3-26-2015

Examining the Return on Investment of Test and Evaluation

Nathan C. Smith

EXAMINING THE RETURN ON INVESTMENT OF TEST AND EVALUATION

THESIS
MARCH 2015

Nathan C. Smith, Captain, USAF

AFIT-ENC-MS-15-M-183

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENC-MS-15-M-183

EXAMINING THE RETURN ON INVESTMENT OF TEST AND EVALUATION

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

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 Cost Analysis

Nathan C. Smith, BS

Captain, USAF

March 2015

AFIT-ENC-MS-15-M-183

EXAMINING THE RETURN ON INVESTMENT OF TEST AND EVALUATION

Nathan C. Smith, BS

Captain, USAF

Committee Membership:

Edward D. White, PhD
Chair

Jonathan D. Ritschel, Lt Col, USAF
Member

Alfred E. Thal, Jr., PhD
Member

AFIT-ENC-MS-15-M-183

Abstract

This research examined the return on investment of Department of Defense test and evaluation. The thesis analyzed the return on investment of the cost avoidance achieved if an issue discovered late in the program had been discovered and corrected during developmental test and evaluation. The methodology utilized two case study examples from the Joint Primary Training Aircraft System to calculate the potential cost avoidance and the potential return on investment if the program had discovered and corrected the issues during developmental test and evaluation. The result of one case was a 9,260% return on investment. The other case results ranged from a -24% to a 153% return on investment. Both cases illustrated the potential return on investment but no statistically significant conclusions can be obtained from the results. Based on the literature's discussion on the value of identifying problems as early as possible and the potential return on investment from these two cases, further research is essential. This research resulted in proposing multiple recommendations to enhance the acquisition process in an attempt to preserve the long term affordability and long term national defense strategy.

I dedicate this work to my beautiful wife and children for remaining patient and supportive throughout this long thesis effort and not divorcing and leaving me.

Acknowledgments

I would like to thank all of my thesis committee members, Dr. Edward D. White, Lt. Col Jonathan D. Ritschel, and Dr. Alfred E. Thal for their help with the thesis and their help preparing me for the thesis through course instruction and coursework. I would also like to give a special thanks to William G. Gruesbeck, T-6 Test and Evaluation Manager, and Michael J. Harman, for taking a real interest in my research and providing me background information and data for the case studies utilized in the thesis.

Nathan C. Smith

Table of Contents

List of Figures

Page

EXAMINING THE RETURN ON INVESTMENT OF TEST AND EVALUATION

I. Introduction

Background

The United States may be rapidly approaching the most financially challenging time in American history. As of Jan 31, 2015, the total U.S. public debt continued its rise over $18.1 trillion, and the total U.S. unfunded liabilities reached $93.7 trillion (U.S. Debt Clock, 2015). In an attempt to limit federal spending, the President and Congress passed the Budget Control Act of 2011. This legislation will continue to place constraints on the Department of Defense (DoD) budget for the foreseeable future. Dr. Frank Kendall, the Undersecretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)), in September 2013 acknowledged, “The budget situation we’re in is pretty much unprecedented. I have not seen this kind of gridlock on Capitol Hill” (Naval Air Station Patuxent River, Maryland, 2013).

As funding diminishes, the DoD must balance risk and uncertainty while managing difficult budget decisions. The DoD is currently experiencing personnel reductions, acquisition program terminations or a reduction in an acquisition program’s production quantity, and requests for Congress to authorize base realignment and closure (BRAC). As these events occur, the DoD continues to investigate innovative ideas to save money or reduce costs through efficiencies. Principal Deputy Assistant Secretary of Defense Darlene Costello, in July 2014 stated, "There are more things out there that the warfighter would like to have that we're not even planning...so anything we can do to make our process more efficient and find some savings would be very beneficial to the whole enterprise” (Lyngaas, 2014).

Conducting early and rigorous test and evaluation (T&E) on DoD acquisition programs supports the DoD in accomplishing its objective of saving money and also reduces uncertainty. DoD program managers (PMs) must confront fiscal realities requiring them to balance risk and uncertainty when formulating budget decisions for T&E. “Ideally, the PM bases all development decisions on test events and not schedules or costs; but in the pragmatic environment of developing systems for the Warfighter, time and cost prove significant drivers in pressuring test activities” (Defense Acquisition University, 2013:Ch 9, 11). “Because these events will occur later anyway, Program Managers (PMs) frequently trade off developmental testing (‘we’ll do that in operational testing’) for near-term buying power” (Hutchison, 2013:133). These tactics often result in programs discovering problems late in the acquisition process that require costly modifications to the system. As DoD appropriations declined, the budgetary culture shifted to pursuing decisions based on what risks could be transferred to the future with the sole purpose of increasing the current budget authority.

“You must spend money to make money,” a phrase first articulated by Plautus, a Roman poet and philosopher, has since been applied throughout the business world (BrainyQuote, 2014). Pertaining to T&E, PMs should consistently scrutinize the program’s life cycle costs (LCC), not just the current budget situation, and spend (invest) money early in the T&E process to make (save) money in the future. The future savings occur by eliminating expensive modifications late in the acquisition process.

In order to convince PMs of the value of T&E investments, defensible, quantitative data and analysis must validate the claim. Currently, a study calculating the return on investment (ROI) of DoD T&E does not exist. However, this research begins the process of collecting and analyzing program data with the aim of laying the groundwork for analyzing the ROI of T&E.

Justification for Research

"In 2010, Congress expressed concern that significant problems with acquisition programs are being discovered during operational testing that: (1) should have been discovered in development testing and (2) should have been corrected prior to operational testing" (Director, Operational Test and Evaluation, 2014:13). Because of Congressional concerns, beginning in its fiscal year (FY) 2011 report, the Director, Operational Test and Evaluation (DOT&E), started reporting significant issues observed in operational testing that "in my view should have been discovered and resolved prior to the commencement of operational testing" (Director, Operational Test and Evaluation, 2011:11). The FY 2013 report expanded the classification of the issues into four types of cases illustrated in Figure 1. Between 2010 and 2013, DOT&E classified 46 DoD programs under its oversight as case 1 problems. Despite the increased scrutiny concerning these issues, the DOT&E FY 2013 report acknowledged, "Unfortunately, each year, operational testing continues to reveal performance problems for a significant number of programs that should have been discovered in developmental testing" (Director, Operational Test and Evaluation, 2014:13).

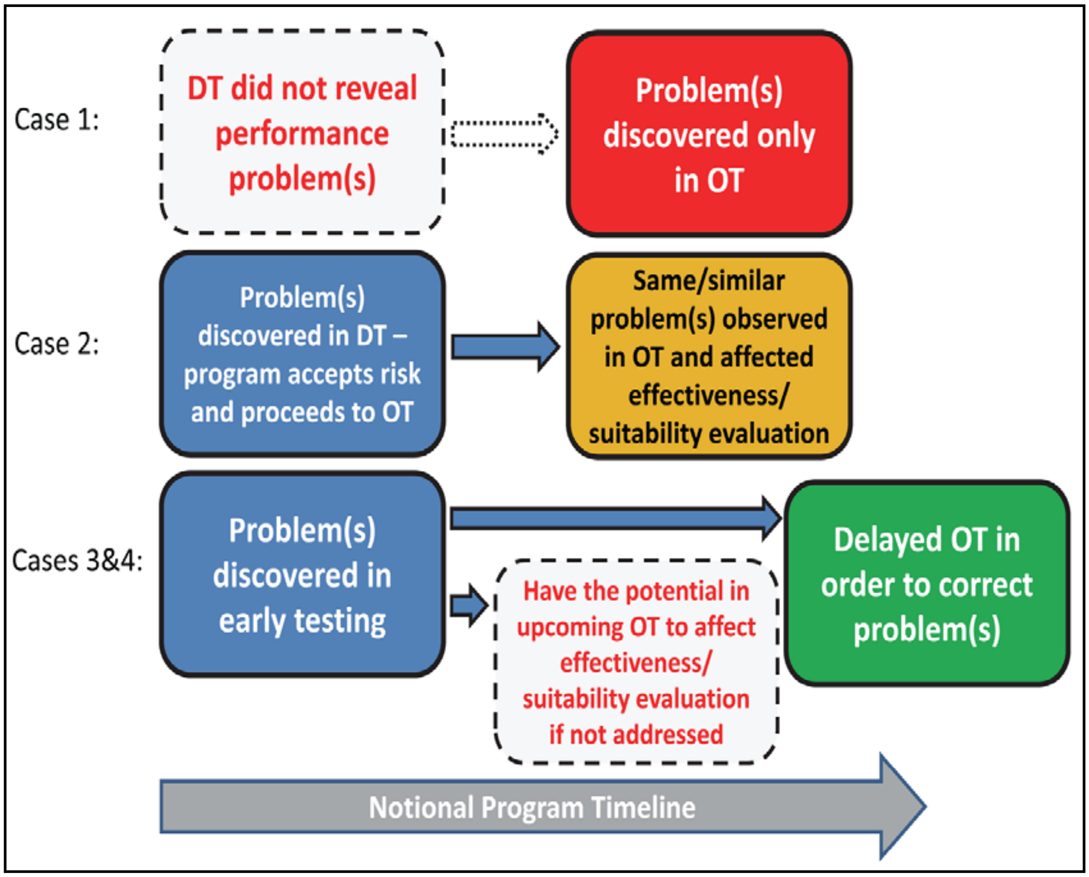

Figure 1. Problem Discovery Cases Observed in DOT&E Oversight Programs (Director, Operational Test and Evaluation, 2014:13)

This research concentrates on case 1, the worst case, and discusses the significance of the consequences of case 1 problems. According to DOT&E,

> The implication is that developmental testing (DT) was not conducted or was not adequate to uncover the problem prior to operational testing (OT). These cases illustrate that when decision makers focus too much on budget and schedule and not enough on the outcomes of testing (and the need to conduct adequate developmental testing), there is an increased likelihood of observing problems in operational testing. (Director, Operational Test and Evaluation, 2014:13)

Numerous Government Accountability Office (GAO), Defense Science Board (DSB), National Research Council (NRC), and Inspector General (IG) reports unanimously agree that issues discovered and corrected early result in less costly modifications. However, quantitative data measuring the savings from early discovery and corrective action remain absent from the literature. A 2000 DSB report concluded,

> The Task Force found that the most significant capability missing in the T&E community is the ability to measure the 'value of testing.' What do you get for what you spend? Is testing worth what we spend? The Task Force found no processes and no metrics to determine the return on investment of the Test and Evaluation process at the Department, Service Headquarters or Test Command Facilities…This Task Force suggests that a serious investigation on the cost to the Government of the failure to test properly be undertaken…The value of this process must be measured and used to justify, defend and intelligently increase funding for this vital activity. (Defense Science Board Task Force, 2000:3; 5; 27)

The recommendations from the 2000 DSB still remain unheeded today.

Issue Investigated

Three problems persist in the T&E process despite decades of studies and reports documenting the issues: late testing, inadequate testing, or, in a number of cases, proceeding to the next acquisition phase despite recommendations from test officials against it. Frequently, these problems result in costly retrofits in addition to increasing the program schedule because of the time required to correct and retest to ensure the issue does not reoccur. Individual PMs retain the decision authority on T&E activities but do not possess quantitative data on the value of T&E. Consequently, the current budget situation often influences trade-offs of T&E resources without a careful consideration of the LCC and the potentially detrimental modification costs in the future if an issue remains undiscovered until late in the program. This thesis examines the value of T&E by analyzing the ROI of the potential cost avoidance achieved if issues discovered late in a program had been discovered and corrected during developmental test and evaluation

(DT&E). The research question examines what is the ROI of the cost avoidance achieved if an issue discovered late in the program had been discovered and corrected during DT&E?

Scope and Limitations

This research intended to include data of case 1 issues, which directly relate to the inquiry from Congress, from the annual DOT&E reports covering the last three FYs. The sponsor of this research, the Scientific Test and Analysis Techniques (STAT) in Test and Evaluation Center of Excellence, utilized its connections within the T&E community to attempt to acquire the data, but unfortunately the data did not become available for this research. Therefore, the joint primary aircraft training system (JPATS) program office, which is not one of the programs under DOT&E oversight, provided the data for this research. The cases are not case 1 issues; however, the two JPATS cases demonstrate the thesis argument: one case exhibits inadequate testing and the other case illustrates the elimination of testing, both requiring costly modifications. Instead of discovering the issues during operational testing, the discovery of the issues occurred during operational use of the aircraft. Thus, the scope consists of two JPATS issues discovered during operational use of the aircraft and not during testing.

Examining a small sample size of issues from only one program creates an obvious limitation. Further, the example cases provided do not match the original intent of case 1 issues (discovered during OT&E but not during DT&E). Because the examples of this research were not discovered during testing, the argument could be made that the issues could not have been discovered during any testing. However, the program office subject matter experts (SMEs) specifically identified these issues that should have been discovered and corrected during DT&E and both case studies provide further background supporting the SME's claims.

Methodology

A case study approach examines two examples of issues discovered late in the JPATS program that, according to program office SMEs, should have been discovered during developmental testing and previously corrected. Historical background of both the JPATS program and the two issues establish the context of why these two particular cases are examined. Then, for both cases, the methodological approach for calculating the ROI is presented. First, the actual costs of correcting the problem are calculated. Next, with the assistance of SMEs, a cost estimate is developed based on the assumption that the issue was discovered and corrected beforehand, during developmental testing, and prior to the start of production. Finally, a comparison of the actual costs with the estimated costs determines the cost avoidance and ROI.

