SYM-AM-18-032

Proceedings
of the
Fifteenth Annual
Acquisition Research Symposium

Wednesday Sessions
Volume I

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

Published April 30, 2018

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

Table of Contents

THIS PAGE INTENTIONALLY LEFT BLANK

SYM-AM-18-032

Proceedings

Of the

Fifteenth Annual Acquisition Research Symposium

Wednesday Sessions
Volume I

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

Published April 30, 2018

THIS PAGE INTENTIONALLY LEFT BLANK

Keynote: Ms. Anne Rung, Director of Government Sector for Amazon Business; Former United States Chief Acquisition Officer

Anne Rung serves as Director of Government Sector for Amazon Business. In this role, she is helping to expand Amazon Business to state, local, and federal governments as well as non-profits.

From 2014 through 2016, Anne served as U.S. Chief Acquisition Officer (CAO) under President Obama. As U.S. CAO, she was responsible for implementing acquisition policies covering $450B in annual federal contract spending. Anne also served under President Obama as Chief Acquisition Officer at the U.S. General Services Administration and Senior Advisor at the U.S. Department of Commerce, overseeing acquisition reform.

Prior to joining the Obama Administration, Anne was Deputy Secretary of Procurement for the State of Pennsylvania. She has also worked for former Vice President Biden when he served as Chairman of the Senate Judiciary Committee.

Anne was raised in Pennsylvania and earned a BA from Pennsylvania State University and an MSc from London School of Economics.

Panel 1. Exploring New Approaches in Ship Design and Construction

Wednesday, May 9, 2018	
9:30 a.m. – 11:00 a.m.	**Chair: Rear Admiral William Galinis, USN,** Program Executive Officer, Ships ***Business Case Valuation of Strategic Flexibility in Ship Building: Justifying and Assessing the Value of Flexible Ships Design Features in New Navy Ship Concepts*** Johnathan Mun, Naval Postgraduate School Thomas Housel, Naval Postgraduate School LCDR Lauren B. Majchrzak, USN ***Designing Out Complexity Early: A Path to Affordable Flexible Warships*** Glen Grogan, Naval Sea Systems Command ***Operational Seakeeping Considerations in LCU Deployment*** Fotis A. Papoulias, Naval Postgraduate School Jarema M. Didoszak, Naval Postgraduate School

Rear Admiral William Galinis, USN—is a native of Delray Beach, FL. He is a 1983 graduate of the U.S. Naval Academy where he received a Bachelor of Science in Electrical Engineering. He holds a Master of Science in Electrical Engineering from the Naval Postgraduate School.

Galinis's tours as a surface warfare officer included damage control assistant aboard USS *Vreeland* (FF 1068) and engineer officer aboard USS *Roark* (FF 1053). He was selected for transfer to the engineering duty officer community in September 1991.

Galinis's initial engineering duty tour was with the supervisor of Shipbuilding, Conversion, and Repair, New Orleans, where he worked on both new construction and repair projects including assignment as the PMS 377 program manager's representative for the LSD (CV) Shipbuilding Program. He subsequently served as the senior damage control inspector for the Board of Inspection and Survey, Surface Trials Board, as well as in a number of program office and staff positions, including the DD 21 and LPD 17 Program Offices, Office of the Chief of Naval Operations in the Requirements & Assessments Directorate (N81), and Office of the Deputy Assistant Secretary of the Navy for Shipbuilding as the chief of staff.

His command assignments included LPD 17 program manager—leading the commissioning of the first four ships of the LPD 17 San Antonio Class, delivering the fifth ship, and starting construction on four additional ships; supervisor of shipbuilding, Gulf Coast overseeing Navy ship construction projects and Foreign Military Sales work in shipyards along the Gulf Coast and Wisconsin; and commanding officer of the Norfolk Ship Support Activity (NSSA), where he led ship maintenance and repair efforts.

Galinis's first flag assignment was deputy commander for surface warfare, Naval Sea Systems Command (NAVSEA 21)/commander, Navy Regional Maintenance Center, responsible for managing critical modernization, maintenance, training, Foreign Military support contracts, and inactivation programs.

Currently, Galinis is serving as program executive officer, Ships, where he is responsible for Navy shipbuilding for surface combatants, amphibious ships, logistics support ships, support craft, and related foreign military sales.

Galinis has received various personal, unit, and service awards including three Navy Battle "E" awards.

Business Case Valuation of Strategic Flexibility in Ship Building: Justifying and Assessing the Value of Flexible Ships Design Features in New Navy Ship Concepts

Johnathan C. Mun—is a research professor at NPS and teaches executive seminars in quantitative risk analysis, decision sciences, real options, simulation, portfolio optimization, and other related concepts. He received his PhD in finance and economics from Lehigh University. He is considered a leading world expert on risk analysis and real options analysis. Dr. Mun has authored 12 books and is the founder and CEO of Real Options Valuation Inc. [jcmun@realoptionsvaluation.com]

Thomas J. Housel—specializes in valuing intellectual capital, knowledge management, telecommunications, information technology, value-based business process reengineering, and knowledge value measurement in profit and non-profit organizations. He is a tenured full professor for the Information Sciences (Systems) Department at NPS. He has conducted over 80 knowledge value added (KVA) projects within the non-profit, Department of Defense (DoD) sector for the Army, Navy, and Marines. Dr. Housel also completed over 100 KVA projects in the private sector. The results of these projects provided substantial performance improvement strategies and tactics for core processes throughout DoD organizations and private sector companies. [tjhousel@nps.edu]

LCDR Lauren B. Majchrzak, USN

Abstract

To successfully implement the Surface Navy's Flexible Ships concept, Program Executive Office–Ships (PEO-SHIPS) requires a new methodology that assesses the total future value of various combinations of Flexible Ships design features and how they will enable affordable warfighting relevance over the ship's full service life. Examples of Flexible Ships design features include decoupling payloads from platforms, standardizing platform-to-payload interfaces, implementing allowance for rapid reconfiguration of onboard electronics and weapons systems, preplanning access routes for mission bays and mission decks, and allowing for sufficient growth margins for various distributed systems. This research analyzes the application of strategic Real Options Valuation (ROV) methodology within the Integrated Risk Management process to assess the total future value of Flexible Ships design features and for use in the Future Surface Combatant Analysis of Alternatives. The current research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Navy in quantifying, modeling, valuing, and optimizing a set of ship design options to create and value a business case for making strategic decisions under uncertainty.

Introduction

The U.S. Navy is tasked with fulfilling its missions globally in environments with rapidly changing threats using an equally rapidly evolving technological base of platform, mission, electronic, and weapon systems. The challenge the U.S. Navy faces is to retain and maintain sufficient military relevance during wartime as well as peacetime, with the added goal of minimizing highly intrusive and costly modernization throughout a ship's service life by incorporating Modular Adaptable Ships (MAS) and Flexible and Adaptable Ship Options (FASO) in the ship design. Pursuing this goal has the added benefit of allowing the Navy to affordably and quickly transform a ship's mission systems over its service life to maintain its required military capabilities (Doerry, 2012).

Historically, naval ship design includes robust features that limit any future capabilities to make requirement changes. For instance, any major requirement changes needed to meet critical operational tasks during wartime would necessitate a major modernization effort or decommissioning the existing ship prior to its end of service life and replacing it with a newly commissioned ship. The concepts of MAS and FASO, if applied correctly, with the optimal options implemented, would reduce the need for costly and lengthy major mid-service-life intrusive modernizations, as well as increase the existing platform's flexibility to adapt to new requirements utilizing a faster and cheaper alternative.

The concept of FASO is not new to the Navy. In fact, benefits of MAS/FASO concepts have been detailed by Jolliff (1974), Simmons (1975), Drewry and Jons (1975), and others. Even as recently as 2015, the Naval Sea Systems Command's (NAVSEA's) Program Executive Office–Ships (PEO-SHIPS) put out a presentation on Flexible Ships, detailing its "affordable relevance over the ship's life cycle" (Sturtevant, 2015). In it, the Director of Science and Technology, Glen Sturtevant, noted that the main current and future challenges confronting the Surface Navy include facing unknown but evolving global threats while managing an accelerated pace of technological changes, coupled with handling rising costs and declining budgets. The analysis found that ships currently cost too much to build and sustain, the ships (Platforms) are too tightly coupled with their capabilities (Payloads), and inflexible and fixed architectures of legacy ships limit growth and capability upgrades or result in lengthy and costly upgrades. The effects of these issues, of course, are compounded by ever-evolving, unknown global threats.

In past speeches, Admiral Greenert (Chief of Naval Operations) and Vice Admiral Rowden (Commander of Naval Surface Forces) echoed the idea that the ability to quickly change payloads and have modularity on ships would maximize the service life of ships and allow faster and more affordable upgrades to combat systems and equipment.

Some examples of MAS and FASO that had been espoused in Navy research literature, such as in Sturtevant (2015); Doerry (2012); Koenig (2009); Koenig, Czapiewski, and Hootman (2008); and others, include Decoupling of Payloads from Platforms, Standardizing Platform-to-Payload Interfaces, Rapid Reconfiguration, Preplanned Access Routes, and Sufficient Service Life Allowance for Growth. These FASO areas can be applied to a whole host of systems such as weapons, sensors, aircraft, unmanned vehicles, combat systems, C4I, flexible infrastructure, flexible mission bays and mission decks, vertical launch systems (VLS) for various multiple missile types, future high-powered surface weapons (laser weapon systems and electromagnetic railguns), and modular payloads (e.g., anti-submarine warfare, special operations, mine warfare, intelligence gathering, close-in weapon systems, harpoon launchers, rigid hull inflatable boats, gun systems, etc.).

The concepts of Adaptability and Flexibility (plug-and-play concepts of rapidly removing and replacing mission systems and equipment pier-side or at sea), Modularity (common design interface and modular components that will greatly simplify adding, adapting, modifying, or modernizing a ship's capabilities), and Commonality/Scalability (capabilities that are built independently of a ship by using standardized design specifications that allow similar systems to be placed across multiple ship platforms) are all concepts of strategic Real Options Valuation (ROV) analytical methodologies. ROV has been used in a variety of settings in industry including pharmaceutical drug development, oil and gas exploration and production, manufacturing, start-up valuation, venture capital investment, information technology infrastructure, research and development, mergers and acquisitions, intangible asset valuation, and others. The current project looks at applying the same flexibility modeling empowered by ROV methods to identify the optimal ship design alternative.

This current research acknowledges that the U.S. Navy has sought out the ability to incorporate FASO and MAS capabilities in its ship design of Future Surface Combatants (FSC). Further, the Navy acknowledges that there is significant value in terms of being able to rapidly upgrade FASO ships at a lower cost while extending the ships' service life, all the while being able to quickly adapt to changes in both external threats and internal new technologies. As such, this current research is not meant to identify said FASO/MAS platforms or payloads per se, but to use previously identified platforms such as the DDG 51 Flight III, where there are opportunities to insert flexible ship features, and we limit the analysis to said surface combatants in the domain of Anti-Submarine Warfare (ASW).

This current research focuses on a series of recommended analytical methodologies to establish a business model or business case analysis that supports strategic decision-making under uncertainty, specific to identifying, modeling, valuing, and optimizing the various strategic real options in flexible ship designs. Currently, there is only a limited set of real-life applications of FASO/MAS in ship design, and they are classified; therefore, actual empirical data is not used in this research. In addition, because the objective of this research is to illustrate in detail the business case modeling process and analytical methodologies such that the method and process can be replicated and used in all future FASO/MAS design decisions, subject matter expert (SME) opinions, publicly available information, and certain basic assumptions or rough order magnitude (ROM) estimates are used. The use of said ROM or SME inputs in no way detracts from the analytical power, efficacy, or applicability of these methods.

In summary, this current research has the explicit goal of proposing a reusable, extensible, adaptable, and comprehensive advanced analytical modeling process to help the U.S. Navy in quantifying, modeling, valuing, and optimizing a set of ship design options to create a business case for making strategic decisions under uncertainty. The process will accomplish the following:

- Identify which FASO/MAS options have a positive return on investment (i.e., in which options the benefits outweigh the costs).
- Model uncertainty and risks (i.e., Monte Carlo Risk Simulations will be applied to simulate hundreds of thousands of possible scenarios and outcomes to model the volatility and ever-changing global threat matrix).
- Frame and value the ship design options (i.e., each design option will be vetted and modeled; options will be framed in context and valued using cost savings [cost savings due to rapid upgrades at lower costs], costs to obtain these options [costs to design and implement these FASO/MAS options], and potential military benefits [using Knowledge Value Added methods to monetize expected military value]).
- Optimize the portfolio of options (i.e., given a set of FASO/MAS design options with different costs, benefits, capabilities, and uncertainties, identify which design options should be chosen given constraints in budget, schedule, and requirements).

Flexible and Adaptable Ship Design

Seventy percent of the world is covered by water. To ensure freedom of navigation, economic independence and national sovereignty, countries must maintain a highly efficient and technologically advanced fleet. With shrinking defense budgets, the current trend is to build fewer warships but maintain the same operational tempo. To continually meet the demands of a larger operational fleet, these new smaller fleets must be built on flexible and adaptable platforms with decoupled payloads that allow the vessel to accomplish a multitude of mission sets. This type of modular design and build “offers an opportunity for a ship to affordably transform its mission systems over its service life to maintain military relevance” (Doerry, 2012). The design characteristics that allow these fleets to flourish are MAS and FASO (Mun & Housel, 2016). MAS- and FASO-incorporated designs provide an economical platform for a sea-going navy to build highly effective warships capable of performing various missions in a multitude of environments.

Flexible and adaptable ship designs are centered around a standard hull with modular mission payloads that offer a wide mission set, affordable scalability, reduced operational downtime, increased availability of the ship, and a reduced total number of mission modules for the fleet (Thorsteinson, 2013). For navies with limited budgets, having a flexible and modular platform allows a vessel to perform at times like a frigate and at other times like a corvette (Paris, Brussels & Fiorenza, 2013). These new fleets of multi-mission vessels are already operational in blue water fleets around the world operated by countries including Denmark, Germany, France, Italy, Australia, and the United States.

Modular build and design has been in use since the mid-20th century. During World War II, Henry Kaiser’s ship yards were able to produce Liberty ships in minimal time due in part to the heavy use of modular construction, and the Germans constructed their Type 21 submarines with modular build principles (Abbott, Levine, & Vasilakos, 2008). Starting in 1979, the German shipyard Blohm + Voss began building modular corvettes and frigates for third world navies using a modular concept known as MEKO. The MEKO concept has continually evolved with time, producing the more mature MEKO A-100, A-200, and now A-400. In 1986, the Royal Danish Navy (RDN) began implementation of a modular concept called STANFLEX for a new class of patrol craft (Abbott et al., 2008) known as the Flyvefisken (SF 300) class. The specific use of modular mission payload within the SF 300s directly translated into the future design and development of the RDN Absalon support ships and Iver Huitfeldt class frigates. The French and Italians have worked together to design a flexible multi-mission frigate known as the FREMM class, while the Australian Royal Navy has the modular Anzac class of frigates and Hobart class of Air-Warfare Destroyers (AWDs).

The U.S. Navy began to look at modular builds in 1975 with the Sea Systems Modification and Modernization by Modularity (SEAMOD) program (Abbott et al., 2008). SEAMOD focused on decoupling “the development of the payload from the development of the platform” (Doerry, 2012). This uncoupling provided two major benefits: it allowed the payload to be developed in parallel with the platform versus in series which allowed the most recent technological systems to be installed onboard at the time the ship was put to sea, and it permitted rapid removal, replacement, or installment of mission payloads, preventing extended maintenance yard periods (Abbott et al., 2008; Doerry, 2012). SEAMOD evolved into the Modular Open Systems Approach (MOSA) and is characterized by “modular design, key interfaces, and the use of open standards for key interfaces where appropriate” (Abbott et al., 2008). These efforts led to the development of the Littoral Combat Ship (LCS) and DDG 1000 for the U.S. Navy (Abbott et al., 2008).

To achieve expected service life, flexible and adaptable ships must be built with payloads that decouple from the platform, be configured with standard interfaces for technical modules, have the ability to reconfigure rapidly, and have allowances for growth margin. Growth margins allow for future technologies to be rapidly implemented into the existing design, preventing the vessel from having to enter into an extended maintenance overhaul period. Growth margins work hand in hand with the parallel development of mature payloads, ensuring that the latest technology can be installed as it is developed because of the standard interfaces.

Over the past 40 years, significant strides have been made by foreign navies with regards to ship designs that incorporated modularity, flexibility, and adaptability. The designs focused heavily on a standard hull with the same machines but offered a variety of modular payloads for specific mission sets. Ultimately, MAS- and FASO-incorporated designs provide an economical platform for a sea-going navy to build powerful, multi-task warships.

Royal Danish Navy

The RDN has been at the forefront of modular ship design since 1987, when the first of 14 Flyvesfisken class or STANFLEX 300 (SF 300) multi-role vessels (MRVs) were commissioned. The design was based on a standard hull that used modular bays to change mission type through use of the Standard Flex (STANFLEX) concept. The Flyvesfisken class was ultimately decommissioned in October 2010 ("Flyvefisken Class," n.d.), but the use of the STANFLEX concept played a fundamental role in the design and development of the larger follow-on modular designs seen in the Absalon class littoral support ships and Iver Huitfeldt class frigates.

Flyvefisken Class (SF 300)

The inception of the Flyvesfisken class and STANFLEX resulted from a feasibility study in 1982. The RDN wanted to replace its fleet of 24 mission-specific ships (eight Fast Attack Craft [FAC], eight patrol boats, and eight mine countermeasure vessels) with a smaller number of multi-role vessels (MRVs; Pike, 2011). The RDN downsized to 14 MRVs and commissioned the SF 300 fleet between 1987 and 1996. To meet the multi-role vessel mission, the SF 300 was built on a standard hull of non-magnetic fiberglass reinforced plastic (FRP) that measured 54 m in length and 9 m in beam, the crew varied between 19 and 29 personnel depending on mission type, and the overall tonnage ranged from 320–485 tons specific to payload installed (Pike, 2011).

STANFLEX design capitalized on mission modularity by incorporating four interchangeable mission containers, one forward and three aft. The stainless-steel containers measured 3 m by 3.5 m by 2.5 m and housed all dedicated machinery and electronic payloads connected by a standard interface panel ("Flyvefisken Class," n.d.). "Each of these units can be (re)configured at a short notice for different roles, simply by installing the right combination of standard-size equipment containers in the four positions" (Pike, 2011). The ability to quickly and efficiently swap payload allowed these MRVs to serve the following mission sets: anti-air warfare (AAW); anti-surface warfare (ASuW); anti-submarine warfare (ASW); electronic warfare (EW); mine countermeasures (MCM); patrol and surveillance; and pollution control (Pike, 2011).

The use of containerized weapon systems permitted the SF 300 to have an open architecture C4I system that allowed "new weapons systems to be added by creating new nodes" ("Flyvefisken Class," n.d.). Major technological upgrades were not required for the ship itself, but merely applied to the appropriate container. Containers could be swapped out in 30–60 minutes pier-side using standard civilian cranes (Pike, 2011), facilitating rapid mission change if necessary. Ultimately, 15 different mission modules were developed for the SF 300, which included weaponized containers for the Mk48 NATO Vertical Launch Sea

Sparrow surface-to-air missile, Boeing's Harpoon Block II surface-to-surface missile, and the 76 mm Oto Melara Super Rapid gun ("Flyvefisken Class," n.d.).

The Flyvefisken class demonstrated that a smaller number of multi-role vessels were capable of meeting the same mission demands of a fleet almost twice its size. STANFLEX and modular payload allowed for containers to be pre-staged for mission flex while simultaneously reducing downtime for upgrades. The success of the SF 300 fleet was the cornerstone for the RDN's development of the Absalon Littoral Combat Ship.

German Navy

At the forefront of modular design for the German Navy is the Blohm + Voss model. The design concept known as *Mehrzweck-Kombination* (MEKO), which translates as "multi-purpose combination," has been utilized in ship construction and design since the 1970s. The success of the MEKO class can be seen in 13 navies worldwide in various corvettes and frigates (Kamerman, 2015). The modular mission payloads in 20-ft standardized ISO containers create adaptability and flexibility and allow navies to rapidly reconfigure mission type based on operational needs. Modules can be rotated for upgrades and maintenance or between ships, which reduces the number of overall payloads required for the fleet. This simple reduction results in significant cost savings in procurement and maintenance over the life cycle of the ship (ThyssenKrupp Marine Systems, n.d.). The MEKO class is comprised of the MEKO A-100 Corvette and the MEKO A-200 Frigate (ThyssenKrupp Marine Systems, n.d.) and is the backbone for the new German frigate class, the Baden-Württemberg (F125).

The German Navy will acquire four Baden-Württemberg class frigates to replace the eight frigates in the Bremen class (F122) commissioned in the 1980s. The Baden-Württemberg frigate design incorporates enhanced survivability capabilities to include floating, moving, and fighting after sustaining damage; to embark and deploy special forces; and to maintain prolonged periods at sea with little maintenance; and incorporates modular mission capabilities (Kamerman, 2015). The F125 is a new hull design drawing from the MEKO A-200 and the German F124. It measures 149.5 m in length with a beam of 18.8 m, displaces 7,300 tons at full load, and will carry a crew of 105–120, but can accommodate up to 190 personnel to include a 20-person aircraft detachment and 50 embarked forces ("Baden-Württemberg," 2017). The first frigate, Baden-Württemberg (F222), will be commissioned in 2017, Nordrhein-Westfalen (F223) in 2018, Sachsen-Anhalt (F224) in 2019, and Rheinland-Pfalz (F225) in 2020 (Pape, 2016).

The F125 class is designed to experience prolonged deployment periods of 24 months and increased hours of operation of 5,000 hr/yr. This extended availability will be accomplished through a two-crew concept with crews swapping every four months in the given operational theater (Kamerman, 2015). Through modernization, automation, and cross-rate training, the crew of the F125 is approximately half the size of the marginally smaller German Sachsen (F124) class frigates that currently deploy for six-month cycles and operate 2,500 hr/yr. The design flexibility of the F125 will double the availability of the current German frigate fleet (Kamerman, 2015) while simultaneously reducing overhead.

The F125 will take advantage of MEKO technology. MEKO designs rely heavily on modularity that increases the speed at which the ship can be built and facilitates faster upgrades and refits. The F125 will feature weapon modules, electronic modules, mast modules, and a modular combat system with standard interfaces (Kamerman, 2015). Given the flexibility in the design, the F125 readily translates into an exportable frigate design within the MEKO family: the MEKO A-400 Generic Evolved MOTs Multi-Role Frigate. The MEKO A-400 will be built on the same class-standard hull with the same machinery as the F125 frigate but offers foreign navies the flexibility to specify any combination of combat systems from any supplier resulting in more than 80% commonality between the two classes of ships (Kamerman, 2015). This commonality creates a larger fleet of ships from which to draw resources, technical knowledge, and maintenance upgrades.

French Navy

Similar to the RDN, the French Navy has made substantial strides over the last decade to replace three separate aging fleets with two smaller, state-of-the-art, flexible and adaptable fleets of frigates. The Frégate Européenne Multi-Mission (FREMM) was a joint venture between the Italian and French navies, built and designed by the Direction des Constructions Navales Services (DCNS, a French naval defense company) and Orizzonte Sistemi Navali with Fincatieri and Finmeccania ("FREMM European," 2017). These highly modular frigate designs allowed the French, Italians, and potential international clients a choice of equipment with regards to weapons and combat systems (Cavas & Tran, 2016). The newer Frégate de Taille Intermédiaire (FTI), specific to the French Navy, was unveiled in October 2016 (Peruzzi, Scott, & Pape, 2016). Designed by DCNS, it promotes modular design with potential international appeal (Cavas & Tran, 2016).

