficial to the acquisition community nor is it absolute in its results. As shown in this research, modeling and simulation is an iterative process building upon the foundation of previous research. Regardless of the initial answers, future work will build upon the previous expanding ERAM’s capabilities to further demonstrate the utility this tool. The final chapter will discuss significant findings uncovered during this research project and aspects of ERAM which warrant additional research.

V. Conclusions

Chapter Overview

The purpose of this research was to utilize computer modeling and simulation to increase the fidelity of the DT&E processes with the goal of gaining new insight into DT&E's role in acquisition. Through increasing the fidelity of ERAM 1.0, the results of ERAM 3.0 supported two previous research conclusions and provided a different conclusion on a third. In addition, based on SME discussions and literature, several potential DT&E delay factors were identified, characterized in ERAM 3.0, and analyzed to analyze their significance with respect to program schedule. Chapter V provides conclusions based on the results and analysis of Chapter IV, areas for future work, and how this research could potentially impact acquisitions.

ERAM Observations

Poor Test Item Quality

Discussion with SMEs identified two primary potential DT&E program schedule delay factors: poor test item quality and a lack of RTO test resources. Interestingly, relevant literature was available on all of the delay factors investigated in this research and senior leadership appears to be well aware of them. Yet, acquisition reform is not a new concept (see to Figures 2-3) and continues to take longer than expected. The apparent ineffectiveness of acquisition reforms may be, in part, due to the DAMS state of causal ambiguity and long program cycle times. These observations further support the underlying concept of this research (and Wirthlin's) that the acquisition community could benefit from simulation model (similar to ERAM) with the capability of quantitatively estimating the impact of acquisition reform initiatives to support senior leadership decision making.

SPO and DT&E Relationship

All of the individuals who participated in this research expressed a genuine desire to improve the DAMS. The fundamental objectives, priorities, and perspectives of the DT&E and SPO communities or organizations are not always the same and at times conflicting. Discussions with SMEs from both communities offered insights into what their respective collective believes are primary factors in program schedule delay. Both communities identified overly optimistic program schedule based on high quality test items as the most significant and common delay factor. The reality is that test deficiencies are always discovered and are generally corrected. However, not all deficiencies are adequately addressed. Pressure to push weapon systems through DAMS drives sub-optimal test program management and test practices. Interestingly, the SPO community also identified the DT&E sub-optimal test methodology as a source of delay. One key document discovered during the literature review investigated the hypothesis that the "Department's developmental and operational test communities' approach to testing drives undue requirements, excessive cost, and added schedule into programs and results in a state of tension between Program Offices and the Testing Community" (Gilmore, 2011). The results of the investigation "found no significant evidence that the testing community typically drives unplanned requirements, cost or schedule into programs" and that "programs are most often delayed because of the results of testing, not the testing itself" (Gilmore, 2011). ERAM 3.0 results supported this conclusion in that the results of testing, such as test item deficiencies due to poor test item quality, are a significant source of program delay and a prime area for future acquisition reform.

RTO Resource Availability

This research investigated two aspects of RTO test resource availability: pre-test execution and test execution RTO test resource availability. The pre-test execution availability of RTO test resources (refer to the Maximum Delay to First Test Mission Intervention) significantly impacted program schedule which, according to SMEs, was believed to be largely due to a lack of RTO test ranges. However, once the program entered the test execution phase of DT&E, the RTO test resources did not significantly impact program schedule (refer to the RTO Test Resource Availability Intervention). The results suggest that there is a program “bottle-neck” located at DT&E, possibly due to the large number of programs attempting to utilize a limited number of test ranges, and a program will experience significant schedule delays here. However, once the program enters test execution phase it will unlikely encounter significant schedule delays due to RTO test resource availability. This was an interesting result because the model did not agree with SME opinion that test resource availability was a significant source of delay both prior to and during testing. This may be in part due to a skewed local perspective where RTO test resource availability does in fact significantly impact DT&E program schedule, but is not significant with respect to schedule to MS C. If the model results are valid, it further supports the basis for the need of a simulation model (like ERAM) to assist in educating the acquisition community on the complex relationships within DAMS and to assist in supporting acquisition reform.

DT&E Silver Bullet

The most substantial improvement from a single intervention was a 14% decrease in the average program schedule (up to MS C) based on providing the RTO additional test items enabling execution of twice as many test missions per day. This intervention decreased the mean time to MS C by approximately 590 days or 1.6 years. Many other interventions were also significant, but were limited to less than 2% reduction in the schedule mean which may by statistically significant, but not practically significant. Because ERAM 3.0 does not take into account the cost of these interventions, it is difficult to conclude their financial feasibility. Future work should investigate integration of the financial domain into the ERAM legacy to broaden its capabilities. Regardless, the ERAM 3.0 demonstrated how modeling and simulation could be utilized to better understand system level impacts through implementing local policy reform.

Program Schedule Confidence Intervals

One of the most interesting results of this research can be seen in Figure 25 which depicts the time required in DT&E to execute a single test mission. The idea that a program could spend over 200 days in DT&E to execute a single test mission is staggering. It would be interesting to observe the differences between the execution of one test mission, progressively increasing the value, and quantifying the point at which the time required to test additional test missions becomes significantly different. The results could indicate that the confidence intervals for executing one, five, ten, or more test missions are statistically the same meaning that on average a small program could plan to execute more test missions and on average incur a statistically insignificant delay. In addition, during the literature review two ERAM research vectors were identified highlighting how the recent research modified ERAM to focus on single program prediction estimates rather than system level performance. As was demonstrated with the execution of a single test mission, DTEM could easily be modified to the single program prediction research vector. By setting the required number of test missions to a programs estimate, the stochastic nature of DTEM combined with Monte Carlo analysis will produce a confidence interval for the a program's DT&E schedule. This could potentially be a valuable tool for both the SPO and RTO communities for estimating DT&E schedule and warrants future research.

Future Research

Model Validation

Modeling and simulation projects are iterative endeavors (Law, 2007: 67) and the several areas of improvement for this research are discussed below which were selected by the author as critical deficiencies in the 3.0 research. ERAM briefings were presented to SAF/AQXC, OUSD (AT&L)/ ARA/OS & FM, DAU, and AFLCMC/AQT who provided input regarding the research methodology, assumptions, and areas of concern. ERAM's validity was the primary concern from these organizations and emanated from the utilization of SME inputs for a majority of the model input parameters. Combined with the small historical MS B-C program schedule sample sizes, these organizations were concerned with ERAMs validity. ERAM was never intended to be utilized in its current configuration as the tool for senior leaders. Its goal was to demonstrate how computer modeling and simulation could be utilized in addressing acquisition reform. If senior DoD leadership desired a tool with the capabilities ERAM demonstrated, then another iteration of ERAM could be executed by a team of acquisition experts who could create a more valid model than a single doctoral candidate and several masters students could. However, if ERAM were to be utilized in its current configuration, efforts should focus on acquiring historical data to replace the SME inputs and collect a larger sample size to improve ERAM's validity.

New DoDI 5000.02

An updated version of the DoDI 5000.02 series was released during the writing of this thesis (USD, 2013). ERAM should be updated to reflect changes in the new DoDI 5000.02 instruction. One of the major changes discussed with SMEs was the ability to tailor the program's acquisition plan. This will result in numerous new possible pathways in ERAM and will undoubtedly impact program schedule. Interestingly, when asked how DT&E would be impacted by the new instruction, many SMEs indicated that at the test execution level there will be no change hinting that as much as DTEM is a valid representation of the current 5000.02 series, it will potentially have the same level of validity in the updated series. However, any actual impacts the new instruction may have on DT&E and acquisitions will take several years for programs to cycle through the DAMS and observe any process changes in reality.

Delay To First Test Mission

The DTEM block "Delay To First Test Mission" was purposefully made, early in the model building phase, as an abstract representation of several delay concepts in order to simplify modeling efforts. This resulted in the confounding of several critical delay factors which were later viewed to have potentially substantial impacts on DT&E program schedules. If another iteration of model building and calibration was possible, the “Delay To First Test Mission” block should be separated into three parallel processes representing delays due to RTO technical reviews, initial test item problems, and RTO test range scheduling conflicts. As was displayed in Chapter IV’s test problem interventions, several areas of the DTEM model were tested separately when in reality there would be some interdependency between the processes. Quantifying the interdependencies between the initial delay due to test item problems and the probability of finding test problems during test execution could result in even improved program schedule performance results and reinforce the idea that test item quality is a significant factor in program schedule delays.

Other MRTFBs

The historical test mission data utilized in this research project was only one of many MRTFB across the country as shown in Figure 26. It would be interesting to analyze test mission data from several MRTFBs and compare how test execution delay factors compared between the MRTFBs. If significant differences were present, it may suggest that program schedule performance could be MRTFB dependent.

Figure 26: Map of DoD MRTFBs (DAG, 2012: 150)

Final Thoughts

The ERAM research has demonstrated how modeling and simulation can provide a powerful analytical capability for supporting acquisition reform. This research improved the fidelity of the ERAM DT&E activities providing additional quantitative evidence supporting new insights into how DT&E impacts major defense acquisition programs. The DAMs is composed of people, process, organizations, cultural, money, politics, technology, and other risks. These aspects and their complex interactions are difficult to completely capture in a simulation model. In an academic setting with restrained resources, a higher fidelity DT&E model (DTEM) was created, increasing the ERAM DT&E construct from 17 to over 80 blocks. No amount of effort will ever produce a 100% exact representation of the DAMS, but this is a known limitation of all simulation. However, the methodology utilized in this research is based on an iterative process where future efforts will identify and correct deficiencies converging to a product capable of supporting acquisition reform. DTEM captured the "essence of the system" (Banks, 2005: 14), supported previous conclusions by Wirthlin, demonstrated a new capability for estimating program DT&E schedules, and further refined acquisition reform analytics. "All models are wrong, but some are useful" (Box, 1987: 424) and this research is a prime example of how abstracted models can clarify complex processes.

Appendix A: Example Discussion Topics

T&E Research Discussion Topics

The focus of this research is the "as is" Air Force T&E processes from Pre-Milestone A to Milestone C. Discussion information will be compiled into an AFIT Master's thesis. Your name, official title, or any identification information will not be used in order to encourage honest responses to the questions and promote discussion. If you would like to have your name included in this research effort, please let me know.

Background Questions

1. What acquisition jobs have you held?
2. What were the ACAT levels of the programs you were involved with?
3. What T&E activities or reviews have you been involved with?

General T&E Questions

4. What are the major T&E activities in acquisitions?
5. What are the major T&E decisions/reviews?
6. What are the critical T&E documents?
7. What non T&E activities or decisions have large impacts on T&E activities or decisions?
8. Are there T&E activities where schedule delays are expected to occur?
    a. If so, why are schedule delays expected to occur here?

T&E Model Specific Questions

Instructions: Accompanying this document is a Visio file containing the current T&E process model. The model is constructed of two types of modeling concepts: activities and decisions. Activities are displayed as rectangles in the flowchart and decisions as diamonds. As you review the model, please consider the following questions:

9. Are the processes in the correct order? Take into account whether the sequence is correct as well as whether the process can occur in parallel or series with respect to other processes.
    a. If not, describe the correct order?
10. Are there any T&E decisions/ activities which may have large impacts on a program's schedule not represented in the model?
    a. If so, describe the activity/ decision and its placement in the model.
11. Are there any areas of the model that can be simplified because they do not significantly impact a
12. Are there any processes in the model that need to be modeled at a lower level fidelity because the lower level activity may have a large impact on a program's schedule?
    a. If so, identify the lower level process and why it can have such a large impact on schedule.
13. Look at each activity. Does the time required to complete the activity or decision probability change depending on the program's ACAT level?
    a. If so, acknowledge this by inputting three triangular distributions next to the appropriate ACAT level in the SME Data Input Excel Sheet.

b. If the process time is the same regardless of ACAT level, input only one distribution and put an "X" in the other two ACAT boxes in the SME Data Input Excel Sheet.

Additional Questions

14. What T&E activities or decisions could you strongly influence?
15. What T&E activities or decisions did you have little influence over?
16. What T&E Phase processes would you concentrate acquisition reform efforts with the goal of addressing schedule/delay challenges?
17. Are there any questions I have not asked that you think I should?
18. Is there anyone specific that you recommend I interview?