Overview of Thesis

This thesis utilizes a four-chapter format. Chapter I introduces the thesis, which includes the background, justification for the research, issues investigated, the scope and limitations of the research, an introduction to the methodology, and an overview of the thesis. Chapter II discusses the literature review, which includes an overview of T&E, incentives driving the acquisition system, historical T&E reports and studies, and prior research methodologies. Chapter III identifies the methodological framework, investigates the background of the JPATS program and the two cases studies, applies the methodology to the two examples, and reports the results. Finally, Chapter IV concludes the research by assessing the findings, providing recommendations for future research, discussing and presenting recommendations on acquisition reform, and describing the significance of the research.

II. Literature Review

The literature review includes four sections. First, an overview of T&E establishes background context by defining the purpose of T&E and discussing the establishment of the offices of the Director, Operational Test and Evaluation (DOT&E) and the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation (DASD(DT&E)). Next, the incentives driving the DoD acquisition process are examined. Then, the historical T&E reports section emphasizes the inadequacy of T&E as reported by a multitude of reports during the last 25 years and further justifies the critical need for this research. Finally, the last section examines the T&E universe of literature for methodologies previously utilized to determine the value of T&E.

Overview of T&E

The subsequent excerpt, from the 2012 DoD T&E Management Guide, depicts the DoD's purpose of T&E as well as brief explanations and differences of DT&E and OT&E.

> The fundamental purpose of T&E is to provide essential information to decision makers, verify and validate performance capabilities documented as requirements, assess attainment of technical performance parameters, and determine whether systems are operationally effective, suitable, survivable, and safe for intended use. During the early phases of development, T&E is conducted to demonstrate the feasibility of conceptual approaches, evaluate design risk, identify design alternatives, compare and analyze trade-offs, and estimate satisfaction of operational requirements. As a system undergoes design and development, the iterative process of testing moves gradually from a concentration on DT&E, which is concerned chiefly with attainment of engineering design goals and verification of technical specifications, to increasingly comprehensive OT&E, which focuses on questions of operational effectiveness, suitability, and survivability. (Department of Defense, 2012:23)

Not only does T&E provide insight and value to multiple customers and the PM, but T&E planning and results play a critical role as part of the Milestone Decision Authority (MDA) review process (Department of Defense, 2012:24).

Congress has demonstrated concerns with the T&E process for over 40 years. Beginning in 1971, Congress required the DoD to report major weapon system's OT&E results to Congress before it would commit production dollars (U.S. General Accounting Office, 1989:2). Congress continued to receive reports from the General Accounting Office (GAO), the DoD Inspector General, and other government agencies detailing the inadequacy of OT&E and decided to enact legislation establishing the office of the Director, Operational Test and Evaluation (DOT&E) (U.S. General Accounting Office, 1994a:1). DOT&E provides independent oversight to the military services, coordinates the military services' planning and execution of operational tests, independently evaluates operational test results, and reports independent and objective evaluations to DoD leadership and Congress (U.S. General Accounting Office, 1989:2).

In 2009, Congress passed the Weapon Systems Acquisition Reform Act (WSARA). The goal of WSARA was to improve DoD's procedures for acquiring major weapon systems. The legislation aimed to establish a sound program foundation by focusing on early weapon systems development activities, which include DT&E (U.S. Government Accountability Office, 2010b:1). One WSARA initiative established the Deputy Assistant Secretary of Defense for Developmental Test and Evaluation DASD(DT&E) (U.S. Government Accountability Office, 2010b:5). DASD(DT&E) acts as a principal advisor to the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)), develops DT&E policy and guidance, reviews and approves

DT&E plans and test activities, and submits an annual report to Congress discussing the year's DT&E activities (U.S. Government Accountability Office, 2010b:8).   Figure 2 depicts the current DoD T&E organizational structure.

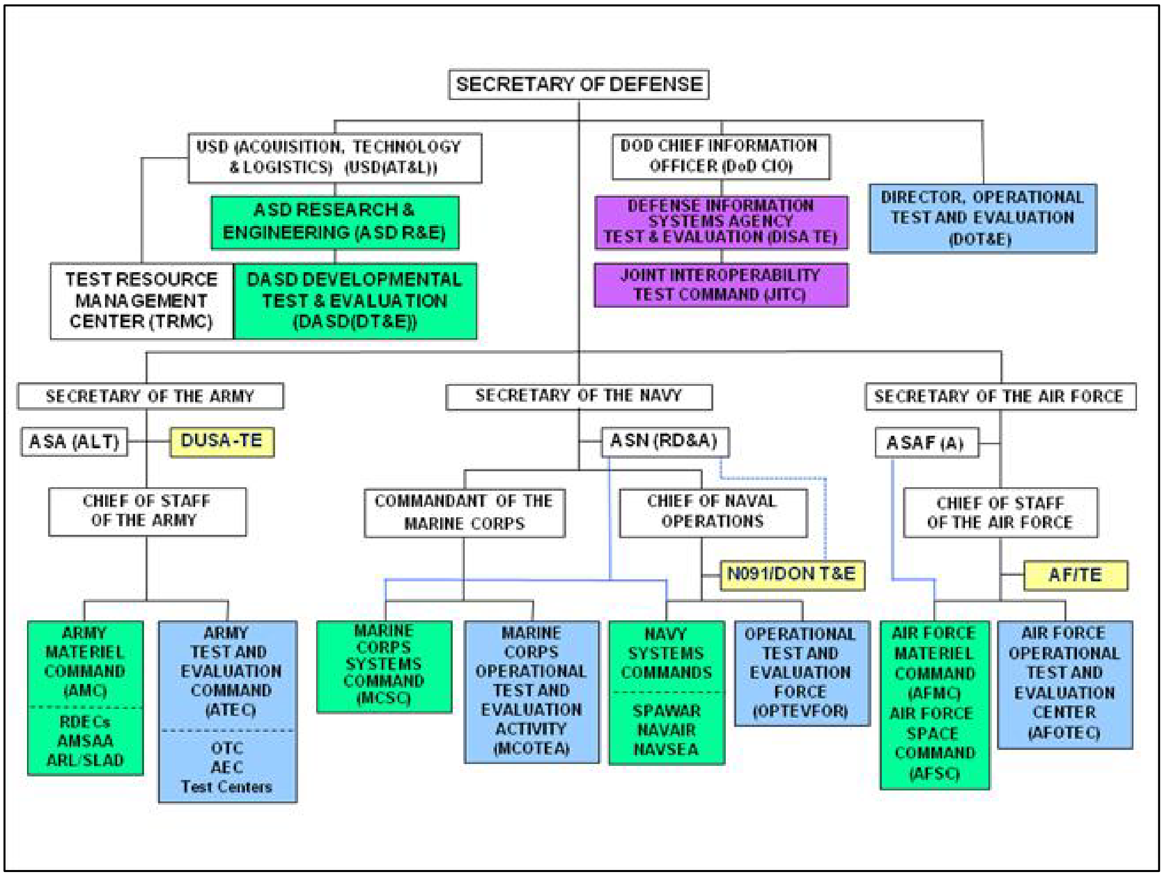

Figure 2. DoD T&E Organizational Structure (Department of Defense, 2012:10)

Examining the Incentives Driving the DoD Acquisition Process

This section examines the incentives that drive the DoD acquisition process. Public choice theory, front loading, and political engineering establish the context for analyzing DoD incentives. First, the fundamentals of public choice theory are examined.

Next, two additional concepts, front loading and political engineering, are reviewed. Finally, the acquisition process is investigated to identify examples of incentives influencing deviations from policy.

Tenets of public choice theory establish the foundation for incentives driving bureaucratic behavior. Public choice theory disputes the traditional belief that portrays bureaucrats as benevolent public servants faithfully executing the will of the people. Instead, it models bureaucratic behavior applying utility maximization and the economic model of rational behavior, which assumes individuals act in a rational, self-interested manner. Bureaucrats strive to advance in their careers and politicians pursue votes to win elections. The motivations of individuals in government are no different than the motivations of individuals in the market economy (Shughart II, 2008).

Two additional concepts, front loading and political engineering, further support the idea of how incentives influence bureaucratic choices. Both concepts were first introduced by Franklin Spinney, a former military analyst for the Pentagon. "Front loading is the practice of planting seed money for new programs while downplaying their future obligations" (Spinney, 1998). Front loading encourages overoptimistic risk, cost, and schedule assumptions to acquire support from skeptics in the Pentagon and Congress. "Political engineering is the strategy of spreading dollars, jobs, and profits to as many important congressional districts as possible. By making voters dependent on government money flows, the political engineers put the squeeze on Congress to support the front-loaded program once its true costs become apparent" (Spinney, 1998). Because a politician's constituents are geographically located, politicians are incentivized to support programs or policies in their home district even if it they are less than ideal for the national interest. The benefits become increasingly favorable when financed by national taxes, mostly from other districts (Shughart, II, 2008). For example, the F-35 Joint Strike Fighter (JSF) provides 32,500 jobs in 46 states, and 18 of the 46 states received an economic impact of over $100M. Additionally, ten other countries are economically impacted by the F-35 (Bender et al., 2014). Both concepts involve controlling money and power.

The following examples illustrate the incentives that cause deviations from policy. A myriad of studies, reviews, and panels over the last few decades repeatedly recommended not initiating a program until demonstrating maturity of the technology by ensuring the technology works as intended. The DoD has incorporated these recommendations into policy. Department of Defense Instruction (DoDI) 5000.02 states,

> Risk Reduction Decision, called Milestone A by DoD, is an investment decision to pursue specific product or design concepts, and to commit the resources required to mature technology and/or reduce any risks that must be mitigated prior to decisions committing the resources needed for development leading to production and fielding. The decision to commit resources to the development of a product for manufacturing and fielding, called Engineering and Manufacturing Development (EMD) by DoD, follows completion of any needed technology maturation and risk reduction … Formally, the development contract award authorized at DoD's Milestone B is the critical decision point in an acquisition program because it commits the organization's resources to a specific product, budget profile, choice of suppliers, contract terms, schedule, and sequence of events leading to production and fielding. (Department of Defense, 2015:7)

The dominant factor that causes deviations from policy is simple and discussed in DoD policy. Funding is the number one incentive driving the acquisition system. According to a 2005 GAO report that interviewed PMs:

> Virtually all program managers we spoke with first defined success in terms of enabling warfighters and doing so in a timely and cost-efficient manner. But when

> the point was pursued further, it became clear that the implied definition for success in DoD is attracting funds for new programs, and keeping funds for ongoing programs. (U.S. Government Accountability Office, 2005:56)

Once the competition for funds starts, the PM is pressured into overly optimistic cost, schedule, and risk assessments and to censor potentially damaging news about the program. It is better to avoid or delay difficult tests that could result in potentially damaging news which could possibly impede program progress and reduce future funding (U.S. Government Accountability Office, 2005:56).

One way to separate a program and attract funding is through differentiation. Differentiation, most often generated through advanced technology, incentivizes the acceptance of immature technology and overly optimistic performance assessments (U.S. Government Accountability Office, 2005:54). If the DoD wants to fund a particular technology to meet a capability requirement, it can attract more funding and ensure commitment to the funding in a formal acquisition program instead of through science and technology activities (U.S. Government Accountability Office, 2005:57). Although acquisition programs and science and technology endeavors both support the acquisition process, both compete for the same acquisition funding. As a result, the incentives encourage accepting immature technologies into a program to both increase and commit to the flow of money despite the increased risks. Unnecessary additional risk is accepted under the assumption the issues will eventually be solved (U.S. Government Accountability Office, 2005:58).

In addition, agencies attempt to justify larger budgets by accepting immature technologies or programs. In a sense, acquisition programs represent both revenue (larger budgets) as well as expenditures (U.S. Government Accountability Office, 2014b:8). Success can often be represented by the size of the budget controlled. In public choice literature, "Budget maximization was assumed to be the bureaucracy's goal because more agency funding translates into broader administrative discretion, more opportunities for promotion, and greater prestige for the agency's bureaucrats" (Shughart II, 2008). This results in an empire building effect whereby agencies or individuals attempt to maximize the budget and power of their empire.

The GAO, in 2005, interviewed PMs both inside and outside the DoD and wrote a report on the importance of supporting PMs to improve acquisition outcomes. "Program managers themselves believe that rather than making strategic investment decisions, DoD starts more programs than it can afford and rarely prioritizes them for funding purposes" (U.S. Government Accountability Office, 2005:5). This initiates the competition for funds at the inception of the acquisition process. DoD PMs identified the following as a few of the chief difficulties they face from the competition of funds: unstable funding, spending a considerable amount of time advocating for the program or preparing and briefing updates for oversight purposes that do not strategically help the program, and accepting additional requirements forced upon the program (U.S. Government Accountability Office, 2005). One DoD PM said, "Unstable funding results in pressure to do aggressive things in order to minimize the impact of budget cuts on schedule and performance. I believe this has been a major factor in recent…program execution problems" (U.S. Government Accountability Office, 2005:40).

Another element critical to successful programs is PM and acquisition executive tenure. The Defense Acquisition Workforce Improvement Act was enacted in 1990 and codified in Title 10, United State Code (USC) Armed Forces 1701 – 1764. Title 10, USC

1734 requires both a PM and deputy PM “be assigned to the position at least until completion of the major milestone that occurs closest in time to the date on which the person has served in the position for four years” (Cornell University Law School, n.d.). This law has been in place 25 years and rarely implemented. A 2007 GAO review discovered “39 major acquisition programs started since March 2001, the average time in development was about 37 months. The average tenure for program managers on those programs during that time was about 17.2 months” (U.S. Government Accountability Office, 2007b:8). Career progression/broadening appear to influence tenure length more than public law and DoD policy.