Aquitaine Class

The Aquitaine class FREMM frigates designed for the French will replace nine D'Estienne d'Orves class avisos (A69 Type Aviso) and nine Tourville and Georges Legues class anti-submarine frigates. The modular design of the FREMM vessels allowed the French Navy to choose between two mission versions: a land attack version with torpedoes, vertical launch system, and cruise missiles or an anti-submarine (ASW) version fitted with torpedoes, vertical launch system, and an active towed array sonar ("FREMM European," n.d.). The French government originally committed to 17 FREMMs, but defense budget cuts reduced the class to 11 and then ultimately eight vessels. The French Navy has committed to building two FREMMs in the land attack configuration and six in the anti-submarine configuration. Aquitaine (D 650) was commissioned in November 2012, Provence (D 652) was commissioned in June 2015, and Languedoc (D 653) was commissioned in March 2016, each configured to ASW (Tomkins, 2016).

The French FREMM is 142 m in length, has a beam of 20 m, displaces 6,000 tons, and carries a crew of 108 ("FREMM European," n.d.). "The frigate's layout has been designed to provide sufficient size for operational effectiveness, maintainability and sustained upgrades. The layout incorporates increased headroom between decks, deeper and longer engine compartments and larger equipment pathways for access and maintenance" ("FREMM European," n.d.).

Both the land attack and anti-submarine versions of the Aquitaine class feature the MBDA Exocet MM40 Block III for anti-ship and littoral attack capability and the MBDA Aster 15 and Aster 30 for air defense. The land attack vessels will also be equipped with MDBA SCALP naval cruise missiles. Additionally, both versions of the frigate boast an aft helicopter hangar and deck encompassing 520 m2 while the land attack frigates "are fitted for a tactical unmanned air vehicle and have the capability to control long-endurance, medium and high-altitude unmanned air vehicles launched from ground sites or from other platforms" ("FREMM European," n.d.).

Similar to the Danish Absalon and Iver Huitfeldt classes, the Aquitaine class Combat Information Center (CIC) features a high-speed data network with an open architecture that will enable future weapon systems to be integrated into the frigates ("FREMM European," n.d.) With external communication equipment compliant with NATO standards, French FREMMs can operate on Link 11, Link 16, Link 22, and JSAT tactical data link ("FREMM European," n.d.). This international NATO co-operability has resulted in the Aquitaine and

Provence participating in joint exercises with the U.S. Navy's Task Force 50 in the Persian-Arabian Gulf (Tomkins, 2016).

The design features of the FREMM have taken into account a flexible and adaptable modular build that allows for future growth in technology at a sustainable cost. Given choices between the various mission sets, growth margins for upgrades, and a relatively small and manageable crew size, FREMM is a viable option for a multitude of foreign navies.

Royal Australian Navy

Currently, the Royal Australian Navy (RAN) utilizes the Anzac class of frigates as its primary anti-submarine warfare platform. Built by Tenix Defense Systems (now part of BAE Systems Australia), eight were commissioned for the RAN between 1996 and 2006, and two were commissioned for the Royal New Zealand Navy in 1997 and 1999 ("Anzac Class Frigate," n.d.). "Anzac frigates are long-range escorts with roles that include air defense, anti-submarine warfare and surveillance" (Kerr, n.d.). The Anzac class displaces 3,600 tons fully loaded, has a length of 118 m with a beam of 14.8 m, and carries a crew of 174 personnel. The design of the Anzac is "based on the Blohm + Voss MEKO 200 modular design which utilizes a basic hull and construction concept to provide flexibility in the choice of command and control, weapons, equipment and sensors" ("Anzac Class Frigate," n.d.). Given the success of the Anzac frigates, the RAN is moving forward with a new class of frigates that will need to incorporate a flexible and adaptable design to meet the growing demand for an efficient, sophisticated, and technologically advanced warship.

The new Future Frigate initiative launched by the Royal Australian Navy is known as the SEA5000 Program. Anticipating an increased military presence in the Asia-Pacific region from both non-state and state actors by 2035, the RAN will need a frigate capable of deterrence and power projection (Goldsmith, 2016). SEA5000 "will oversee the acquisition of nine high-capability Future Frigates and these major surface combatants will be capable of Anti-Air Warfare (AAW), Anti-Surface Warfare (ASuW), with a strong emphasis on Anti-Submarine Warfare (ASW)" (Goldsmith, 2016).

FASO/MAS at PEO-SHIPS: Flexibility on Guided Missile Destroyers

DDG 51 Flight III

The Arleigh Burke class of guided missile destroyers (DDG) is the U.S. Navy's first class of destroyer built around the Aegis Combat System and the SPY-1D multi-function passive electronically scanned array radar. The class is named for Admiral Arleigh Burke, the most famous American destroyer officer of World War II and later Chief of Naval Operations. The class leader, USS *Arleigh Burke*, was commissioned during Admiral Burke's lifetime (Office of the Director, Operational Test and Evaluation [ODOT&E], 2013).

The DDG class ships were designed as multi-mission destroyers to fit the AAW role with their powerful Aegis radar and surface-to-air missiles; the ASW role with their towed sonar array, anti-submarine rockets, and ASW helicopter; the ASUW role with their Harpoon missile launcher; and the strategic land strike role with their Tomahawk missiles. With upgrades to AN/SPY-1 phased radar systems and their associated missile payloads, as part of the Aegis Ballistic Missile Defense System, members of this class have also begun to demonstrate some promise as mobile anti-ballistic missile and anti-satellite weaponry platforms. Some versions of the class no longer have the towed sonar or Harpoon missile launcher (ODOT&E, 2013).

The DDG 51 class destroyers have been designed to support carrier strike groups, surface action groups, amphibious groups, and replenishment groups. They perform primarily AAW with secondary land attack, ASW, and ASUW capabilities. The MK 41 vertical launch system has expanded the role of the destroyers in strike warfare, as well as their overall performance.

The U.S. Navy will use the DDG 51 Flight III Destroyer equipped with the Aegis Modernization program and AMDR to provide joint battlespace threat awareness and defense capability to counter current and future threats in support of joint forces ashore and afloat.

Step 1: Identification of FASO/MAS Options

The following provides two high-level examples of identifying and framing strategic flexibility options in the DDG 51 and DDG1000 environments. These are only notional examples with rough order magnitude values to illustrate the options framing approach.

Power Plant Options

This real options example illustrates the implications of the standard LM2500 GE Marine Gas Turbines for DDG 51 FLT III ships versus the Rolls-Royce MT30 Marine Gas Turbine Engines for the Zumwalt DDG 1000, where the latter can satisfy large power requirements in warships. The LM2500 provides 105,000 shaft hp for a four-engine plant. In comparison, the MT30 can generate upwards of 35.4 MW, and its auxiliary RR4500 Rolls-Royce turbine generators can produce an added 3.8 MW, and each DDG1000 carries two MT30s and two RR4500s. This means that the combined energy output from the Zumwalt can fulfil the electricity demands in a small- to medium-sized city. In contrast, two LM2500 gas turbines can only produce a total of 95.2 kW, which is approximately 0.12%, or 1/825, of the power the Zumwalt can produce. Manufacturer specifications indicate that the LM2500 has an associated Cost/kW of energy of $0.34 and the MT30 Cost/kW is $0.37. In addition, the MT30 prevents warships from running off balance when an engine cannot be restarted until it has cooled down, as is the case in the LM2500.

Figure 1 illustrates a real options strategy tree with four mutually exclusive paths. Additional strategies and pathways can be similarly created, but these initial strategies are sufficient to illustrate the options framing approach. Path 1 shows the As-Is strategy, where no additional higher capacity power plant is used; that is, only two standard LM2500 units are deployed, maintain zero design margins for growth, and only the requirements for the current ship configuration are designed and built. Medium and large upgrades will require major ship alterations, with high cost and delayed schedule. Path 2 implements the two required LM2500 units with additional and sufficient growth margins for one MT30 power plant but currently only with a smaller power plant incorporated into the design. Sufficient area or modularity is available where parts of the machinery can be removed and replaced with the higher energy production unit if needed. Upfront cost is reduced, and future cost and schedule delays are also reduced. Path 3 is to have two prebuilt MT30s and RR4500s initially. While providing the fastest implementation pathway, the cost is higher in the beginning, but total cost is lower if indeed higher energy weapons will be implemented. Path 4 is an option to switch whereby one LM2500 is built with one MT30 unit. Depending on conditions, either the LM2500 or MT30 will be used (switched between units). When higher-powered future weapons such as electromagnetic railguns (E.M. Rail Guns) or high-intensity lasers (H. I. Lasers) as well as other similarly futuristic weapons and systems are required, the MT30 can be turned on.

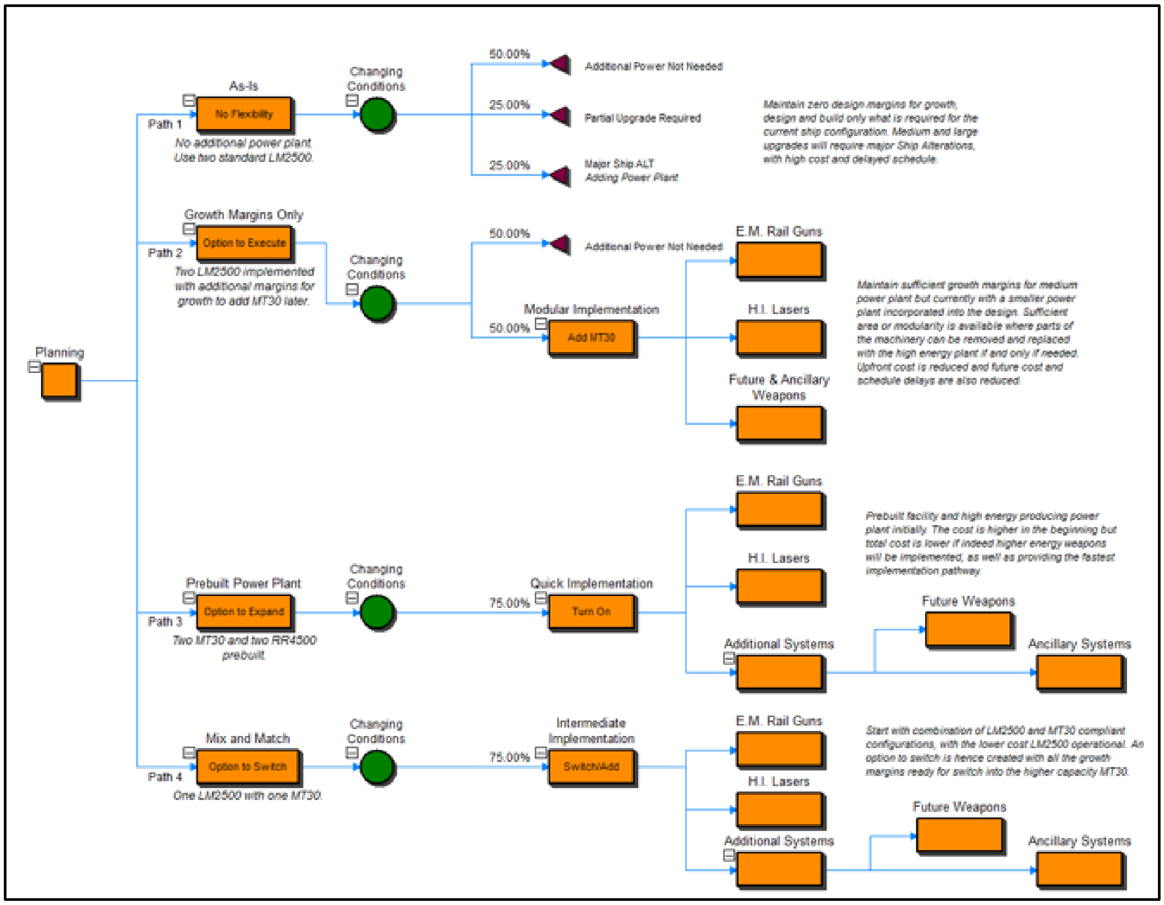


Figure 1. Options Framing on Power Generation

Having a warship flexibility with two LM2500s (As-Is base case) allows the Navy a savings of $31.76 million by deferring the option of the other two additional LM2500s. Therefore, having a flexible ship, the Navy can invest later in one LM2500 and attach another MT30 (preventing any engine off-balance effects when the engines cannot be restarted due to excessive heat) and can save $34.58 million. The usage of options to defer/invest that combine gas turbine specifications allows the Navy to prevent high sunk costs, properly adjusting the true kW requirements, and allows different combinations of propulsion and energy plants. This analysis can be further extended into any direction as needed based on ship designs and Navy requirements.

Step 2: Cost Analysis and Data Gathering

Once the various FASO/MASO options are framed and modeled, as shown in the previous step, the modeling process continues with additional data gathering activities. Figure 2 shows some examples of shadow revenues (i.e., cost savings from lowered cost of future upgrades and technology insertions; costs mitigated by reducing the need for alternative equipment and lower spare parts; and other costs deferred by reducing the need for maintenance and operating costs) or costs savings, additional direct and indirect costs of implementing the new option, and capital requirements.

Year	2016	2017	2018	2019	2020	2021	2022	2023	...	2041	2042	2043
Revenues	$1,742.51	$11,737.14	$225,850.13	$225,850.13	$225,850.13	$225,850.13	$225,850.13	$225,850.13	...	$235,437.44	$235,437.44	$235,437.44
Cost Savings (Future Upgrades and Insertion)	$1,132.63	$7,629.14	$146,802.58	$146,802.58	$146,802.58	$146,802.58	$146,802.58	$146,802.58	...	$153,034.34	$153,034.34	$153,034.34
Cost Mitigated (Alternative Equipment)	$522.75	$3,521.14	$67,755.04	$67,755.04	$67,755.04	$67,755.04	$67,755.04	$67,755.04	...	$70,631.23	$70,631.23	$70,631.23
Cost Deferred (Maintenance and Operations)	$87.13	$586.86	$11,292.51	$11,292.51	$11,292.51	$11,292.51	$11,292.51	$11,292.51	...	$11,771.87	$11,771.87	$11,771.87
Direct Costs	$1,141.09	$1,141.09	$26,392.75	$26,392.75	$26,392.75	$26,456.81	$27,888.82	$27,888.82	...	$32,021.41	$32,021.41	$32,021.41
Direct Expenses	$1,110.26	$1,110.26	$24,896.68	$24,896.68	$24,896.68	$24,896.68	$24,896.68	$24,896.68	...	$25,961.75	$25,961.75	$25,961.75
Operational Costs	$18.50	$18.50	$414.95	$414.95	$414.95	$453.38	$829.89	$829.89	...	$1,730.79	$1,730.79	$1,730.79
Maintenance	$12.33	$12.33	$25.62	$25.62	$25.62	$51.25	$51.25	$51.25	...	$106.87	$106.87	$106.87
Direct Expenses	$0.00	$0.00	$1,055.50	$1,055.50	$1,055.50	$1,055.50	$2,111.00	$2,111.00	...	$4,222.00	$4,222.00	$4,222.00
Gross Profit (Operating Income)	$601.42	$10,596.05	$199,457.38	$199,457.38	$199,457.38	$199,393.32	$197,961.31	$197,961.31	...	$203,416.03	$203,416.03	$203,416.03
Indirect Expenses (General & Administrative)	$799.42	$3,073.28	$9,212.61	$9,212.61	$9,212.61	$9,212.61	$9,212.61	$10,877.49	...	$12,259.92	$9,465.41	$9,465.41
Training and Administrative	$0.00	$31.00	$703.00	$703.00	$703.00	$703.00	$703.00	$703.00	...	$733.00	$733.00	$733.00
Contracts and Bidding	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Operations	$0.00	$0.00	$1,248.07	$1,248.07	$1,248.07	$1,248.07	$1,248.07	$1,248.07	...	$1,248.07	$1,248.07	$1,248.07
Maintenance	$799.42	$2,997.82	$4,758.48	$4,758.48	$4,758.48	$4,758.48	$4,758.48	$6,423.36	...	$7,733.14	$4,938.63	$4,938.63
Parts and Service	$0.00	$0.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	$1,506.00	...	$1,506.00	$1,506.00	$1,506.00
Miscellaneous	$0.00	$44.46	$997.06	$997.06	$997.06	$997.06	$997.06	$997.06	...	$1,039.71	$1,039.71	$1,039.71
EBITDA: Earnings Before Interest, Taxes, Depreciation, and Amortization	($198.00)	$7,522.77	$190,244.77	$190,244.77	$190,244.77	$190,180.71	$188,748.70	$187,083.82	...	$191,156.11	$193,950.62	$193,950.62
Depreciation	$0.00	$9,874.00	$39,827.00	$39,074.00	$38,161.00	$37,206.00	$36,172.00	$35,223.00	...	$24,502.00	$23,977.00	$23,444.00
Amortization	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
EBIT: Earnings Before Interest and Taxes	($198.00)	($2,351.23)	$150,417.77	$151,170.77	$152,083.77	$152,974.71	$152,576.70	$151,860.82	...	$166,654.11	$169,973.62	$170,506.62
Interest	$0.00	$6,779.32	$25,892.66	$22,767.15	$19,224.35	$15,842.53	$13,062.00	$12,303.79	...	$653.99	$666.90	$667.48
EBT: Earnings Before Taxes	($198.00)	($9,130.55)	$124,525.11	$128,403.62	$132,859.42	$137,132.18	$139,514.70	$139,557.03	...	$166,000.12	$169,306.72	$169,839.14
Corporate Taxes	($56.43)	($2,602.21)	$35,489.66	$36,595.03	$37,864.93	$39,082.67	$39,761.69	$39,773.75	...	$47,310.03	$48,252.42	$48,404.15
NET INCOME	($141.57)	($6,528.34)	$89,035.45	$91,808.59	$94,994.49	$98,049.51	$99,753.01	$99,783.28	...	$118,690.09	$121,054.30	$121,434.99
Total Noncash Expense Items	$0.00	$9,874.00	$39,827.00	$39,074.00	$38,161.00	$37,206.00	$36,172.00	$35,223.00	...	$24,502.00	$23,977.00	$23,444.00
Change in Net Working Capital	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Capital Expenditures	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Other Noncash Expenses	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
Total Gross Invested Operating Capital	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
CAPITAL INVESTMENTS	$250,000.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	...	$0.00	$0.00	$0.00
NET OPERATING PROFIT AFTER TAXES (NOPAT)	($141.57)	($1,681.13)	$107,548.71	$108,087.10	$108,739.90	$109,376.92	$109,092.34	$108,580.49	...	$119,157.69	$121,531.14	$121,912.23
NET CASH FLOW (NCF)	($141.57)	$3,345.66	$128,862.45	$130,882.59	$133,155.49	$135,255.51	$135,925.01	$135,006.28	...	$143,192.09	$145,031.30	$144,878.99
OPERATING CASH FLOW (OCF)	($141.57)	$8,192.87	$147,375.71	$147,161.10	$146,900.90	$146,582.92	$145,264.34	$143,803.49	...	$143,659.69	$145,508.14	$145,356.23
FREE CASH FLOW (FCF)	($141.57)	$8,192.87	$147,375.71	$147,161.10	$146,900.90	$146,582.92	$145,264.34	$143,803.49	...	$143,659.69	$145,508.14	$145,356.23

Figure 2. Financial and Economic Cost Savings and Cost Averted Cash Flow Model

Step 3: Financial Modeling

The *Discounted Cash Flow* section, shown in Figure 3, is at the heart of the input assumptions for the analysis. Additional time-series inputs are entered in the data grid as required, while some elements of this grid are intermediate computed values.

Year	2016	2017	2018	2019	2020	2021	2022	2023	2024	...	2040	2041	2042	2043
EARNINGS BEFORE INT, TAX, DEP, AMORT (EBITDA)	-198	7,522.77	190,244.77	190,244.77	190,244.77	190,180.71	188,748.70	187,083.82	188,393.60	...	193,775.67	191,156.11	193,950.62	193,950.62
EARNINGS BEFORE INTEREST AND TAXES (EBIT)	-198	-2,351.23	150,417.77	151,170.77	152,083.77	152,974.71	152,576.70	151,860.82	153,915.60	...	168,829.67	166,654.11	169,973.62	170,506.62
NET INCOME (NI)	-141.57	-6,528.34	89,035.45	91,808.59	94,994.49	98,049.51	99,753.01	99,783.28	101,776.23	...	120,236.60	118,690.09	121,054.30	121,434.99
NET OPERATING PROFIT AFTER TAXES (NOPAT)	-141.57	-1,681.13	107,548.71	108,087.10	108,739.90	109,376.92	109,092.34	108,580.49	110,049.65	...	120,713.21	119,157.69	121,531.14	121,912.23
NET CASH FLOW (NCF)	-141.57	3,345.66	128,862.45	130,882.59	133,155.49	135,255.51	135,925.01	135,006.28	136,254.23	...	145,182.60	143,192.09	145,031.30	144,878.99
OPERATING CASH FLOW (OCF)	-141.57	8,192.87	147,375.71	147,161.10	146,900.90	146,582.92	145,264.34	143,803.49	144,527.65	...	145,659.21	143,659.69	145,508.14	145,356.23
FREE CASH FLOW (FCF)	-141.57	8,192.87	147,375.71	147,161.10	146,900.90	146,582.92	145,264.34	143,803.49	144,527.65	...	145,659.21	143,659.69	145,508.14	145,356.23
RETURN ON INVESTED CAPITAL (ROIC)	-0.11%	-1.31%	83.68%	84.09%	84.60%	85.10%	84.88%	84.48%	85.62%	...	93.92%	92.71%	94.55%	94.85%
ECONOMIC VALUE ADDED (EVA)	-12,994.67	-14,534.23	94,695.61	95,234.00	95,886.80	96,523.82	96,239.24	95,727.39	97,196.55	...	107,860.11	106,304.59	108,678.04	109,059.13
TIMES INTEREST EARNED (TIE)	0	-0.35	5.81	6.64	7.91	9.66	11.68	12.34	13.3	...	253.27	254.83	254.87	255.45
NET PROFIT MARGIN (NPM)	-8.12%	-55.62%	39.42%	40.65%	42.06%	43.41%	44.17%	44.18%	45.06%	...	51.07%	50.41%	51.42%	51.58%
OPERATING PROFIT MARGIN (OPM)	-8.12%	-14.32%	47.62%	47.86%	48.15%	48.43%	48.30%	48.08%	48.73%	...	51.27%	50.61%	51.62%	51.78%
EARNINGS PER SHARE (EPS)	0	0	0	0	0	0	0	0	0	...	0	0	0	0
INVENTORY TURNOVER (IT)	2.58	17.35	333.8	333.8	333.8	333.8	333.8	333.8	333.8	...	347.97	347.97	347.97	347.97
DAYS SALES OUTSTANDING (DSO)	841.22	124.89	6.49	6.49	6.49	6.49	6.49	6.49	6.49	...	6.23	6.23	6.23	6.23
TOTAL ASSET TURNOVER (TAT)	1.19%	7.99%	153.74%	153.74%	153.74%	153.74%	153.74%	153.74%	153.74%	...	160.27%	160.27%	160.27%	160.27%
BASIC EARNING POWER (BEP)	-0.13%	-1.60%	102.39%	102.91%	103.53%	104.13%	103.86%	103.38%	104.78%	...	114.93%	113.45%	115.71%	116.07%
PRICE TO EARNINGS RATIO (PE)	-215,961,301.34	-4,683,215.98	343,387.27	333,015.05	321,846.49	311,818.41	306,493.42	306,400.46	300,400.60	...	254,279.00	257,592.21	252,561.37	251,769.63
RETURN ON ASSET (ROA)	-0.10%	-4.44%	60.61%	62.50%	64.67%	66.75%	67.90%	67.93%	69.28%	...	81.85%	80.80%	82.41%	82.66%
RETURN ON EQUITY (ROE)	-0.20%	-9.26%	126.24%	130.17%	134.69%	139.02%	141.43%	141.48%	144.30%	...	170.48%	168.28%	171.64%	172.17%

Figure 3. Financial and Economic Performance Ratios

Static Portfolio Analysis and Comparisons of Multiple Projects

Analysts can also identify and create the various options, and compute the economic and financial results such as net present value (NPV), internal rate of return (IRR), modified internal rate of return (MIRR), profitability index (PI), return on investment (ROI), payback period (PP), and discounted payback (DPP). This is shown in Figure 4, complete with various charts, cash flow ratios and models, intermediate calculations, and comparisons of the options within a portfolio view, as illustrated in the figure. As a side note, the term Option is used to represent a generic analysis option, where each project can be a different asset, project, acquisition, investment, research and development, or simply variations of the same investment (e.g., different financing methods when acquiring the same firm, different market conditions and outcomes, or different scenarios or implementation paths). Therefore, the more flexible terminology of *Project* is adopted instead.