Appendix B: Final Conceptual Model

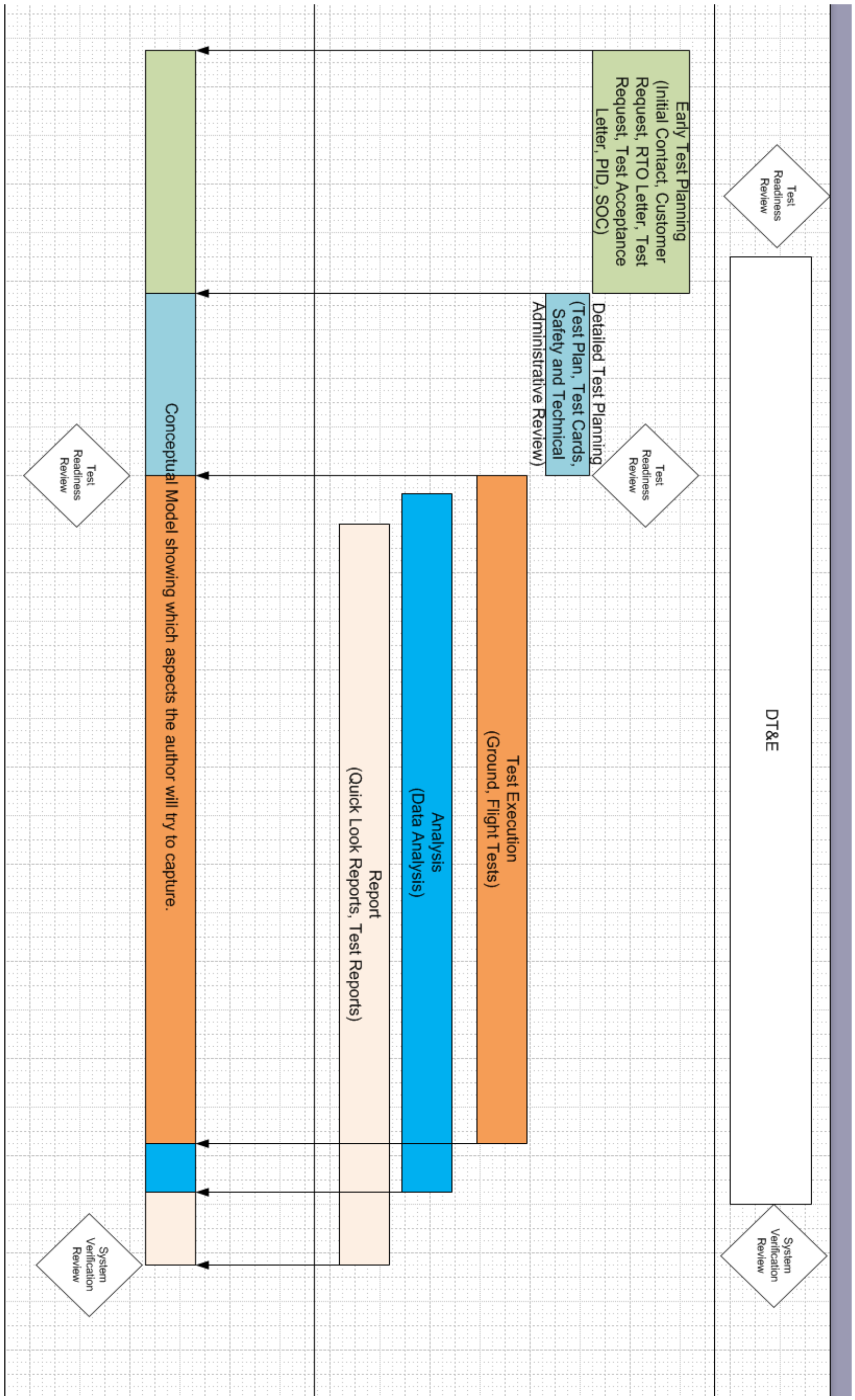

Figure 27: DTEM Conceptual Model

Appendix C: Configuration Control Document

Enterprise Requirements Acquisition Model

Configuration Management Worksheet

This form provides a listing of the development and the changes done on the ERAM Simulation Model. Use the table below to provide the simulation software used (Arena or ExtendSim), the new version number, the name of the author and corresponding organization, the date of revision and the description and purpose of changes.

Simulation Software	Source Version Number	New Version Number	Implemented By	Org	Date	Description of Change	Purpose of Change
Arena	1.0	3.0	Sutherlin	United States Air Force Institute of Technology	03/27/14	- Integrated DTEM model into ERAM 1.0 replacing the following blocks: - Test Readiness Review - Check TRR looping condition -Determine TRR delay -TRR Delay PreC -Determine Cost and schedule penalties for TRR Delays -Developmental system testing and Live Fire test and Operational Assessment testing -Make Trades? Check looping condition -Determine trades delay -Trades Delay PreC Determine cost and schedule penalties for trades delays -Combined Testing -Assign Set close to end SDD contract condition -System Verification Review Set SVR rework -SVR rework and delay -Set SVR delay cost and schedule penalties -	Improved fidelity of ERAM 1.0 DT&E activities to enable investigation of DT&E delay factors

Appendix D: Acronym List

ACAT	Acquisition Category
ADDM	Acquisition Document Development Model
AF	Air Force
AFSPC/A5	Air Force Space Command's Directorate of Requirements
APM	Acquisition Process Model
DAE	Defense Acquisition Executive
DAMS	Defense Acquisition Management System
df	degrees of freedom
DoD	Department of Defense
DT&E	Developmental Test and Evaluation
DTEM	Developmental Test and Evaluation Model
ERAM	Enterprise Requirements and Acquisitions Model
FOUO	For Official Use Only
GAO	Government Accountability Office
JCIDS	Joint Capabilities Integrations and Development System
MRTFB	Major Range and Test Facility Base
MS	Milestone
PEO	Program Executive Officer
PM	Program Manager
PPBE	Planning, Programming, Budget, and Execution
RAMP	Requirements and Acquisitions Management Plan
RTO	Responsible Test Organization
SAE	Service Acquisition Executive
SMART	Systems Metric and Reporting Tool
SME	Subject Matter Expert
SVR	Systems Verification Review
TRR	Test Readiness Review
US	United States

Appendix E: DTEM Construct and Input Parameters

The following pages will step through the DTEM and explain in detail the various blocks, distributions, and model logic. DTEM was created as a separate model with the intent to integrate it into the ERAM 1.0. In order to accomplish this, key interface blocks and variables in ERAM 1.0 are present in DTEM which have no impact on the model if run separately from ERAM but were purposefully retained to support integration efforts. DTEM may be simulated as a stand-alone model or it may be incorporated into ERAM 1.0 with an adjustment to the “Assign ACAT Level and Number of Required Test Missions” block which will be discussed later. Unless stated otherwise, the inputs for the decision blocks will be presented as the percent true. The process time triangular distributions will be expressed in the order of minimum, mean, and maximum value. The model is divided into zones in order to provide a readable figure of the model.

Zone 1 (in Figure 28) displays the initial phase of the DTEM. The first activity block is the “Delay to TRR” block which has a time distribution of 0, 14, and 180. This block represents the delay period before a program meets the TRR. The block inputs were provided by SMEs.

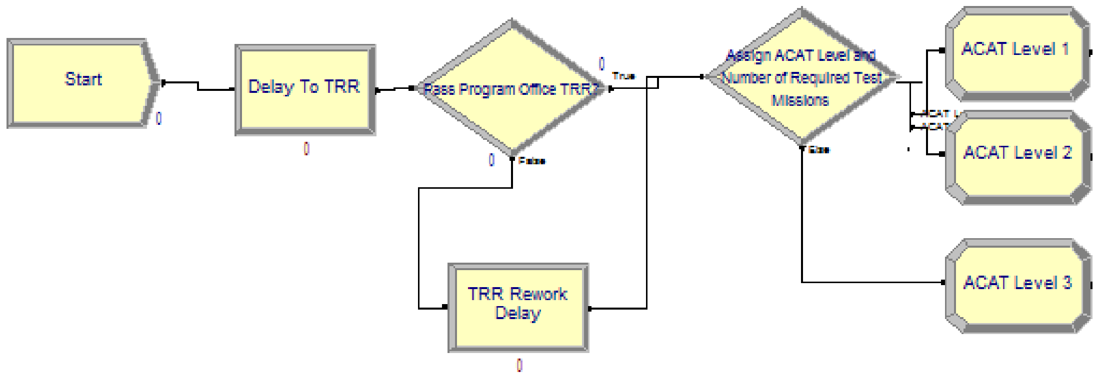

Figure 28: Zone 1

The next activity is the “Pass Program Office TRR” (Figure 28) which has a probability of %90. The system level TRR evaluates a program’s preparedness for testing. If the program fails the review, it will proceed to the “TRR Rework Delay” block which has a triangular distribution of 17.5, 42.5, 70. This block represents the amount of time required for the program office to address issues identified during the TRR that caused the review failure. The input parameters were provided by SMEs.

The “Assign ACAT Level and Number of Required Test Missions” block (Figure 28) randomly selects the program ACAT level. The probability of ACAT selection is 24 % for ACAT I, 14% ACAT II, and 62% ACAT III. These probabilities are historical data collected by Wirthlin (2009; 127).

The “Assign ACAT Level and Number of Required Test Missions” decision block will direct the program to one of the three assignment blocks: “ACAT Level 1”, “ACAT Level 2”, or “ACAT Level 3” (refer to Figure 28). Each block contains an ACAT specific distribution which randomly assigns the baseline number of required test missions needed to accomplish DT&E. The variable “Total Number of Missions

Required" is utilized to hold the program in DT&E until the total number of test missions required is achieved. The distribution were constructed by SMEs and are 88, 175, 385 for ACAT I, 36, 58, 117 for ACAT II, and 25, 41, 93 for ACAT III. It is important to note interesting phenomena occurred when asking SMEs to estimate these distributions. The question asked was, "For an ACAT III program, what is the min, average, and maximum number of test missions required to successfully complete DT&E?" Anticipating that the answers would vary, this same question was asked to the same SME on different occasions. Different answers were received. For example, the same SME provided three estimates during three different discussions (approximately one week apart) for the minimum required test missions for an ACAT III program as 1, 30, and 75. In addition, because this question was referring to the number of test missions completed at the end of a program, the SMEs were taking into account the test mission growth due to cancellations, aborts, test mission effectiveness, and other factors which impacted the number of test missions required. It was necessary to reduce the distribution inputs by the test mission growth factor the SME was taking into account. SMEs identified that on average, a program would experience a 30% growth based on the original estimate. The final values populating the "Total Number of Test Missions Required" are an average of SME inputs after subtracting 30% for test mission growth.

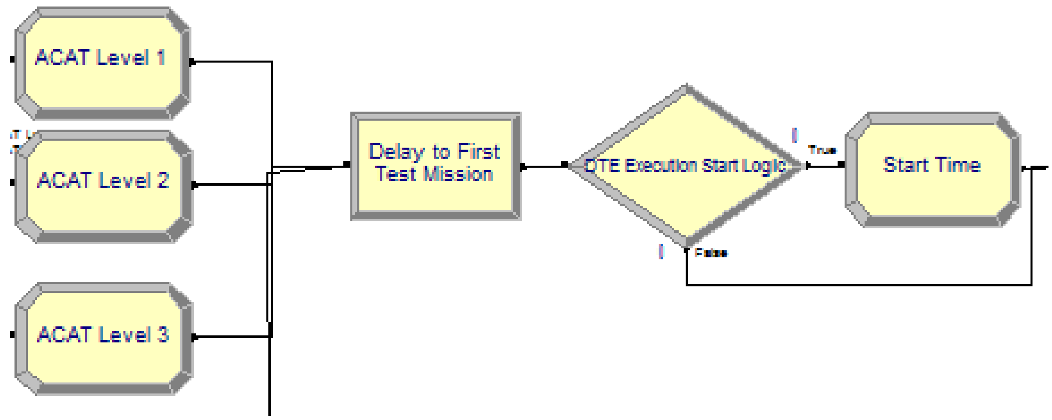

Figure 29: Zone 2

The "Delay to First Test Mission" (refer to Figure 29) is the delay a program experiences after passing the TRR to execution of the first test mission. The block has a distribution of 1, 30, 365 and was provided by SMEs. This abstract block represents several potential activities and delays for a program. SMEs discussed that for large programs, once the test item is delivered to the RTO there may be test item issues prohibiting test mission execution. These issues must be addressed before the test item may be operated. Other delays captured by this block are the RTO technical and safety reviews. The reviews last for several hours and occur at weekly intervals. Exceptions are made for higher priority programs or special circumstances. A program with a high quality test item which accomplished the RTO technical reviews in parallel with the TRR could potentially execute the first test mission one day after a successful TRR. On the opposite spectrum, a poor quality test item may take up to a year to correct test item deficiencies before the item is capable of test mission execution.