Historical T&E Reports (GAO/DSB/DOT&E)

This section emphasizes the chronological documentation of the inadequacy of T&E from 1989 to 2014. Several different organizations including the GAO, DSB, and DOT&E authored the reports. A summary of each report’s key topics applicable to this research follows.

A 1989 United States General Accounting Office (GAO) report entitled *Adequacy of Department of Defense Operational Test and Evaluation* reported the prepared statement of Frank C. Conahan, Assistant Comptroller General, National Security and International Affairs Division. Frank Conahan discussed the inadequacy of OT&E and the inconducive environment for thorough OT&E created by concurrent development. The conclusion from this report and over 50 GAO reports since 1970 remained that “testing has not been comprehensive, realistic or rigorous…sound and independent testing is needed if systems are to avoid costly redesign and modification after production or deployment” (U.S. General Accounting Office, 1989:1). Too often trade-offs occur between testing and possible delays in fielding. The report identifies possible causes of the trade-offs including “such factors as urgency of the requirement and the cost of building prototypes may…outweigh the need to identify and correct performance shortcomings identified through operational testing and evaluation” (U.S. General Accounting Office, 1989:1).

Frank Conahan also discussed his concern with concurrent acquisition programs achieving performance objectives and the possibility of cost growth. Five concurrent programs including Air Launch Cruise Missile, B-1B bomber, Sergeant York Air Defense Gun, F/A-18 aircraft, and the AGM-88A High Speed Antiradiation Missile failed to obtain critical OT&E results prior to the start of production despite the programs plan to attain the test results before making a production decision (U.S. General Accounting Office, 1989:7). The DoD IG also reported the C-17 and SINGCARS programs failed to complete any OT&E before the production of a substantial quantity of the systems. The GAO strongly encouraged programs to obtain OT&E results before committing to production (U.S. General Accounting Office, 1989:11).

A 1994 United States General Accounting Office (GAO) report entitled *Role of Test and Evaluation in System Acquisition Should Not Be Weakened* reported the prepared statement of Louis J. Rodrigues, Director for Systems Development and Production Issues, National Security and International Affairs Division. Louis Rodrigues discussed T&E legislation proposals including GAO’s assessment of the proposals and low rate initial production (LRIP) beginning before operational testing occurs. Mr.

Rodrigues identified several issues leading to the legislation proposals, which attempted to decrease T&E requirements and discipline.

The program office frequently regarded the start of production as the most important aspect of the program regardless of the uncertainty of whether or not the system worked as intended; consequently, the program office reduced the length of the testing process in an attempt to reduce the length of the overall acquisition process and start production as soon as possible (U.S. General Accounting Office, 1994a:6). Also, the acquisition community viewed testing as a requirement imposed on them instead of a tool to reduce technical risks and increase the chance of success for the program (U.S. General Accounting Office, 1994a:5). In particular, developers expressed frustration from delays and expenses imposed by conducting a rigorous testing program; however, the test and evaluation master plan (TEMP), which included developers' inputs, determined the testing to be accomplished (U.S. General Accounting Office, 1994a:6). In GAO's experience, programs did not become delayed because of testing but because of poor test performance, and acquisition schedules poorly forecasted the time required to resolve any issues discovered during testing (U.S. General Accounting Office, 1994a:9). Developers must demonstrate the promised capabilities and should not become frustrated by the thorough testing needed to prove the capabilities (U.S. General Accounting Office, 1994a:6).

DoD programs persisted in starting and continuing LRIP based on schedule considerations and not on the system's technical maturity; furthermore, LRIP legislation permitted and even encouraged LRIP before any operational testing occurred. Frequently, systems entering LRIP prematurely encountered issues with effectiveness and suitability in operational testing that required costly modifications. The C-17, T-45A, B1-B defensive avionics, Advanced Medium Range Air-to-Air Missile, and many electronic warfare systems all required design changes and costly modifications due to poor test results (U.S. General Accounting Office, 1994a:10).

The GAO routinely recommended less concurrent development and production and completing all possible operational testing before production to reduce the risk of discovering issues after production begins (U.S. General Accounting Office, 1994a:10). Despite these recommendations, GAO found "defense system acquisition programs continue to enter and proceed well into production before being put under serious scrutiny…there should be very few cases where there is a need to assume the additional risks inherent in a highly concurrent acquisition strategy" (U.S. General Accounting Office, 1994a:11).

"In light of the problems that we continue to find in the acquisition of defense systems, the priority given to T&E should increase, not decrease" (U.S. General Accounting Office, 1994a:1). The DoD should strengthen the "fly-before-buy" principle and ensure the demonstration of requirements before making major commitments to the program (U.S. General Accounting Office, 1994a:1). "Much more attention needs to be focused on identifying and addressing problem areas earlier…because early fixes are less expensive, easier to implement, and less disruptive to the program" (U.S. General Accounting Office, 1994a:8).

The FY 2000 Defense Authorization Act established a Defense Science Board (DSB) Task Force to review the DoD's T&E capabilities. The report discussed the value and quality of T&E within the DoD. It also emphasized T&E's importance in the acquisition process because of the essential information T&E provides decision makers (Defense Science Board Task Force, 2000:ES-1).

The first and most important topic discussed was the value of T&E. "The Task Force found that the most significant capability missing in the T&E community is the ability to measure the 'value of testing.' What do you get for what you spend? Is testing worth what we spend?" (Defense Science Board Task Force, 2000:3). The task force did not find a single process or metric within the DoD to measure the return on investment of T&E (Defense Science Board Task Force, 2000:4).

Acquisition reformers repeatedly pressured program managers to reduce the test program and program offices viewed T&E as a hurdle to progress to the next milestone (Defense Science Board Task Force, 2000:4, 5). Historically, T&E accounted for only 3-4% of the total system cost, yet attempts to reduce T&E kept reoccurring. "With the vital issues at stake, the minimal cost and the incredible value (return on test cost investment) suggests we should maximize testing to discover any weaknesses or flaws as early as possible" (Defense Science Board Task Force, 2000:3). The task force recommended creating a methodology to determine the value of testing and utilizing the methodology to "justify, defend and intelligently increase funding for this vital activity" (Defense Science Board Task Force, 2000:5).

The report also discussed the quality of T&E. Continuous pressure on programs to reduce costs without impacting the schedule caused programs to "decrease the number of test articles in the program, omit steps in the testing process, use more Modeling and Simulation (M&S) even if the M&S is not truly representative of the subject system, arrange for waivers to simplify testing and avoid trouble spots, etc." (Defense Science

Board Task Force, 2000:19). Each circumstance degraded the quality of testing. In several instances, the task force found developmental testing lacked the robustness needed to discover flaws in designs. Also, programs cut corners in the T&E process and advanced systems to the next acquisition phase prior to being ready (Defense Science Board Task Force, 2000:26).

The MV-22 program, one example cited in the report, severely cut the developmental testing program to save money and recover from schedule slips (Defense Science Board Task Force, 2000:27). An investigation into the MV-22B Osprey crash on 8 April 2000 that killed 19 marines cited testing that was severely curtailed (Defense Science Board Task Force, 2000:28). “Despite the rhetoric about early involvement of testers in programs, about testing for learning, or about discovering design and operational problems early-on, we are not allocating sufficient funds early enough to avoid costly redesigns, modifications or deferrals late in a program’s life” (Defense Science Board Task Force, 2000:27). The task force recommended a reform of the acquisition process to ensure adequate and robust T&E occur early in the acquisition process (Defense Science Board Task Force, 2000:20).

The United States General Accounting Office (GAO) published a report in 2000 entitled *A More Constructive Test Approach Is Key to Better Weapon System Outcomes* after a request by the Chairman and the Ranking Minority Member, Subcommittee on Readiness and Management Support, Senate Committee on Armed Services. The report examined “(1) how the conduct of testing and evaluation affects commercial and DoD program outcomes, (2) how best commercial testing and evaluation practices compare with DoD’s, and (3) what factors account for the differences in these practices” (U.S.

General Accounting Office, 2000:4). The following paragraphs compare and contrast the T&E process of the DoD and commercial firms as presented by the GAO report.

Discovering issues during the development process is normal; however, the implementation of T&E, the most successful tool for identifying problems, vastly differed between commercial firms and the DoD (U.S. General Accounting Office, 2000:4). One firm employed the phrase “late-cycle churn” to explain the scramble that ensued after T&E identified a major problem late in the development stage that required further money, time, and effort to correct (U.S. General Accounting Office, 2000:17). The commercial companies GAO reviewed encountered late-cycle churn in the past, but now utilize T&E to avoid late-cycle churn while creating products “in less time, with higher quality, and at a lower cost” (U.S. General Accounting Office, 2000:23). For example, Boeing employed extensive T&E and delivered the 777-200 aircraft with a 60% reduction in errors and rework (U.S. General Accounting Office, 2000:23).

In contrast, late discovery and late-cycle churn persist in DoD programs. The DoD too often waited and tested a full system, such as a missile launch or flying an aircraft, in order to discover problems, instead of previously testing subsystems to discover problems earlier in the development process. For example, multiple failures in flight tests of the Theater High Altitude Area Defense (THAAD) system could have been discovered during ground testing. Another example occurred in 1993 when the army entered into a contract to purchase cargo trailers without first testing the trailers to ensure they met requirements; 6,700 purchased truck trailers could no longer be used due to safety concerns and damage to the trucks (U.S. General Accounting Office, 2000:17).

The companies GAO reviewed applied T&E to validate a product's maturity and ensure the product worked as intended (U.S. General Accounting Office, 2000:26). Three maturity levels comprised the validation process as shown in Figure 3. "The key to minimizing surprises late in development is to reach the first two levels in such a way as to limit the burden on the third level" (U.S. General Accounting Office, 2000:26). To accomplish this, challenging tests occurred early to uncover design flaws; AT&T described the process as their "break it big early" philosophy and Boeing as "move discovery to the left" (U.S. General Accounting Office, 2000:28, 29). The successful element common to all the firms was reducing the burden during system testing in the late stages of development (U.S. General Accounting Office, 2000:26).

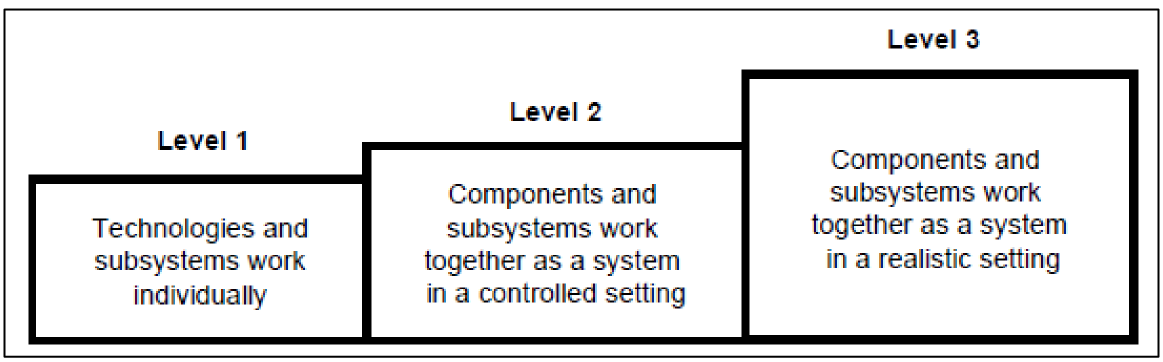

Figure 3. Product Maturity Levels Commercial Firms Seek to Validate (U.S. General Accounting Office, 2000:27)

In comparison, the DoD placed a disproportionate share of system validation on maturity level 3 and attempted to reach all three levels of maturity late in development (U.S. General Accounting Office, 2000:34). "Product knowledge was validated later, with system level testing—such as flight testing—carrying a greater burden of discovery and at a much higher cost than found in leading commercial firms" (U.S. General

Accounting Office, 2000:26). Both the THAAD and DarkStar deferred testing of the first two product maturity levels until maturity level 3. Program officials admitted taking shortcuts and expected to acquire the necessary knowledge during flight testing (U.S. General Accounting Office, 2000:34). Both programs experienced multiple flight test failures which should have been discovered during standard tests conducted before flight testing (U.S. General Accounting Office, 2000:37).

In addition to the previous differences, commercial programs and DoD programs operate under different incentives. Before taking office as the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)), Dr. Jacques Gansler identified the following differences.

> In the commercial world, the reason for testing and evaluating a new item is to determine where it will not work and to continuously improve it…By contrast, testing and evaluation in the Department of Defense has tended to be a final exam, or an audit, to see if a product works…This rather perverse use of testing causes huge cost and time increases on the defense side, since tests are postponed until the final exam and flaws are found late rather than early. (U.S. General Accounting Office, 2000:41)

A successful commercial product launch requires identifying and solving unknown factors as early as possible. Commercial managers view T&E as constructive because it identifies and eliminates the unknown factors and consider testers to be valued assets to the success of the product. Testers remain involved throughout the entire development process and their credibility influences critical decisions (U.S. General Accounting Office, 2000:41). Managers encourage and reward testers for discovering flaws as early as possible (U.S. General Accounting Office, 2000:44). Consequently, all the firms GAO reviewed made commitments to executing disciplined validation methods and providing abundant time and funding to accomplish them (U.S. General Accounting Office, 2000:41).