Figure 4 illustrates the *Portfolio Analysis* of multiple *Projects*. This Portfolio Analysis returns the computed economic and financial indicators such as NPV, IRR, MIRR, PI, ROI, PP, and DPP for all the projects combined into a portfolio view (these results can be stand-alone with no base case or computed as incremental values above and beyond the chosen base case). The *Economic Results* (Level 3) subtabs show the individual project's economic and financial indicators, whereas this Level 2 *Portfolio Analysis* view shows the results of all projects' indicators and compares them side by side. There are also two charts available for comparing these individual projects' results. The *Portfolio Analysis* is used to obtain a side-by-side comparison of all the main economic and financial indicators of all the projects at once. For instance, analysts can compare all the NPVs from each project in a single results grid. The bubble chart on the left provides a visual representation of up to three chosen variables at once (e.g., the y-axis shows the IRR, the x-axis represents the NPV, and the size of the bubble may represent the capital investment; in such a situation, one would prefer a smaller bubble that is in the top right quadrant of the chart).

Economic Results	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Option 9	Option 10
Net Present Value (NPV)	$608,388.34	$427,132.76	$40,765.22	$41,613.74	($10,610.44)	($23,774.85)	$728,339.38	$554,258.99	$31,837.41	$46,377.25
Net Present Value (NPV) with Terminal Value	72648877.00%	54046710.00%	7033594.00%	6615455.00%	1525722.00%	2718633.00%	112457959.00%	74402419.00%	16876439.00%	12973950.00%
Internal Rate of Return (IRR)	29.31%	11.17%	16.21%	19.21%	8.55%	6.76%	11.20%	13.74%	9.29%	14.77%
Modified Internal Rate of Return (MIRR)	0.15	0.10	0.12	0.13	0.09	0.08	0.10	0.11	0.07	0.10
Profitability Index (PI)	343.00%	114.00%	137.00%	154.00%	90.00%	86.00%	109.00%	129.00%	134.00%	176.00%
Return on Investment (ROI)	2.43	0.14	0.37	0.54	-0.10	-0.14	0.09	0.29	0.34	0.76
Payback Period (PP)	3.80	10.82	6.38	5.45	10.32	9.78	9.98	7.87	9.19	7.30
Discounted Payback Period (DPP)	$4.80	$22.81	$9.80	$7.84			$22.35	$13.79	$12.18	$9.21
Show on Charts										

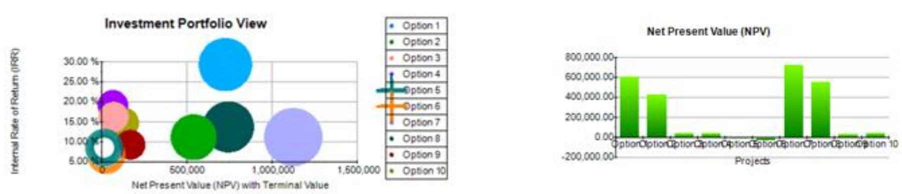


Figure 4. Static Portfolio Analysis

Step 4: Tornado and Sensitivity Analytics

Figure 5 illustrates the *Applied Analytics* results, which allows analysts to run *Tornado Analysis* and *Scenario Analysis* on any one of the projects previously modeled—the analytics cover all the various projects and options. We can, therefore, run tornado or scenario analyses on any one of the projects or options. Tornado analysis is a static sensitivity analysis of the selected model's output to each input assumption, performed one at a time, and ranked from most impactful to least impactful. We can start the analysis by first choosing the output variable to test.

We used the default sensitivity settings of ±10% on each input assumption to test and decide how many input variables to chart (large models with many inputs may generate unsightly and less useful charts, whereas showing just the top variables reveals more information through a more elegant chart). The sensitivity run was based on the input assumptions as unique inputs, but the inputs can also be grouped as a line item (all individual inputs on a single line item are assumed to be one variable), or the analysis can be run as variable groups (e.g., all line items under *Revenue* will be assumed to be a single variable).

		Option 1: Net Present Value (NPV)	Base Value:	608,388.34				
% UP	% DOWN	INPUT PRECEDENCE	Output Downside	Output Upside	Effective Range	Input Downside	Input Upside	Base Case Value
10.00%	10.00%	Revenues	471,501.71	745,274.96	273,773.25	5,419,480.71	6,623,809.75	6,021,645.23
10.00%	10.00%	DCF | Discount Rate (%)	694,674.49	533,487.10	161,187.39	0.09	0.11	0.10
10.00%	10.00%	DCF | Marginal Tax Rate (%)	642,603.81	574,172.86	68,430.96	0.26	0.31	0.29
10.00%	10.00%	DCF | CAPITAL INVESTMENTS	633,388.34	583,388.34	50,000.00	225,000.00	275,000.00	250,000.00
10.00%	10.00%	DCF | Depreciation	629,216.94	587,559.73	41,657.22	726,039.90	887,382.10	806,711.00
10.00%	10.00%	Direct Costs	625,471.60	591,305.07	34,166.54	700,308.60	855,932.74	778,120.67
10.00%	10.00%	DCF | Interest	615,236.58	601,540.09	13,696.49	140,594.88	171,838.18	156,216.53
10.00%	10.00%	Indirect Expenses	614,676.01	602,100.66	12,575.34	249,962.98	305,510.31	277,736.65
10.00%	10.00%	DCF | Change in Net Working Capital	608,388.34	608,388.34	0.00	0.00	0.00	0.00
10.00%	10.00%	CFR | Accounts Receivables	608,388.34	608,388.34	0.00	3,614.40	4,417.60	4,016.00
10.00%	10.00%	CFR | Total Inventories	608,388.34	608,388.34	0.00	608.95	744.27	676.61
10.00%	10.00%	CFR | Long-Term Operating Assets	608,388.34	608,388.34	0.00	102,685.50	125,504.50	114,095.00
10.00%	10.00%	CFR | Current Liabilities	608,388.34	608,388.34	0.00	16,533.00	20,207.00	18,370.00
10.00%	10.00%	CFR | Current Asset	608,388.34	608,388.34	0.00	29,525.40	36,086.60	32,806.00
10.00%	10.00%	CFR | Shares Outstanding	608,388.34	608,388.34	0.00	1,019,121,381.00	1,245,592,799.00	1,132,357,090.00
10.00%	10.00%	CFR | Stock Price Per Share	608,388.34	608,388.34	0.00	24.30	29.70	27.00
10.00%	10.00%	CFR | Total Assets	608,388.34	608,388.34	0.00	132,210.90	161,591.10	146,901.00
10.00%	10.00%	DCF | Capital Expenditures	608,388.34	608,388.34	0.00	0.00	0.00	0.00
10.00%	10.00%	DCF | Amortization	608,388.34	608,388.34	0.00	0.00	0.00	0.00
10.00%	10.00%	CFR | Common Equity	608,388.34	608,388.34	0.00	63,477.00	77,583.00	70,530.00
10.00%	10.00%	Growth Rate	608,388.34	608,388.34	0.00	0.02	0.02	0.02
10.00%	10.00%	DCF | Other Noncash Expenses	608,388.34	608,388.34	0.00	0.00	0.00	0.00
10.00%	10.00%	CFR | Total Debt	608,388.34	608,388.34	0.00	52,200.90	63,801.10	58,001.00
10.00%	10.00%	DCF | Total Gross Invested Operating Capital	608,388.34	608,388.34	0.00	0.00	0.00	0.00
10.00%	10.00%	CFR | Total Net Operating Capital	608,388.34	608,388.34	0.00	115,677.90	141,384.10	128,531.00

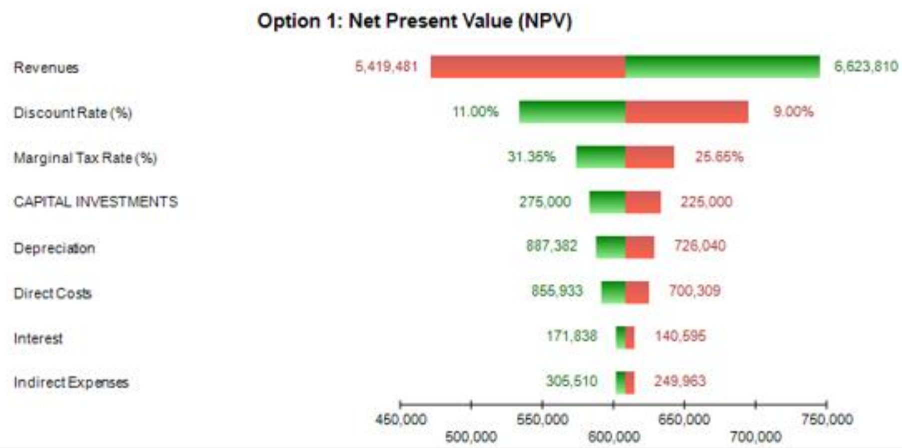


Figure 5. Applied Analytics—Tornado Analysis

Step 5: Monte Carlo Risk Simulation

Figures 6 and 7 illustrate the *Risk Simulation* analysis, where Monte Carlo risk simulations can be set up and run. Analysts can set up probability distribution assumptions on any combinations of inputs, run a risk simulation tens to hundreds of thousands of trials, and retrieve the simulated forecast outputs as charts, statistics, probabilities, and confidence intervals to develop comprehensive risk profiles of the projects.

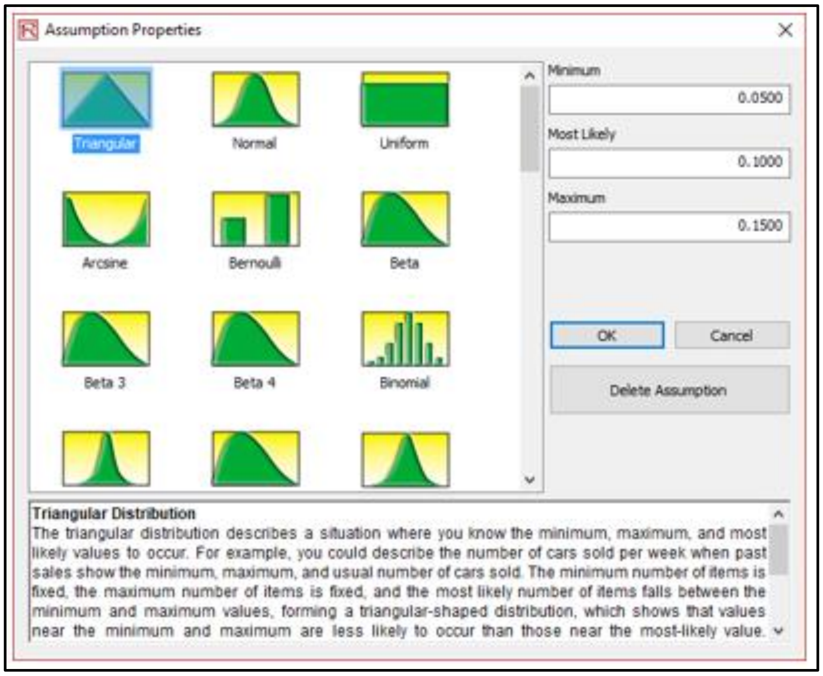


Figure 6. Risk Simulation Input Assumptions

Statistics/Percentiles	Value
Trials	10,000.00
Mean	1,264,569.20
Median	1,223,025.65
Stdev	323,440.89
CV	0.26
Skew	0.59
Kurtosis	0.07
Minimum	528,515.81
Maximum	2,690,456.43
Range	2,161,940.62
0.00%	528,515.81
5.00%	806,158.76
10.00%	873,477.43
20%	$975,191
30%	1,066,998.73
40.00%	1,146,536
50.00%	1,223,025.65
60.00%	1,310,131.64
70.00%	1,408,675.46
80.00%	152992680.50%
90.00%	1,711,752.31
95.00%	1,871,099.08
100.00%	2,690,456.43

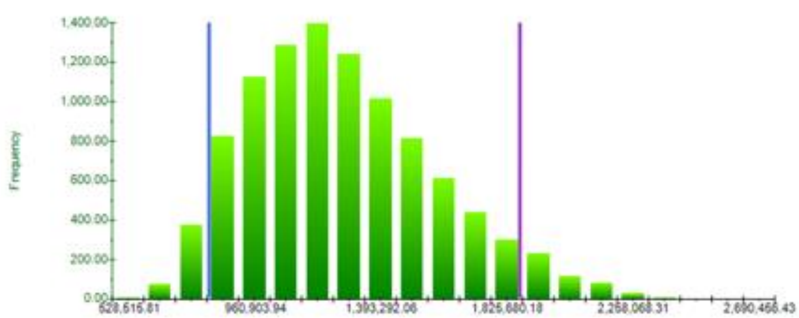


Figure 7. Risk Simulation Results

Analysis of Alternatives and Dynamic Sensitivity Analysis

Figure 8 illustrates the *Analysis of Alternatives* results. Whereas the *Overlay Results* shows the simulated results as charts (PDF/CDF), the *Analysis of Alternatives* shows the results of the simulation statistics in a table format as well as a chart of the statistics such that one project can be compared against another. The standard approach is to run an analysis of alternatives to compare one project versus another, but analysts can also choose to analyze the results on an *Incremental Analysis* basis.

OPTIONS	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Option 7	Option 8	Option 9	Option 10
Acronym										
Mean	1,264,569.20	-129,548.71	115,517.52	104,069.16	186,548.19	50,886.00	982,026.93	2,222,255.83	120,139.43	115,842.87
Median	1,223,025.65	-269,130.82	111,946.33	101,647.69	182,533.06	48,903.31	809,951.34	2,125,984.32	118,821.08	113,873.98
Stdev	323,440.89	756,980.40	33,246.01	27,242.55	42,751.66	24,260.40	1,841,083.18	774,387.88	22,748.49	26,884.24
CV	25.58%	584.32%	28.78%	26.18%	22.92%	47.68%	187.48%	34.85%	18.94%	23.21%
Skew	0.59	0.83	0.50	0.46	0.46	0.41	0.48	0.64	0.31	0.38
Kurtosis	0.07	0.42	-0.06	-0.02	-0.05	-0.06	0.18	0.29	-0.06	-0.02
Minimum	528,515.81	-1,612,389.19	33,915.58	34,781.88	72,360.45	-13,544.30	-4,062,019.59	409,940.95	53,294.46	45,687.22
Maximum	2,690,456.43	3,186,736.54	238,613.77	208,832.14	361,534.33	150,216.66	8,105,902.62	5,492,144.74	213,754.98	234,987.45
Range	2,161,940.62	4,799,125.72	204,698.19	174,050.27	289,173.87	163,760.96	12,167,922.22	5,082,203.79	160,460.52	189,300.23
0% Percentile	528,515.81	-1,612,389.19	33,915.58	34,781.88	72,360.45	-13,544.30	-4,062,019.59	409,940.95	53,294.46	45,687.22
5% Percentile	806,158.76	-1,126,810.37	66,833.21	63,341.55	123,647.00	14,407.25	-1,721,563.12	1,118,422.25	85,021.02	74,880.17
10% Percentile	873,477.43	-985,321.30	75,292.44	70,536.07	134,523.19	21,061.66	-1,234,281.88	1,308,170.97	91,480.18	82,478.56
20% Percentile	975,191.16	-787,408.29	86,592.52	80,551.23	149,044.20	29,596.95	-575,471.45	1,553,575.46	100,319.83	92,352.05
30% Percentile	1,066,998.73	-607,802.26	95,258.85	87,689.17	160,430.73	36,309.09	-88,078.13	1,756,505.72	107,072.89	100,314.71
40% Percentile	1,146,536.48	-435,048.39	103,478.90	94,871.67	171,896.55	42,716.18	366,415.18	1,937,629.66	113,197.08	106,958.95
50% Percentile	1,223,025.65	-269,130.82	111,946.33	101,647.69	182,533.06	48,903.31	809,951.34	2,125,984.32	118,821.08	113,873.98
60% Percentile	1,310,131.64	-70,640.29	120,924.23	108,733.13	193,556.04	55,280.95	1,260,274.23	2,321,587.17	124,955.05	121,112.91
70% Percentile	1,408,675.46	168,159.10	130,549.63	116,448.61	206,334.42	62,431.52	1,795,294.68	2,549,885.15	131,348.03	128,934.06
80% Percentile	1,529,926.80	474,006.49	143,040.93	127,054.19	222,193.22	71,595.89	2,464,689.40	2,848,749.31	139,239.82	138,456.24
90% Percentile	1,711,752.31	956,928.86	161,722.53	141,005.97	245,083.83	83,760.91	3,475,890.85	3,275,508.70	150,031.84	151,426.79
95% Percentile	1,871,099.08	1,333,426.70	177,165.94	153,294.40	263,658.89	93,827.83	4,368,679.78	3,668,268.40	159,781.91	163,466.89
100% Percentile	2,690,456.43	3,186,736.54	238,613.77	208,832.14	361,534.33	150,216.66	8,105,902.62	5,492,144.74	213,754.98	234,987.45

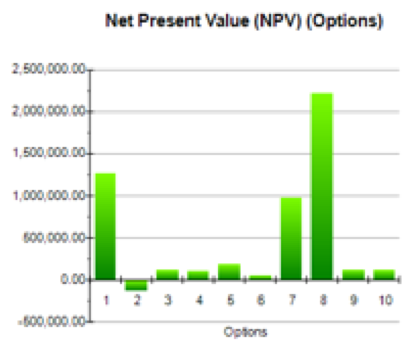


Figure 8. Simulated Analysis of Alternatives

Figure 9 illustrates the *Dynamic Sensitivity Analysis* computations. Tornado analysis and scenario analysis are both static calculations. Dynamic sensitivity, in contrast, is a dynamic analysis, which can only be performed after a simulation is run. Analysts start by selecting the desired project's economic output. Red bars on the *Rank Correlation* chart indicate negative correlations, and green bars indicate positive correlations for the left chart. The correlations' absolute values are used to rank the variables with the highest relationship to the lowest, for all simulation input assumptions. The *Contribution* to *Variance* computations and chart indicate the percentage fluctuation in the output variable that can be statistically explained by the fluctuations in each of the input variables.

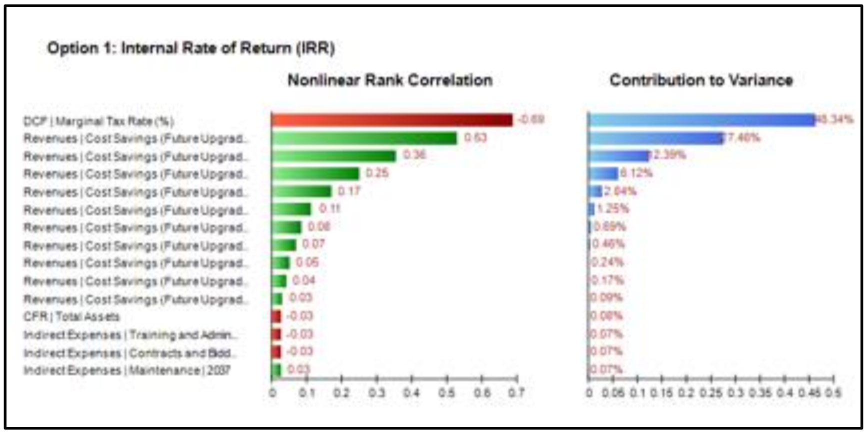


Figure 9. Simulated Dynamic Sensitivity Analysis

Step 6: Strategic Real Options Valuation Modeling

Figure 10 illustrates the *Options Valuation* and the *Strategy View.* This part of the analysis performs the calculations of ROV models. Analysts must understand the basic concepts of real options before proceeding.

American: Option to Abandon	450,355.44	
		Result:
Asset Value (Present Value of Net Benefits):	445,625.18	450,355.44
Volatility (Annualized Risk %):	0.22	
Maturity (Total Years to Option Expiration):	5.00	
Risk-Free Rate (Riskless Discount Rate %):	0.04	
Dividend Rate (Opportunity Cost %):	0.00	
Lattice Steps (Typically 100 to 1000):	100.00	
Salvage:	250,000.00	

Figure 10. Options Valuation

Step 7: Portfolio Optimization

Figure 11 illustrates the *Portfolio Optimization's Optimization* settings and assumptions. In the Portfolio Optimization analysis, the individual projects can be modeled as a portfolio and optimized to determine the best combination of projects for the portfolio. In today's competitive global economy, companies are faced with many difficult decisions. These decisions include allocating financial resources, building or expanding facilities, managing inventories, and determining product-mix strategies. Such decisions might involve thousands or millions of potential alternatives. Considering and evaluating each of them would be impractical and may even be impossible. A model can provide valuable assistance in incorporating relevant variables when analyzing decisions and in finding the best solutions for making decisions. Models capture the most important features of a problem and present them in a form that is easy to interpret. Models often provide insights that intuition alone cannot. An optimization model has three major elements: decision variables, constraints, and an objective. In short, the optimization methodology finds the best combination or permutation of decision variables (e.g., which products to sell or which projects to execute) in every conceivable way such that the objective is maximized (e.g., revenues and net income) or minimized (e.g., risk and costs) while still satisfying the constraints (e.g., budget and resources).

Analysts start by deciding on the optimization method (Static or Dynamic Optimization). Then they select the decision variable type of *Discrete Binary* (choose which Project or Options to execute with a Go/No-Go Binary 1/0 decision) or *Continuous Budget Allocation* (returns % of budget to allocate to each *option* or *project* as long as the total portfolio is 100%), select the *Objective* (Max NPV, Min Risk, etc.), set up any *Constraints* (e.g., budget restrictions, number of projects restrictions, or create customized restrictions), select the options or projects to optimize/allocate/choose (default selection is *all options*), and when completed, run the Optimization.

Figure 11 illustrates the *Optimization Results*, which return the results from the portfolio optimization analysis. The main results are provided in the data grid, showing the final *Objective Function* results, final *Optimized Constraints*, and the allocation, selection, or optimization across all individual options or projects within this optimized portfolio. The typical optimization results chart illustrates the final objective function. The chart will only show a single point for regular optimizations, whereas it will return an investment efficient frontier curve if the optional *Efficient Frontier* settings are set (min, max, step size).