For large programs, the RTO technical and safety reviews occur regularly for each stage of testing. These phases of testing occur simultaneously depending upon priority, technical risk, safety, and other criteria. This method of test execution allows for improved control of program schedule when test deficiencies are discovered. Each stage of testing will be divided into focus areas containing similar testing requirements able to be executed during a single mission. For example, an aircraft weapon system's test plan may include several stages of high and low speed flight test. After completion of several high speed flight test stages and a test item issue is discovered, it may be reasonable to execute other low speed or ground test missions in order to minimize schedule delay while a fix is implemented for the high speed issue. The RTO will attempt to execute these reviews in parallel with testing of other phases in order to minimize program schedule delay. Thus, only the first RTO review is accounted for in the model because the following reviews occur in parallel with testing and are already accounted for. This model logic was supported by SMEs.

The "DTE Execution Start Time Logic" (refer to Figure 29) routes programs which fail the "Pass System Verification Review" block (discussed later) around the "DTE Start Time" block. This keeps the model entity from resetting the "DTE Start Time" variable set in the proceeding "Start Time" assign block.

The model executes DT&E test missions based on a projected number of test days executed in a single week (refer to Figure 30). The block "Start Week and Assign Number of Days Attempt to Execute Test Missions for 1 Week" randomly selects the number of test missions the RTO will attempt to execute in a single week. This decision directs the entity towards one of the next five assign blocks based on the probability that one (1%), two (90%), three (8%), four (0.5%), or five (0.5%) days of testing will be executed in one week. SMEs provided the probabilities and indicated that programs will plan to execute at least one test missions every week eliminating the possibility of attempting zero. In addition, the probability of executing six or more times a week is not practical due to manning requirements, work load, data analysis time, and other factors for that RTO and not included in the model. This aspect of the model is dependent upon the RTO resources available and could be tailored to a specific organization.

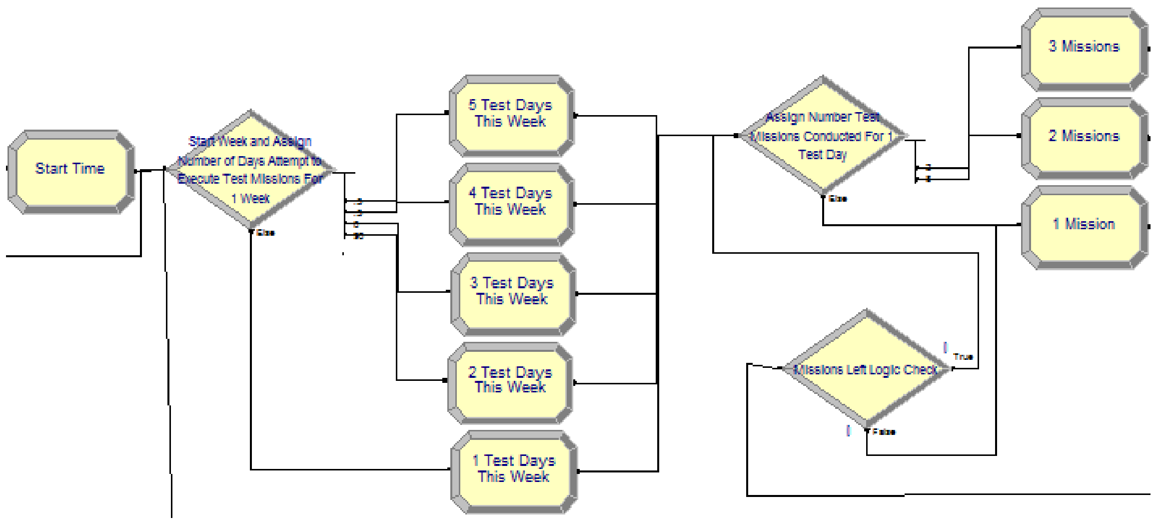

Figure 30: Zone 3

After assigning the number of test days executed in a week, the delay due to non test days for that week is calculated by the “Days Not Testing Delay” shown in Figure 30. For example, if a program executed two days of testing in one week, five days of non-test days occurred, and the program experienced seven days of total delay.

The “Assign Number Test Missions Conducted For 1 Test Day” (refer to Figure 30) randomly selects the number of test missions the RTO will attempt to execute on a single day. The probability of executing one test mission is 90%, two is 8%, and three is 2%. These values were constructed from SME input. These probabilities are program/

RTO resource dependent and representative of aircraft ground/ flight test missions where there is only one test aircraft.

Once the number of test missions for a single day of test is randomly selected, the entity will pass through one of the three assign blocks labeled “3 Missions” or “2 Missions” or “1 Mission” as shown in Figure 30. In this block the variable “Missions Per Day” will track how many test missions are executed in one day. If three test missions per day is selected, the entity will progress to the “Create 3 Missions” block and three entities are created representing three test missions. From this point, the model logic is easier to understand if the program flowing through the model is viewed as a test mission entity. Each test mission entity will pass independently through DTEM until the “Combine 1 Days Worth of Testing” block discussed later.

The test mission entity will then progress through test mission cancel blocks as shown in Figure 31. Each block represents the probability that a test mission is canceled the day of test mission execution but before test mission execution begins. If a test mission cancelation occurs, the mission will not contribute towards the total number of test missions required to complete DT&E. The penalty for a cancelation depends on the cancel factor which his discussed in the next paragraph. The test mission cancelation data are based on FOUO historical data and not presented in the research paper.

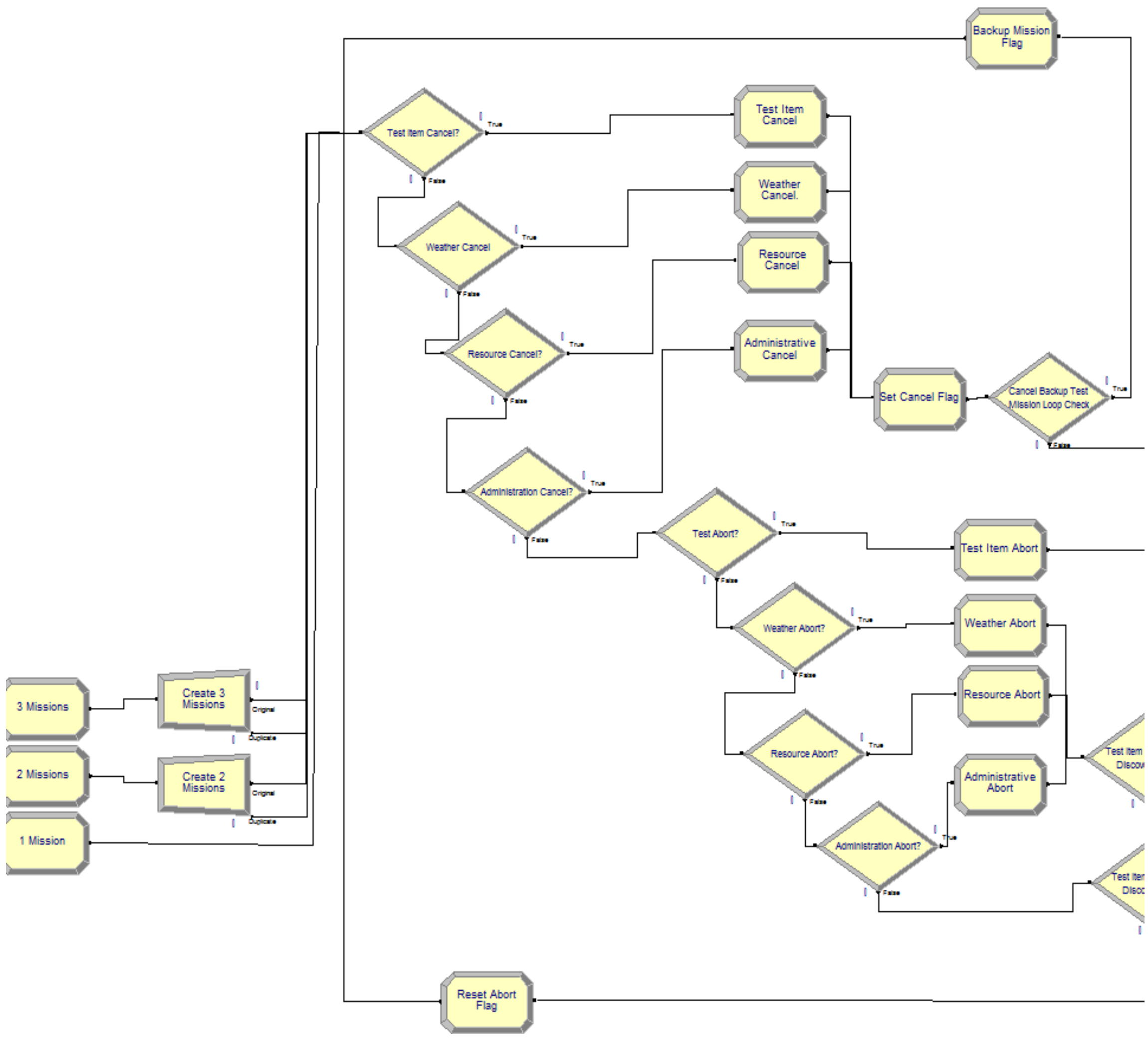

Figure 31: Zone 4

The "Test Item Cancel" factor represents problems due to poor quality test articles. If this block is true, the program will incur a penalty to the total number of test missions required. The penalty is based on a distribution of 0, 1, 3 test missions which will be added to the original baseline total number of test missions required variable. SMEs indicated that test item issues can generally be addressed in parallel with other testing resulting in no schedule delay. This is the reasoning for a test mission cancelation occurring but no program delay is experienced (represented by the minimum value of zero in the triangular distribution).

The “Weather Cancel” block represents the probability of a test mission cancellation due to weather. This delay factor does not result in a penalty.

The “Resource Cancel” refers to the cancelation of a test mission attributed to non-availability of RTO test resources. These resources may include test aircraft, test personnel, test ranges, ground instrumentation, and other test infrastructure. Programs are assigned a priority number which is one method utilized to decide which programs receive resource support. In DTEM, there is no penalty associated with a successful result in this block.

The “Administrative Cancel” largely represents the concept of scheduling primary and secondary test missions. For every test mission an RTO plans to execute, a backup mission is also scheduled as a risk mitigation technique in case the primary mission is canceled or less effective than required. If the primary mission is a success, the secondary mission is purposefully canceled by the RTO. The historical data indicated that the purposeful cancelation of backup test missions by the RTO represents an overwhelming majority of this block. However, other minor aspects accounted for include: the possibility of cancelation by senior RTO leadership due to observed safety issues, unanticipated support of civilian or military events, or other instances where RTO leadership cancels a test mission. There is no penalty associated with a successful result of this block.

If a mission is canceled, the entity will progress through one of the respective four cancel assign blocks in Figure 31. These blocks are used to assign delays and for statistical analysis. After the assign block, the entity will attempt to execute a backup test mission. For each canceled test mission, a single backup test mission is attempted. The entity will pass through the "Set Cancel Flag" block (utilized for model analysis) and continue to the "Cancel Backup Test Mission Loop Check" which will direct the entity based on whether the test mission has already attempted a backup test mission. The "Backup Mission Flag" block sets the "Backup Test Mission" variable which tracks if a backup mission has previously been attempted for this particular entity. There is no schedule penalty associated with executing a backup mission due to the assumption that the backup test mission is executed the same day as the primary mission.

If a test mission is not canceled, it will proceed to the test mission abort area of the model, shown in Figure 31, which operates according to similar logic as the test mission cancelation area. A test mission abort is defined as a test mission that started test execution but did not finish the mission due to one of four abort factors. The abort factor decision blocks are populated with FOUO historical data and not presented in this report. If a mission is aborted, it will proceed to one of the "Test Item Abort," "Weather Abort," "Resource Abort," or "Administrative Abort" assign blocks which are utilized for statistical analysis and delay calculation. The "Test Item Abort" block results in a delay of 0, 1, 3 test missions if true. The other abort assign blocks do not result in a penalty. The "Set Abort Flag" assign block is utilized for model analysis.