DoD PMs viewed T&E and testers completely opposite to commercial firms. PMs perceive T&E as less constructive and just an obstacle to overcome to acquire funding or progress to the next milestone (U.S. General Accounting Office, 2000:42). This creates an adversarial relationship between program managers and the test community which significantly limits the influence of testers on the program (U.S. General Accounting Office, 2000:48). GAO found that test officials repeatedly voiced serious concerns, but PMs fixated on cost and schedule deadlines overruled them (U.S. General Accounting Office, 2000:49). Commercial firms required testing become a centerpiece of the development process; however, schedule and funding dedicated to testing contribute only a trivial portion of the development process for the DoD (U.S. General Accounting Office, 2000:51).

Overall, the DoD T&E process was vastly inferior to commercial firms. Because of the fierce competition for funding among programs, several issues arose (U.S. General Accounting Office, 2000:41-49).

1. The necessity of estimates to fall within forecasted available funding led to overly optimistic estimates.
2. The pressure to distinguish itself from other programs encouraged the inclusion of differentiating capabilities utilizing less mature technology and encouraged PMs to accept increases in technical unknowns and risk.
3. Problems revealed during T&E could jeopardize future funding which caused PMs to delay challenging tests and limit communication of poor results.

4. Testing methods were degraded and funding was cut for other priorities so the program could maintain low advertised costs.
5. T&E became an afterthought instead of a focal point of development.
6. Few incentives existed for discovering an issue early.

The consequence of the previously mentioned issues resulted in postponing validation until late in development, which often caused late-cycle churn (U.S. General Accounting Office, 2000:42). PMs preferred results showing the minimum progress needed to continue the program instead of testing against criteria, which could possibly expose system limitations (U.S. General Accounting Office, 2000:48).

> Instead of using testing, especially in the early stages, as a vital learning mechanism and an opportunity to expand product knowledge, testing is often used as a basis for withholding funding, costly rescheduling, or threats of cancellation…distrust remains between the development and test communities, noting that some program offices have been reluctant to involve these communities early in an attempt to maintain control of the early test results. (U.S. General Accounting Office, 2000:49)

The DoD T&E process and incentives need an overhaul to correct these failures and reach the superior T&E capabilities utilized by commercial firms.

In the summer of 2007, the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD(AT&L)) established a Defense Science Board (DSB) Task Force to investigate the causes of the large proportion of programs completing IOT&E with a final evaluation of "not operationally effective and/or suitable." Of the programs completing IOT&E since 2000, almost 50% received an evaluation of "not operationally effective and/or suitable" with issues of suitability dominating and reliability failings representing the main deficiency (Defense Science Board Task Force, 2008:13). The report focused on reliability, availability, and maintainability (RAM) issues and particularly on reliability issues because they account for 50% of the root causes of suitability failures (Defense Science Board Task Force, 2008:23).

The report's findings identified several issues in the T&E process as factors for poor suitability evaluations. First, after the events of September 11, 2001, the Combatant Commanders desired new capabilities delivered quickly to deploy against adapting threats. This desire resulted in sacrificing rigorous T&E to meet the schedule demands of the commanders (Defense Science Board Task Force, 2008:15). Next, budgetary pressures influenced a reduction of the DT&E portion of the total research, development, test and evaluation (RDT&E) budget. For example, the Air Force reduced the DT&E portion of the RDT&E budget from 9.8% in 1996 to 7.3% in 2005 (Defense Science Board Task Force, 2008:19). Finally, reliability growth processes where "a system is continually tested from the beginning of development, reliability problems are uncovered, and corrective actions are taken as soon as possible" were discontinued in the mid-1990s (Defense Science Board Task Force, 2008:21).

The report also explains that Army studies indicate "almost 90% of the in-service costs are directly correlated with the reliability of the system" (Defense Science Board Task Force, 2008:22). Consequences resulting from poor reliability include reduced performance in the field and LCC increases. The V-22 program required over $1B in additional funding to solve its suitability problems. Because of the substantial sustainment costs during the life cycle of a system, reliability investments can result in a significant ROI (Defense Science Board Task Force, 2008:22).

Finally, the DSB presented an example of the ROI of reliability. A Logistics Management Institute (LMI) study on reliability, discussed more thoroughly in the next section of the chapter, concluded "an investment in total program reliability equal to twice the average production unit cost would yield an approximate 35% reduction in support costs" (Defense Science Board Task Force, 2008:23). Minimal investments in reliability will successfully impact both operational availability and the LCC of the system. "This additional increase in reliability usually requires finding failure modes through continuous testing" (Defense Science Board Task Force, 2008:22). One of the primary recommendations from the task force included improving DT&E to discover and correct suitability deficiencies early which improves the chance of success during IOT&E (Defense Science Board Task Force, 2008:13).

DOT&E submits a report to Congress annually to comply with statutory requirements. "In 2010, Congress expressed concern that significant problems with acquisition programs are being discovered during operational testing that: (1) should have been discovered in development testing and (2) should have been corrected prior to operational testing" (Director, Operational Test and Evaluation, 2014:13). Over the last three FY reports (FY11-FY13), DOT&E started reporting significant issues discovered during operational testing that "in my view should have been discovered and resolved prior to the commencement of operational testing" (Director, Operational Test and Evaluation, 2011:11). The three reports identified 46 programs (17 in 2010-2011, 17 in 2012, and 12 in 2013) with significant issues. In addition, 33 programs between FY12 and FY13 experienced over 400 cybersecurity vulnerabilities in which 90% should have been corrected earlier during system development (Director, Operational Test and Evaluation, 2014:14).

After the implementation of WSARA, DOT&E started receiving assessments of operational test readiness (AOTRs), in which the DASD(DT&E) makes recommendations on a system's readiness to enter IOT&E. Since 2009, DOT&E received six AOTRs recommending against the system continuing to IOT&E. All six programs proceeded with IOT&E despite the recommendation. Five of six (83%) of the programs performed poorly and experienced significant issues during IOT&E. "The trend is that major discrepancies are being discovered and raised to the Service leadership, but decisions to enter IOT&E are not being affected by these AOTRs" (Director, Operational Test and Evaluation, 2011:11).

The most recent GAO report on selected weapon programs was published in March 2014. The GAO has been recommending multiple knowledge-based practices since the inception of its first report assessing selected weapon programs in 2003. The following examples, from the 2014 report on selected weapon programs, examine the DoD's current activities as compared to GAO's longstanding recommendations in regards to technology demonstration and testing. The examples illustrate the continued practice of delaying testing until late in the acquisition process.

Two of the knowledge-based practices recommend to demonstrate all critical technologies in a realistic environment and to test an early integrated prototype prior to the critical design review (CDR). Three programs conducted a CDR in 2013 and none of the three programs completed either the demonstration of critical technologies or the testing of an early prototype. The Joint Light Tactical Vehicle conducted system prototype testing seven months after its CDR, the KC-46 Tanker program plans to start 18 months after its CDR, and the Warfighter Integrated Network-Tactical Increment 3 plans to start 22 months after its CDR. The report also assessed 30 other programs that held a CDR prior to 2013. Six of the 30 programs demonstrated all critical technologies prior to the CDR. Only three of the 25 non-ship programs tested an early integrated prototype with the other 21 non-ship programs starting an average of 33 months after the CDR (U.S. Government Accountability Office, 2014a:32).

Another knowledge-based practice, as well as DoD policy, recommends to demonstrate a production-representative prototype works as intended in its planned environment. One of two programs that started production in 2013 previously tested a production-representative prototype in its intended environment. Sixteen programs that held production decisions prior to 2013 were assessed and six programs actually tested a production-representative prototype prior to the start of production. Five of 14 programs with future production decisions plan to have tested a production-representative prototype prior to the production decision (U.S. Government Accountability Office, 2014a:35).

The report also evaluated the extent of concurrent DT and production among programs currently in production and programs that will start production in the next few years. Starting with programs currently in production, 15 out of 18 plan to or have already completed more than 30% of DT concurrent with production. Five out of eight programs currently executing concurrent test and production also plan to have greater than 10% of the procurement quantities under contract prior to the completion of DT. “The F-35 program in particular plans to have 530 aircraft, more than 20 percent of its total procurement quantity, under contract at a cost of approximately $57.8 billion before developmental testing is completed in 2017" (U.S. Government Accountability Office, 2014a:46-47). Of the 12 programs GAO assessed that will have a production decision in the next few years, half of them intend to conduct more than 30% of DT concurrent with production. Two of the six plan to procure more than 10% of the total procurement quantity prior to the completion of DT (U.S. Government Accountability Office, 2014a:46-47).

The JSF is a prime example of concurrent test and production and has been controversial because of its history of cost growth. The JSF, as planned, will be the most expensive acquisition program in DoD history. The JSF program, from October 2001 to August 2013, already had total program cost growth of $107.5 billion in FY 2014 dollars or a 47.8% increase in total program cost and unit cost growth of 72.5% due to a reduction in the planned procurement quantity of 14.3% (U.S. Government Accountability Office, 2014a:69).

The JSF program began development in 2001 and started production in 2007 with all three variants not expected to start flight testing until two years later and fully integrated flight testing not expected until four years later (U.S. Government Accountability Office, 2007a:89). In 2007, the DoD decreased test aircraft and flight test hours to preserve schedule and cost plans (U.S. Government Accountability Office, 2008:105). Despite flight testing only 2% complete in November 2008 and a fully integrated, capable aircraft not expected to be available for at least four years, the program decided to accelerate the production of an additional 169 aircraft between FYs 2010 and 2015 (U.S. Government Accountability Office, 2009:94). As of December

2009, only four of the planned 13 developmental aircraft had flown; flight testing was merely 3% complete and a fully integrated, capable aircraft was not expected until 2012 (U.S. Government Accountability Office, 2010a:84). The 2014 GAO reported observed several issues the JSF continues to confront including: four critical technologies are still not mature, design changes continue, developmental testing is far from complete and may drive further design and manufacturing changes in the future, and only 25% of critical manufacturing processes are mature and capable of consistent production quality (U.S. Government Accountability Office, 2014a:69-70).

The previously summarized reports from the Government Accountability Office (GAO), the Defense Science Board (DSB), and the Director, Operational Test and Evaluation (DOT&E) maintain consistent themes and language dating back to 1989. For well over 25 years, these same themes have been documented by several different agencies, yet they continue to occur. The following list highlights the critical takeaways from these reports.

1. The central theme repeatedly emphasized in every study is the DoD should maximize T&E effort and funding as early as possible to discover problems early in the program when modifications cost significantly less, are easier to implement, and cause less of a disruption to the program.
2. Multiple pressures placed upon the PM such as the urgency of the requirement, the competition for available funding, and schedule demands outweigh the need to identify and correct deficiencies as early as possible.
3. Pressures in 2.) result in multiple T&E issues including:
    a. T&E becomes an afterthought and not a focal point of development.

b. Trade-offs occur between testing and other priorities.

c. Programs cut corners weakening the T&E process.

d. PMs accept increased technical unknowns and risks from utilizing less mature technology because of the necessity to differentiate its capabilities from other programs to receive more funding.

e. Programs view the start of production as the most important aspect of the program, regardless if the system works as intended, and attempt to reduce the testing process to start production as soon as possible.

f. PMs prefer results showing the minimum progress required to move the program forward and delay challenging tests out of fear of jeopardizing future funding if testing reveals problems.

g. Test officials repeatedly voice serious concerns to leadership, but leadership overrules them.

4. Concurrent development creates an inconducive environment for thorough T&E and the GAO routinely recommends less concurrent development and production.
5. Programs identify and resolve issues during OT&E or late in the program that should have been discovered and corrected during DT&E
6. The most significant capability missing from T&E is the ability to measure the ROI of testing.

How do commercial firms apply T&E with greater success than the DoD? As previously mentioned, leading commercial companies experienced late issues in the past; however, by utilizing T&E early and effectively, the companies now experience far fewer issues and create products faster, cheaper, and of higher quality. The firms purposefully schedule difficult tests early in development to discover problems early and avoid significant issues creeping up late in product creation. Regardless of the testing tools applied, the one successful strategy common to all the leading companies includes validating products at increasing maturity levels by testing the technology, components, and subsystems individually and together before testing a complete system in a realistic environment. In contrast, the DoD, because of the variety of pressures previously mentioned, too often cancels or postpones difficult tests until late in the development when it tests the whole system together.

Methodologies Applied in Previous Research

This section examines methodologies applied in previous research and compares them with this research. First, two recent reports discussing reliability and LCC are explored because they utilize a similar methodology. Finally, several sources that address different methods of determining the value of T&E are analyzed.

Logistics Management Institute (LMI) Government Consulting published a report in 2007 entitled *Empirical Relationships between Reliability Investments and Life-Cycle Support Costs*. "Test results since 2001 show that roughly 50 percent of DoD's programs are unsuitable at the time of initial operational test and evaluation (IOT&E), because they do not achieve reliability goals" (Long et al., 2007:iii). Reliability plays a substantial role in determining LCC. DOT&E, concerned with the potential consequences of poor reliability testing, solicited LMI to "study the cost of not achieving adequate levels of operational suitability by investigating the empirical relationships between reliability investment and life-cycle support costs" (Long et al., 2007:iii).

LMI created two overarching constructs to approach the problem. The first construct stated "reliability is a function of reliability goal setting, maturity of technology, and investment in reliability effort," and the second construct explained "support cost is a function of utilization, primarily density and operational tempo (OPTEMPO); product design, for example, reliability and maintainability; and support process design, particularly repair cycle time" (Long et al., 2007:iii).

The report analyzed six case studies: Predator Unmanned Aerial Vehicle (UAV), Global Hawk UAV, MH-60S Fleet Combat Support Helicopter, CH-47F Improved Cargo Helicopter (ICH), Force XXI Battle Command, Brigade-and-Below (FBCB2) system, and a complex vehicle electronics system (Long et al., 2007:iv). For each case study, Long et al. (2007:1-2) utilized the Cost Analysis Strategy Assessment (CASA) model to estimate the life cycle support costs for reliability demonstrated early in the program and to estimate the life cycle support costs using the most current reliability information. The data from the case studies helped develop two relationships: "the relationship between investment in reliability and reliability improvement" and "the relationship between reliability improvement and support cost reduction" (Long et al., 2007:iv).