Objective Function	6.1286	6.7465	6.9478	6.9478	6.9478
Frontier Variable	2000000.0000	2500000.0000	3000000.0000	3500000.0000	4000000.0000
Optimized Constraint	1978818.0000	2487042.0000	2718646.0000	2718646.0000	2718646.0000
Option1	1.00	1.00	1.00	1.00	1.00
Option2	0.00	1.00	1.00	1.00	1.00
Option3	1.00	1.00	1.00	1.00	1.00
Option4	1.00	1.00	1.00	1.00	1.00
Option5	1.00	0.00	1.00	1.00	1.00
Option6	0.00	0.00	1.00	1.00	1.00
Option7	0.00	0.00	0.00	0.00	0.00
Option8	1.00	1.00	1.00	1.00	1.00
Option9	0.00	0.00	1.00	1.00	1.00
Option10	0.00	1.00	1.00	1.00	1.00

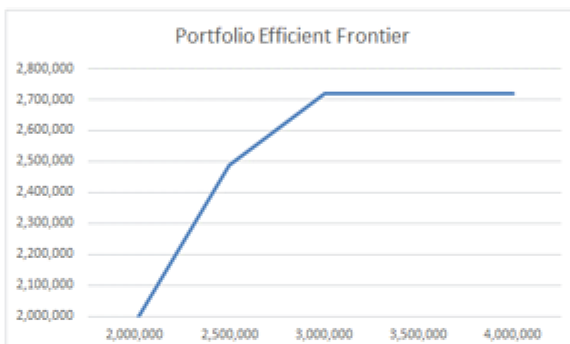


Figure 11. Portfolio Optimization Results

Figures 11 and 12 provide examples of the critical results for decision makers as they allow flexibility in designing their own portfolio of options. For instance, Figure 11 shows an efficient frontier of portfolios, where each of the points along the curve are optimized portfolios subject to a certain set of constraints. In this example, the constraints were the number of options that can be selected in a ship, and the total cost of obtaining these options are subject to a budget constraint. The colored columns on the right in Figure 12 show the various combinations of budget limits and maximum number of options allowed. For instance, if a program office in the Navy only allocates $2.5 million (see the Frontier Variable located on the second row) and no more than four options per ship, then only options 3, 7, 9, and 10 are feasible, and this portfolio combination would generate the highest bang for the buck while simultaneously satisfying the budgetary and number of options constraints. If the constraints were relaxed to, say, five options and a $3.5 million budget, then option 5 is added to the mix. Finally, at $4.5 million and no more than seven options per ship, options 1 and 2 should be added to the mix. Interestingly, even with a higher budget of $5.5 million, the same portfolio of options is selected. In fact, the Optimized Constraint 2 shows that only $4.1 million is used. Therefore, as a decision-making tool for the budget-setting officials, the maximum budget that should be set for this portfolio of options should be $4.1 million. Similarly, the decision-maker can move backwards, where, say, if the original budget of $4.5 million was slashed by the U.S. Congress to $3.5 million, then the options that should be eliminated would be options 1 and 2.

Model	Model 1	Model 2	Model 3	Model 4	Model 5	Count
Objective	1,408,735.73	51.16	53.56	48.10	53.56	
Budget Constraint	3,800,000	4,000,000	4,000,000	3,750,000	4,000,000	
Program Constraint	6	7	7	6	7	
Option 1	1.00	1.00	1.00	0.00	1.00	4
Option 2	0.00	0.00	0.00	0.00	0.00	0
Option 3	1.00	1.00	1.00	1.00	1.00	5
Option 4	0.00	1.00	1.00	0.00	1.00	3
Option 5	1.00	1.00	1.00	1.00	1.00	5
Option 6	0.00	1.00	1.00	1.00	1.00	4
Option 7	1.00	0.00	0.00	0.00	0.00	1
Option 8	0.00	1.00	1.00	1.00	1.00	4
Option 9	1.00	0.00	0.00	1.00	0.00	2
Option 10	1.00	1.00	1.00	1.00	1.00	5

Figure 12. Multi-Criteria Portfolio Optimization Results

While Figure 11 shows the efficient frontier where the constraints such as number of options allowed and budget were varied to determine the efficient portfolio selection, Figure 12 shows multiple portfolios with different objectives. For instance, the five models shown were to maximize the financial bang for the buck (minimizing cost and maximizing value while simultaneously minimizing risk), maximizing OPNAV value, maximizing KVA value, maximizing Command value, and maximizing a Weighted Average of all objectives. This capability is important because depending on who is doing the analysis, their objectives and decisions will differ based on different perspectives. Using a multiple criteria optimization approach allows us to see the scoring from all perspectives. Options with the highest count (e.g., 5) would receive the highest priority in the final portfolio, as it satisfies all stakeholders' perspectives and would hence be considered first, followed by options with counts of 4, 3, 2, and 1.

Conclusions and Recommendations

Key Conclusions and Next Steps

Strategic ROV provides the option holder the right, but not the obligation, to hold off on executing a certain decision until a later time when uncertainties are resolved and when better information is available. The option implies that flexibility to execute a certain path exists and was predetermined or predesigned in advance. Based on the research performed thus far, we conclude that the methodology has significant merits and is worthy of more detailed follow-on analysis. It is therefore recommended that the ROV methodology be applied on a real case facing the Navy, applied with actual data, and the project's outcomes tracked over time.

References

Abbott, J. W., Levine, A., & Vasilakos, J. (2008). Modular/open systems to support ship acquisition strategies. *ASNE Day,* 23–25. Retrieved from http://navalengineers.net/Proceedings/AD08/documents/paper32.pdf

Anzac class frigate, Australia. (n.d.). Retrieved from http://www.naval-technology.com/projects/anzac

Baden-Württemberg (type 125) class. (2017, January 17). Retrieved from https://janes.ihs.com.libproxy.nps.edu/FightingShips/Display/1357283

Cavas, C., & Tran, P. (2016, October 18). France unveils new FTI frigate designed for the French navy and export. Retrieved from http://www.defensenews.com/articles/france-unveils-new-fti-frigate-ship-is-designed-for-the-french-navy-and-for-exportFlyvefisken

Doerry, N. H. (2012, August). Institutionalizing modular adaptable ship technologies. Journal of *Ship Production and Design, 30*(3), 126–141.

Drewry, J. T., & Jons, O. P. (1975, April). Modularity: Maximizing the return on the Navy's investment. *Naval Engineer's Journal, 87*(2), 198–214.

Flyvefisken class (SF 300), Denmark. (n.d.). Retrieved from http://www.naval-technology.com/projects/fly/

FREMM European multimission frigate, France/Italy. (n.d.). Retrieved from http://www.naval-technology.com/projects/fremm/

Goldsmith, S. (2016, May 6). SEA5000 CEP: Critical capability considerations for the future frigates. Retrieved from http://navalinstitute.com.au/sea5000-cep-critical-capability-considerations-for-the-future-frigates/

Jolliff, J. V. (1974, October). Modular ship design concepts. *Naval Engineers Journal*, 11–30.

Kamerman, J. (2015). *Meeting the future: The German experience* [PowerPoint slides]. Retrieved from https://www.aspi.org.au/__data/assets/pdf_file/0016/26503/Kamerman-The-German-experience-slides.pdf

Kerr, J. (n.d.). *Frigate rivals told to think local*. Retrieved from http://specialreports.theaustralian.com.au/541255/frigate-rivals-told-to-think-local/

Koenig, P. C. (2009, April). *Real options in ship and force structure analysis: A research agenda.* Paper presented at the American Society of Naval Engineers ASNE Day.

Koenig, P. C., Czapiewski, P. M., & Hootman, J. C. (2008). Synthesis and analysis of future naval fleets. *Ships and Offshore Structures, 3*(2), 81–89.

Mun, J. C., & Housel, T. J. (2016). *Flexible and adaptable ship options: Assessing the future value of incorporating flexible ships design features in Navy ship concepts*. Monterey, CA: Naval Postgraduate School.

Office of the Director, Operational Test and Evaluation (ODOT&E). (2013). *DDG 51 flight III destroyer/air and missile defense radar (AMDR)/Aegis modernization*. Retrieved from http://www.dote.osd.mil/pub/reports/FY2014/pdf/navy/2014ddg51.pdf

Pape, A. (2016, April 12). Germany's first type 125 frigate begins sea trials. Retrieved from http://www.janes.com/article/59438/germany-s-first-type-125-frigate-begins-sea-trials

Paris, C. M., Brussels, N. F., & Fiorenza, N. (2013, March 25). Complex tradeoffs between specialized and modular combat ships. Retrieved from http://aviationweek.com/awin/complex-tradeoffs-between-specialized-and-modular-combat-ships

Peruzzi, L., Scott, R., & Pape, A. (2016, October 20). Euronaval 2016: French Navy's new frigate design unveiled. Retrieved from http://www.janes.com/article/64763/euronaval-2016-french-navy-s-new-frigate-design-unveiled

Pike, J. (2011, November 11). Flyvefisken-class STANFLEX 300 ships. Retrieved from http://www.globalsecurity.org/military/world/europe/hdms-flyvefisken.htm

Simmons, J. L. (1975, April). Design for change: The impact of changing threats and missions on system design philosophy. *Naval Engineers Journal,* 120–125.

Sturtevant, G. (2015). Fle*xible ships: Affordable relevance over the ship's life cycle.* Retrieved from http://www.asne-tw.org/asne/events/presentation/2015-01Sturtevant.pdf

Thorsteinson, J. (2013). Modular warships. *Canadian Naval Review, 8*(4), 29–30. Retrieved from http://www.navalreview.ca/wp-content/uploads/public/vol8num4/vol8num4art7.pdf

ThyssenKrupp Marine Systems. (n.d.). Retrieved from https://www.thyssenkrupp-marinesystems.com/en/

Tomkins, R. (2016, March 18). Third FREMM frigate delivered for French Navy. Retrieved from http://www.upi.com/Defense-News/2016/03/18/Third-FREMM-frigate-delivered-for-French-Navy/4091458325168/

Operational Seakeeping Considerations in LCU Deployment

Jarema M. Didoszak—is a Research Assistant Professor in the Department of Mechanical and Aerospace Engineering at the Graduate School of Engineering and Applied Sciences, Naval Postgraduate School, in Monterey, CA. [jmdidosz@nps.edu]

Fotis Papoulias—is an Associate Professor in the Department of Systems Engineering at the Graduate School of Engineering and Applied Sciences, Naval Postgraduate School, in Monterey, CA. [papoulias@nps.edu]

Abstract

The current class of Landing Craft Utility (LCU) has been in service in the U.S. Navy since the 1960s. Primarily used to land heavy vehicles, equipment, personnel, and cargo ashore in an amphibious assault, its basic design has served well over the last half century of use. However, certain loading combinations impacted by weight creep of particular cargoes have recently come to challenge established operational stability limits. The stability criteria currently employed came from traditional open ocean stability studies and therefore may not be optimal for the typical coastal transits of these specialized vessels. This study examines the intact transverse static and dynamic stability of the LCU in order to determine more appropriate criteria for short-range transits close to shore. The analysis mainly uses the Program of Ship Salvage Engineering (POSSE) software and the standard Ship Motion Program (SMP) to model a stochastic sea state, simulate the LCU's loading conditions, and predict the craft's dynamic response in various sea state conditions. The LCU's static transverse stability is derived by the POSSE software in terms of righting arm diagrams for different loading conditions, while the SMP software determines the dynamic transverse stability. The SMP analysis is based on seakeeping theory, using sea spectra model techniques to determine the LCU's roll angle dynamic responses. Based upon these simulation results, the study evaluates the current stability criteria and arrives at several dynamic stability recommendations and operational limits for loading conditions of interest.

Introduction and Method of Analysis

Motivation

Once thought of as being outdated by the introduction of the Landing Craft Air Cushion (LCAC), the Landing Craft Utility (LCU) 1610 class has continued to be a mainstay in U.S. Navy and Marine Corps amphibious operations through the present day (Schmitz, 2001). Of the 72 LCUs built in the 1960s and 1970s for the Navy, over 30 are still in service today (Colton, 2015). Recently, the contract for detailed design and construction of the LCU 1700, a newer version of the current LCU which was designed in the late 1950s, was awarded (Eckstein, 2018). While plans are to have this newer version of the familiar LCU provide a one-for-one replacement of those vessels currently in service, the continued need for full operational performance of current LCUs has not diminished and will need to be sustained over the next decade during this changeover period.

The LCU, a small displacement craft, is primarily used in amphibious operations to transport troops and military equipment, such as wheeled vehicles and tanks, and other types of cargo ashore. Launched from amphibious assault ships, its primary objective is to safely and expeditiously land and/or transport these items along the beachfront. While capable of long range transits and sustained operations of over a week, typically these vessels are deployed in limited duration coastal operations wherein short crested waves and not long period swells dominate the local seagoing environment (Bottleson, 2001).

While the cargo carrying capacity of the LCU 1610 has not significantly changed over the years—as often is the case due to loss of design margin in conventional warships over their life cycle resulting from ship alterations, modifications, and upgrades—its utility is impacted by another issue related to weight (Pedatzur, 2016). Impressive weight creep, occurring in some of its primary cargoes, such as the M1A1 Main Battle Tank, directly affects desired loading requirements by maxing out payload capacity. This weight gain, stemming from up-armoring of tanks and other vehicles, or replacing "designed for" cargo with more hefty versions of their predecessors, has reportedly pushed the limits of operational performance with respect to seakeeping at certain loading conditions (Eckstein, 2016).

With respect to this, ship stability and associated seakeeping considerations for the safe operation of the current LCU 1610 in coastal waters under normally occurring conditions are investigated herein.

Background

Ship stability is a basic principle of naval architecture. In general, the broader topic can be divided into transverse stability and longitudinal stability. The first of these, transverse stability, which can be described as the ship's ability to return to an equilibrium position when perturbed by an external force that generates a moment about the centerline axis of the vessel, is of typically greater interest to ship designers due to the greater length to beam ratio of most ships. Longitudinal stability, the ability of the ship to resist trim, the difference in forward and aft drafts, is generally of a secondary order.

A particular ship's stability is influenced by many internal factors, including displacement, load distribution, and underwater volume, as well as additional external influences such as wind speed, sea state conditions, turning angle, and speed. These factors, expressed as numerical parameters, contribute to the generation of the ship's stability curves. Stability curves describe the ship's transverse stability over a wide range of heeling angles and provide information about the required righting arm and moments in order to return the ship to the initial equilibrium state when it has been disturbed by a particular heeling angle. These curves are then used to derive the stability criteria of a ship for a particular set of parameters, such as the heeling angle, the righting arm, and the area under the curves, which are expressed in terms of mathematical limitations of the associated parameter values. Thus these stability criteria provide a ship or vessel with an operational guide based on the environmental and other varying inputs, such as cargo driven changes in displacement, that influence these curves. Compliance with such criteria ensures a ship's positive stability (i.e., the ship's ability to restore itself to its initial position), in contrast to negative stability, which refers to the ship's tendency to overturn.

The stability criteria currently used for the LCU are mainly based on the *Procedures Manual for Stability Analysis of U.S. Navy Small Craft* (Koelbel, 1977, pp. 11–40). This manual provides a transverse dynamic stability analysis for small displacement vessels based on a partially empirical procedure, which makes use of stability curves, and provides stability criteria by focusing on the ship's restoring moment. The ship restoring moment is the moment produced by the misalignment of the gravity and buoyancy forces and contributes to the ship's return to initial equilibrium position.

A primary assumption used in the analysis is of vessels making open ocean transits. These conditions are associated with higher wind velocities, yet the majority of LCU missions occur in coastal waters exhibiting much lower wind velocities (Joint Chiefs of Staff, 2009). This then raises the question of whether or not the current stability criteria are in fact optimal for use with the LCU 1610 in its predominately coastal missions at desired loading conditions. Therefore, it is postulated that these criteria may be overly conservative, resulting in a negative impact on LCU operational envelopes. Specific stability criteria for the LCU missions are not currently documented, but it has been reported that as a result of perceptions in potential reduced stability conditions, operational limitations are in place for the LCU.

Objectives

The main objective of this study is to investigate the suitability of stability criteria currently used for the LCU through rigorous analysis. This analysis examines the intact stability of the craft in conditions experienced during coastal missions. More specifically, this analysis focuses on LCU performance in short-range coastal transits from amphibious assault ships to the beach carrying different equipment loads and personnel. A further objective is to contribute to a guideline for the entire LCU fleet based on the conditions and characteristics of its typical coastal missions. One key aspect in this is the weight creep of the primary cargo load. Additionally, since existing stability criteria have been primarily developed based on data obtained from larger ships operating in different environments, they may not be adequate in addressing current operational concerns for this class of vessel and must be reviewed for suitability.

Tasks

The following tasks were undertaken in a systematic way in order to address the objectives of this work:

- Determination of the ship's static stability and dynamic response
- Evaluation of the LCU seakeeping performance stability on the basis of the obtained simulation results
- Development of recommended operational envelopes for the LCU deployment during typical coastal water missions
- Categorization of the currently used LCU stability criteria

Assumptions

This study followed three basic assumptions. The assumptions listed as follows were deemed both necessary in order to proceed with the analysis and reasonable so that the results could be used as a basis for further refinement:

- The ship's center of gravity does not change with changes in the angle of heel.
- The ship's center of buoyancy is always defined as the geometric centroid of the ship's underwater hull area.
- The shape of the ship's underwater hull area will continue to change with changing angles of heel.

Approach

The approach used in this work was comprised of the following steps. For brevity, we present here only the steps taken and follow with the fundamental results. Further technical details can be provided upon request and are documented in a recently completed Naval Postgraduate School master's thesis and related technical report (Roussopoulos, 2017).

Data Analysis

The first step was to analyze the existing static stability criteria. Following a classic approach, five criteria were considered, namely,

- Wind action and rolling
- Lifting heavy weights over the side
- Crowding of personnel to one side
- High-speed turning
- Topside ice (Sarchin & Goldberg, 1963, pp. 429–433)

As expected, some of these criteria were not applicable to this case and, thus, resulted in no additional usable information. From the static stability criteria that were found to be applicable, wind action and rolling was the most limiting. In all realistic loading cases and with environmental conditions considered, the LCU passed the criterion.

The next step was to initiate the analysis for dynamic stability. To that end, it was decided to use a six degree of freedom motions program, Standard Ship Motion Program (SMP), with several loading conditions and a variety of sea states (Herbert-ABS, 2012, pp. 10–34; Conrad, 2015, pp. 1–16).

The following loading conditions were used in the modeling:

- Lightship
- LCU with half cargo deadweight
- LCU with full cargo deadweight

Several sea spectra models, as indicated here, were used in the investigation:

- Pierson-Moskowitz spectrum model
- Bretschneider spectrum model
- JONSAWP spectrum model
- Ochi-Hubble spectrum model

In addition to long-crested or unidirectional seas, we also considered short-crested seas which are better suited for operations in coastal regions and the littorals (IHS Markit, 2017). Figures 1 and 2 illustrate sample input and response parameters used in portions of the SMP analyses. Detailed results at varying load conditions and wave spectra are found in the appendix.

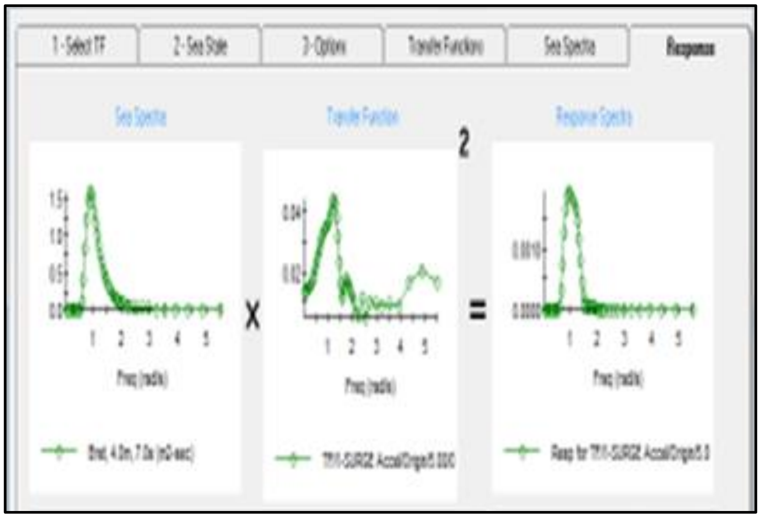


Figure 1. Sample Input for SMP Ship Response Calculator

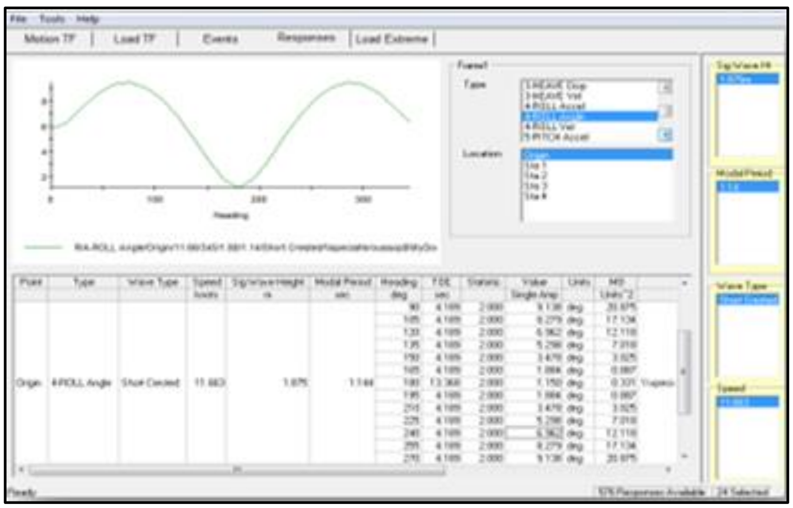


Figure 2. Sample of SMP Ship Response Results

Ship Model

This study used the LCU 1644 model, as shown in Figure 3, as a representative case for study of the LCU. Some basic characteristics of this particular version of the LCU 1610 class craft are as follows:

LCU 1627 General Characteristics (McCreight, 1998, pp. 8–9, 12–13):

- Length (Overall): 41.1 m
- Length (Between Perpendiculars): 40.84 m
- Beam: 8.8 m
- Depth 2.44 m
- Maximum Speed: 5.66 m/s (11 knots)
- Maximum Range: 2,222.4 Km (1200 Nautical Miles)
- Economic Speed: 4.12 m/s (8 knots)
- Maximum Load: 127 Metric Tones

- Crew Members: 16
- Propulsion system: 4 Detroit 6-71 diesels 519.007 KW (696 hp)

Figure 3. LCU 1644 Awaiting Loading by Beachmaster Unit
(U.S. Navy Photo)

Engineering drawings and detailed characteristics necessary to accurately model the LCU 1644 and its loading conditions using the Program of Ship Salvage Engineering (POSSE) were provided by Naval Sea Systems Command (Herbert-ABS, 2013, pp. 75–79, 115–119, 124–125). Figures 4 and 5 depict the LCU hull geometry and sample loading case as modeled in the POSSE program.