The "Test Item Deficiency Discovered #1" block (refer to Figure 32) represents the probability a test item deficiency is discovered during a test mission. This probability was provided by SMEs and has a value of 90%. If a deficiency is discovered, the probability it results in a delay is calculated by the "Additional Test Missions Required

#1" block and has a probability of 15%. SMEs indicated that a majority of test deficiencies are addressed in parallel to other testing efforts to minimize schedule impact. If a test item deficiency is selected to cause a delay, the "Assign Test Item Issues Missions Delay #1" block will calculate the additional test missions required based on a triangular distribution of 0.25, 1, 3. These inputs are SME estimates. The logic and block values are the same for test missions that do not abort progressing through the "Test Item Deficiency Discovered #2," "Additional Test Missions Required #2," and "Assign Test Item Issues Missions Delay #2" blocks.

The "Abort Mission Effective?" block (refer to Figure 32) represents the probability that an aborted mission accomplished any test requirements before the mission abort occurred. This probability is based on FOUO historical data and not presented in the report. If the aborted test mission was effective, it will pass through the "Abort Mission Effectiveness Level" which will randomly select one of five assign blocks based on its probability of occurrence: "75% Effective" (10%), "50% Effective" (75%), and "25% Effective" (15%). This model construct was supported and estimated by SMEs. These blocks represent the reality that a test mission may be executed and test requirements accomplished before the mission aborted. Test mission effectiveness may be measured in the number of test points completed compared to the original number of points planned. For example, if a test mission was executed that planned on executing ten test points, but only five were executed, the test mission was 50% effective. Thus 0.5 effective test missions were completed and contributed towards the total number of test missions required to pass DT&E. Each test mission initially has the potential to contribute one effective test mission to the total number of test missions required to pass

DT&E. By definition, an aborted test mission did not complete all the test requirements and cannot be 100% effective.

A test mission that does not cancel or abort will also progress through model logic to calculate test mission effectiveness. In Figure 32, test mission that are not canceled or aborted will proceed through the "Mission Effectiveness Level?" block. This block operates the same as the "Abort Mission Effectiveness Level?" block but with adjusted effectiveness levels and probabilities: "100% Effective" (10%), "75% Effective" (75%), and "25% Effective" (10%). Test missions that do not cancel or abort are assumed to be greater than 0% effective.

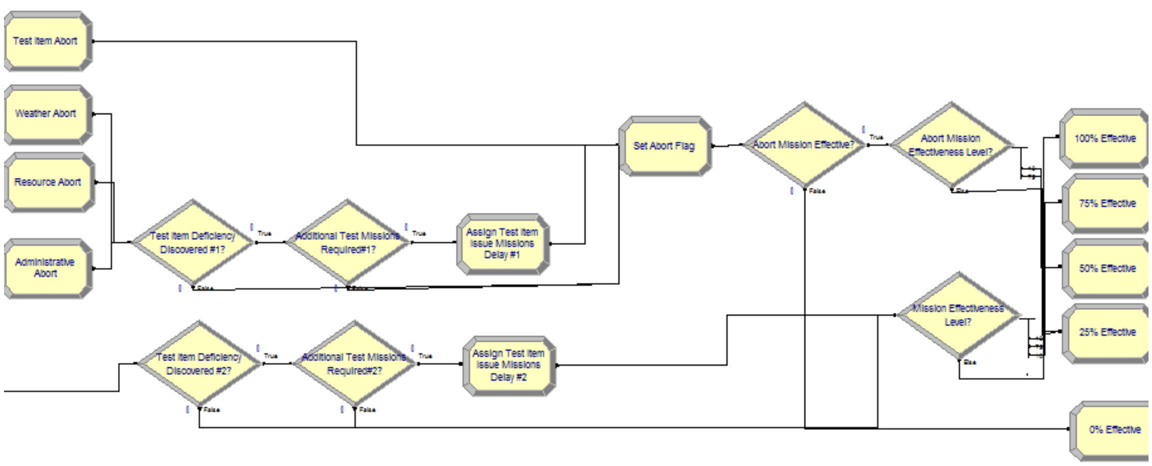

Figure 32: Zone 5

If a test mission was 75% effective, it will progress to the "75% Effective Make Trades?" block (refer to Figure 33). SMEs indicated that for test missions which were not 100% effective, the SPO may decide that the data acquired are suitable for their analysis and not execute additional test missions to collect the rest of the data. This concept was referred to as making trades. SMEs provided estimates for these blocks: "75% Effective Make Trades?" (75%), "50% Effective Make Trades?" (50%), and "25% Effective Make

Trades?" (25%). If a program is 0%effective, a make trade situation is not possible. If a trade is able to be made, the "Update Mission Effectiveness Variable" block will assigned a value of one to the test mission effectiveness variable which will contribute one count towards the total number of test missions completed. If a make trade situation is not possible, the test mission entity retains the test mission effectiveness value. The "Update Total Missions Completed Variable" block updates the total number of effective test missions completed through the "Total Number of Test Missions Completed" variable.

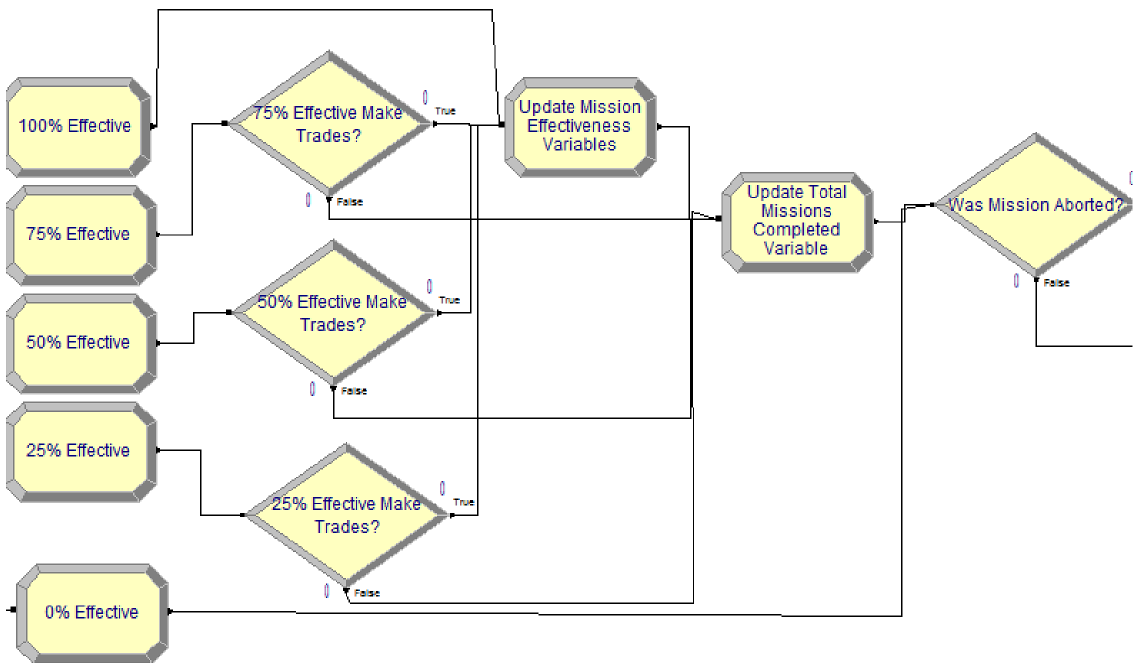

Figure 33: Zone 6

The "Was Mission Aborted" and "Backup Mission Available?" and "Backup Test Mission Effectiveness Check" (refer to Figure 34) direct the entity based on whether a test mission was aborted, less than 60% effective, and has not previously executed a backup mission. If these criteria are met, a single backup mission is attempted by looping through the model. The requirement of a test mission effectiveness level greater than 60% was provided by SMEs.

The "Reset Flags" (refer to Figure 34) assign block resets the backup test mission, cancel, and abort flags. These flags control possible entity pathways based on what events have occurred for that test mission.

If two test missions for a single day was selected in the "Assign Number Test Missions Conducted For 1 Test Day" (refer to Figure 30), each test mission will progress independently through the model (starting at the "Create 2 Test Mission" block) until the "Combine 1 Days Worth of Testing" block (refer to Figure 34). After each test mission has been canceled, aborted, or successfully completed, it will remain at this location until all test missions for that day also arrive.

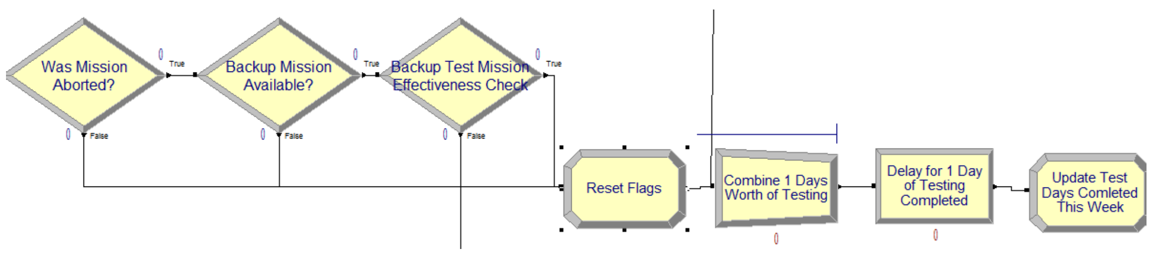

Figure 34: Zone 7

After all test missions are executed for a single day of testing, one day of program schedule delay occurs in the "Delay for 1 Day of Testing Completed" block. The block "Update test Days Completed This Week" tracks how many days of testing are completed each week and will route the entity through the model until all test missions for a week are completed.

When the total number of test missions completed equals the total number of test missions required, the block "Test missions Completed vs Required" (refer to Figure 35) will direct the program out of test execution and into the final activities in DTEM. If the total number of completed test missions is less than the total number of required test missions, the entity will proceed to the “End of 1 Week” decision block. This block compares the number of completed test days with the number of test days assigned for one week. If the number of test days for one week of testing is less than the number assigned, the entity will loop through the model passing through the “Update Variables” block (which updates the number of test days completed) and the “Missions Left Logic Check” block. Once less than one test missions is required to complete the test execution phase of DTEM (difference between the number of test missions completed and number of test missions required), the “Missions Left Logic Check” block will assign one test mission for a single day of test. For example, if 300 test missions are required and 299.5 test missions have been completed, DTEM will assign a maximum of one test missions to a single day of test. This logic prohibits executing two or three test missions to accomplish 0.5 test missions and potentially skewing the number of test missions completed. It is possible to complete more test missions than are required due to the model logic, but by a value less than one. Once the number of test days completed in a week equals the number of test days assigned for a single week, the block “End of 1 Week” will direct the entity through the “Days Not Testing Delay” and “Update test days Completed and Assigned This Week” to the “Start Week and Assign Number of Days Attempt to Execute Test Missions For 1 Week” block. The entity will loop through the model as previously discussed until the number of test missions completed is equal to the number of test missions required.

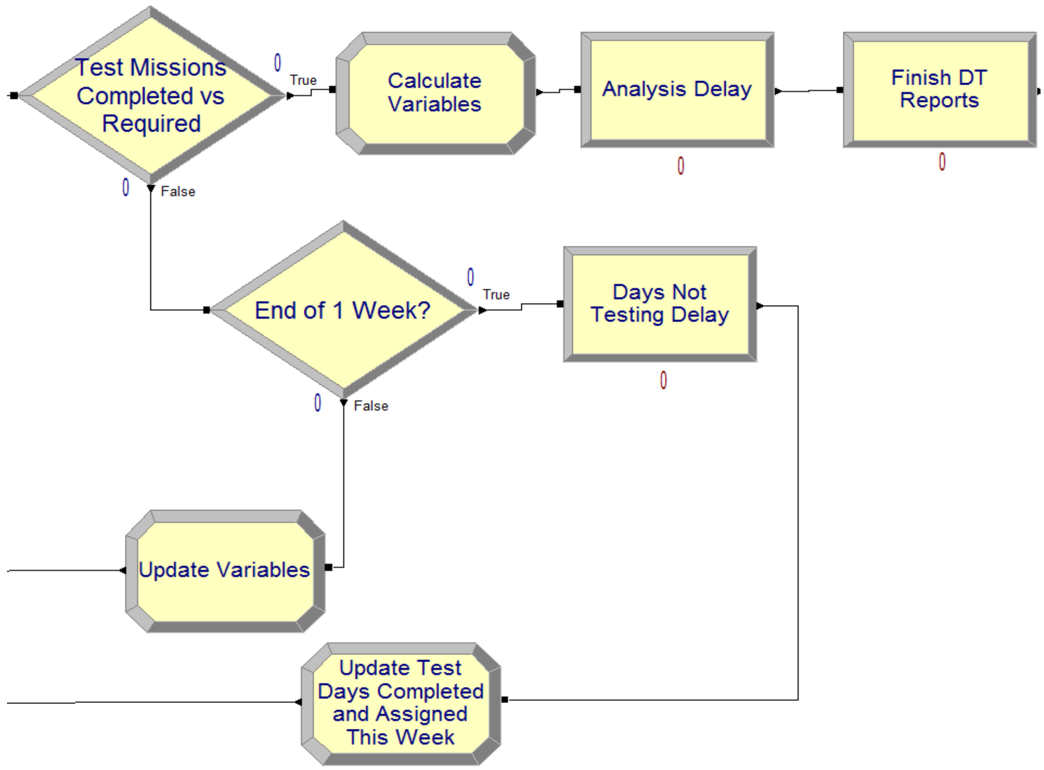

Figure 35: Zone 8

Once the total number of test missions completed equals the total number of test missions required, the “Test Missions Completed vs Required” block will direct the entity to the “Calculate Variables” block which identifies the finish time for test execution and updates other model variables.