LMI results indicated reliability improvements ranging from 23.6% to 674.5% for the five fielded systems and concluded the results are likely system and technology independent (Long et al., 2007:3-1;2-32). Further, the authors reported the following ROI ratios and reductions in 20-year support costs (2003 dollars) for four fielded systems: Predator UAV ROI of 22.7:1 and support cost reduction of $887.2M or 60.6%, Global Hawk UAV ROI of 5:1 and support cost reduction of $588.6M or 23.1%, MH-60S ROI of 49:1 and support cost reduction of $319.9M or 83.2%, FBCB2 ROI of 128:1

and support cost reduction of $11,179.6M or 85.6% (Long et al., 2007). The authors stressed, "The relationship between investment in reliability and support cost reduction is almost certainly system and technology dependent…should not be generalized (Long et al., 2007:2-35).

LMI emphasized two critical findings: "reliability goals, although established and articulated in operational requirements documents, do not appear to be driving either management or engineering effort" and "under-investment in reliability may be large" (Long et al., 2007:3-1). The authors criticized the quality and lack of data and highlighted several data issues in the report. However, LMI concluded, "While recognizing the limitations flowing from a limited sample and the less-than-ideal data, the preliminary results indicate that it is possible to estimate the reduction in support cost as a function of reliability investment" (Long et al., 2007:vi).

A year after the LMI report, the Institute for Defense Analysis (IDA) published a report in 2008 entitled *Cost of Unsuitability: Assessment of Trade-offs Between the Cost of Operational Unsuitability and Research, Development, Test and Evaluation (RDT&E) Costs*. "Between 1984 and 2006, 36 out of the 136 systems that underwent operational test and evaluation (OT&E) were evaluated as unsuitable" (Lo et al., 2008:S-1). DOT&E requested the IDA conduct a study on unsuitability with two specific questions: "When a system is found to be operationally unsuitable, what are the associated costs?" and "To what extent can such costs be avoided by addressing unsuitability issues during the System Development and Demonstration (SDD) phase?" (Lo et al., 2008:S-1).

Operational suitability consists of a system's safety, interoperability, availability, maintainability, and reliability; however, to ensure the scope of the report remained manageable, Lo et al. (2008:S-1) limited the characterization of unsuitability to just the aspect of substandard reliability. Substandard reliability, measured by low mean time between maintenance (MTBM), low mean time between failures (MTBF), and other factors, was chosen because the "associated costs are large, readily identifiable, and calculable using validated methods" (Lo et al., 2008:S-1). The authors described the cost of unsuitability as the additional LCC occurring from maintenance personnel, replacement parts, repairs, and initial spares (Lo et al., 2008:S-4).

The report examined three aircraft (F-22, MV-22, and C-17), which addressed substandard reliability with different approaches. Both the F-22 and MV-22 received unsuitable evaluations during IOT&E and then attempted to resolve the unsuitable reliability through additional investment in re-design, re-engineering, and retrofit of fielded units (Lo et al., 2008:S-1). In contrast, the C-17 wanted to avoid failure at IOT&E after early flight testing revealed several reliability metrics, including the primary reliability metric (PRM), remained below contractually specified growth curves; therefore, the C-17 program invested heavily and early in reliability improvements during SDD (Lo et al., 2008:S-2). The authors applied four steps to analyze the three aircraft:

> First, we projected the system's primary reliability metric (PRM) at maturity both with and without additional reliability investment. Second, we identified the system's additional reliability investment. Third, we estimated the reduction in the system's LCC that resulted from the investment-driven increase in reliability. Finally, we compared the reliability investment to the LCC reduction it produced. (Lo et al., 2008:S-2)

The IDA study utilized previously validated simulation models, cost estimating relationships (CERs), and demand curves to calculate the reduction in LCC (Lo et al., 2008:S-4).

Gross LCC savings in 2007 dollars and ROI ratios for each system include: F-22 $0.8B and 2.8, MV-22 $5.0B and 5.7, C-17 $16.1B and 18.3 (Lo et al., 2008:S-7). Because the programs exhibited vastly different LC flying hours, Lo et al. (2008:S-7) standardized the data by dividing the ROI by the total LC flight hours, resulting in these adjusted ROI figures: F-22 2.3, MV-22 2.0, and C-17 3.5. "Even the adjusted ROIs show that the C-17's strategy of investing to improve substandard reliability during SDD produced substantially greater returns than those of the F-22 or MV-22" (Lo et al., 2008:S-7).

The authors suggested two plausible reasons for the C-17's superior ROI: system configuration changes during SDD, when the changes are easier to accomplish, resulted in "proportionally larger increases in reliability for a given amount of investment" and because contractor development resources (capital and labor) were already available during SDD, reliability improvement projects cost less (Lo et al., 2008:S-7). Overall, the findings of Lo et al. (2008:41) indicate that investing in reliability during any acquisition phase provides value and significantly reduces LCC. The IDA study concluded, "While the results of the study are only illustrative of the optimality of suitability investment during SDD, it may not be feasible to generate statistical confidence to that effect" (Lo et al., 2008:42).

The latest DOT&E annual report published in January of 2014 illustrated the lack of improvement in reliability. From FY97 to FY13, only 75 of 135, or 56%, of systems that conducted initial operational testing met or exceeded reliability thresholds (depicted in Figure 4), compared to 64% of systems between FY85 and FY96 (Director, Operational Test and Evaluation, 2014:vi). Reliability thresholds include such factors as mean time between failure and mean time between maintenance.

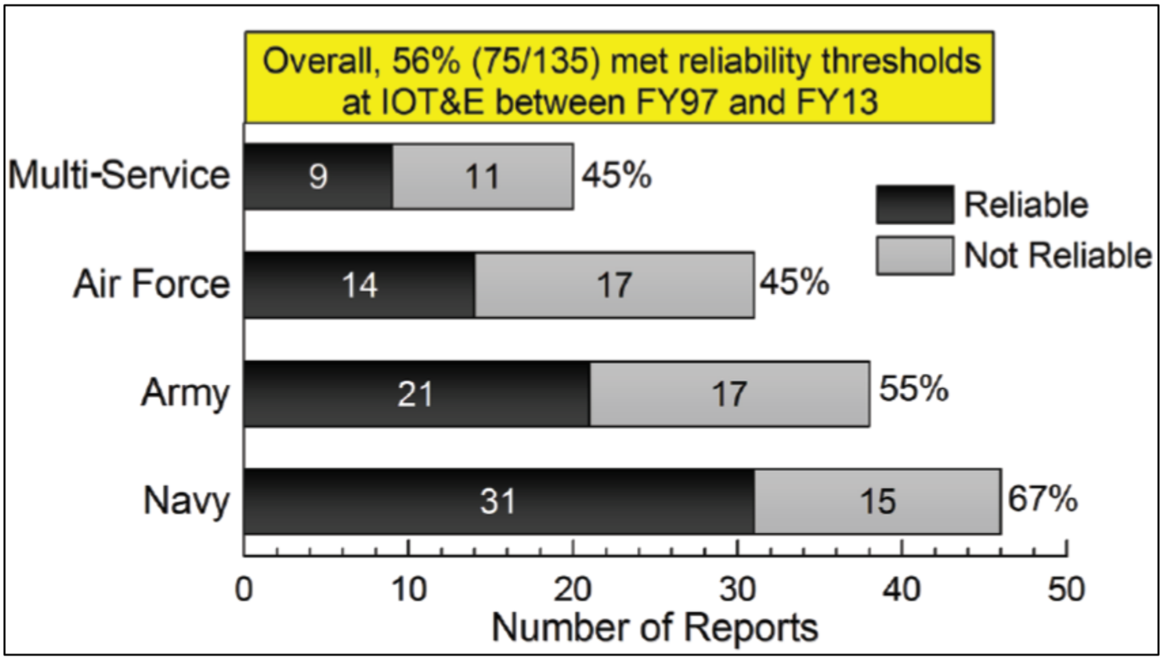

Figure 4. Fraction of DOT&E Oversight Programs Meeting Reliability Thresholds at IOT&E (Director, Operational Test and Evaluation, 2014:vi)

The two reliability studies support this research because they apply similar methodologies utilizing cost avoidance, and both reliability and T&E investments are critical to the success of each other. To improve reliability, the reliability issues must be discovered through testing, corrected, and then tested again to ensure the reliability improved. Both reliability reports highlighted two critical issues also faced in testing: DoD programs, despite the rhetoric and literature emphasizing the importance, inadequately invest in reliability and reliability deficiencies corrected early incur substantially less costs.

Other literature sources focus mostly on measuring the value of T&E through risk reduction. Bjorkman et al. (2013:541) estimate uncertainty reduction using Shannon's information entropy and apply the uncertainty reduction as a direct measure of test value; this enables a decision maker to optimize the allocation of test resources among a test portfolio based on the value the tests provide by using cost as the only constraint. Browning (2003:53) explains that in its simplest form, the ratio of the benefits to cost represents the value. Browning (2003:53) developed a risk value method by measuring the benefits (value) as the reduction of risk.

Deonandan et al. (2010) continue to develop the Prescriptive and Adaptive Testing Framework (PATFrame) with the focus of their research on unmanned and autonomous system of systems (SoS). Through a combination of surveys, interviews, and working group meetings with the DoD T&E community, Deonandan et al. (2010) identified significant cost drivers applicable to T&E. "Number of systems, integration complexity, number of requirements, technology maturity, synchronization complexity, requirements changes test complexity and diversity are all rated very high in their impacts on effort for SoS testing" (Deonandan et al., 2010). The authors describe testing as risk mitigation, and by using a risk-based approach they identified the risks that need to be mitigated and suggest making testing decision priorities based on the identified risks (Deonandan et al., 2010).

These literature sources provide several key insights applicable to DoD testing. The DoD test community should optimize the value of a portfolio of tests and not just each individual test. However, the value of a test should not just directly measure uncertainty reduction except in circumstances where safety represents the critical consideration of a test. The costs of the consequences of failure must be considered as well to properly compare the benefits to the costs. If safety is not an issue and the costs of the consequences remain small, then there is little incentive to conduct testing to reduce uncertainty.

PMs also benefit by focusing T&E considerations on maximizing value instead of focusing only on reducing costs. T&E activities do not just provide value in the information acquired from one particular test; the maximization of value occurs through the sequencing and coordination of the whole T&E process so that the right information reaches the right organization at the right time resulting in the right decision. Deonandan et al. (2010) remain in the preliminary stages of developing a cost and risk model for T&E, but, if successful, the model may develop into a much needed addition to both the T&E and cost estimating communities. The fact that the authors focus on both risk and cost is imperative. By reducing the uncertainty of the most significant costs drivers, savings throughout the LCC of the system occur.

Summary

This chapter provided an overview of T&E, examined the incentives that drive the acquisition process, presented a significant sample of historical reports documenting the inadequacy of T&E throughout the last four decades, and explored prior methodologies utilized to determine the value of T&E. The historical documentation provides a convincing argument for both the recurring inadequacy of T&E and the vital need of this research. Prior methodologies focused primarily on the reduction in uncertainty as a measure of the value of testing. However, the cost of the consequences of failure must be taken into account as well to accurately compare the benefits to costs and calculate a return on investment. The next chapter studies two cases from the JPATS program. Both indicate insufficient T&E results in costly modifications when the issues are finally discovered in the future. One case demonstrates the inadequacy of T&E and the other case illustrates the elimination of testing by the PM.

III. Methodology and Results

Chapter II discusses two reports measuring the ROI of early investments in reliability. Both reports calculated the ROI of reliability based on the cost avoidance in the LCC if the programs invested in and improved reliability earlier in the program. This research applies a similar methodology by calculating the ROI of the cost avoidance if the program discovered and corrected an issue early, during developmental testing and before the start of production, as opposed to the program discovering and correcting the issue late in the program. Two cases from the JPATS program are utilized to demonstrate the methodology. The next section explains the methodology framework. Finally, the remainder of Chapter III investigates the background of the JPATS program and the two cases, delves further into the application of the methodology to each case, and reports the results.

Methodology Framework

The methodology utilizes a case study approach. Both cases involve an issue discovered late in the program that should have, according to program office SMEs, been discovered and corrected during DT&E. The methodology framework consists of four steps applied to each case:

1. Calculate the actual costs incurred by the systems program office (SPO) to correct the issue.
2. Estimate the costs incurred by the SPO if the issue had been identified and corrected during DT&E and before the start of production.

3. Calculate the cost avoidance by subtracting the estimated costs from the actual costs.
4. Calculate the ROI by dividing the cost avoidance by the estimated initial investment needed to identify and correct the issue during DT&E.

The JPATS program provided the firm-fixed price contracts required to correct each issue. All costs are converted from constant year dollars to base year 2014 dollars using the Office of the Secretary of Defense (OSD) inflation calculator in Microsoft Excel®. SMEs from both within the JPATS program and outside the program are consulted to assist in the cost estimate had the issue been identified and corrected in DT&E. In order to capture the uncertainty in the SME's estimate, they provide three estimates: low, most likely, and high. The differences of the actual costs and estimated costs are calculated for each of the three estimates and divided by each of the respective estimated costs to compute a low, most likely, and high ROI for each issue.