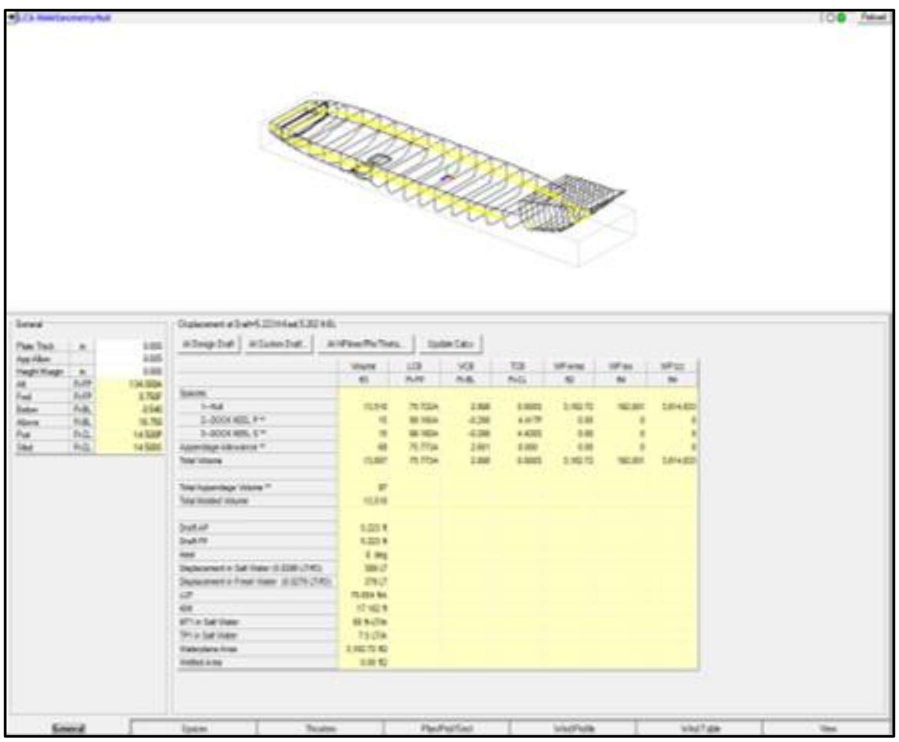


Figure 4. LCU Hull Geometry in POSSE

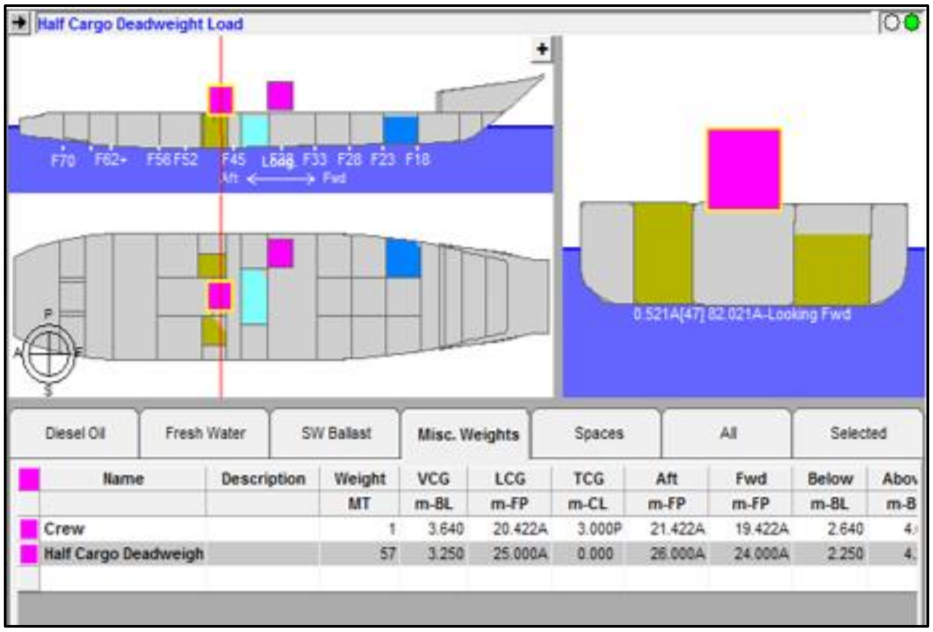


Figure 5. Lightship With Half Cargo Deadweight Loading Condition

Results and Discussion

Findings

Static Stability Assessment

The static stability of the LCU was considered adequate in all loading cases and all sea states. Calculations were performed for the lightship condition as well as LCU with half cargo deadweight and LCU with full cargo deadweight cases. The corresponding total displacement of these cases was 257, 314 and 371 metric tons, respectively. Static stability was based on the Navy's wind-rolling criterion, which was met throughout the operational trade space of the LCU. It was found that the LCU meets the minimum requirements of the reserve area as well as the maximum wind heeling arm. Numerical results were compared with analytical predictions and the agreement was found to be excellent.

Dynamic Stability Assessment

Dynamic stability of the LCU was studied for a variety of random seas and for all loading conditions as for the static stability case. The full range of LCU operational speeds was considered. Random seas results were obtained with Bretschneider two-parameter spectra and the Ochi-Hubble six-parameter spectral family. The Bretschneider family is used extensively in most standard dynamic stability and seakeeping studies and is typically found in deep waters. Two parameters, the significant wave height and the spectral peak, were utilized in order to provide coverage from developing to decaying seas. The Ochi spectral family is more common in coastal areas where a local wind-driven sea is superimposed to a long range swell. In addition, both long-crested and short-crested formulations were used for all spectra. Short crested seas are more realistic models of actual conditions in real life and are composed of a collection of long-crested unidirectional seas with a standard cosine-squared spreading. Care is taken to ensure that the total energy contained in the seaway is preserved in order to get meaningful comparisons.

Due to the large number of sea state conditions, ship speeds, loading conditions, and headings considered, we summarize the results in a set of operational recommendations as shown in Table 1. It should be emphasized that these results are preliminary and further studies are needed to arrive at more conclusive recommendations. Such sensitivity studies along with specific operational recommendations will be presented in follow on studies and accompanying technical reports.

Table 1. Ship Speed and Heading Recommendations for Typical Loading Conditions at Various Sea States

Given Variables \ Selected Variables		Sea Heading (Degrees)	Ship Speed (Knots)	
Lightship, and Half Cargo Deadweight Loading Cases			Sea Heading 0–120 and 240–360	Sea Heading 120–240
	Sea State 2	-	-	-
	Sea State 4	Avoid sea headings 60 – 90 and 270 – 300	Reduce or maintain speed close to 4	Increase or maintain speed close to 11
	Sea State 6	Avoid sea headings 60 – 90 and 270 – 300	Reduce or maintain speed close to 4	Increase or maintain speed close to 11
Full Cargo Deadweight Loading Case			Sea Heading 0–135 and 225–360	Sea Heading 135–225
	Sea State 2	-	-	-
	Sea State 4	Avoid sea headings 60 – 90 and 270 – 300	Reduce or maintain speed close to 4	Increase or maintain speed close to 11
	Sea State 6	Avoid sea headings 60 – 90 and 270 – 300	Reduce or maintain speed close to 4	Increase or maintain speed close to 11

The operational recommendations for the various loading conditions and sea states presented in Table 1 can be visualized via polar diagrams provided as Figures 6 and 7. These figures show schematically the recommended actions by the LCU operators for different loading conditions in sea states 4 and 6, and for the full range of sea headings in 15 degree increments. Such diagrams need to be refined and superimposed to specific geographical areas of operations and expected sea states in order to arrive at a recommended route within adequate safety constraints and the operational requirements at hand.

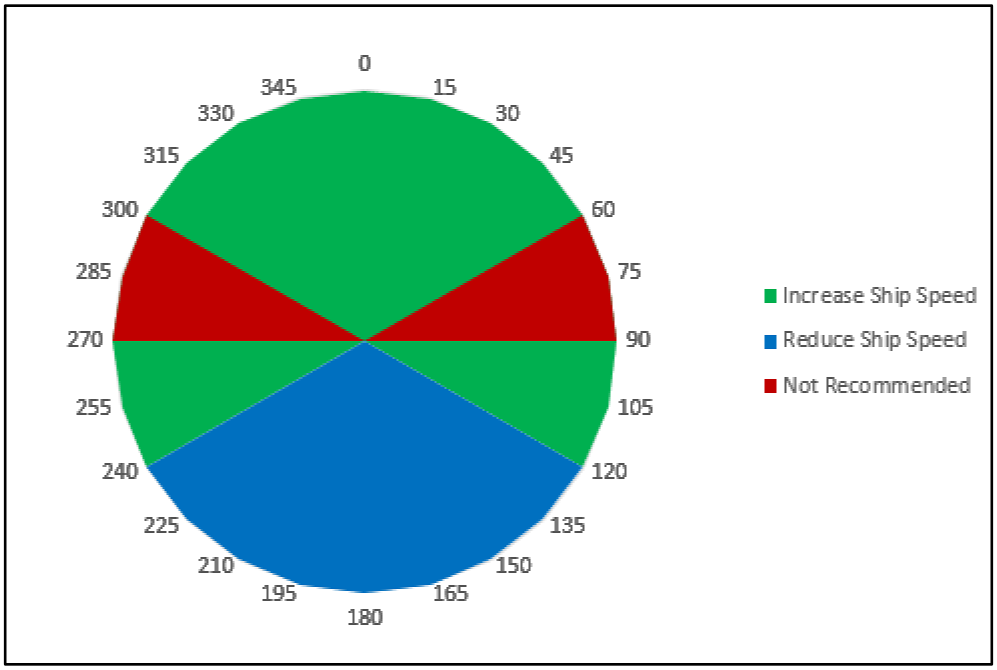


Figure 6. Sample Operational Diagram for LCU Lightship and Half Cargo Deadweight Loading Conditions in Sea States 4 and 6

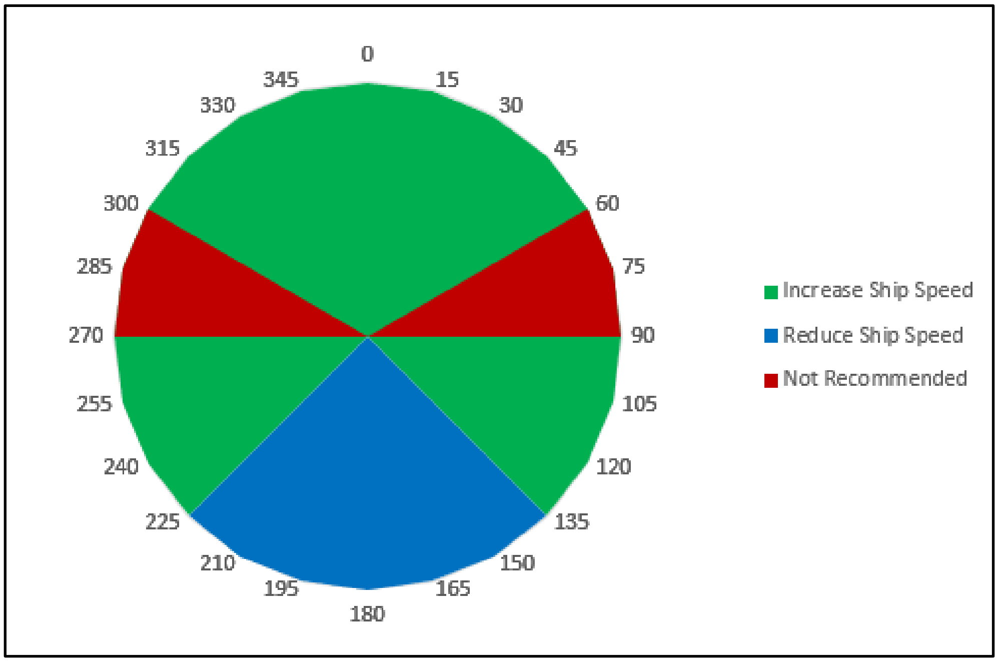


Figure 7. Sample Operational Diagram for LCU Full Cargo Deadweight Loading Conditions in Sea States 4 and 6

References

Bottleson, J. D. (2001). *Landing craft utility as a force multiplier in the littorals*. Quantico, VA: United States Marine Corps Command and Staff College. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a401240.pdf

Colton, T. (2015, May 5). Landing Craft, Utility—LCU. Retrieved from http://shipbuildinghistory.com/smallships/lcu.htm

Conrad, R. (2005). SMP95: *Standard Ship Motions Program user manual* (Rep. NSWCCD-50-TR-2005/074). Bethesda: MD: Naval Surface Warfare Center.

Eckstein, M. (2016, January 14). Navy to begin LCU affordability talks with industry after procurement accelerated. *U.S. Naval Institute News*. Retrieved from https://news.usni.org/2016/01/14/navy-to-begin-lcu-affordability-talks-with-industry-after-procurement-accelerated

Eckstein, M. (2018, April 4). NAVSEA picks Swiftships LLC to design, build LCU replacement in $18M contract award. *U.S. Naval Institute News*. Retrieved from https://news.usni.org/2018/04/04/navsea-awards-18m-lcu-1700-contract-louisiana-based-swiftships-llc

Herbert-ABS. (2012). *Ship Motions Program*. Alameda, CA.

Herbert-ABS. (2013). *HECSALV/POSSE user manual*. Alameda, CA.

IHS Markit. (2017, February 21). Landing Craft Utility 1600 Class. Retrieved from https://janes.ihs.com.libproxy.nps.edu/Janes/Display/jfs_a509-jfs

Joint Chiefs of Staff. (2009, August). *Amphibious operations* (Joint Publication 3-02). Washington, DC: Author.

Koelbel, G. (1977). *Procedures manual, dynamic stability analysis for U.S. Navy small craft* (Rep. 23095-1). Norfolk, VA: Combatant Craft Engineering Naval Ship Engineering Center. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a047493.pdf

McCreight, K. (1998). *A note on selection of wave spectra for design evaluation* (Rep. CRDKNSWC-HD-974-02). Bethesda, MD: Naval Surface Warfare Center.

Pedatzur, O. (2016). Weight design margins in naval ship design: A rational approach. *Naval Engineers Journal, 128*(2), 57–64.

Roussopoulos, K. (2017). *Stability criteria analysis for Landing Craft Utility* (Master's thesis). Monterey, CA: Naval Postgraduate School. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a445093.pdf

Sarchin, T., & Goldberg, L. (1963). *Stability and buoyancy criteria for U.S. Naval surface ships*. New York, NY: Society of Naval Architects and Marine Engineers.

Schmitz, K. L. (2001). *LCAC versus LCU: Are the LCAC worth the expenditure?* Quantico, VA: United States Marine Corps Command and Staff College. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a516264.pdf

Appendix

This appendix contains some of the detailed results supporting the conclusions described in the main body of the paper.

Bretschneider Spectrum Roll Angle Response for Lightship Loading Case

Table 2. Roll Angle Responses in Bretschneider (Tm = 7 sec) Short-Crested Sea Waves for LCU Lightship

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.43	0.56	0.75	0.88	2.65	3.49	4.65	5.48	7.02	9.19	12.19	14.36
15	0.51	0.65	0.82	0.95	3.20	4.02	5.13	5.95	8.44	10.55	13.42	15.56
30	0.70	0.83	0.99	1.12	4.36	5.16	6.17	6.96	11.34	13.38	16.04	18.13
45	0.90	1.01	1.16	1.28	5.56	6.28	7.20	7.95	14.26	16.14	18.60	20.62
60	1.06	1.15	1.28	1.38	6.53	7.13	7.92	8.60	16.57	18.19	20.38	22.26
75	1.16	1.23	1.32	1.41	7.15	7.59	8.20	8.77	18.03	19.28	21.08	22.71
90	1.20	1.23	1.29	1.36	7.36	7.60	8.01	8.44	18.53	19.33	20.63	21.91
105	1.16	1.16	1.18	1.23	7.16	7.18	7.36	7.63	18.05	18.33	19.05	19.88
120	1.06	1.02	1.01	1.03	6.54	6.35	6.30	6.41	16.61	16.34	16.44	16.80
135	0.90	0.84	0.80	0.79	5.58	5.19	4.96	4.92	14.13	13.51	13.06	13.00
150	0.71	0.62	0.56	0.54	4.39	3.86	3.51	3.37	11.41	10.16	9.13	8.95
165	0.52	0.42	0.36	0.33	3.24	2.63	2.23	2.04	8.53	6.97	5.95	5.45
180	0.43	0.33	0.26	0.23	2.70	2.04	1.64	1.45	7.13	5.43	4.37	3.87

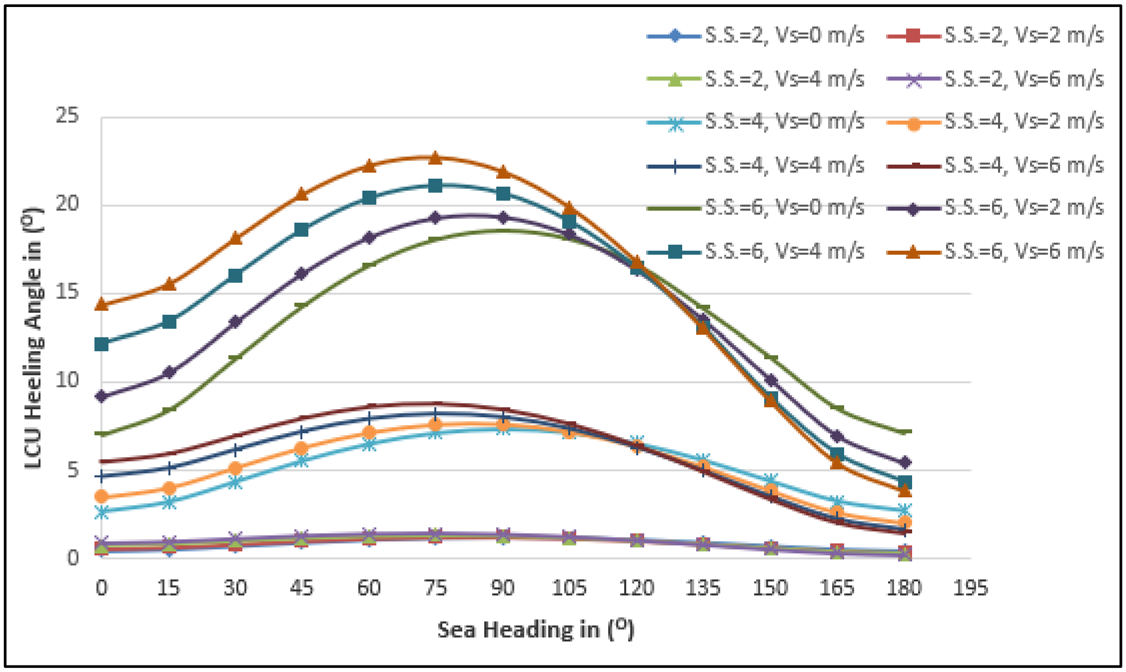


Figure 8. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 7 sec) Short-Crested Sea Waves for LCU Lightship

Table 3. Roll Angle Responses in Bretschneider (Tm = 15 sec) Short-Crested Sea Waves for LCU Lightship

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.16	0.19	0.24	0.28	0.99	1.18	1.47	1.74	2.64	3.15	3.92	4.64
15	0.18	0.21	0.26	0.30	1.12	1.31	1.59	1.86	2.98	3.49	4.24	4.94
30	0.23	0.26	0.30	0.34	1.41	1.60	1.86	2.11	3.75	4.25	4.95	5.63
45	0.28	0.30	0.34	0.38	1.73	1.90	2.14	2.38	4.59	5.05	5.69	6.32
60	0.32	0.34	0.38	0.41	1.99	2.14	2.34	2.55	5.30	5.68	6.23	6.79
75	0.35	0.36	0.39	0.42	2.17	2.27	2.43	2.60	5.76	6.04	6.46	6.93
90	0.36	0.37	0.38	0.40	2.23	2.28	2.38	2.51	5.93	6.06	6.34	6.69
105	0.35	0.35	0.35	0.37	2.17	2.17	2.21	2.29	5.77	5.76	5.89	6.10
120	0.32	0.31	0.31	0.31	2.00	1.94	1.93	1.96	5.31	5.17	5.14	5.23
135	0.28	0.26	0.25	0.25	1.73	1.63	1.58	1.57	4.60	4.35	4.21	4.19
150	0.23	0.21	0.19	0.19	1.41	1.29	1.21	1.18	3.76	3.44	3.23	3.14
165	0.18	0.16	0.15	0.14	1.12	0.99	0.91	0.87	3.00	2.64	2.42	2.32
180	0.16	0.14	0.13	0.12	1.00	0.87	0.79	0.75	2.66	2.31	2.10	2.00

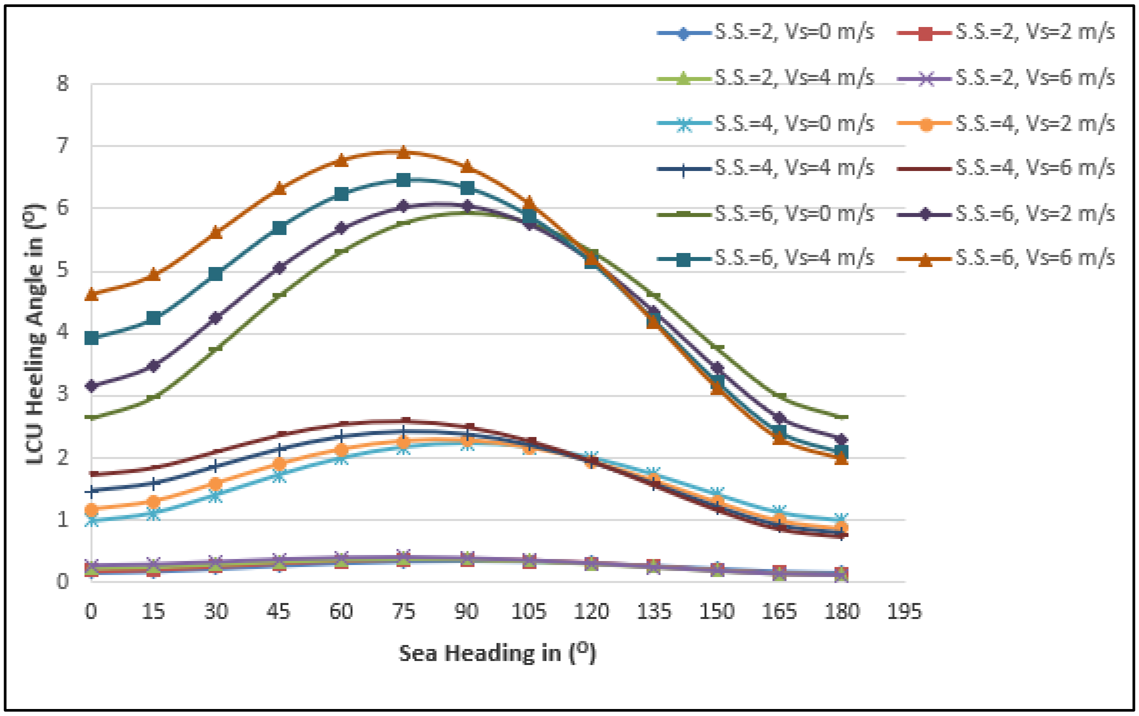


Figure 9. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 15 sec) Short-Crested Waves for LCU Lightship

Bretschneider Spectrum Roll Angle Response for LCU Plus Half Cargo Deadweight Loading Case

Table 4. Angle Responses in Bretschneider (Tm = 7 sec) Short-Crested Sea Waves LCU Carrying Half Cargo Deadweight

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.51	0.72	0.98	1.16	3.14	4.48	6.06	7.14	8.13	11.36	15.32	18.11
15	0.62	0.83	1.08	1.25	3.85	5.13	6.65	7.70	9.80	12.89	16.71	19.48
30	0.87	1.06	1.29	1.45	5.30	6.50	7.91	8.92	13.06	15.98	19.63	22.40
45	1.11	1.29	1.50	1.65	6.73	7.82	9.12	10.09	16.19	18.89	22.38	25.16
60	1.32	1.46	1.64	1.77	7.86	8.79	9.94	10.83	18.57	20.97	24.22	26.89
75	1.45	1.55	1.69	1.80	8.56	9.28	10.22	10.99	20.04	22.03	24.88	27.29
90	1.49	1.55	1.64	1.72	8.81	9.26	9.93	10.51	20.54	22.02	24.29	26.26
105	1.45	1.45	1.49	1.54	8.57	8.72	9.08	9.44	20.06	20.92	22.45	23.83
120	1.32	1.27	1.26	1.27	7.88	7.71	7.75	7.87	18.61	18.76	19.43	20.12
135	1.12	1.03	0.98	0.96	6.76	6.30	6.04	5.97	16.24	15.63	15.42	15.46
150	0.87	0.75	0.67	0.64	5.34	4.63	4.17	3.97	13.14	11.81	10.87	10.44
165	0.64	0.49	0.40	0.36	3.91	3.05	2.49	2.23	9.94	8.00	6.59	5.94
180	0.52	0.37	0.27	0.22	3.22	2.28	1.67	1.39	8.32	6.03	4.44	3.71