The “Analysis Delay” represents the final stages of test mission data analysis which will occur at the end of test execution. After a test mission, collected data require analysis. SMEs indicated that data analysis will occur in parallel with other test efforts and between test missions. If a specific test mission data require analysis before execution of the next test mission for that phase of testing, the RTO will attempt analyze the data between test missions or execute other phases of testing as allowed by priority, technical risk, and safety risk in order to minimize schedule delay. Because the data analysis occurs in parallel with testing, it is included through representation in the test mission execution time. However, once the last test mission is executed, the time required to analyze the data must be accounted for which is done in the "Analysis Delay" block. This block has inputs of 1, 10.5, 90, and was provided by SMEs.

The RTO will create DT&E program reports at regular intervals which provide the SPO with program performance. These reports incorporate data analysis results. RTO SMEs indicated that it is standard policy to be allowed up to three months to compile and finish the final program report after completion of data analysis of the last test mission. This finalization of the program DT&E report is represented by the "Finish DT Reports" and has SME inputs of 14, 30, 90.

Next the entity will progress to the "Assign Set Close to end SDD contract Condition." This block is from ERAM 1.0 and included in DTEM for integration purposes. The entity then enters the "Pass System Verification Review" block and has a probability of 95% of passing the review (based on SME input). SME consensus was that the likelihood of not passing a SVR is very small because any deficiencies found in DT&E should have been fixed by this point. It not, the deficiency is usually passed to next phase of DT&E.

If a program does not pass SVR, it will progress to the "Check SVR Loop" decision block which observes the number of times a program has failed SVR. SMEs suggested the probability of failing two SVRs is highly unlikely and excluded from the model. The "Check SVR Loop" prevents programs from failing the SVR a second time. If a program has not previously failed the SVR, the entity will progress to the "Update

Total Number of Missions Required" block. This block calculates a penalty due to executing additional test missions in order to address the issues which caused the program to fail SVR. This penalty is determined by a percent of the original total number of required test missions and is added to the "Total Number of Test Missions Required" variable. The distribution is 10%, 25%, 50% of the "Initial Total Number of Test Missions Required." None of the SMEs were able to provide estimates for this distribution and the values are author estimates. After a test mission penalty is assigned, the entity will then proceed to the "Delay to First Test Mission" block previously discussed where the entity will loop through the model until completing all the required number of test missions.

If a program does not fail the SVR, the entity will proceed to the "Set DTE Finish Flag" which is used for statistics collection. The entity will then exit the model and one DTEM simulation replication is complete. DTEM records a single observation of the user requested statistics to data files which are utilized for data analysis. Because of the stochastic nature of DTEM, each replication will result in a different schedule time. Utilizing Monte Carlo techniques, thousands of programs are executed in DTEM. The ability to conduct analysis of the compilation of these data is discussed in Chapter IV.

Appendix F: Research Team Contact Information

Capt. Jason Sutherlin

jason.sutherlin@us.af.mil

Dr. John Colombi

john.colombi@afit.edu

Lt. Col. Kyle Oyama

kyle.oyama@us.af.mil

Mr. Jerry Vandewiele

jerry.vandewiele@dau.mil

Lt. Col. Joseph Wirthlin

joseph.wirthlin@us.af.mil

Appendix G: Statistical Analysis of ERAM 1.0 and 3.0 DT&E Schedule

For the purpose of this analysis section, DT&E time is defined as the time from entering the "Test Readiness Review" to passing the "System Verification Review" block in ERAM 1.0 which was instrumented with additional assign blocks and variables in order to gather the required data. ERAM 1.0 was executed and the respective number of ACAT programs which progressed though DT&E activities was utilized as the number of replications for DTEM to ensure an accurate comparison between the two models. It is important to note that DTEM represents the time programs in ERAM 3.0 will spend in DT&E. Regardless of whether the data was collected from ERAM 3.0 or DTEM, the results would be the same. However, due to the research timeline, DTEM was chosen to be run due to a drastically reduced simulation run time. In addition, a two-sample Kolmogorov-Smirnov (KS) Test was utilized for this analysis because it is sensitive to differences in sample distribution characteristics including mean, dispersion, and skewness (Siegel, 1988: 144). The two-sample KS test compares the maximum absolute difference between the cumulative distribution functions (CDFs) for each sample. If the maximum deviation is greater than the KS critical value, the null hypothesis that the two samples come from the same population is rejected. For large sample sizes (greater than 25), the critical test statistic is calculated from equation (3) where m and n are the respective sample sizes.

$$\text{KS } statistic = 1.36\sqrt{\frac{m+n}{mn}} \tag{3}$$

For each ACAT grouping, a histogram and CDF of the time spent in DT&E for each model is presented followed by a data table with the differences between the descriptive statistics of the models. Lastly, the results of the KS test and differences between model descriptive statistics are discussed. The All ACAT grouping data is analyzed first in Figures 36-37 and Table 23

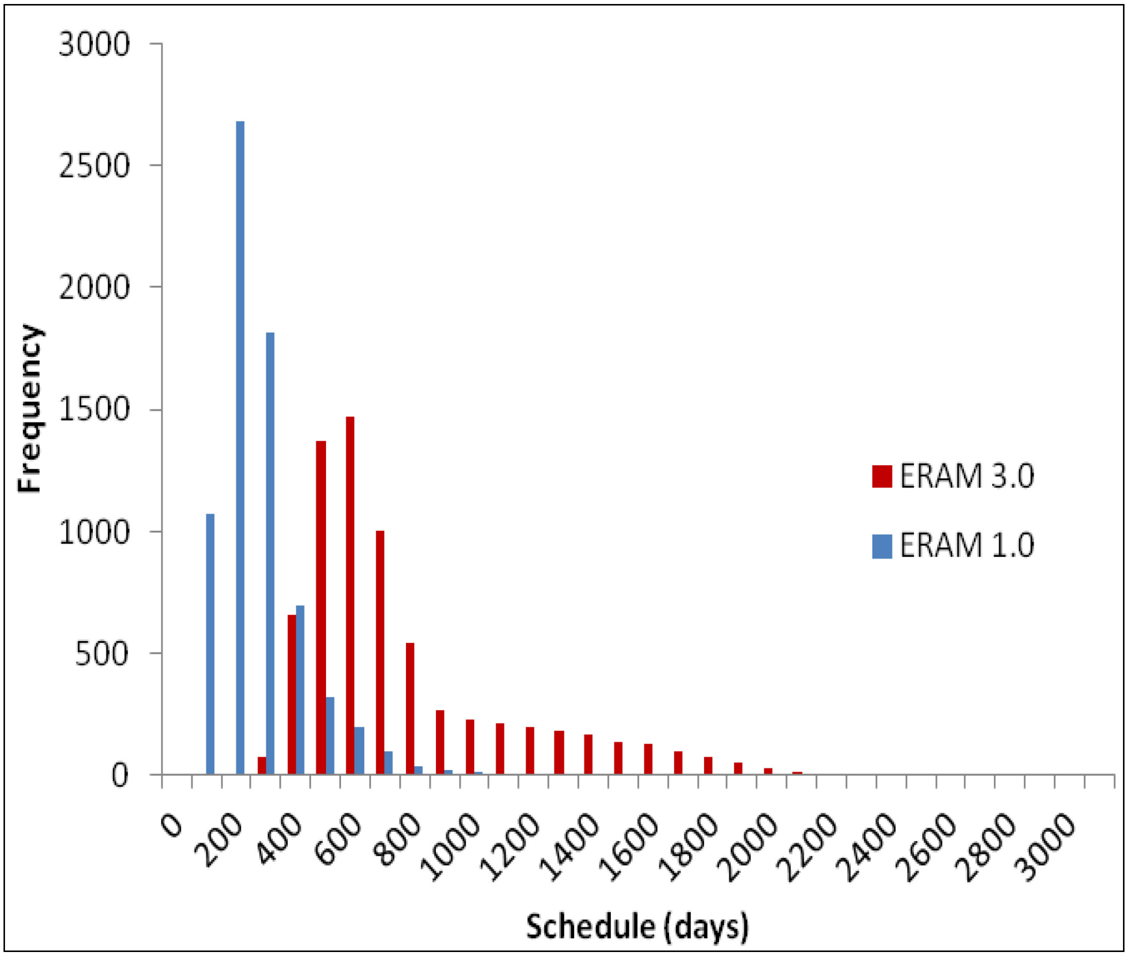

Figure 36: Histogram ERAM 1.0 and 3.0 All ACATs DT&E Time

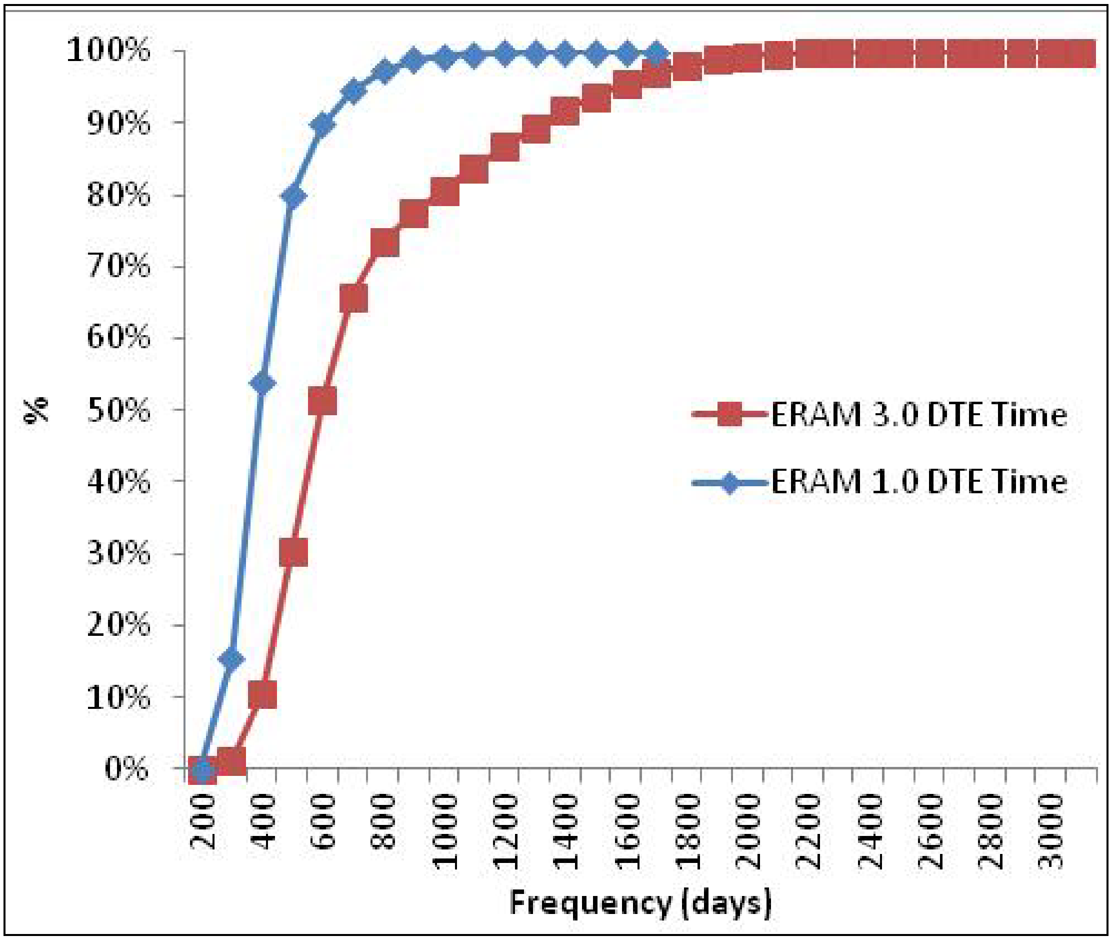

Figure 37: CDFs of ERAM 1.0 and ERAM 3.0 All ACATs DT&E Time

Table 23: ERAM 1.0 and 3.0 All ACATs DT&E Normalized Statistics

Exit at MS C	ERAM 1.0	ERAM 3.0	% Difference
Mean (days)	222.21	725.92	226.68
Standard Error	1.73	4.54	161.95
Median (days)	189.66	593.93	213.15
Standard Deviation (days)	144.59	378.74	161.95
Sample Variance	20905.12	143443.88	586.17
Kurtosis	5.93	2.38	-59.81
Skewness	1.93	1.60	-16.83
Range (days)	1403.63	2845.11	102.70
Minimum (days)	32.34	212.87	558.34
Maximum (days)	1435.97	3057.98	112.96
Program Count	6967.00	6967.00	0.00

The KS Test calculated an absolute maximum deviation b 0.048 with a critical statistic of 0.023 resulting in a rejection of the null hypothesis. The ACAT I data are presented in Figures 38-39, and Table 24.