JPATS Program Background

In 1989, the Congressional Armed Services Committees directed the DoD to submit a procurement plan for Air Force and Navy training aircraft for the 21st century. The DoD consolidated Air Force and Navy requirements and strategies into a single trainer aircraft plan. The strategy included the joint acquisition of a primary aircraft training system (Stockman et al., 2011:129). JPATS consists of three elements: T-6 Texan II, ground based training system, and contractor logistics support (Kinzig and Bailey, 2010:50). It replaced the AF T-37B and the Navy's T-34C (Stockman et al., 2011:129).

Just before the JPATS program began, the Federal Acquisition Streamlining Act (FASA) of 1994 was passed. The Pentagon's acquisition reform office wanted low risk programs with a high probability to succeed to become Defense Acquisition Pilot Programs (DAPP) to demonstrate FASA's innovative commercial practices and persuade the DoD to implement FASA initiatives. JPATS served as one of the initial DAPPs (Stockman et al., 2011:129-130).

Because of the JPATS DAPP designation, JPATS was specified a commercial based program and sought an aircraft with an existing Federal Aviation Administration (FAA) certification. In 1995, Raytheon Beech Aircraft won the contract award with its proposed Pilatus PC-9 commercial aircraft. However, by the time development was completed and the aircraft missionized, the final product comprised few commonalities with the original design. Further, the FAA certification required testing the AF and Navy did not require, only allowed FAA certified pilots to fly the testing requirements instead of AF test pilots, and resulted in additional cost and schedule slip which provided little benefit to the AF and limited the time the AF could test (Stockman et al., 2011:131-132). "FAA testing was given number one priority with Government tests occurring as time permitted" (Kinzig and Bailey, 2010:43).

A 2000 DOT&E report noted that although a Milestone III production decision was already scheduled, contractor developmental testing was still not complete and future testing still included both fatigue and durability testing. The same report also stated that aircraft delivery to the user occurred prior to the completion of developmental and operational testing and concluded "delivery of any system to the user prior to completion of appropriate testing is never a good situation. The process by how a system is chosen to be a commercial acquisition candidate should be reviewed" (Director, Operational Test and Evaluation, 2001:V-108).

The Air Force Operational Test and Evaluation Center conducted two OT&Es, one in 2001 and the other in 2003, both with the same result; they concluded the T-6 Texan II was operationally effective with numerous limitations and deficiencies but not suitable because of maintenance and support issues (Stockman et al., 2011:133). The DOT&E sent a letter to the Secretary of the Air Force in August of 2001 highlighting his concerns about initiating student pilot training and entering full rate production before the safety and suitability issues identified during OT&E are corrected. The decision to continue with the program was implemented, despite the DOT&E concerns, and student pilot training began at Moody AFB in October of 2001 and initial operational capability officially started in July 2002 (Kinzig and Bailey, 2010:50). The two following cases illustrate issues that resulted from limited testing and ignoring DOT&E recommendations.

Case I: Control Stick Lever Replacement

The first case involves the T-6 control stick. The control stick for the T-6 was originally an aluminum cast component. During development, a fatigue test was performed in March 2001 on the entire flight control system for two lifetimes. At that time, no cracking issues were identified with the control stick.

The T-6 aircraft began experiencing several failures with the control stick casting beginning in 2011. All Navy and AF aircraft were grounded until the control stick successfully passed inspections (Department of Defense, 2011:5). After the control stick failures, a recommendation was made to examine the original control stick previously fatigue tested. Utilizing non-destructive inspection (NDI) techniques not previously used, a crack on the control stick was identified.

After identifying the cracking issues, the Air Force Research Lab (AFRL) Materials Integrity Branch conducted several studies to determine the problem. The AFRL concluded the fractures were preceded by fatigue cracking. AFRL tested the fatigue crack growth rates revealing a faster growth rate than the NASGRO® database (software for fatigue crack growth analysis), which the manufacturer used for its analysis (Ware, 2012).

The control stick links the pilot's control inputs with the flight control surfaces. Fractures of the control stick can seriously compromise the pilot's ability to operate the aircraft's ailerons and elevator, possibly resulting in a loss of aircraft (Ware, 2012). The JPATS program office decided to replace all of the control stick levers after recommendation from the AFRL. The redesigned control stick is a wrought aluminum lever which at higher loads did not crack after 10,000,000 cycles whereas the cast aluminum lever previously on the aircraft showed cracking in as few as 5,000 cycles. The new control lever component dramatically extends the service life (Jacobs et al., 2013).

The JPATS program office provided the two firm-fixed price contracts to resolve the issue. The two contracts were for engineering change proposal (ECP) 156 which modified contract number FA8617-07-D-6151 0015. The first contract resulted in a cost of $2,407,648 in FY 2013. The second contract resulted in a cost of $1,677,329 in FY

2014. Utilizing the OSD inflation calculator, the total cost is $4,121,092 in FY 2014. Both control stick levers on 789 Air Force and Navy T-6 Texan II aircraft were replaced.

It would not be cost effective to perform an NDI on every component. However, all safety of flight or fracture critical components should receive an NDI. If the control stick would have originally received the safety of flight classification, as AFRL later argued and the control stick did eventually receive, then an NDI would have been performed and the crack discovered. Therefore, the only test not originally executed that would have needed to be done to discover the cracking is the NDI. The cost, according to the AFRL, to prepare and complete an NDI is $445 in FY 2014.

The difference between the actual cost and the additional investment in testing represents the cost avoidance which equals $4,120,647. The cost savings divided by the additional investment in testing calculates the ROI which equals a ROI percentage of 9,260%. The difference in cost between originally using cast versus wrought aluminum is negligible and not included in the estimate.

This case represents an example of insufficient testing. By not originally performing an NDI, the JPATS office now faces this costly situation today. One issue with the control stick involved the control stick not receiving safety of flight classification. According to Hawker Beechcraft Defense Company (HBDC), the control stick was not fracture critical and received a Grade B casting per MIL-A-21180 (Ware, 2012). However, AFRL argued, “the lever assembly is critical to flight safety and is also a highly stressed component with margins of less than 10 percent in select locations. According to MIL-A-21180 and JSSG-2006, this component should be classified as fracture critical, Grade A (highly stressed)" (Ware, 2012). After the control stick issues, the control stick attained reclassification as a safety of flight component.

Case II: Nose Landing Gear Friction Collar Retrofit

The second case involves the T-6 landing gear. According to the program office, the PM decided to cut landing gear testing to save money. In April 2007, the program office identified the nose landing gear (NLG) shimmy as an area of interest. A NLG shimmy consists of a rapid and violent left and right oscillation of the nose wheel and can occur during landing or takeoff, but primarily during landing. The NLG shimmy can cause damage or deterioration to aircraft components. Control of the aircraft may be compromised which can result in runway departure, loss of aircraft, and injury to pilots.

Both the Navy and AF continued reporting shimmy events with 1,326 reported through June 2009. In October 2007, a severe NLG shimmy occurred resulting in assembly component damage. The FAA deemed the NLG unsafe and in December 2007 directed HBDC to investigate the root cause and develop a solution. Furthermore, HBDC concluded the NLG shimmy events initiate cracks in the NLG upper strut housing. The cracks required increased maintenance inspections and the shortage of spare struts resulted in grounded aircraft which reduced aircraft availability for the mission. This example demonstrates how one issue can easily lead to multiple issues with negative effects.

HBDC designed a NLG friction collar as the solution to preventing the shimmy events and received FAA certification for it in September 2011. The JPATS program office provided the firm-fixed price contract to resolve the issue. The contract was for

ECP 151 which modified contract number FA8617-07-D-6151 0015. The contract resulted in a cost of $1,129,896 in FY 2013. Utilizing the OSD inflation calculator, retrofitting the T-6 Texan II aircraft with the NLG friction collar resulted in a cost of $1,146,844 in FY 2014. The SME providing the estimate works at the Air Force Material Command Landing Gear Test Facility, which performs full-gear failure and fatigue and wear testing on landing gear. The SME estimates for performing complete landing gear testing, including fatigue testing include: low estimate of $500K, most likely estimate of $750K, and a high estimate of $1.5M. The increased range from the most likely to high estimate is due to the uncertainty in follow-on testing required if initial testing identifies issues.

The difference between the actual cost and the additional investment in testing represents the cost avoidance which equals $646,844 for the low, $396,844 for the most likely, and a loss of $353,156 for the high. The cost savings divided by the additional investment in testing calculates the ROI which equals a ROI percentage of 129% for the low, 53% for the most likely, and a negative return of 24% for the high.

Summary

Chapter III details the methodology applied to each case study issue. A brief background and discussion of the issue supplements each case study. The methodology framework contains four steps applied to each case to calculate the ROI for each particular issue. Each case study includes the ROI results of that case. The next and final chapter discusses the implications for PMs and the acquisition community. This chapter makes recommendations for action and future research along with significance of the research.

IV. Conclusions and Recommendations

Conclusions of Research

Both JPATS cases illustrate the potential savings and ROI resulting from discovering and correcting issues early. Only with the substantially high landing gear estimate did the ROI actually result in a potentially negative ROI of 24%. The positive results ranged from 53% to 129% ROI for the landing gear and 9,260% ROI for the control stick.

Because of the limited data of only two cases, no statistically significant conclusions can be obtained from the results. Based on the literature's discussion on the value of identifying problems as early as possible and the potential ROIs from these two cases, further research is essential. Finally, these cases only quantify the costs of the material, labor, and overhead of the contractor and do not account for qualitative factors that potentially result in greater costs than just the contract costs and would further increase the ROI if eliminated.

Other factors more qualitative in nature also need to be considered when discussing the consequences of not discovering issues until late in development or after the deployment of the weapon system. This research focused on cost and the ROI, but PMs must also consider qualitative factors when making T&E investment decisions. The first and most important is the life of a military member. The failure of a system could result in the loss of a service member's life and safety considerations should never be overlooked. Another imperative factor is mission readiness. The discovery of a critical issue could substantially reduce or eliminate mission readiness by preventing the use of the system until a suitable solution is achieved. The entire T-6 fleet was grounded due to the landing gear issues. Further, if the FAA deems an aircraft unsafe, production can be halted until the contractor finds a suitable solution which can alter both future cost and mission capability.

Finally, the opportunity cost characterizes the most important and often overlooked cost because of the difficulty attempting to quantify it. Countless time and effort are expended to find solutions to these issues. The two T-6 examples discussed are still implementing solutions today and have already been ongoing for over 3 and 7 years. The numerous hours exhausted investigating and implementing solutions to issues that should never have occurred in the first place could have been applied to more productive activities elsewhere.

Recommendations for Further Research

Further research on the ROI of T&E must undoubtedly be pursued. Future research should match the original intent of this thesis by utilizing case 1 data (issues identified during OT&E that should have been discovered and corrected during DT&E) from the DOT&E programs identified in its annual reports. Utilizing these issues proves the problems could be discovered during the T&E process with SMEs concluding the problems should have been previously identified and corrected. Case 1 issues correspond directly with the 2010 Congress inquiry.

Programs continually disregard DOT&E recommendations and proceed with additional risk. ROI denotes an easily understood metric DOT&E can employ to demonstrate to PMs the value of early discovery and the costly consequences of advancing the program without first ensuring the entire weapon system operates as intended. To accomplish the research, DOT&E should start requiring each of the identified program offices in its annual report to collect the data and make it available for research. An independent organization, such as one previously mentioned in this research (GAO, DSB, IDA, LMI), should conduct the study to avoid program office biases, ensure independence, and because of the considerable effort it will require to complete.

Discussion and Recommendations for Acquisition Reform

This research focused on examining the ROI of T&E; however, when combining this research with previous research and philosophies discussed in Chapter II, recommendations emerged for the much broader topic of acquisition reform. The perception and criticism of the DoD acquisition process is that it follows a "build it now, Band Aid™ it later" approach to acquisition (Hutchison, 2014:16). Frank Kendall, current USD(AT&L), criticized the acquisition process when he proclaimed, "Putting the F-35 into production years before the first test flight was acquisition malpractice" (Majumdar, 2012). Steven Hutchison, former acting DASD(DT&E) claimed, "Permitting development problems to become the warfighter's problems is the real definition of acquisition malpractice" (Hutchison, 2015:8). How can the DoD reform the acquisition process to defend itself from criticism and prevent acquisition malpractice?

Acquisition reform efforts have appeared with regularity over the last four decades. The GAO's high risk list has included the DoD's acquisition of major weapon systems since 1990 and the GAO continues to observe the same issues that lead to the

DoD's first appearance on the list (U.S. Government Accountability Office, 2013:1). "Reforms that focus on the methodological procedures of the acquisition process are only partial remedies because they do not address incentives that deviate from sound practices" (U.S. Government Accountability Office, 2013:1). It is not necessarily unsuccessful policy causing ineffective acquisition outcomes, but the incentives that motivate deviations from policy (concurrent testing and production, optimistic assumptions, and delayed testing) as multiple examples in Chapter II illustrated. "The fact that programs adopt practices that run counter to what policy and reform call for is evidence of the other pressures and incentives that significantly influence program practices and outcomes" (U.S. Government Accountability Office, 2013:7).