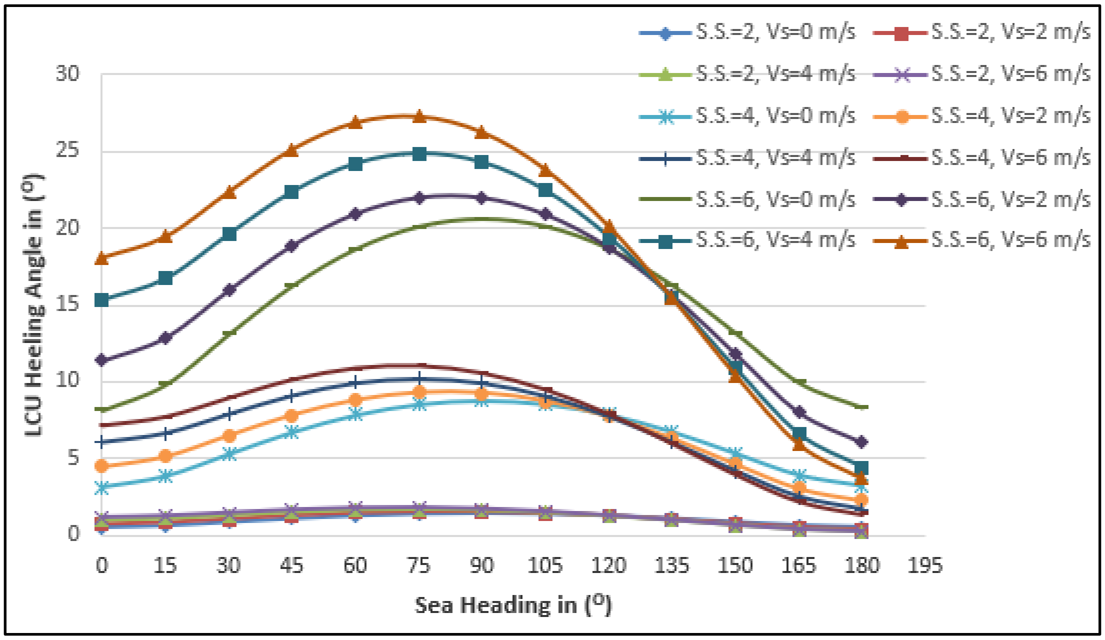


Figure 10. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 7 sec) Short-Crested Waves for LCU Carrying Half Cargo Deadweight

Table 5. Roll Angle Responses in Bretschneider (Tm = 15 sec) Short-Crested Sea Waves for LCU Carrying Half Cargo Deadweight

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.17	0.22	0.30	0.36	1.07	1.39	1.84	2.24	2.86	3.70	4.88	5.96
15	0.20	0.25	0.32	0.38	1.24	1.55	1.98	2.38	3.28	4.12	5.26	6.31
30	0.26	0.31	0.37	0.43	1.60	1.90	2.31	2.68	4.23	5.03	6.11	7.10
45	0.32	0.36	0.42	0.48	1.98	2.26	2.63	2.99	5.24	5.97	6.96	7.89
60	0.37	0.41	0.46	0.51	2.31	2.54	2.86	3.18	6.06	6.68	7.54	8.39
75	0.41	0.43	0.47	0.51	2.52	2.69	2.94	3.20	6.59	7.05	7.75	8.46
90	0.42	0.43	0.46	0.49	2.59	2.68	2.85	3.05	6.78	7.04	7.53	8.07
105	0.41	0.41	0.42	0.44	2.52	2.53	2.61	2.73	6.60	6.65	6.91	7.25
120	0.37	0.36	0.36	0.37	2.31	2.24	2.24	2.29	6.07	5.92	5.95	6.09
135	0.32	0.30	0.29	0.29	1.99	1.85	1.79	1.78	5.25	4.92	4.76	4.73
150	0.26	0.23	0.21	0.20	1.61	1.43	1.32	1.28	4.25	3.79	3.51	3.40
165	0.20	0.17	0.15	0.14	1.25	1.05	0.93	0.88	3.31	2.80	2.49	2.36
180	0.17	0.14	0.12	0.12	1.09	0.89	0.78	0.73	2.89	2.37	2.07	1.95

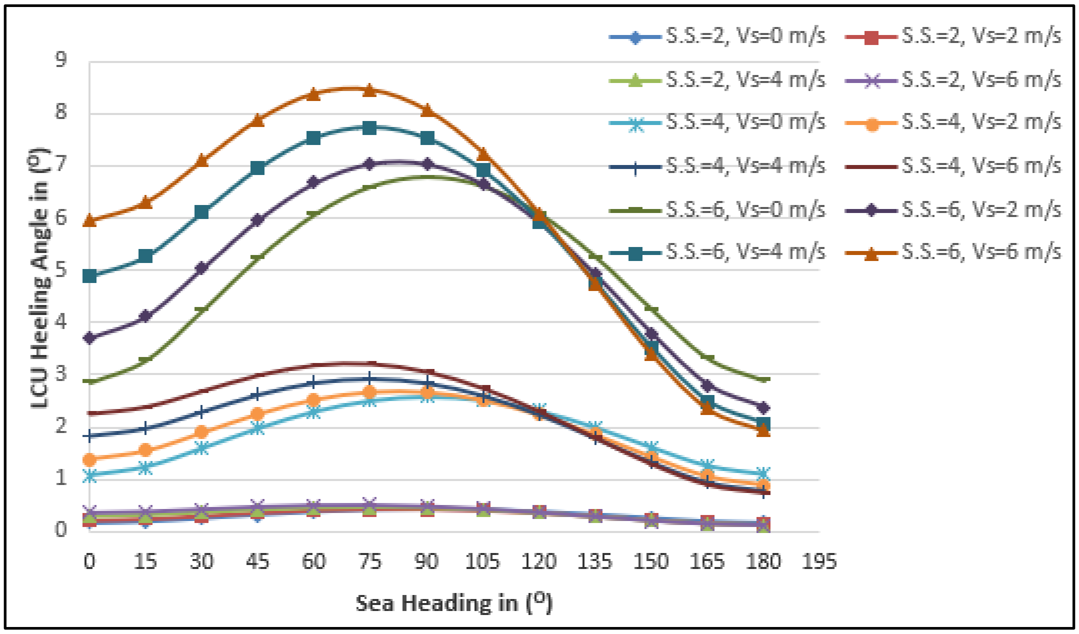


Figure 11. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 15 sec) Short-Crested Waves for LCU Carrying Half Cargo Deadweight

Roll Angle Responses in Bretschneider Spectrum for LCU Plus Full Cargo Deadweight Loading Case

Table 6. Roll Angle Responses in Bretschneider (Tm = 7 sec) Short-Crested Sea Waves for LCU Carrying Full Cargo Deadweight

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.62	0.94	1.21	1.39	3.76	5.67	7.31	8.44	9.28	13.49	17.68	20.65
15	0.77	1.07	1.32	1.52	4.64	6.40	7.98	9.19	11.07	15.00	19.14	22.29
30	1.08	1.35	1.58	1.80	6.33	7.90	9.39	10.73	14.44	18.06	22.17	25.60
45	1.39	1.63	1.83	2.06	7.92	9.34	10.76	12.18	17.53	20.90	24.99	28.59
60	1.64	1.84	2.00	2.23	9.12	10.37	11.68	13.09	19.81	22.88	26.85	30.45
75	1.81	1.95	2.06	2.27	9.86	10.90	12.00	13.32	21.19	23.88	27.53	30.92
90	1.86	1.94	1.99	2.17	10.12	10.86	11.68	12.81	21.65	23.85	26.95	29.95
105	1.81	1.82	1.82	1.95	9.87	10.28	10.73	11.59	21.21	22.79	25.11	27.51
120	1.65	1.60	1.54	1.62	9.14	9.17	9.21	9.76	19.85	20.68	22.01	23.63
135	1.40	1.29	1.19	1.22	7.96	7.60	7.25	7.45	17.60	17.56	17.76	18.47
150	1.09	0.94	0.82	0.79	6.39	5.69	5.03	4.88	14.55	13.60	12.71	12.50
165	0.79	0.61	0.47	0.41	4.74	3.76	2.93	2.55	11.26	9.43	7.68	6.75
180	0.64	0.45	0.29	0.23	3.90	2.75	1.84	1.41	9.55	7.13	4.88	3.75

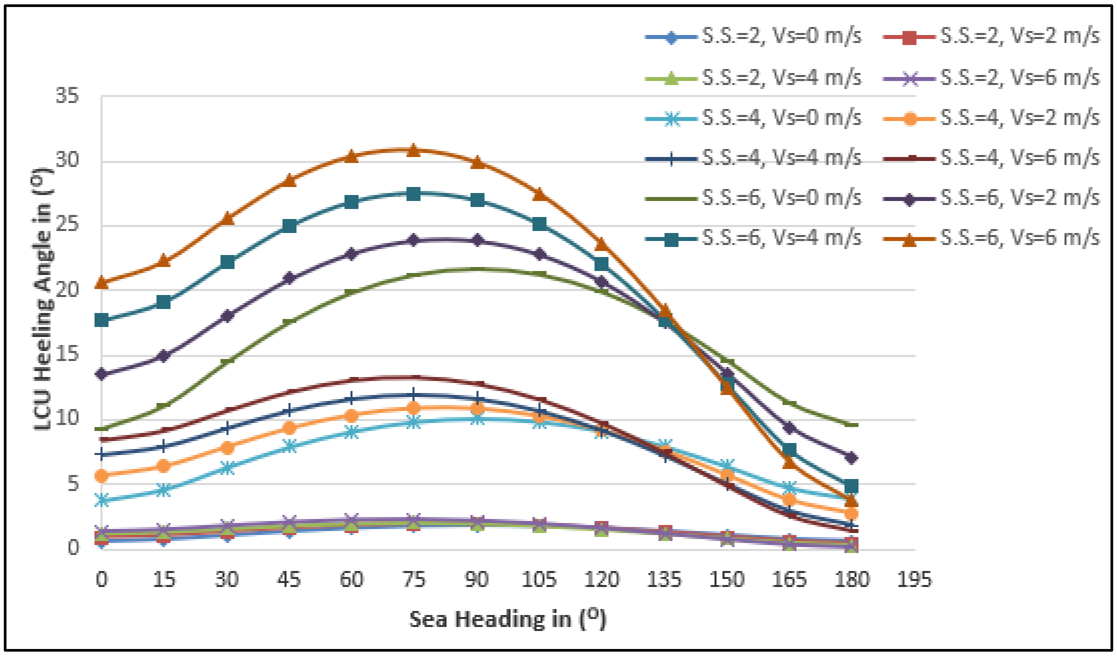


Figure 12. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 7 sec) Short-Crested Waves LCU Carrying Full Cargo Deadweight

Table 7. Roll Angle Responses in Bretschneider (Tm = 15 sec) Short-Crested Sea Waves for LCU Carrying Full Cargo Deadweight

Sea Heading (Degrees)	LCU Heeling Angle (Degrees)											
	Sea State 2				Sea State 4				Sea state 6			
	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s	Vs=0 m/s	Vs=2 m/s	Vs=4 m/s	Vs=6 m/s
0	0.19	0.28	0.36	0.45	1.20	1.71	2.26	2.77	3.18	4.49	5.94	7.28
15	0.23	0.31	0.39	0.47	1.42	1.90	2.43	2.95	3.72	4.98	6.37	7.74
30	0.30	0.38	0.45	0.54	1.87	2.32	2.80	3.35	4.88	6.04	7.33	8.73
45	0.38	0.45	0.51	0.60	2.35	2.75	3.19	3.73	6.06	7.10	8.28	9.67
60	0.45	0.50	0.56	0.64	2.74	3.08	3.44	3.97	6.99	7.87	8.91	10.25
75	0.49	0.53	0.57	0.64	2.99	3.24	3.52	3.99	7.57	8.26	9.11	10.31
90	0.51	0.53	0.55	0.61	3.08	3.23	3.40	3.79	7.77	8.23	8.83	9.83
105	0.49	0.49	0.50	0.54	3.00	3.03	3.10	3.38	7.58	7.77	8.08	8.82
120	0.45	0.43	0.42	0.45	2.75	2.67	2.64	2.81	7.00	6.91	6.92	7.39
135	0.38	0.36	0.33	0.34	2.36	2.20	2.08	2.14	6.08	5.73	5.49	5.66
150	0.31	0.27	0.24	0.23	1.89	1.66	1.49	1.46	4.92	4.38	3.96	3.89
165	0.23	0.19	0.16	0.15	1.44	1.18	1.00	0.93	3.78	3.13	2.67	2.47
180	0.20	0.15	0.13	0.12	1.23	0.96	0.79	0.72	3.25	2.55	2.11	1.93

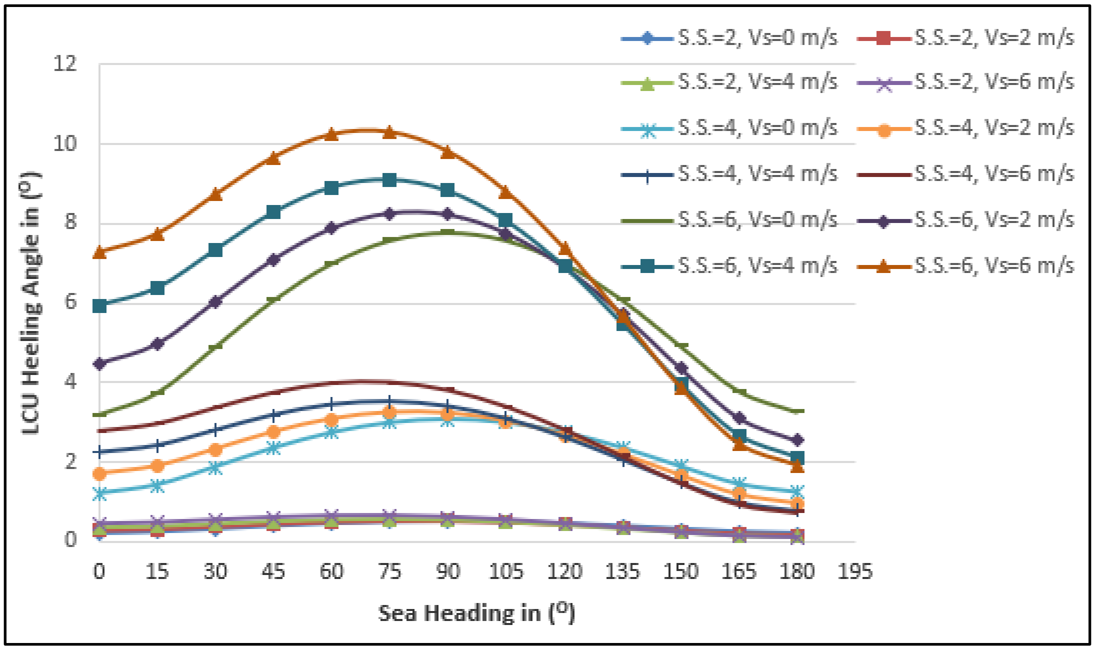


Figure 13. Heeling Angle Versus Sea Heading in Bretschneider (Tm = 15 sec) Short-Crested Waves LCU Carrying Full Cargo Deadweight

Panel 2. A Multidimensional Approach to Acquisition Workforce Management

Wednesday, May 9, 2018	
9:30 a.m. – 11:00 a.m.	**Chair: Joe Everling,** Chief of Staff, Navy Director, Acquisition Career Management ***Modeling the Department of Navy Acquisition Workforce With System Dynamics*** David Ford, Texas A&M University Altyn Clark, Chief Solutions Officer, Transformation Systems, Inc. ***Embracing Structure and Discipline to Manage Challenges in the Department of Navy Acquisition Workforce*** Altyn Clark, Chief Solutions Officer, Transformation Systems, Inc. Elizabeth S. Rosa, Office of Navy Director, Acquisition Career Management ***Adding Agility to Department of Navy Acquisition Workforce Management in Digital Collaboration Centers*** Altyn Clark, Chief Solutions Officer, Transformation Systems, Inc.

Modeling the Department of Navy Acquisition Workforce With System Dynamics

David N. Ford—PE,is the Beavers Charitable Trust/William F. Urban '41 Professor in Construction Engineering and Management in the Zachry Department of Civil Engineering at Texas A&M University and a Research Associate Professor of Acquisition in the Graduate School of Business and Public Policy at the Naval Postgraduate School. He researches acquisition management, community recovery from disasters, managerial real options, and development project dynamics. Before entering academia, Dr. Ford designed and managed the development of constructed facilities in industry and government. He received his PhD from MIT and his master's and bachelor's degrees from Tulane University. [davidford@tamu.edu]

Altyn Clark—is a troubleshooter whom leaders call to diagnose ambiguity and craft a way ahead and to address wicked problems. He uses the engineering design process, coupled with industrial psychology and methods of large scale change, to assess and advance organizational systems. As Chief Solutions Officer of Transformation Systems Inc (TSI), Dr. Clark invents practical and useful models, methods, approaches, and tools. Dr. Clark brings over 30 years of subject matter expertise in engineering, quality control and assurance, program and personnel management, research, logistics, human capital, and strategic management. Dr. Clark received his PhD, MS, and BS from Virginia Tech. [ac@transformationsystems.com]

Abstract

Acquiring effective and efficient materiel solutions that support naval missions is critical to meeting Department of the Navy (DoN) objectives. Maintaining the readiness of the current Navy to fight and win, accelerating the delivery of warfighting capability for the next Navy, and researching and transitioning to new technologies for the Navy after next all require that the DoN maintain a healthy acquisition workforce that is large enough and qualified to be smart buyers over 30+ year time horizons. The naval acquisition workforce faces losses of experience and capacity as the current workforce ages and retires, as knowledge half-life diminishes the relevance of current skills and experience, and as a tightening labor market draws government employees to the private sector. Leaders throughout the DoN are challenged to identify and implement actionable levers to sustain required workforce capability and capacity. This study developed a realistic simulation model of a portion of the naval acquisition workforce and demonstrated its potential use in workforce planning and management.

Introduction

Challenges Facing the Department of the Navy Acquisition Workforce

The Department of the Navy (DoN) acquisition enterprise exists to put capability in the hands of warfighters so that when necessary they can fight and win. The DoN acquisition workforce manages the planning, design, procurement, manufacturing and construction, testing, and deployment of materiel solutions and services to fulfill the Navy's mission and support operations. Doing so requires thousands of contracts, millions of contract actions, and billions of dollars each year. The DoN acquisition workforce must be effective and have adequate capacity to fulfill the demand for naval acquisition work.

The DoN faces several challenges in providing an adequate acquisition workforce. First, the demands placed on the acquisition workforce are changing. The naval fleet is growing toward a target of more than 300 ships. Materiel solutions are becoming increasingly complex as threats and technologies evolve. Pacing the threats requires that the acquisition workforce have deeper, more varied knowledge and skills than were required in the past.

In addition to these demand-side challenges, the DoN faces challenges in maintaining the capability and capacity of its acquisition workforce. The current acquisition workforce is relatively old. Therefore the workforce is currently losing, and will soon lose more, experience and capacity as members retire or seek employment elsewhere. Maintaining capability and capacity will require the DoN to recruit and train new acquisition personnel.

The acquisition workforce (AWF) obligates over $300 billion annually to acquire goods and services. The GAO has reported on the need for ensuring that the AWF is adequately sized, trained, and equipped to meet the needs of the DoD. To help address some of the challenges, Congress created the Defense Acquisition Workforce Development Fund (DAWDF). The fund has been applied to a variety of uses, including increasing the size of the workforce. The Better Buying Power (BBP) initiatives (DoD, 2010–2016) also addressed acquisition workforce needs, including improvements in recruiting and hiring, training and development, and retention and recognition. These efforts have generally increased the certification rates of the acquisition workforce (see Figure 1). However, a further challenge in maintaining an adequate AWF is the uncertainty around which of these potential levers addressed by DAWDF or BBP will have the greatest positive impact on AWF performance. Also of concern are temporal lag factors between interventions and effects, and the near certainty of unintended consequences for any AWF change initiative.

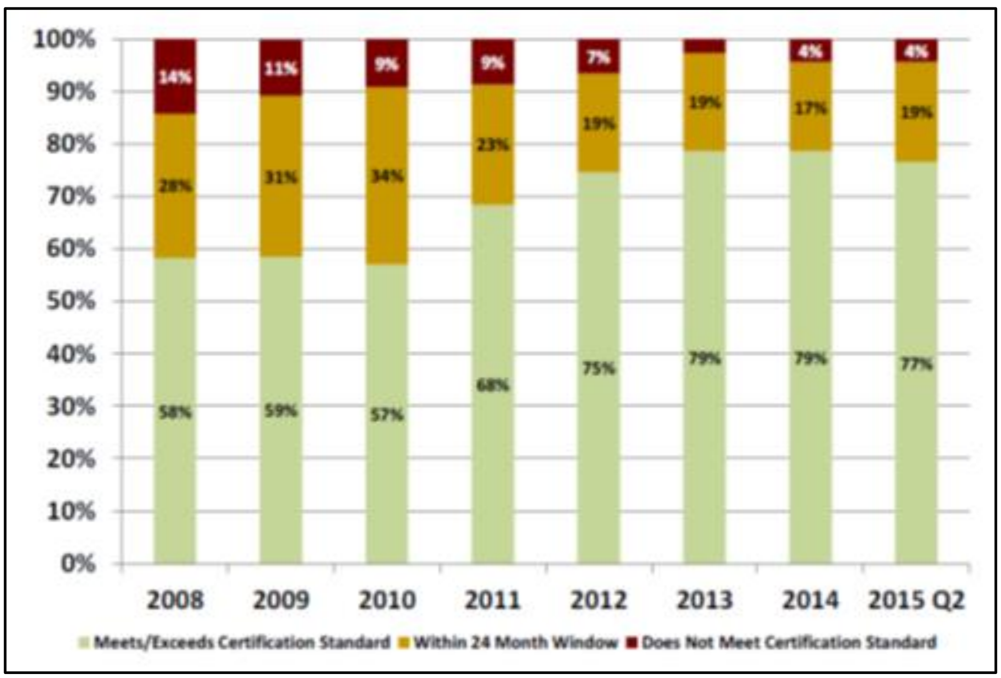


Figure 1. Acquisition Workforce Meeting Certification Standards (2008–2015Q2)
(DoD, 2015, Figure 4-1)

The critical role of the temporal dimension is the most important feature shared by the acquisition workforce challenges described previously. Insights and solutions require an understanding of both short-term and long-term impacts of system designs and policies. Given the complexity of the system and the difficulty of conducting experiments in the actual system, realistic models for DoN AWF design and policy development would be useful.

Background and Problem Description

At a macro level, warfighting system demand signals articulated in U.S. National Strategy documents, such as the DoN 30-year shipbuilding plan, generate a certain amount of acquisition work that must be accomplished to get capability into the hands of warfighters. The DoN has allocated various domains of acquisition work among several acquisition system commands (SYSCOMs). Each SYSCOM uses a tailored approach to translate warfighting system demand signals in their domain into a sequenced volume of work to be accomplished. Each SYSCOM's tailored approach estimates how many people (with appropriate knowledge and experience levels) are needed to perform that volume of work. Work accomplished by the DoN AWF enables industry partners in the value stream to construct and deliver warfighting systems. Delivering acquisition, modernization, and maintenance of warfighting capability to warfighters is the aim of the civilian-military-industrial enterprise. The outcome is readiness to fight and win. Changes in national strategy over time have caused cyclical shrinkage and swelling among the ranks of the AWF.

The DoN acquisition workforce system includes many diverse parts, processes, and stakeholders that interact over time in a wide variety of ways. Attempts to improve individual parts of the system (e.g., training, assignment rotation) or aspects of the system (e.g., economics, experience levels) cannot be completely successful if enacted separately. Addressing DoN AWF challenges requires a systems perspective and systems level solutions. The tools and methods that facilitate that perspective and those solutions must be able to integrate the numerous and diverse aspects of the workforce (e.g., specialization, training, experience, assignment rotation and advancement, location) and measures of workforce performance (e.g., capabilities and capacities).

Understanding the interactions among workforce components is critical to developing improved policies. Developing that understanding is not intuitive or obvious, largely because the workforce and its performance are dynamic: they evolve in response to system structure, current conditions, and current and future policies. Those interactions create causal feedback loops, unintended side effects, delays, and resistance to otherwise well-designed policies. Improving acquisition workforce understanding and developing effective and efficient policies requires tools and methods that can capture the systemic, dynamic feedback in the system; can reflect current and future policies; and can reflect their impacts on workforce performance.

Research Methodology

The research developed a dynamic simulation model that can be used to illustrate DoN acquisition workforce challenges, explain the structural causes of those challenges, and communicate the nature and degree of those challenges to policy-makers. The ultimate goal is a set of tools that can be used by policy-makers to better understand DoN acquisition workforce challenges and design effective and efficient policies.