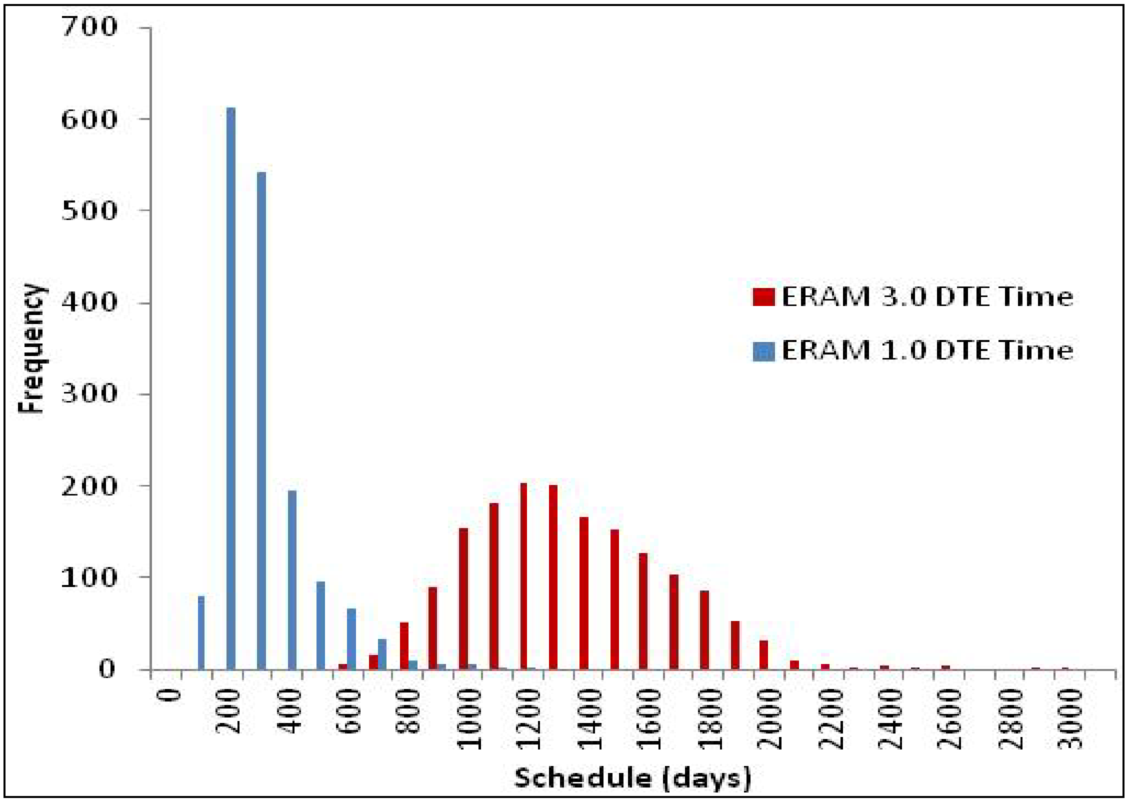

Figure 38: Histogram ERAM 1.0 and 3.0 ACAT I DT&E Time

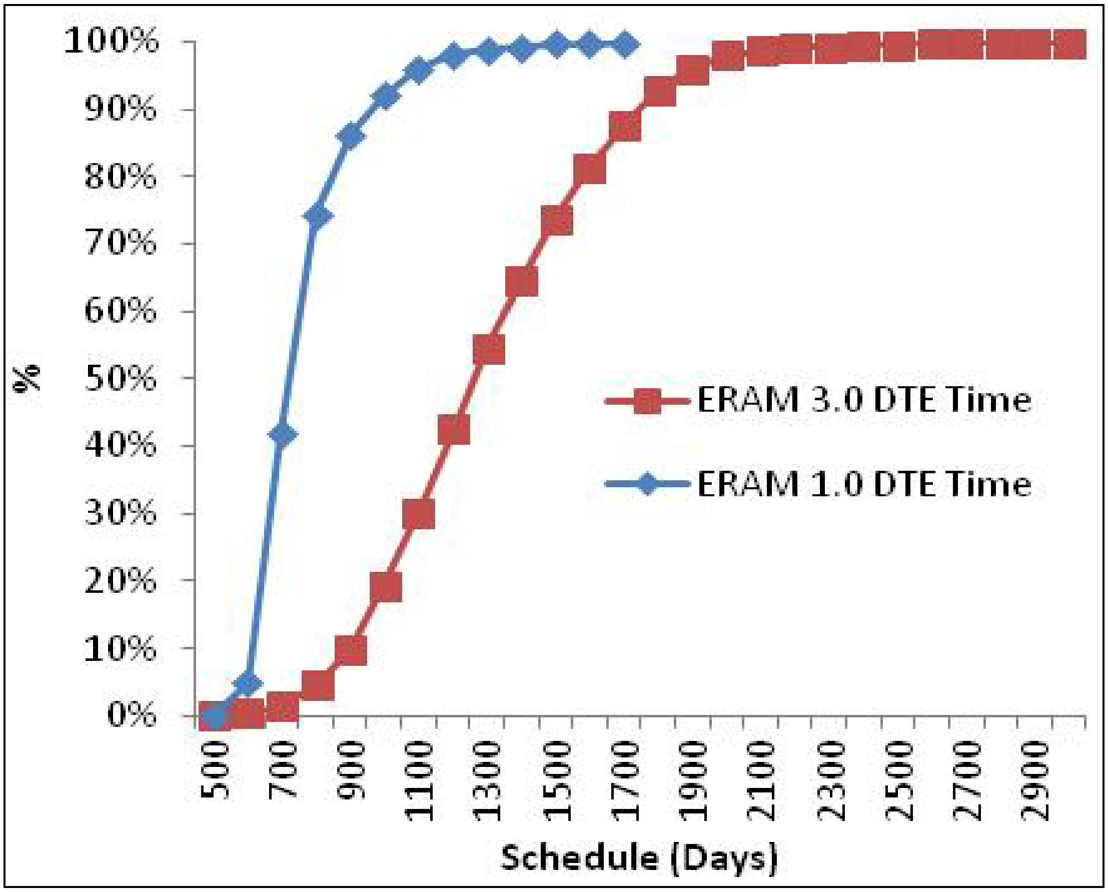

Figure 39: CDFs of ERAM 1.0 and ERAM 3.0 ACAT I DT&E Time

Table 24: ERAM 1.0 and 3.0 ACAT I DT&E Statistics

Exit at MS C	ERAM 1.0	ERAM 3.0	% Difference
Mean (days)	259.21	1295.05	399.61
Standard Error	3.68	8.14	121.28
Median (days)	223.86	1263.55	464.43
Standard Deviation (days)	149.94	331.79	121.28
Sample Variance	22482.66	110086.85	389.65
Kurtosis	5.45	0.61	-88.76
Skewness	1.96	0.57	-70.69
Range (days)	1133.55	2389.97	110.84
Minimum (days)	36.97	511.30	1282.93
Maximum (days)	1170.52	2901.27	147.86
Program Count	1660.00	1660.00	0.00

The KS test results in a maximum deviation 0.1085 with a critical KS statistic of 0.0472 and the null was rejected. The ACAT II analysis is presented in Figures 40-41, and Table 25.