When PMs and acquisition executives fight to fund capabilities that enhance national security and improve military safety, they almost certainly do so with sincere intentions. "While individual participants see their needs as rational and aligned with the national interest, collectively, these needs create incentives for pushing programs and encouraging undue optimism, parochialism, and other compromises of good judgment" (U.S. Government Accountability Office, 2013:8). National security and military safety comprise the primary mission of the DoD. How can anyone argue against rushing the delivery of cutting-edge technologies and defense systems to the field? Rushing cutting-edge capabilities to the military enhances national security and saves lives. "Pressure to make exceptions for programs that do not measure up are rationalized in a number of ways: an urgent threat needs to be met; a production capability needs to be preserved; despite shortfalls, the new system is more capable than the one it is replacing; or the new system's problems will be fixed in the future" (U.S. Government Accountability Office, 2013:9). The sooner, the better, right? In the short term, this probably holds true. However, an assessment of the long term may reveal that national security and military safety become compromised in the future if the military is driven to reduce the size of the force, accept fewer capabilities into the field, average system age escalates, reliability diminishes, and some systems do not work as intended because of the deficiency in long term affordability caused by an ineffective investment strategy and an inefficient acquisition system.

In fact, even in the short term, lives may be lost when the capabilities do not work as intended or suffer reliability issues in the field. The investigation into the MV-22B Osprey crash on 8 April 2000 that killed 19 marines, disclosed testing requirements that were severely curtailed (Defense Science Board Task Force, 2000:28). The program limited developmental testing requirements to save money and stay on schedule. Is national security enhanced and more lives saved from rushing capabilities into the field or ensuring the long term affordability of the national security strategy? According to the DoD website, the most important resource is "not tanks, planes or ships, it's... People. We will never compromise on the quality of our most important resource: the people" (Department of Defense, n.d.). However, the future unaffordability of the entire acquisition system results in fewer tanks, planes, ships, and people.

Two major decisions ultimately drive a program: the decision to initiate a program and the decision to start production. Advancing a program prematurely, especially at these decision points, leads to increased risk, cost growth, and schedule growth. The resulting recommendations concentrate on improving the acquisition process through an investigation of the incentives that ultimately drive unsuccessful results and countering those incentives by simplifying PM responsibilities and applying rigorous T&E throughout the acquisition process. By first investigating the incentives, motivations, and rationales that result in premature decisions, then recommendations can be formulated to counter the premature decisions.

The first major decision involves the decision to initiate a program. Thomas Christie, former DoD Director, OT&E from 2001 – 2005, delivered the keynote address at the 2009 International Test and Evaluation (ITEA) Symposium in which he presented an insightful view of the Defense Acquisition Board (DAB) processes he participated in. Thomas Christie acknowledged,

> Time and again I sat in program review meetings, including numerous DABs, where I was struck by the lack of credible information concerning the status or the results of development testing to date. In case after case, Pentagon decision-makers acquiesced in programs entering EMD and even low-rate initial production before technical problems were identified, much less solved; before credible independent cost assessments were accomplished and included in program budget projections; before critical technologies were shown to be sufficiently mature; and even before the more risky requirements were demonstrated in testing. (Christie, 2009)

Too often, PMs must start a program with a fatally flawed business case (U.S. Government Accountability Office, 2014b:7). How can the DoD ensure technology maturity so that a program is established with an executable foundation? The determination of technology maturity is vague and overoptimistic assumptions about the risk and maturity of the technology are encouraged through incentives for funding. Despite noble intentions to reform policy and processes, the status quo process continuously confronts inefficient acquisition outcomes caused by accepting too many programs that are unaffordable, competition for funding, immature technology, and unstable support from DoD senior leaders and Congress.

Recommendation 1: An independent DoD test agency should test, validate, and then officially certify a particular technology is mature and works as intended before the technology can be accepted into an acquisition program.

Recommendation 2: Separate the competition for funding between science and technology projects and acquisition programs by dedicating a portion of the acquisition budget to the research and development of technology.

Recommendation 3: The DoD should accept fewer acquisition programs into the acquisition process by making strategic investments in capability needs, and not capability wants, that support the long term defense strategy. Specifically, trade-offs must be formulated between long-term wants and short-term needs. Recommendation 1 should assist in limiting the number of acquisition programs through constraints on technology maturity.

Recommendation 4: Once a program is initiated, DoD senior leaders and Congress should fully support a program as long as the program remains relevant to the long term defense strategy and the original business case that resulted in the investment in the program has not changed.

Recommendation 5: Congress should enforce PM and acquisition executive tenure laws already established, particularly during the crucial stage of development.

Multiple benefits stem from recommendations 1 – 5. Acquiring a weapon system through the acquisition process is a complex and daunting task for anyone. By first ensuring the technology is mature through certification by the testing community, PMs can focus on executing the program without also needing to develop technology. Although testing will identify issues that will need be corrected, the risk of issues directly related to technology readiness will be substantially reduced thus relieving PMs from also resolving technology issues. Dedicating a portion of the acquisition budget to science and technology provides an equitable balance between technology maturation and program maturation.

Accepting fewer programs into the acquisition system and fully committing to programs already accepted, PMs can spend less time fighting for funding or advocating the relevance of the program which permits the PM to execute the program's objectives. "Program managers themselves believe that rather than making strategic investment decisions, DoD starts more programs than it can afford and rarely prioritizes them for funding purposes" (U.S. Government Accountability Office, 2005:5). This initiates the competition for funds at the inception of the acquisition process because it positions the DoD acquisition system in a continuous state of unaffordability with too many systems within the process and not enough money to afford all of them at the original intended quantity. Obtaining full support diminishes the adversarial relationship that causes PMs to censor potentially damaging news and provides the foundation for desirable open communication.

Jim Cramer, former hedge fund manager and host of CNBC's "Mad Money", describes the financial asset investment process by advocating to research first and make sure the investment has a strong business case before initiating the investment. Once initiated, the process does not stop there; an investor must continue researching (possibly on a quarterly or annual basis) to ensure the original business case that led to the decision has not changed. The investor must avoid allowing fluctuations of the market to influence the sell decision because the only reason to sell the investment is if the original business case changes.

The same should hold true for DoD investments. "With an investment strategy, senior leaders will be better positioned to formally commit to a business case that assures new programs fit in with priorities, that they begin with adequate knowledge about technology, time, and cost, and that they will follow a knowledge-based approach as they move into design and production" (U.S. Government Accountability Office, 2005:63). Even though various setbacks will definitely occur, the DoD and Congress should fully support the program unless the national defense strategy or the original business case changes.

Several of the recommendations correspond with commercial practices. Technology development is deliberately detached from a commercial PM's responsibilities because technology does not progress into a program unless mature and proven to work as intended. The commercial PM receives full support from leadership thus eliminating the advocacy role and encouraging open communication with leadership to discuss and implement solutions to issues. "Program managers we spoke with for this review specifically cited this process as an enabler for their own success … it did not require them to perform "heroic" efforts to overcome problems resulting from large gaps between wants and resources, such as technology challenges or funding shortages" (U.S. Government Accountability Office, 2005:23-26). DoD PMs deserve the same support as their commercial counterparts. Figure 5 summarizes the keys differences of commercial and DoD programs.

	Commercial companies	**DOD**
Success	Sale to customer.	Attracting funds.
Means to success	Strategic planning/prioritizing.	Competition for funds.
	Realism and candor.	Optimism and unknowns.
	Early testing.	Late testing.
	Early redlights, greenlights based on demonstration.	Early greenlights; late redlights.
	Collaboration and trust.	Oversight and distrust.
	Senior leaders are program advocates. Corporate research departments are technology developers. Program manager is executor.	Program manager is often the advocate, technology developer, and executor.
	Single program manager is accountable for delivery.	Multiple program managers are accountable for continuation.

Figure 5. Key Differences in Definition of Success and Resulting Behaviors (U.S. Government Accountability Office, 2005:55)

Finally, ensuring PMs and acquisition executives remain in their positions for the timeframe established by law is critical to improving accountability and incentivizing a long-term prospective. This enables PMs and acquisition executives to implement change and achieve their planned objectives that are now detailed in a program manager agreement signed by the PM. Currently, career progression/broadening appear to influence tenure length more than public law and DoD policy. How is a PM expected to maintain a long-term perspective and accomplish program objectives when the average tenure in less than 18 months?

The following recommendations now concentrate on the decision to start production. The single most detrimental practice in the acquisition process, the way it currently operates, is Low-Rate Initial Production (LRIP). DoD policy states, without specific details, OT&E should be conducted throughout the acquisition process; however, LRIP has no definitive OT&E requirements for validating the system works as intended before LRIP begins, which results in multiple harmful consequences (U.S. General Accounting Office, 1994b:21). As a result, many programs fail to start OT&E until after LRIP has already begun.

In the 1980s Congress discovered the DoD procuring significant quantities of weapon systems through LRIP without successfully completing OT&E. In response, Congress attempted to prevent the situation by enacting public law 101 – 189. According to the law "LRIP was defined as the minimum quantity needed to (a) provide production-representative articles for OT&E, (b) establish an initial production base, and (c) permit orderly ramp-up to full-rate production upon completion of OT&E" (U.S. General Accounting Office, 1994b:13). The law, although well-intentioned, has been ineffective in preventing the LRIP process from producing significant quantities of weapon systems under the facade of LRIP. "In the conference report for the act [public law 101 – 189], the conferees indicated that they did not condone the continuous reapproval of LRIP quantities that eventually total a significant percentage of the total planned procurement" (U.S. General Accounting Office, 1994b:13).

For example, the Global Hawk program started both development and limited production at the same time in 2001, and by the end of 2013 the program procured all 45 aircraft through LRIP and never held a full rate production review (U.S. Government Accountability Office, 2014a:116). In May 2011, DOT&E reported the Block 30 variant was not operationally effective or suitable (U.S. Government Accountability Office, 2012:77). The program has experienced three Nunn-McCurdy breaches and the DoD and Air Force proposed retiring the block 30 system to reduce program costs which would affect half of the Global Hawk fleet of aircraft (U.S. Government Accountability Office, 2014a:116).

Two production decisions exist with the full-rate production (FRP) decision representing the major decision as far as quantity. Consequently, legislation focused on the entry criteria to start FRP and completely disregarded any entry criteria for starting LRIP. Because LRIP does not require any OT and the FRP decision requires completion of IOT&E, the testing paradigm was altered. Testing activities are delayed until late in the acquisition process and the focus on IOT&E does not occur until after LRIP has already begun. Political engineering almost guarantees that after a program starts LRIP few circumstances can interrupt production. Therefore, in GAO's view, the LRIP decision often becomes the de-facto FRP decision. "LRIP is often continued, despite the evidence of technical problems, well beyond that needed to provide test articles and to establish an initial production capability. As a result, major production commitments are often made during LRIP" (U.S. General Accounting Office, 1994b:20). Technical problems may delay the FRP decision, but LRIP is rarely halted or significantly slowed down (U.S. General Accounting Office, 1994b:20).

According to 10 USC 2399, a program shall not proceed beyond low-rate initial production (BLRIP) until IOT&E has been completed and a BLRIP report submitted to the Secretary of Defense, the Under Secretary of Defense for Acquisition, Technology, and Logistics, and the congressional defense committees (Cornell University Law School, n.d.). However, no requirement exists necessitating successful completion of OT&E. The BLRIP report is just one of multiple criteria considered prior to making the FRP decision and an unfavorable designation of not operationally effective and/or suitable fails to prevent the start of FRP. In fact, the BLRIP appears to have little, if any, influence on the FRP decision. Thomas Christie, former Director, OT&E from 2001 – 2005, affirmed:

> Speaking from my own experience as the DOT&E from 2001 to early 2005, my office was responsible for producing roughly 30 Beyond Low-Rate Initial Production, or BLRIP, reports to the Secretary of Defense and Congress. By law, these reports are a prerequisite for any full-rate production decision. These reports assessed over half of these systems to be either not operationally effective or not operationally suitable, or both. In not one case was one of these programs stopped as a result of the information available in the reports or presented at the production DAB…some systems with serious reliability and maintenance problems found in development and operational testing have been waived through the decision process into production and deployment…What is disturbing about these failures is that most of these programs should not have been cleared to enter OT&E in the first place. They clearly had not completed development testing successfully – they had either failed to meet effectiveness or suitability requirements in DT&E or, in some cases, had truncated planned DT&E in order to stay on schedule or to stay within costs. (Christie, 2009)

In a perfect acquisition process, DASD(DT&E) and DOT&E perform integrated T&E throughout development, correcting issues as discovered, and IOT&E should be nothing more than a final confirmation that production is ready to begin. How can the DoD decrease the risk of entering production prematurely?

Recommendation 6: Integrate DASD(DT&E) and DOT&E into a single agency that conducts all independent oversight testing (Hutchison, 2015:10).

Recommendation 7: An independent DoD test agency must test, validate, and then officially certify the system exceeds all key performance parameters, IOT&E has been completed with the system verified as operationally effective and suitable, and there is minimal risk of any further design changes before the start of production.

Recommendation 8: Congress should penalize noncompliant acquisition programs by reducing or eliminating funding.

Integrating DASD(DT&E) and DOT&E can do more than just enhance efficiency. The critical purpose of integrating is to prevent the thought process that one is more critical than the other or that they are two separate activities. Both are interdependent and need to be applied thoroughly during the entire LC of the system. Combined into an integrated product team, both can work together to develop and execute the TEMP so that the sequencing of test activities collects the data needed for informed decision making.

As previously mentioned, LRIP currently permits too many unintentional consequences. Before committing to FRP, IOT&E should be completed with the system verified as operationally effective and suitable, and there should be minimal risk of any further design changes. LRIP can then be used for its intended purpose of slowly ramping up production while ensuring the manufacturing process is in statistical control. Once the manufacturing process has been tested and in statistical control, then FRP can start. LRIP should not be ongoing while also continually updating design changes. The sole purpose is to decrease the risk the weapon system enters production prematurely and to prevent deficiencies that lead to major and costly modifications.