This initial model demonstrates the potential of a dynamic simulation model to help improve policy-makers' and the workforce's understanding of the system and the impacts of potential policy changes. The model could thereby play a central role in educating and communicating with policy-makers about challenges and possible solutions. To investigate the potential of the model to facilitate meeting these goals, the research addressed the following question: "How can a dynamic simulation model be used to investigate, explain, and communicate DoN acquisition workforce challenges and potential solutions?"

The research applied the system dynamics modeling methodology. The system dynamics methodology combines a broad perspective of systems with a control theory approach to improve the design and management of complex human systems. System dynamics combines servo-mechanism thinking with computer simulation to allow the analysis of systems in ways that are not possible with human reasoning alone. It is one of several established and successful approaches to systems analysis and design (Flood & Jackson, 1991; Jackson, 2003; Lane & Jackson, 1995). Forrester (1961) developed the methodology's philosophy, and Sterman (2000) specified the modeling process with examples and described numerous applications. When applied to engineered systems such as the defense acquisition workforce, system dynamics focuses on how performance evolves in response to interactions within the causal structure of the system (e.g., retirement rates, development of knowledge and experience), development and management policies (e.g., training developed in specialty areas), and conditions (e.g., capacity levels, budget constraints). System dynamics is appropriate for modeling the acquisition workforce because of its ability to explicitly model the diverse set of critical features, characteristics, and relationships that drive behavior and performance.

System dynamics has been applied to military systems, including planning and strategy (Bakken & Vamraak, 2003; Duczynski, 2000; McLucas, Lyell, & Rose, 2006; Melhuish, Pioch, & Seidel, 2009), workforce management (Bell & Liphard, 1978), technology (Bakken, 2004), command and control (Bakken & Gilljam, 2003; Bakken, Gilljam, & Haerem, 2004), operations (Bakken, Ruud, & Johannessen, 2004; Coyle & Gardiner, 1991), logistics (Watts & Wolstenholme, 1990), acquisition (Bartolomei, 2001; Ford & Dillard, 2008, 2009a, 2009b; Homer & Somers, 1988), and large system programs (Homer & Somers, 1988; Lyneis, Cooper, & Els, 2001). Coyle (1996) also provided a survey of applications of system dynamics to military issues.

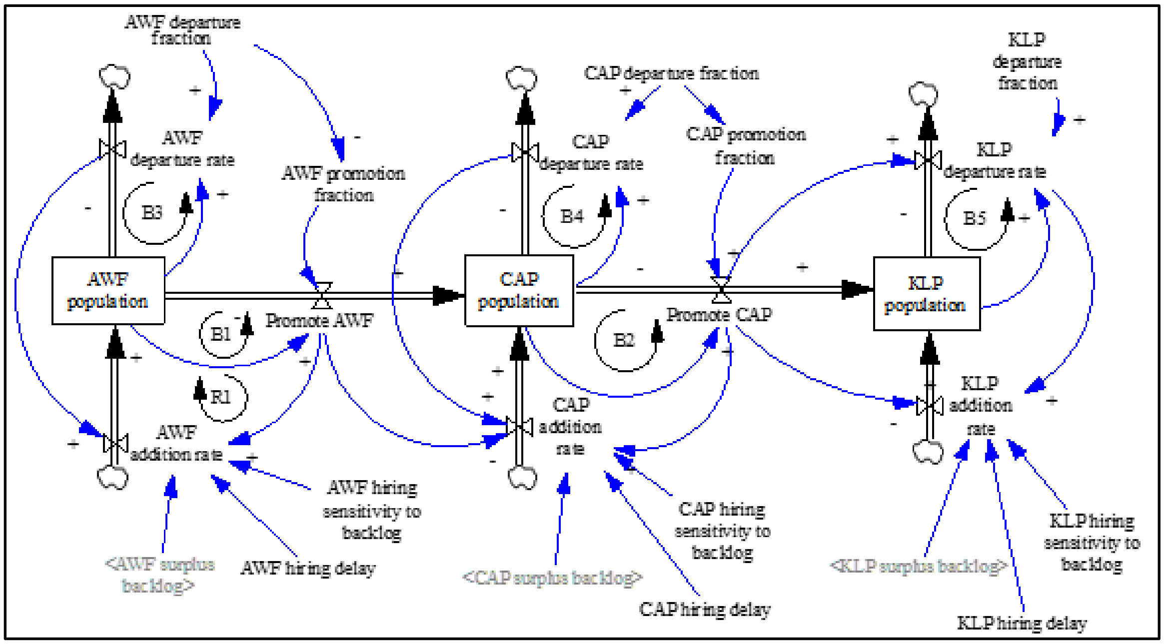
The nature and extent of the acquisition workforce challenges faced by the DoN were identified and described to develop a specific focus for the research. The key variables that best describe the important concepts to be considered in the research were identified. The problem description was used to identify the data required for model development. Available literature was used to build a basic understanding of the core components and processes within the DoN acquisition workforce. Discussions with the DoN Director of Acquisition Career Management (DACM) personnel were used to collect additional information for model development. The problem description was also used to develop a conceptual model of the DoN acquisition workforce using established system dynamics model structures. The conceptual model was formalized into a computer simulation model that uses difference equations to describe the components and relationships in the system. The model structure and policies were fully specified and documented. Estimated parameter values were used to calibrate the formal model. The model was tested to improve the model structure and to develop confidence in the model's ability to reflect DoN acquisition workforce issues. Standard structural and behavior tests for system dynamics models (Sterman, 2000) were used. Structural tests included testing the model's structural similarity to the DoN workforce system and unit consistency tests. Behavior tests included extreme conditions testing and similarity of simulation results to the reference modes. The tested model was used to illustrate DoN acquisition workforce challenges and the impacts of a variety of policies.

Model Structure

The model described herein represents the DoN acquisition workforce system for a single career field. Primary data for the model were developed at a meeting with the DoN director of Acquisition Career Management (DACM) on May 23, 2017. The model is based on the parallel accumulations and flows of three critical parts of the acquisition workforce system: (1) people, (2) knowledge and experience, and (3) work. The model is purposefully simpler than the actual system to facilitate describing and understanding the workforce-related drivers of acquisition performance. For instance, the model depicts only the government acquisition workforce and does not include the government research and development or readiness and sustainment workforces, nor does it include industry contractors.

The Acquisition Workforce

At any point in time, each person in the system is assumed to be a member of one of three populations, each hereafter referred to as a category: the acquisition workforce (AWF), the Critical Acquisition Positions workforce (CAP), or the Key Leadership Positions workforce (KLP). Each of these populations is modeled as a stock (boxes in Figure 2). A category's population size, combined with its productivity, is taken as a surrogate for the resources applied to perform work. After an average period of time as a member of the AWF, a portion of those persons are promoted to CAP (arrow from AWF to CAP and loop B1). Similarly, a portion of the CAP population is promoted to KLP after a characteristic time as a CAP (arrow from CAP to KLP and loop B2). This creates an aging chain of people moving through the system over time. Each stock is also drained by departures from the SYSCOM/career field (arrows above boxes and loops B3, B4, and B5) and increased by "additions" (hiring and transfer of persons into that category, arrows below boxes). Departures are modeled as a fraction of the population per month. In steady state, AWF additions are assumed to include the replacement of AWF promotions (loop R1); additions to each population equal the net effect of promotions and departures; and KLP departures are assumed to include promotions from CAP.



Legend of Feedback Loops [1]

B1 & B2 – Promotion controls the workforce population
B3, B4, & B5 – Departures control the workforce population
R1 – AWF promotions increases AWF hiring which increases AWF population and promotions

Figure 2. System Structure Diagram of People in the Acquisition Workforce System Dynamics Model

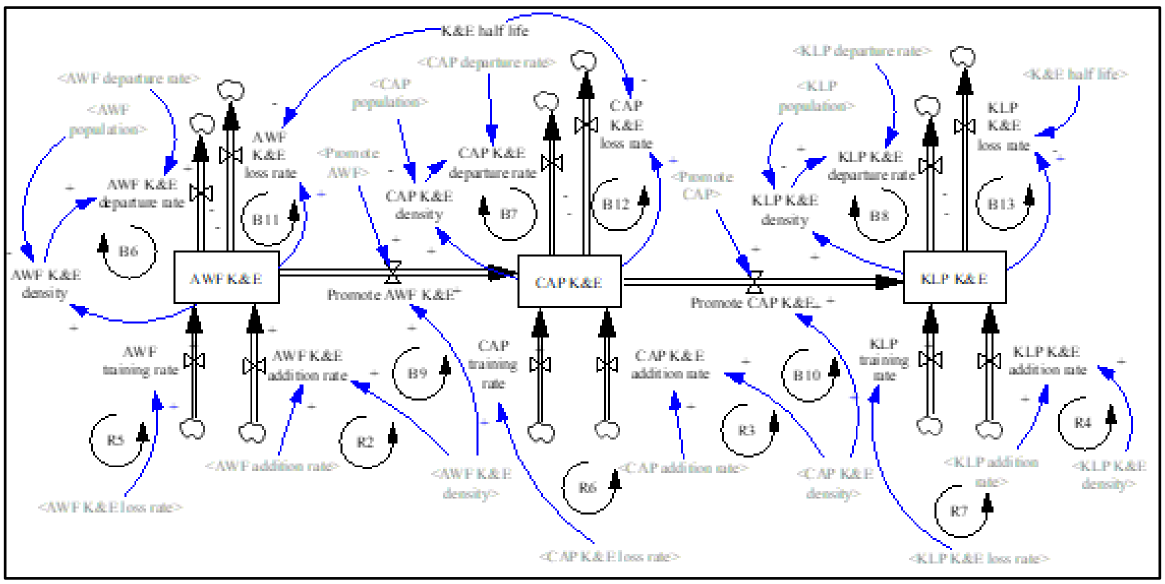
A simple workforce management policy is modeled. The workforce managers are assumed to respond to surplus backlog over a target amount of backlog by adding people to the workforce. People are added through hiring or transfers into each of the three workforce categories (AWF, CAP, KLP) through the “addition rate” inflow to each workforce stock. The size of the addition rate at any point in time is based on three factors: (1) the size of the surplus backlog over the target backlog, (2) the sensitivity of the workforce manager to surplus backlog, and (3) the hiring/transfer delay between the increase in the backlog surplus and when the new staff become productive.

Some levers for improving DoN acquisition system performance increase the numbers of people in some or all of the three categories. For example, retention efforts that are focused on the KLP population reduce the departure rate from the stock of KLP persons (but not the AWF or CAP stocks). People can be added through increased hiring and transfers from outside the SYSCOM or career field. Other levers, such as the average time spent in a category (and thereby the promotion rates), limits on the sizes of category populations, and hiring policies (e.g., hiring for attrition) also impact the inflows to and outflows from the three population stocks.

[1] Feedback loops that are required purely to start the model in steady state conditions are not fully described here for brevity but are included in the formal model.

Knowledge and Experience (K&E)

The accumulations and flows of the total knowledge and experience of each category of persons are modeled in an aging chain (see Figure 3) that mimics the structure of the aging chain of the groups of people described previously. A category's K&E reflects its capability to perform work. Knowledge and experience is measured in generic "K&E units," which include educational degrees, person-years of experience, training, certifications achieved, impacts of mentoring and other on-the-job training, and so forth. For each of the three categories, the average K&E of a person in the category (i.e., the K&E density) is estimated as the stock of K&E divided by the number of persons in that category (loops B6, B7, and B8). When the average density of K&E increases, more K&E is lost through promotions from the category the person is promoted from (loops B9 and B10) and gained by the category that the person is promoted to. Similarly, K&E is lost through departures (loops B11, B12, and B13) and gained through hiring and transfers (loops R2, R3, and R4) based on the K&E density and those flows of persons. K&E is also lost from each stock due to people forgetting and knowledge and experience becoming obsolete (right outflows above stocks in Figure 3 and loops B11, B12, and B13), based on the half-life of the knowledge and experience in that SYSCOM and career field.



Legend of Feedback Loops

B6, B7, & B8 – If the amount of K&E and therefore the K&E density increases, more K&E is lost when people depart, thereby limiting the amount of K&E

B9 & B10 – If the amount of K&E and therefore the K&E density increases, more K&E is lost when people are promoted, thereby limiting the amount of K&E

B11, B12, & B13 – If the amount of K&E increases, more K&E is lost (e.g., through obsolescence), thereby limiting the amount of K&E

R2, R3, & R4 – If the amount of K&E and thereby the K&E density increases, the amount of K&E added when people join the population also increases. This increases the amount of K&E.

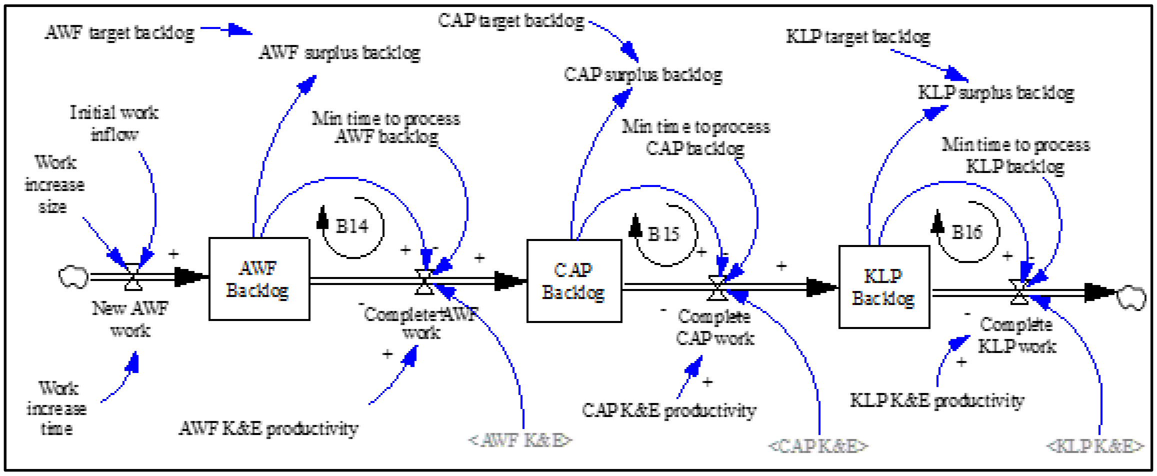
R5, R6, & R7 – K&E losses are replaced by training, which increases the "churn" of K&E

Figure 3. System Structure Diagram of Knowledge and Experience in the Acquisition Workforce System Dynamics Model

To describe steady state conditions, the model assumes that K&E losses are replaced by the addition of equal amounts of K&E through training (loops R5, R6, and R7), where “training” is taken as a shorthand label for any development activities that increase capability or capacity to perform the work required, such as education, structured training, on-the-job training, job experience, mentoring, and so forth. Numerous levers can influence the quantity, quality, availability, and impacts of education, experience, and training opportunities, and thereby the K&E inflows and outflows to the three K&E stocks.

Work

Each category of persons performs work that decreases their backlog of work to be completed (i.e., the category’s backlog plus work in progress, referred to hereafter as their backlog). Work is measured in small fungible information packets (info packet). An info packet may be taken as a standard work package unit that represents typical acquisition work products such as design documents, contract actions, engineering changes orders, test plans, and so forth. The three backlogs are modeled as stocks with the completion of work moving that work out of the category’s backlog and into the backlog of its downstream category (see Figure 4). This creates the third parallel aging chain of the model. The completion rates increase with the amount of K&E in the category and the average productivity of the category. These flows are also limited by the size of the backlogs to model the constraint that work cannot be completed if the backlog is empty (loops B14, B15, and B16). New work is assumed to enter the most upstream backlog (AWF Backlog) based on the acquisition demand signal. When modeling steady state conditions, the amount of new work added to the AWF backlog is equal to the completion of AWF work. This simplified model assumes that all AWF work must be reviewed and approved or processed sequentially by CAPs and then KLPs, which is not necessarily true in the real system. Further, in reality both CAPs and KLPs may have New Work inflows that do not originate with New AWF work. Also, the model ignores quality of work, error rates, and rework.



Legend of Feedback Loops

B14, B15, & B16 – More backlog increases completion, controlling backlogs (prevents negative backlogs)

Figure 4. System Structure Diagram of Work in the Acquisition Workforce System Dynamics Model[2]

A simple depiction of management monitoring the workforce for decision-making is modeled. It is assumed that the workforce manager monitors the current backlog (AWF, CAP, and KLP) and compares the current backlog to a target backlog size to estimate the surplus backlogs. In general, larger surplus backlogs indicate worse performance.

Increasing the productivity of applying the K&E in each category can improve DoN acquisition system performance. For example, changes in working conditions (e.g., telecommuting) and schedules (e.g., flexible work schedules) can increase productivity. Performance-based incentives can also increase productivity. Note that increasing productivity does not increase the quantity of K&E (modeled in the K&E sector), but how effectively the K&E is applied.

Performance Measures and Model Calibration

Thirteen traditional measures of aging chain performance are included in the model: three category sizes and the total workforce size, three category work backlogs and the total backlog of work, average time that work stays in the three backlogs and in the whole system, and annual workforce cost.

The model was initially calibrated to steady state conditions to facilitate distinguishing between model adjustment from initial conditions to steady state and model responses to exogenous inputs. Calibrated conditions were estimated by the modelers based upon a generic but realistic single career field in a single SYSCOM.

---

[2] Feedback loops that are required purely to start the model in steady state conditions are not fully described here for brevity but are included in the formal model.

Illustration of Model Use

For a demonstration of typical behavior and an illustration of model use, the inflow of work to the AWF backlog was assumed to increase by 25% at month 40 and remain at the higher level. As shown in Figure 5, the AWF initially increases. This is because the current AWF workforce cannot complete more work. However, the increased backlog sends a signal that additional resources are required to the AWF workforce manager in the form of the AWF backlog surplus.

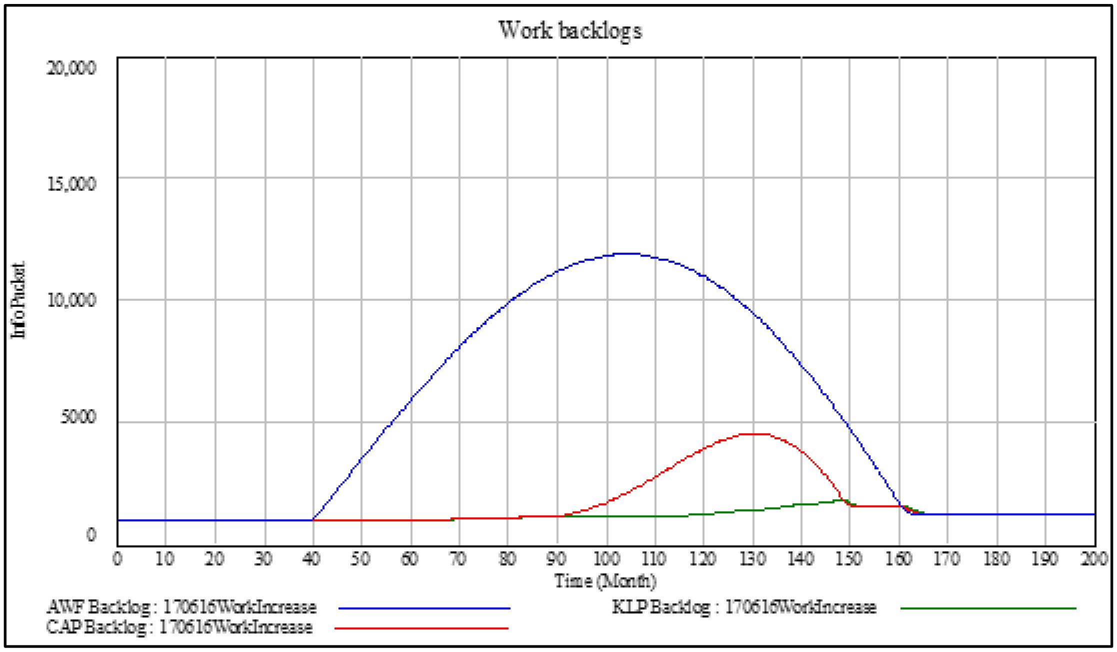


Figure 5. Changes in Work Backlogs Due to an Increase in Workflow

After the AWF hiring/transfer delay (assumed to be six months), the number of people in the AWF workforce increases (see Figure 6, months 0–170).

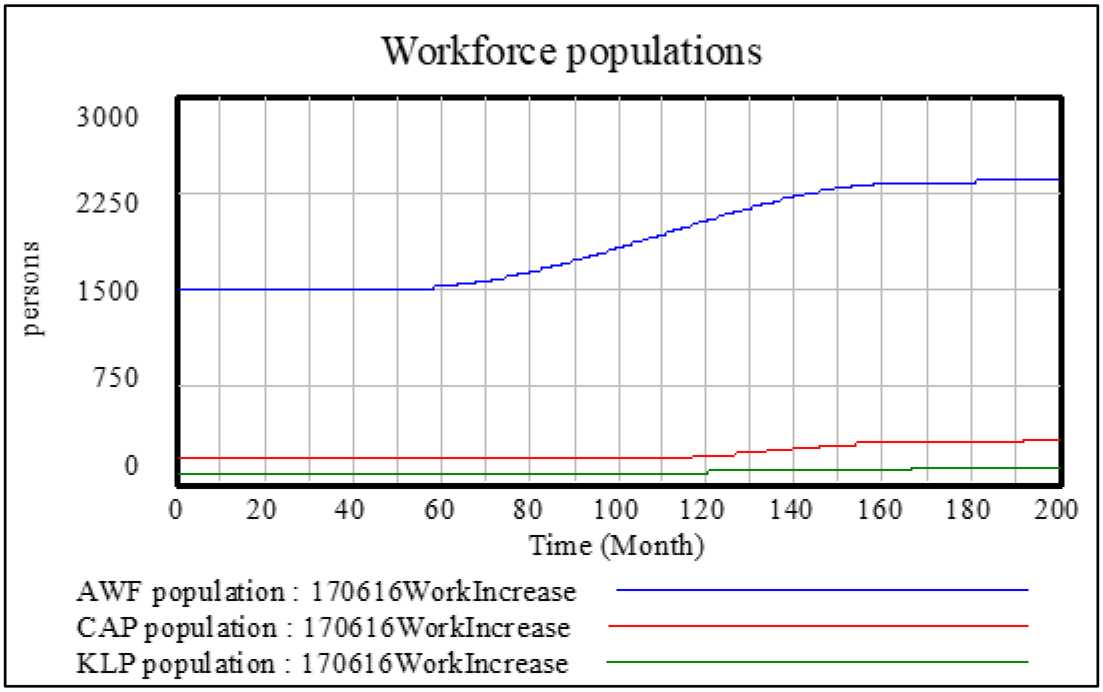


Figure 6. Changes in Workforce Sizes Due to an Increase in Workflow

The increase in the AWF population increases the knowledge and experience available to apply to the AWF backlog (see Figure 7, months 60–170). This increased capacity reduces the rate of rise in the AWF backlog (see Figure 5, months 75–105) and then starts to decrease that backlog (see Figure 5, months 105–165). The resulting decrease in the backlog surplus decreases the AWF hiring/transfer rate until a new steady state AWF workforce is achieved (see Figure 6, months 170+), which stabilizes the knowledge and experience in the AWF workforce (see Figure 7, months 170+).