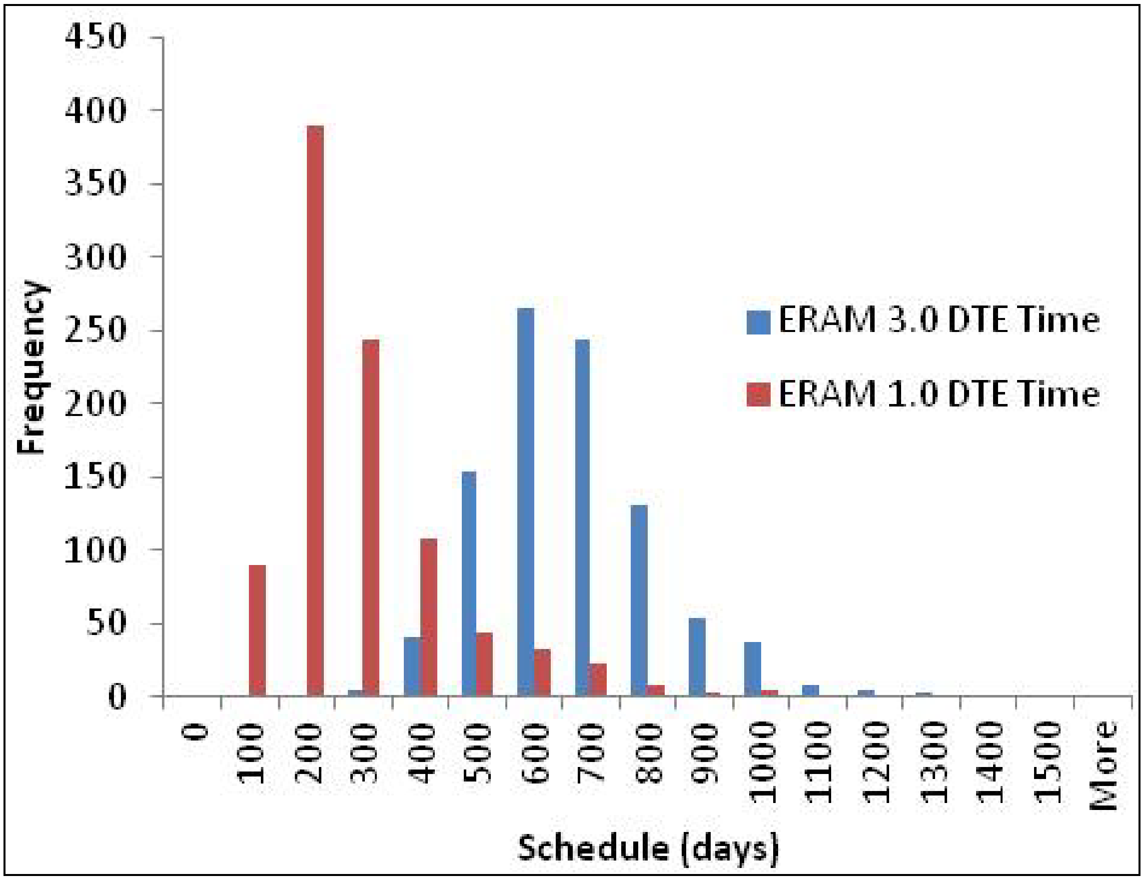

Figure 40: Histogram ERAM 1.0 and 3.0 ACAT II DT&E Time

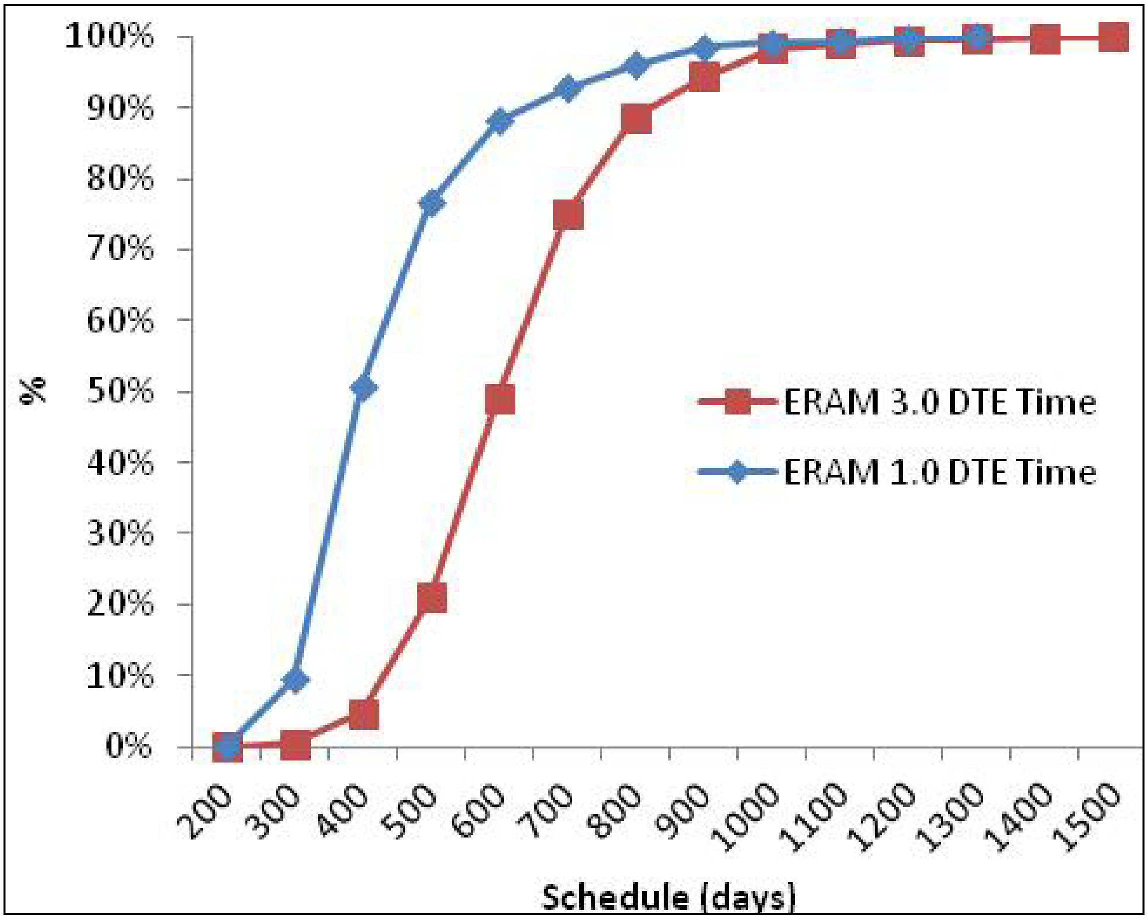

Figure 41: CDFs of ERAM 1.0 and ERAM 3.0 ACAT II DT&E Time

Table 25: ERAM 1.0 and 3.0 ACAT II DT&E Statistics

Exit at MS C	ERAM 1.0	ERAM 3.0	% Difference
Mean (days)	238.64	620.90	160.18
Standard Error	4.82	5.00	3.69
Median (days)	197.31	603.86	206.05
Standard Deviation (days)	148.24	153.72	3.69
Sample Variance	21974.91	23628.85	7.53
Kurtosis	4.21	1.75	-58.35
Skewness	1.82	0.90	-50.79
Range (days)	965.23	1166.29	20.83
Minimum (days)	47.07	258.23	448.65
Maximum (days)	1012.30	1424.52	40.72
Program Count	944.00	944.00	0.00

The null was rejected based on a calculated KS statistic of 0.0626 and a critical statistic of 0.1342. Figure 42-43 and Table 26 are the results of the ACAT III analysis.

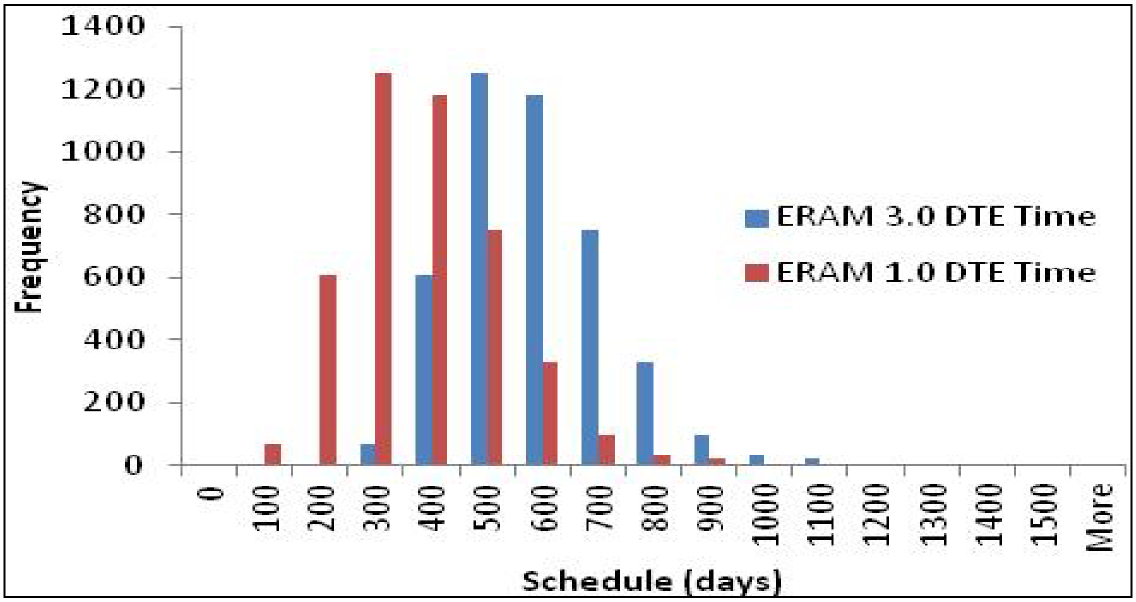

Figure 42: Histogram ERAM 1.0 and 3.0 ACAT III DT&E Time

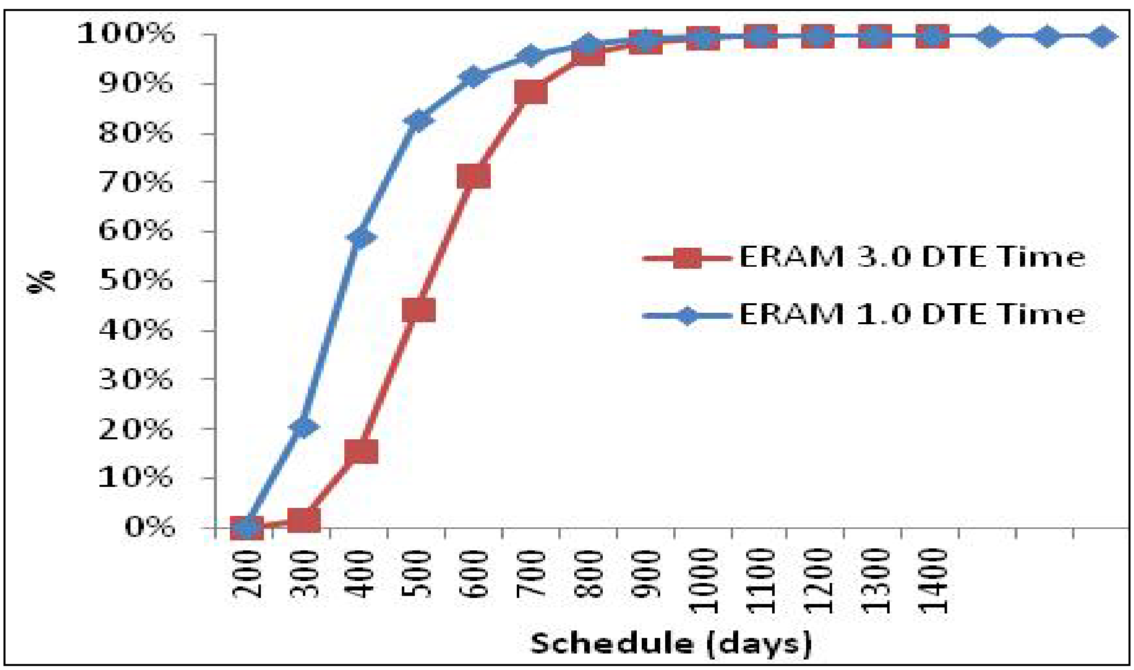

Figure 43: CDFs of ERAM 1.0 and ERAM 3.0 ACAT III DT&E Time

Table 26: ERAM 1.0 and 3.0 ACAT III DT&E Statistics

Exit at MS C	ERAM 1.0	ERAM 3.0	% Difference
Mean (days)	204.58	534.97	161.49
Standard Error	2.10	2.08	-0.81
Median (days)	174.78	518.19	196.47
Standard Deviation (days)	138.54	137.42	-0.81
Sample Variance	19193.29	18883.33	-1.61
Kurtosis	6.96	1.27	-81.78
Skewness	2.01	0.79	-60.62
Range (days)	1403.63	1100.31	-21.61
Minimum (days)	32.34	212.87	558.34
Maximum (days)	1435.97	1313.18	-8.55
Program Count	4363.00	4363.00	0.00

The absolute maximum deviation between the CDFs was 0.1442. This value was larger than the critical KS statistic of 0.0291 and the null hypothesis was rejected. A summary of the differences between the model's descriptive statistics between ERAM 1.0 and 3.0 is provided in Table 27.

Table 27: Summary of Percent Differences Between ERAM 1.0 and ERAM 3.0

Exit at MS C	All ACAT % Difference	ACAT I % Difference	ACAT II % Difference	ACAT III % Difference
Mean (days)	227	400	160	161
Standard Error	162	121	4	-1
Median (days)	213	464	206	196
Standard Deviation (days)	162	121	4	-1
Sample Variance	586	390	8	-2
Kurtosis	-60	-89	-58	-82
Skewness	-17	-71	-51	-61
Range (days)	103	111	21	-22
Minimum (days)	558	1283	449	558
Maximum (days)	113	148	41	-9

A summary of the KS Test results is presented in Table 28 indicating that the ERAM 1.0 and 3.0 are different with respect to all ACAT groupings.

Table 28: KS Test Results Summary

ACAT Group	KS Test Result
All	Reject H_o
I	Reject H_o
II	Reject H_o
III	Reject H_o

Appendix H: Additional Intervention Results Analysis

Table 29: TRR Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4243	0
Standard Error	201	21	0
Median (days)	3953	3945	0
Standard Deviation (days)	1693	1699	0
Sample Variance	2867719	2885263	1
Kurtosis	0.47	0.45	-4
Skewness	0.92	0.92	0
Range (days)	9189	9134	-1
Minimum (days)	1344	1320	-2
Maximum (days)	10534	10455	-1
Program Count	6582	6592	0

Table 30: SVR Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4233	0
Standard Error	21	21	0
Median (days)	3953	3932	-1
Standard Deviation (days)	1693	1695	0
Sample Variance	2867719	2874276	0
Kurtosis	0.47	0.44	-8
Skewness	0.92	0.91	-1
Range (days)	9189	9134	-1
Minimum (days)	1345	1320	-2
Maximum (days)	10534	10455	-1
Program Count	6582	6594	0

Table 31: RTO Test Resource Availability Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4247	0
Standard Error	21	21	0
Median (days)	3953	3952	0
Standard Deviation (days)	1693	1697	0
Sample Variance	2867719	2881313	0
Kurtosis	0.47	0.43	-9
Skewness	0.92	0.91	-1
Range (days)	9189	9005	-2
Minimum (days)	1345	1345	0
Maximum (days)	10534	10350	-2

Table 32: 110% Additional Test Missions Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4284	1
Standard Error	21	21	1
Median (days)	3953	3991	1
Standard Deviation (days)	1693	1715	1
Sample Variance	2867719	2941144	3
Kurtosis	0.47	0.49	5
Skewness	0.92	0.93	2
Range (days)	9189	9259	1
Minimum (days)	1345	1391	3
Maximum (days)	10534	10650	1
Program Count	6582	6596	0