Finally, by Congress penalizing noncompliant programs, it sends a clear message that noncompliance is no longer acceptable. "It is the funding approvals that ultimately define acquisition policy" (U.S. Government Accountability Office, 2013). As long as Congress continues to fund noncompliant programs, more and more programs will continue to defy the law and DoD policy because approving funding for noncompliant programs implies noncompliance is acceptable.

James Madison realized the fault of human nature and knew checks and balances were needed to counter ulterior motives. "This policy of supplying, by opposite and rival interests, the defect of better motives, might be traced through the whole system…where the constant aim is to divide and arrange the several offices in such a manner as that each may be a check on the other" (Madison, 1788). The recommendations proposed utilize independent testing as the check against the milestone decision authority (MDA). The test community and MDA incentives and responsibilities counteract each other. The former is responsible for ensuring the weapon system works as intended to prevent the military from receiving a deficient system while also attempting to minimize the cost of future retrofits and repairs. The latter desires to acquire the weapon system to provide to the military as quickly as possible at minimum cost. Because of these differences in incentives, an independent test agency is the ideal authority to certify the program is ready to proceed at program initiation and the start of production. “While independent, we [test community] also are a partner because we share the goal of ensuring that development problems do not become the warfighter’s problems” (Hutchison, 2015:11).

Significance of Research

The United States and DoD continue to confront challenging financial times as the U.S. debt expands and DoD funding shrinks. This research advocates for early and rigorous T&E and proposes multiple recommendations to enhance the acquisition process in an attempt to preserve the long term affordability and long term national defense strategy. David Packard, former Deputy Secretary of Defense and Chairman of the Packard commission, once recognized, “We all know what needs to be done. The question is why aren’t we doing it?” (U.S. Government Accountability Office, 2013:7). By counteracting the incentives that cause deviations from law and policy, the DoD can impact the root causes that influence deviations from policy and achieve a sustainable transformation of the acquisition system.

Bibliography

Bender, Jeremy, Armin Rosen, and Skye Gould. "This Map Shows Why The F-35 Has Turned Into A Trillion-Dollar Fiasco," *Business Insider*, 20 August 2014. 20 Feb 2015 http://www.businessinsider.com/this-map-explains-the-f-35-fiasco-2014-8

Bjorkman, Eileen A., Shahram Sarkani, and Thomas A. Mazzuchi. "Test and Evaluation Resource Allocation Using Uncertainty Reduction," *IEEE Transactions on Engineering Management*, 60(3):541-551 (August 2013). 26 July 2014 http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6392237

BrainyQuote. 2014. 13 July 2014 http://www.brainyquote.com/quotes/authors/p/plautus_2.html

Browning, Tyson R. "On Customer Value and Improvement in Product Development Processes," *Systems Engineering*, 6(1):49-61 (2003). 26 July 2014 http://sbuweb.tcu.edu/tbrowning/Publications/03-1-SE--Cust%20Value%20in%20PD.pdf

Christie, Thomas. "Yet Another Dose of Acquisition Reform for the T&E Community – What Does the Past Tell Us About the Future?" *Project On Government Oversight*, 29 September 2009. 22 February 2015 http://pogoblog.typepad.com/files/2009itea.pdf

Cornell University Law School. *Legal Information Institute*, n.d. 21 February 2015 https://www.law.cornell.edu/uscode/text/10/1734

Defense Acquisition University. *Defense Acquisition Guidebook*, 2013. 22 July 2014 https://dag.dau.mil/Pages/Default.aspx

Defense Science Board Task Force. *Test and Evaluation Capabilities*, 2000. 23 June 2014 http://www.acq.osd.mil/dsb/reports/TECapabilities_Dec2000.pdf

-----. *Developmental Test and Evaluation*, 2008. 23 June 2014 http://www.acq.osd.mil/dsb/reports/ADA482504.pdf

Deonandan, Indira, Jo Ann Lane, Ricardo Valerdi, and Filiberto Macias. "Cost and Risk Considerations for Test and Evaluation of Unmanned Autonomous Systems of Systems," *2010 5th International Conference on System of Systems Engineering*. 14 July 2014 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5544062&tag=1

Department of Defense. *Selected Acquisition Report for Joint Primary Aircraft Training System.* Defense Acquisition Management Information Retrieval, 31 December 2011. 5 January 2015 http://www.dod.mil/pubs/foi/logistics_material_readiness/acq_bud_fin/SARs/DEC_2011_SAR/JPATS-SAR_31_DEC_2011.pdf

-----. *DoD 101 An Introductory Overview of the Department of Defense*, n.d. 24 February 2015 http://www.defense.gov/about/dod101.aspx

-----. *Test & Evaluation Management Guide*, 2012. 2 August 2014 https://acc.dau.mil/temg

-----. *Instruction 5000.02*, 2015. 19 February 2015 http://www.dtic.mil/whs/directives/corres/pdf/500002p.pdf

Director, Operational Test and Evaluation. *FY 2000 Annual Report*, 2001. 10 January 2015 http://www.dote.osd.mil/pub/reports/FY2000/

-----. *FY 2011 Annual Report*, 2011. 28 June 2014 http://www.dote.osd.mil/pub/reports/FY2011/

-----. *FY 2013 Annual Report*, 2014. 28 June 2014 http://www.dote.osd.mil/pub/reports/FY2013/

Hutchison, Steve. "Shift Left!" *ITEA Journal of Test & Evaluation: Official Publication of the International Test and Evaluation Association, 34*(2):133-137 (June 2013).

-----. "Whatever Happened to Good Old-Fashioned DT&E?" *ITEA Journal of Test & Evaluation: Official Publication of the International Test and Evaluation Association*, 35(1):16-26 (March 2014).

-----. "Test and Evaluation Myths and Misconceptions," *Defense Acquisition University*, 2015. 24 February 2015 http://www.dau.mil/publications/DefenseATL/DATLFiles/Jan-Feb2015/Hutchison.pdf

Jacobs, Nicholas J., Robert H. Ware, and Steven R. Thompson. Air Force Research Lab. *Comparative Fatigue Testing of Cast and Wrought T-6 Control Stick Levers (Material Evaluation)*. Report No. AFRL/RXS 13-010, March 2013.

Kinzig, Bill and Dave Bailey. *T-6 Texan II Systems Engineering Case Study*. Center for Systems Engineering at the Air Force Institute of Technology, WPAFB, OH, 2010. 5 January 2015 http://www.dtic.mil/get-tr-doc/pdf?AD=ADA538810

Lo, Tzee-Nan K., Harold S. Balaban, Waynard C. Devers, Christopher S. Wait, and Kristen M. Guerrera. Institute for Defense Analysis. *Cost of Unsuitability: Assessment of Trade-offs Between the Cost of Operational Unsuitability and*

Research, Development, Test and Evaluation (RDT&E) Costs. IDA Paper P-4330, 2008. 22 July 2014 www.dtic.mil/get-tr-doc/pdf?AD=ADA493879

Long, E. Andrew, James Forbes, Jing Hees, and Virginia Stouffer. LMI Government Consulting. *Empirical Relationships between Reliability Investments and Life-Cycle Support Costs.* Report SA701T1, 2007. 18 July 2014 http://www.dote.osd.mil/pub/reports/SA701T1_Final%20Report.pdf

Lyngaas, Sean. "DOD Stresses Testing, Evaluation Improvements," *FCW The Business of Federal Technology*, 23 July 2014. 24 July 2014 http://fcw.com/articles/2014/07/23/dod-stresses-testing.aspx?admgarea=TC_Management

Madison, James. "The Federalist No. 51 The Structure of the Government Must Furnish the Proper Checks and Balances Between the Different Departments," *Independent Journal,* 6 February 1788. 23 February 2015 http://www.constitution.org/fed/federa51.htm

Majumdar, Dave. "Kendall: Early F-35 Production Acquisition Malpractice," *Defense News*, 6 February 2012. 23 February 2015 http://www.defensenews.com/article/20120206/DEFREG02/302060003/Kendall-Early-F-35-Production-8216-Acquisition-Malpractice-8217-

Naval Air Station Patuxent River, Maryland. "Official Predicts Bleak Budget Picture for Fiscal 2014," *American Forces Press Service*, 5 September 2013. 12 July 2014 http://www.defense.gov/news/newsarticle.aspx?id=120726

Shughart II, William F. "Public Choice," *Library of Economics and Liberty*. Liberty Fund, Inc., 2008. 24 February 2015 http://www.econlib.org/library/Enc/PublicChoice.html

Spinney, Franklin C. "Defense Power Games," *Project On Government Oversight*, 1998. 23 February 2015 http://www.dnipogo.org/fcs/def_power_games_98.htm

Stockman, William, Milt Ross, Robert Bongiovi, and Greg Sparks. *Successful Integration of Commercial Systems A Study of Commercial Derivative Systems*. PESystems, Inc. and Dayton Aerospace, Inc., 2011. 5 January 2015 http://daytonaero.com/wp-content/uploads/2012/07/Successful-Integration-of-Commercial-Systems-A-Study-of-Commercial-Derivative-Aircraft-June-2011.pdf

U.S. Debt Clock. 2015. 31 January 2015 http://usdebtclock.org/

U.S. General Accounting Office. *Adequacy of Department of Defense Operational Test and Evaluation*. T-NSIAD-89-39, 1989. 23 June 2014 http://www.gao.gov/products/T-NSIAD-89-39

-----. *Role of Test and Evaluation in System Acquisition Should Not Be Weakened*. T-NSIAD-94-124, 1994a. 23 June 2014 http://www.gao.gov/products/T-NSIAD-94-124

-----. *Low-Rate Initial Production Used to Buy Weapon Systems Prematurely*. GAO/NSIAD-95-18, 1994b. 18 February 2015 http://gao.gov/products/NSIAD-95-18

U.S. Government Accountability Office. *Better Support of Weapon System Program Managers Needed to Improve Outcomes.* GAO-06-110, 2005. 19 February 2015 http://gao.gov/products/GAO-06-110

-----. *Assessments of Selected Weapon Programs.* GAO-07-406SP, 2007a. 18 February 2015 http://gao.gov/products/GAO-07-406SP

-----. *Department of Defense Actions on Program Manager Empowerment and Accountability*. GAO-08-62R, 2007b. 13 February 2015 http://gao.gov/products/GAO-08-62R

-----. *Assessments of Selected Weapon Programs.* GAO-08-467SP, 2008. 18 February 2015 http://gao.gov/products/GAO-08-467SP

-----. *Assessments of Selected Weapon Programs.* GAO-09-326SP, 2009. 18 February 2015 http://gao.gov/products/GAO-09-326SP

-----. *Assessments of Selected Weapon Programs.* GAO-10-388SP, 2010a. 18 February 2015 http://gao.gov/products/GAO-10-388SP

-----. *DoD Needs to Develop Performance Criteria to Gauge Impact of Reform Act Changes and Address Workforce Issues.* GAO-10-774, 2010b. 23 June 2014 http://gao.gov/products/GAO-10-774

-----. *Assessments of Selected Weapon Programs.* GAO-12-400SP, 2012. 18 February 2015 http://gao.gov/products/GAO-12-400SP

-----. *Where Should Reform Aim Next?* GAO-14-145T, 2013. 19 February 2015 http://gao.gov/products/GAO-14-145T

-----. *Assessments of Selected Weapon Programs.* GAO-14-340SP, 2014a. 18 February 2015 http://gao.gov/products/GAO-14-340SP

-----. *Addressing Incentives is Key to Further Reform Efforts.* GAO-14-563T, 2014b. 19 February 2015 http://gao.gov/products/GAO-14-563T

Ware, Robert H. Air Force Research Lab. *T-6 Control Stick Lever Arm Cracking (Failure Analysis)*. Report No. AFRL/RXS 11-005, January 2012.

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)* 26-03-2015	**2. REPORT TYPE** Master's Thesis				**3. DATES COVERED** *(From – To)* September 2013 – March 2015
TITLE AND SUBTITLE Examining the Return on Investment of Test and Evaluation					**5a. CONTRACT NUMBER**
					5b. GRANT NUMBER
					5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Smith, Nathan C., 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 Mathematics and Statistics (AFIT/ENC) 2950 Hobson Way, Building 641 WPAFB OH 45433-8865					**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT-ENC-MS-15-M-183
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Scientific Test and Analysis Techniques in Test and Evaluation Center of Expertise 2950 Hobson Way, Building 646 WPAFB OH 45433-8865 / darryl.ahner@afit.edu ATTN: Dr. Darryl Ahner					**10. SPONSOR/MONITOR'S ACRONYM(S)** STAT in T&E COE
					11. SPONSOR/MONITOR'S REPORT
12. DISTRIBUTION/AVAILABILITY STATEMENT **DISTRUBTION 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 This research examined the return on investment of Department of Defense test and evaluation. The thesis analyzed the return on investment of the cost avoidance achieved if an issue discovered late in the program had been discovered and corrected during developmental test and evaluation. The methodology utilized two case study examples from the Joint Primary Training Aircraft System to calculate the potential cost avoidance and the potential return on investment if the program had discovered and corrected the issues during developmental test and evaluation. The result of one case was a 9,260% return on investment. The other case results ranged from a -24% to a 153% return on investment. Both cases illustrated the potential return on investment but no statistically significant conclusions can be obtained from the results. Based on the literature's discussion on the value of identifying problems as early as possible and the potential return on investment from these two cases, further research is essential. This research resulted in proposing multiple recommendations to enhance the acquisition process in an attempt to preserve the long term affordability and long term national defense strategy.					
15. SUBJECT TERMS Return on Investment, Test and Evaluation, Value, Acquisition Reform, Cost Avoidance					
16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Edward D. White, PhD , AFIT/ENC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	UU	83	(937) 255-3636, ext 4540 (edward.white@afit.edu)

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