Table 33: 115% Additional Test Missions Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4306	2
Standard Error	21	21	2
Median (days)	3953	4019	2
Standard Deviation (days)	1693	1726	2
Sample Variance	2867719	2980340	4
Kurtosis	0	1	15
Skewness	1	1	2
Range (days)	9189	9617	5
Minimum (days)	1345	1371	2
Maximum (days)	10534	10987	4
Program Count	6582	6604	0

Table 34: 45 Day Maximum Delay to Execution of First Test Mission Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4148	-2
Standard Error	21	21	0
Median (days)	3953	3855	-2
Standard Deviation (days)	1693	1689	0
Sample Variance	2867719	2854192	0
Kurtosis	0.47	0.44	-6
Skewness	0.92	0.91	0
Range (days)	9189	9148	0
Minimum (days)	1345	1343	0
Maximum (days)	10534	10491	0
Program Count	6582	6574	0

Table 35: 182.5 Day Maximum Delay to Execution of First Test Mission Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4185	-1
Standard Error	21	21	0
Median (days)	3953	3900	-1
Standard Deviation (days)	1693	1688	0
Sample Variance	2867719	2850115	-1
Kurtosis	0.47	0.47	0
Skewness	0.92	0.91	0
Range (days)	9189	9634	5
Minimum (days)	1345	1290	-4
Maximum (days)	10534	10924	4
Program Count	6582	6601	0

Table 36: 100% Decrease Test Mission Deficiencies Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4134	-2
Standard Error	21	20	-2
Median (days)	3953	3836	-3
Standard Deviation (days)	1693	1655	-2
Sample Variance	2867719	2737611	-5
Kurtosis	0.47	0.30	-36
Skewness	0.92	0.87	-5
Range (days)	9189	8915	-3
Minimum (days)	1345	1281	-5
Maximum (days)	10534	10196	-3
Program Count	6582	6574	0

Table 37: 50% Decrease Test Mission Deficiencies Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4183	-1
Standard Error	21	21	-2
Median (days)	3953	3885	-2
Standard Deviation (days)	1693	1668	-2
Sample Variance	2867719	2781795	-3
Kurtosis	0.47	0.4	-14
Skewness	0.92	0.9	-2
Range (days)	9189	9012	-2
Minimum (days)	1345	1263	-6
Maximum (days)	10534	10275	-2
Program Count	6582	6587	0

Table 38: 37.5% Decrease Test Mission Deficiencies Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	4195	-1
Standard Error	21	21	-1
Median	3953	3895	-1
Standard Deviation (days)	1693	1673	-1
Sample Variance	2867719	2800061	-2
Kurtosis	0.47	0	-14
Skewness	0.92	1	-2
Range (days)	9189	9077	-1
Minimum (days)	1345	1354	1
Maximum (days)	10534	10430	-1
Count	6582	6608	0

Table 39: Increase Test Item Quantity Intervention Results

Exit at MS C	Baseline	Intervention	% Difference
Mean (days)	4237	3654	-14
Standard Error	21	19	-10
Median (days)	3953	3318	-16
Standard Deviation (days)	1693	1519	-10
Sample Variance	2867719	2306163	-20
Kurtosis	0.47	-0.17	-135
Skewness	0.92	0.75	-18
Range (days)	9189	7812	-15
Minimum (days)	1345	1137	-15
Maximum (days)	10534	8949	-15
Program Count	6582	6604	0

Table 40: Aggregate Intervention Results

Exit at MS C	**Baseline**	**Intervention**	**% Difference**
Mean (days)	4237	3613	-15
Standard Error	21	19	-11
Median (days)	3953	3280	-17
Standard Deviation (days)	1693	1517	-10
Sample Variance	2867719	2300524	-20
Kurtosis	0.47	-0.17	-135.88
Skewness	0.92	0.75	-18.34
Range (days)	9189	8232	-10
Minimum (days)	1345	1113	-17
Maximum (days)	10534	9345	-11
Program Count	6582	6603	0

Appendix I: Research Methodology

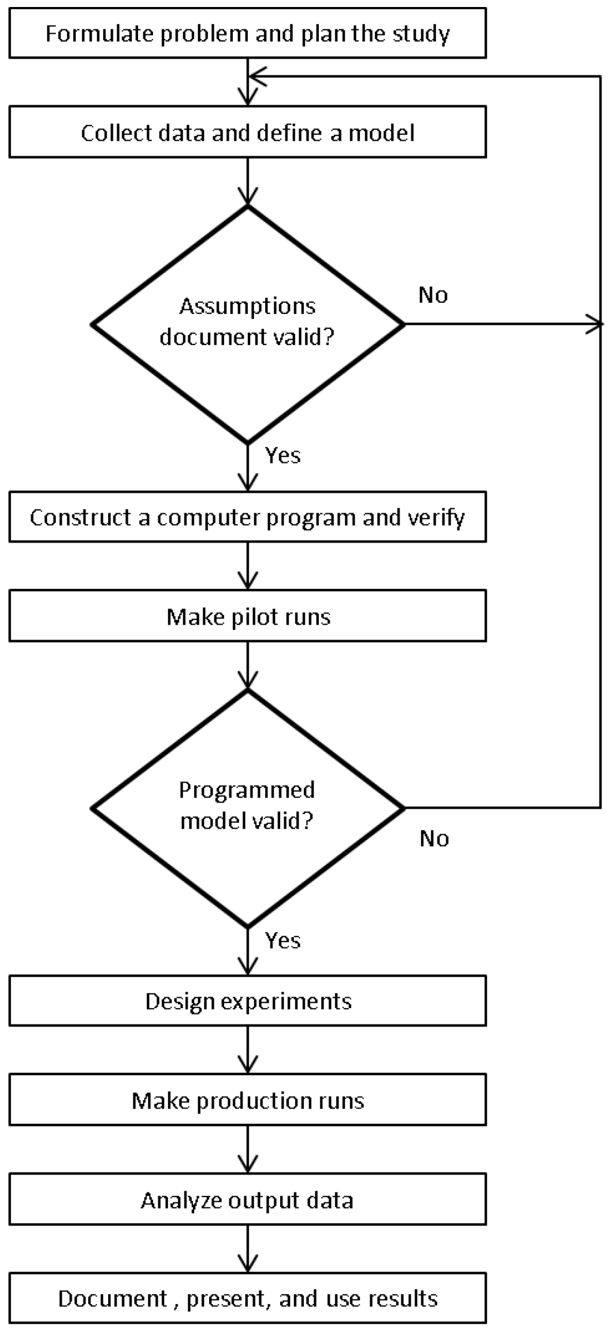

Figure 44: Methodology for a Simulation Study (Law, 2007: 67)

Bibliography

ACPO. *Acquisition Process Model*. 2011. http://acpo.crethq.com/acpo.htm. 8 Aug, 2013.

ASC/XRCC. *Acquisition Document Development and Management (ADDM) Overview*. Briefing, Wright-Patterson AFB, OH: ASC/XR, 2010.

Baldus, Lars, and others. *Modeling Space Launch Process Delays To Improve Space Vehicle Acquisition Timelines.*MS Graduate Research Paper, AFIT/ISE/ENV/13-J-01. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 2013.

Banks, Jerry, and others. *Discrete-Event System Simulation (4th ed.)*. Pearson Education, Inc, 2005.

Banks, Jerry and Randall R. Gibson. "Don't Simulate When…10 Rules for Determining when Simulation is Not Appropriate," *IIE Solutions*, (September 1997).

Box, George E. and Norman R. Draper. *Empirical Model-Building and Response Surfaces*. Wiley, 1987.

DAU. *Defense Acquisition Guidbook*. Washington: 2013.

DAU. *Integrated Defense Acquisition, Technology, and Logistics Life Cycle Management System Chart*. 8 May 2013. https://ilc.dau.mil/.

Eide, K. Col., USA, and others, *The More Things Change, Acquisition Reform Remains the Same*. 2012, ARJ.

Gilchrist, Michael. *Requirements and Acquisition Management Plan Initiative.* Briefing, Peterson AFB, CO: HQ AFSPC/A5X, 2011.

Gilmore, Michael J., *Key Issues Causing Prgram Delays in Defense* Acquisitions. 2011, ITEA.

Government Accountability Office. *A Case for Change GAO-06-257T.* Washington D.C.: Government Accountability Office, 2005.

Government Accountability Office. *Where Should Reform Aim Next? GAO-14-145T* Washington D.C.: Government Accountability Office, 2013.

Hagel, Chuck, Secretary of Defense. "Speech". Address to National Defense University. National Defense University, Washington D.C.. 3 April 2013.

Kadish, T., Lieutenant General, USA (Ret), and others, *A Report by the Assessment Panel of the Defense Acquisition Performance Assessment for the Defense Acquisistion Performance Assessment Project*. 2005, DAPA: Washington, D.C.

Law, Averill M. *Simulation Modeling & Analysis (4th ed.)*. McGraw-Hill Companies, Inc, 2007.

Gilmore, Michael and Ashton Carter. *Test and Evaluation of Defense Programs*. Office of Secretary of Defense, Washington D.C.: 3 June 2011.

Milam, Donna. ASC/XR. Personal Corresponence. March 2012.

Montgomery, Jonathon A. *Expansion of Department of Defense Enterprise Requirements and Acquisition Model.* MS Graduate Research Paper, AFIT/ISE/ENV/12-J03. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 2012.

Obama, P. B. (2009, May 22). The Weapons Systems Acquisition Reform Act. Washington, D.C.: Office of the Press Secretary.

Searle, Chad T. and David A. Leach. *Department of Defense Enterprise Requirements and Acquisition Model.* MS Graduate Research Paper, AFIT/ISE/ENV/11-J01. School of Engineering and Management, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, June 2011.

Siegel, Sidney, and others. *Nonparametric Statistics For The Behavioral Sciences(2^{nd} ed.)*. McGraw-Hill, Inc,1988.

USD (AT&L). *Interim DoD Instruction 5000.02 Operation of the Defense Acquisition System*. Washington D.C.: 2013.

Wirthlin, J. Robert. *Identifying Enterprise Leverage Points in Defense Acquisition Program Performance.* PhD dissertation. Massachusetts Institute of Technology, Massachusetts, 2009.

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 an 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)* 27-03-2014			**2. REPORT TYPE** Master's Thesis			**3. DATES COVERED** *(From – To)* Oct 2012 – Mar 2014
4. TITLE AND SUBTITLE Improving the Enterprise Requirements and Acquisition Model's Developmental Test and Evaluation Process Fidelity						**5a. CONTRACT NUMBER**
						5b. GRANT NUMBER
						5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Sutherlin, Jason W., Captain, USAF						**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/EN) 2950 Hobson Way Wright-Patterson AFB OH 45433-7765 Colombi, John M Civ USAF AETC AFIT/ENV john.colombi@afit.edu						**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENV-14-M-60
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Acquisition Excellence & Change Office Chief Process Officer Dr. Robert Pollock 1060 Air Force Pentagon Washington DC 20330 Ferguson, Brad E CIV USAF SAF/AQXC brad.ferguson@pentagon.af.mil						**10. SPONSOR/MONITOR'S ACRONYM(S)** SAF/AQXC
						11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A. Approved for Public Release; Distribution Unlimited.						
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.						
14. ABSTRACT Shrinking budgets and dynamic military conflicts have driven Department of Defense (DoD) leadership to reform how the military acquires weapon systems with the goal of decreasing program schedules and costs, while maximizing performance. With fifty years of reform, the DoD has been unable to adequately control program schedule objectives. Previous research attempted to support acquisition reform through modeling and simulation. The Enterprise Requirements and Acquisition Model (ERAM) captures a program's progression through the Defense Acquisition Management System to gain insight into significant delays that impact program schedule and probability of completion. A past unexpected result included the insignificant impact that Developmental Test and Evaluation (DT&E) had to a program's overall schedule. This research improves the fidelity of the DT&E activities through data collection, subject matter expert feedback, modeling and simulation, and Monte Carlo analysis. Interventions included modifying the probability of passing the Test Readiness Review, System Verification Review, decreasing the maximum delay to a program's first test mission, improvements in Responsible Test Organization resource availability, test item quality, and test item quantity. Several interventions significantly reduced major program schedule by 15% (21 months). The research demonstrates a methodology for quantitatively supporting acquisition reform interventions by characterizing DT&E activities and delays.						
15. SUBJECT TERMS acquisition reform, schedule performance, Enterprise Requirements and Acquisition Model, Developmental Test and Evaluation, modeling and simulation						
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** John M. Colombi, AFIT/ENV	
REPORT U	**ABSTRACT** U	**c. THIS PAGE** U	UU	126	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636 x3347 john.colombi@afit.edu	

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