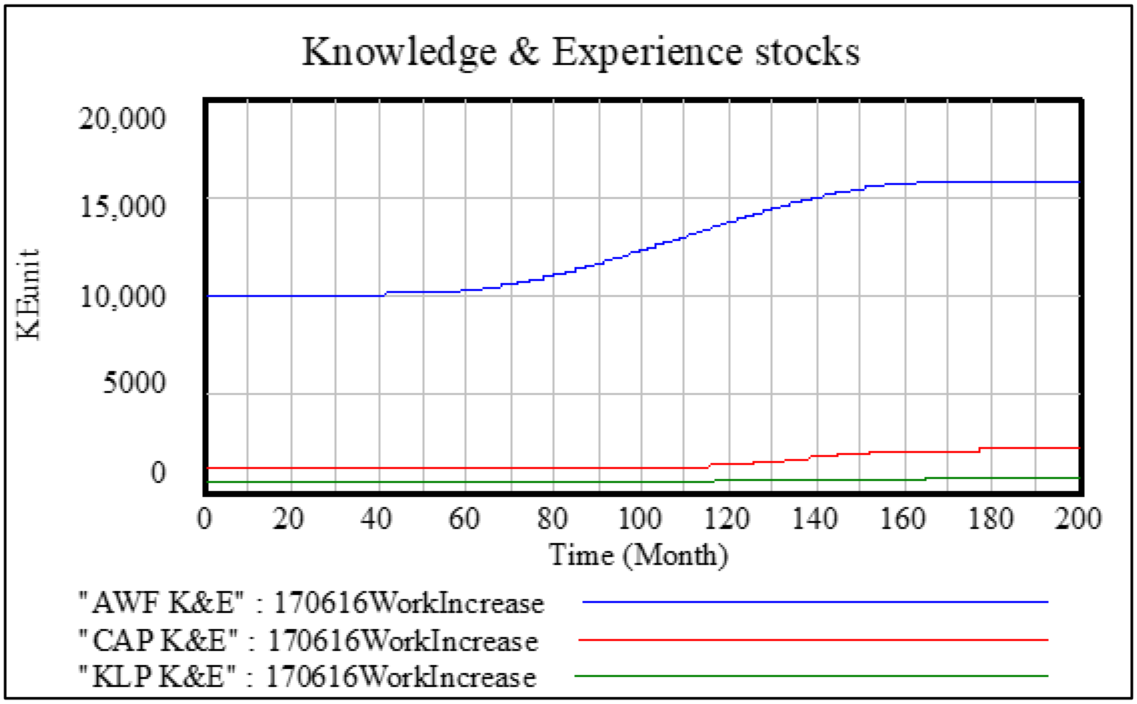


Figure 7. Changes in Knowledge and Experience Due to an Increase in Workflow

The increase in the workflow propagates through the AWF workforce, knowledge and experience, and backlogs into the CAP and KLP workforces, knowledge and experience, and backlogs. In turn, each of those two downstream workforces experience an increase in backlog (see Figure 5, months 90–145) and surplus backlog as their current workforce is unable to increase their completion rates. After the CAP and KLP hiring/transfer delays (assumed to be nine and 12 months, respectively) their workforces increase (see Figure 6, months 115–195), which increases their knowledge and experience (see Figure 7, months 115–195) and eventually their production rates, which reduces their backlogs (see Figure 5, months 130–165) to near their target levels.

The model also simulates system performance measures such as changes in the total workforce size (see Figure 8, grey line), average time required for a piece of work to be processed (see Figure 9), and monthly workforce costs (see Figure 10).

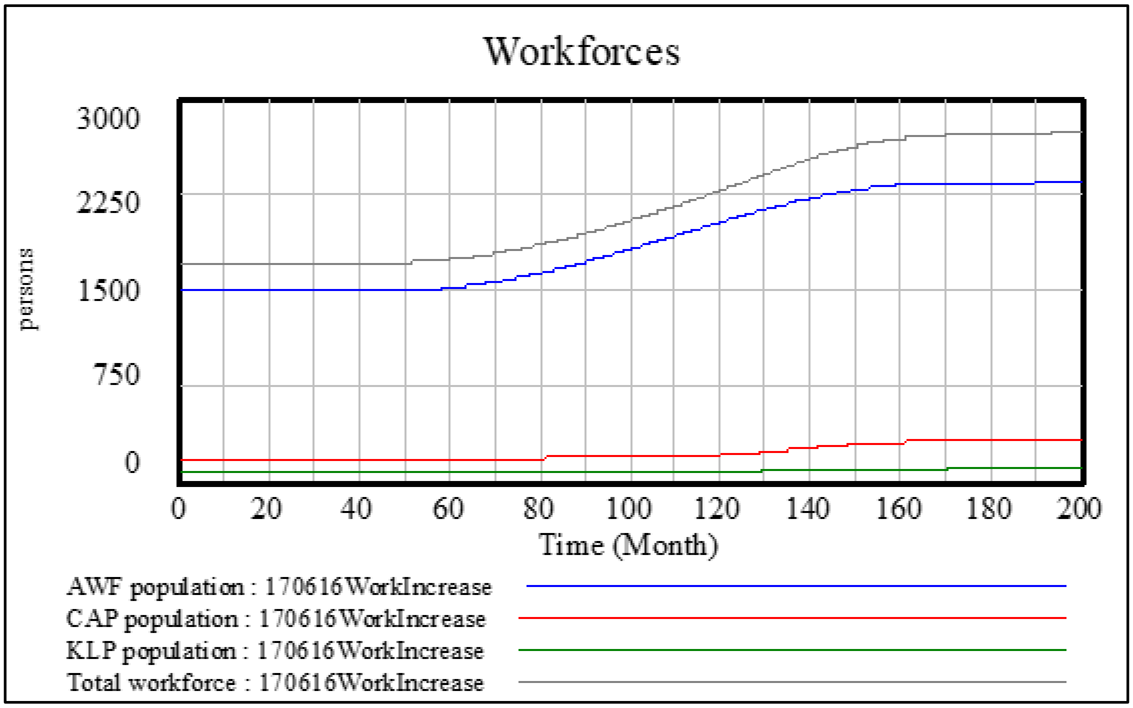


Figure 8. Changes in Workforce Sizes Due to an Increase in Workflow

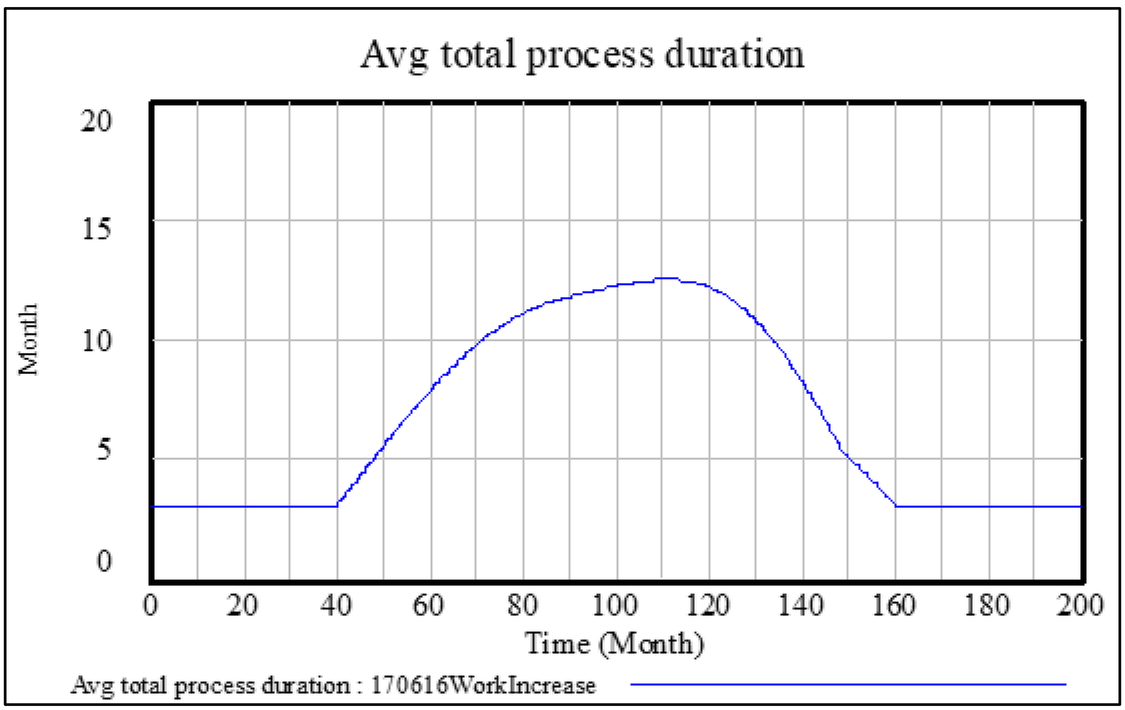


Figure 9. Changes in Average Time Required for a Piece of Work to Be Processed Due to an Increase in Workflow

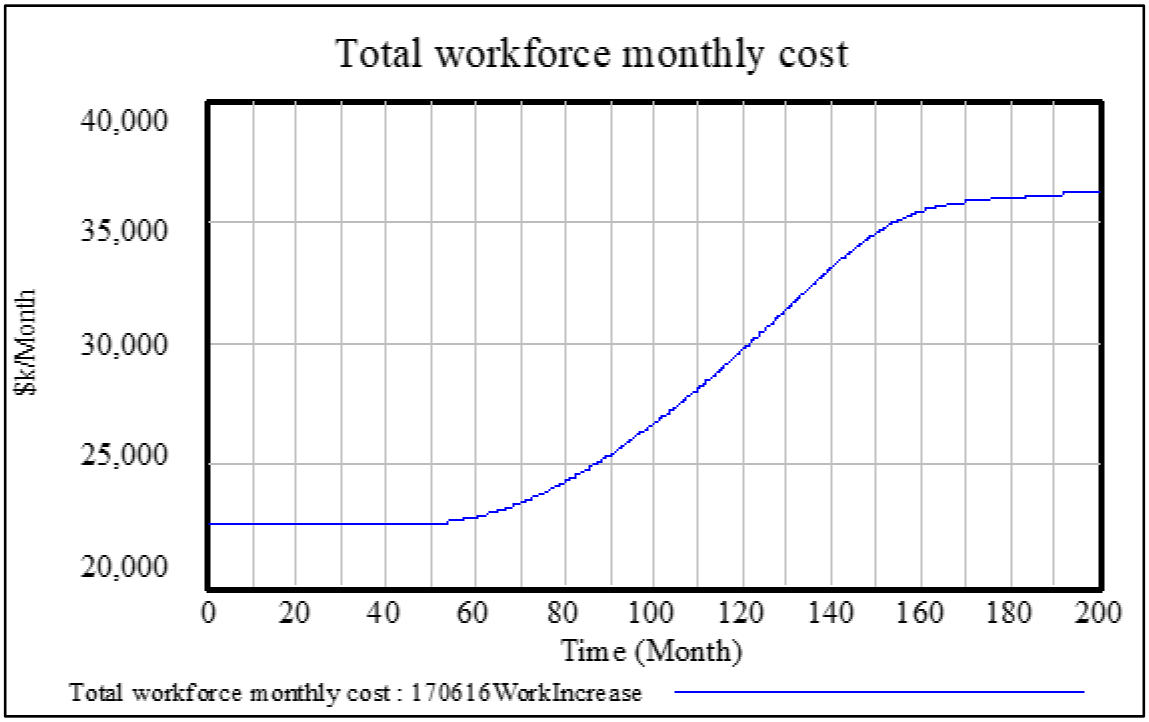


Figure 10. Changes in Monthly Workforce Cost Due to an Increase in Workflow

Planned Model Development

The progression of system descriptors from input (policies, etc.) to output (performance measures) is perceived as progressing through four types of model parameters:

Levers → Mediating Factors → Acquisition Processes → Performance Measures

Example levers, mediating factors, acquisition processes, and performance measures that may be useful in the model include the following:

Levers

- Acquisition Category I (ACAT I) versus ACAT II reporting requirements
- Rotation & job swapping
- Defense Acquisition Workforce Improvement Act (DAWIA) certification requirements
- Advanced degrees
- Leadership Development Programs (LDPs)
- Mentoring
- Hiring and retention bonuses
- University of North Carolina (UNC) and University of Virginia (UVA) Darden
- Recruiting strategies
- Various training opportunities

Mediating Factors

- Departure fractions and departure rates / Fraction (AWF, CAP, KLP)
- Average time in category/promotion fraction (AWF, CAP, KLP)
- Training and education of the workforce (AWF, CAP, KLP)
- Productivity of applying knowledge and experience (AWF, CAP, KLP)
- Half-life of knowledge and experience
- Hire rates (AWF, CAP, KLP)
- Transfer rates (AWF, CAP, KLP)

Acquisition Processes

- Completion rates (AWF, CAP, KLP)

Performance Measures

- Dashboard metrics
- Work backlog (AWF, CAP, KLP)
- Total work backlog
- Workforce size (AWP, CAP, KLP)
- Total workforce size
- Workforce cost per month (AWF, CAP, KLP)
- Total workforce cost per month
- Cumulative workforce cost
- Average process duration (time work stays in backlog + work in progress) (AWF, CAP, KLP)
- Average total process duration (time work stays in backlog + work in progress)

Future work can add data on the practices and potential DACM actions within one or more specific categories (i.e., career field in a SYSCOM). This can be used to further develop the model to reflect active and future changes in the workforce system. Data from the categories can be used to calibrate and validate the model. Simulations using the validated model can then be used to analyze current and potential future actions to improve the acquisition workforce. The results can be the basis for recommendations to the DoN DACM.

References

Bakken, B. E. (2004). The Atlantic defense technology gap: Will it be closed? In *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England: System Dynamics Society.

Bakken, B. E., & Gilljam, M. (2003). Dynamic intuition in military command and control: Why it is important, and how it should be developed. *Cognition, Technology and Work, 5,* 197–205.

Bakken, B. T., Gilljam, M., & Haerem, T. (2004). Perception and handling of complex problems in dynamic settings: Three cases of relevance to military command and crisis management. In *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England: System Dynamics Society.

Bakken, B. T., Ruud, M., & Johannessen, S. (2004). The system dynamics approach to network centric warfare and effects based operations: Designing a “learning lab” for tomorrow. In *Proceedings of the 22nd International Conference of the System Dynamics Society*. Oxford, England: System Dynamics Society.

Bakken, B. T., & Vamraak, T. (2003). Misperception of dynamics in military planning: Exploring the counter-intuitive behaviour of the logistics chain. In *Proceedings of the 21st International Conference of the System Dynamics Society*. New York, NY: System Dynamics Society.

Bartolomei, J. (2001). A system dynamics model of government engineering support during the development phase of a military acquisition program. *In Proceedings of the 19th International Conference of the System Dynamics Society*. Atlanta, GA: System Dynamics Society.

Bell, J. W., Jr., & Liphard, R. E. (1978*). A system dynamics model of the Department of Defense enlisted force for investigation of alternative retirement proposals* (GPO Report No. ADA065970). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Coyle, R. G. (1996, August). System dynamics applied to defense analysis: A literature survey. *Defense & Security Analysis, 212*(2), 141–160.

Coyle, R. G., & Gardiner, P. A. (1991). A system dynamics model of submarine operations and maintenance schedules. *Journal of the Operational Research Society, 42*(6), 453–462.

DoD. (n.d.). What is Better Buying Power? Retrieved November 3, 2015, from Better Buying Power website: http://bbp.dau.mil/

DoD. (2015). *Performance of the defense acquisition system: 2015 annual report.* Retrieved from Better Buying Power website: http://bbp.dau.mil/docs/Performance-of-Defense-Acquisition-System-2015.pdf

DoN. (2015, March). *A cooperative strategy for 21st century seapower*. Retrieved from http://www.navy.mil/local/maritime/150227-CS21R-Final.pdf

Duczynski, G. (2000). Profiler: An “effects-based” military capability manager. In *Proceedings of the 18th International Conference of the System Dynamics Society* (pp. 59–72). Bergen, Norway: System Dynamics Society.

Flood, R. L., & Jackson, M. C. (1991). *Creative problem solving: Total systems intervention.* Chichester, UK: Wiley.

Ford, D. N., & Dillard, J. T. (2008). Modeling the integration of open systems and evolutionary acquisition in DoD programs. In *Proceedings of the Fifth Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Ford, D. N., & Dillard, J. T. (2009a). Modeling open architecture and evolutionary acquisition in ARCI with applications to RCIP. In *Proceedings of the Sixth Annual Acquisition Research Symposium.* Monterey, CA: Naval Postgraduate School.

Ford, D. N., & Dillard, J. T. (2009b). Modeling the performance and risks of evolutionary acquisition. *Defense Acquisition Review Journal, 16*(2), 143–158.

Forrester, J. W. (1961). *Industrial dynamics*. Waltham, MA: Pegasus Communications.

GAO. (2012). *Defense acquisition workforce: Improved processes, guidance, and planning needed to enhance use of workforce funds* (GAO-12-747R). Retrieved from http://www.gao.gov/products/GAO-12-747R

Homer, J. B., & Somers, I. (1988). Defense program lifecycle management: A dynamic model for policy analysis. In *Proceedings of the 1988 International System Dynamics Conference*. La Jolla, CA: International System Dynamics Society.

Jackson, M. C. (2003). *Systems thinking: Creative holism for managers*. Chichester, UK: Wiley.

Lane, D. C., & Jackson, M. C. (1995). Only connect! An annotated bibliography reflecting the breadth and diversity of systems thinking. *Systems Research and Behavioral Science, 12*(3), 217–228.

Lyneis, F., Cooper, K., & Els, S. (2001). Strategic management of complex projects: A case study using system dynamics. *System Dynamics Review, 17*(3), 237–260.

McLucas, A. C., Lyell, D., & Rose, B. (2006). Defence capability management: Introduction into service of multi-role helicopters. In *Proceedings of the 24th International Conference of the System Dynamics Society* (pp. 92–110). Nijmegen, The Netherlands: System Dynamics Society.

Melhuish, J., Pioch, N., & Seidel, A. (2009). Improving military strategy using predictive agents with embedded mental simulation models. In *Proceedings of the 27th International Conference of the System Dynamics Society*. Albuquerque, NM: System Dynamics Society.

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. New York, NY: McGraw-Hill.

Watts, K. M., & Wolstenholme, E. F. (1990). The application of a dynamic methodology to assess the benefit of a logistics information system in defense. In *Proceedings of the 1990 International System Dynamics Conference* (pp. 128–136). Chestnut Hill, MA: International System Dynamics Society.

Embracing Structure and Discipline to Manage Challenges in the Department of Navy Acquisition Workforce

Altyn Clark—is a troubleshooter whom leaders call to diagnose ambiguity and craft a way ahead to address wicked problems. He uses the engineering design process, coupled with industrial psychology and methods of large scale change, to assess and advance organizational systems. As Chief Solutions Officer of Transformation Systems Inc. (TSI), Dr. Clark invents practical and useful models, methods, approaches, and tools. He brings over 30 years of subject matter expertise in engineering, quality control and assurance, program and personnel management, research, logistics, human capital, and strategic management. Dr. Clark received his PhD, MS, and BS all from Virginia Tech. [ac@transformationsystems.com]

Elizabeth S. Rosa—has supported several U.S. Navy programs in the following roles: technical direction agent, in-service engineering agent, reliability lead, risk manager, contract specialist, engineering project manager, and acquisition manager. She uses this diverse background of experience to bring a unique perspective to acquisition workforce management. Rosa received her BS and BA from the University of Rhode Island.

Introduction

The Department of Navy (DoN) Acquisition Workforce Strategic Plan Fiscal Year (FY) 2016–2022 set three major goals for leaders of the Acquisition Workforce (AWF): (a) energize the workforce, (b) ensure there is a focus on professional and technical excellence, and (c) reinforce responsibility and accountability. Ensuring these goals are met, through a structured and disciplined process to build a workforce to compete and win, optimizes the DoN's ability to develop smart buyers who understand that every person matters, every day matters, and every dollar matters as we meet warfighter needs for affordable, agile, lethal capability.

The Naval Acquisition Workforce Is Itself a Major Acquisition Program

The DoN is committed to establishing and maintaining a specialized, professional, world class, agile, motivated AWF that consistently makes smart business decisions, acts in an ethical manner, and delivers timely and affordable capabilities to the warfighter. The workforce is comprised of over 60,000 Navy and Marine Corps civilian and military employees, in 14 career fields and more than 15 major commands. They are located worldwide in system command (SYSCOM) offices, research labs, industrial complexes, and test ranges. The DoN invests roughly $7 billion per year in salaries and benefits for the AWF. This workforce is highly technical and is responsible for approximately $60 billion of the Navy's budget per year. Within a context of legislation, regulation, and policy, they innovate, design, build, sustain, modernize, and maintain complex ships, aircraft, and vehicles with associated equipment, combat systems, weapons, and ordnance to support the DoN's military missions. Many of the major acquisition programs have life cycles exceeding 50 years and often span more than a single workforce generation.

Downsizing pressure and the focus on speed and cost control are greater than ever before. Recent sequestrations, pay freezes, and furloughs, along with associated budgetary turmoil, have had a negative effect on the AWF. With an operating environment characterized by constrained budgets, increasing threat profiles, increasing system complexity, limited competition, a shrinking industrial base, and cyber security challenges, maintaining our technological edge will require an even more innovative, astute, proactive, and responsive AWF. To operate effectively in this environment, the DoN must better understand the cost, productivity, and risk associated with AWF staffing to responsibly manage this workforce throughout its life cycle (recruiting, hiring, developing, retaining, and retiring). Our challenge is to manage the AWF as a major acquisition program. Five enduring major acquisition program themes apply to the AWF when considering the workforce as a major program itself:

- Getting the Requirements Right—Understand the operational and technical requirements and the drivers of our workforce requirements over time.
- Performing to a Stable Plan—Hire and grow strategically, aligned to National Strategy and Naval Strategy documents.
- Making Every Dollar Count—Make every dollar count across the AWF life cycle phases. Leverage best practices and investments.
- Relying on Experienced AWF—Better define the career paths within the career fields to create experts with the proper education, training, and experience (hands-on, job rotations, industry knowledge).
- Fostering a Healthy Industrial Base—Understand the right balance of knowledge and experience between industry and government.

The Structure of the Naval Acquisition Workforce Begins With Chain of Command and U.S. Code

The DoN acquisition chain of command, responsibility, authority, and accountability flows from the Assistant Secretary of the Navy for Research Development and Acquisition (ASN[RD&A]) to program executive officers (PEOs) and to the program managers (PMs) supported by the systems commands (SYSCOMs). The chain of command serves several important purposes in the accomplishment of the DoN's acquisition mission. It defines responsibilities and identifies accountability, provides direction and clear communications, and promotes efficiency and effectiveness. Ensuring that the chain of command is carried through to all levels enables informed and sound business judgment in acquisition and empowers the DoN AWF to meet future challenges with resilience.

The DoN AWF must have technical and professional excellence to ensure a technological edge for our military, while balancing cost, schedule, and risk, in a complex legislative, regulatory, and dynamic policy environment. Within the DoN, strategies to strengthen the capability and capacity of this workforce continue to be a major element of emphasis, and significant efforts are on track to shape and improve productivity and quality with a focus on having the right people, in the right job, at the right time.

The Defense Acquisition Workforce Improvement Act (DAWIA; 1990) and 10 U.S.C., Chapter 87 provide the foundation for a system of policies and processes used to effectively manage the AWF. The DoN AWF is represented by 14 career fields worldwide, listed in Table 1.

Table 1. DoN Acquisition Career Fields

DoN Acquisition Career Fields	
Business-Cost Estimating (BUS-CE)	Life Cycle Logistics (LCL)
Business-Financial Management (BUS-FM)	Production, Quality and Manufacturing (PQM)
Contracting (CON)	Program Management (PM)
Engineering (ENG)	Purchasing (PUR)
Facilities Engineering (FE)	Science and Technology Manager (S&TM)
Industrial/Contract Property Management (IND)	Small Business (SB)
Information Technology (IT)	Test and Evaluation (T&E)

The Mission of the Naval Acquisition Workforce Is Vital to National Security

The foundation of any organization is a set of values that connect women and men working together with focus on their mission. There is no more noble mission than to support the men and women who are protecting and serving our nation. The AWF is grounded in the values of integrity, trust, diversity, teamwork, dedicated service, and excellence. Having highly educated, highly skilled, and experienced acquisition professionals is key to providing the warfighters the products and services they need for success. For the DoN AWF, having the right people, in the right job, at the right time will translate to effective and efficient execution, delivering the finest warfighting capability in the world at an affordable price.

DoN Acquisition is a team sport. It requires us to be innovative and holistic in our thought. We recognize and embrace diversity as we recruit, develop, and retain the best and brightest to provide knowledge, oversight, and stewardship to our acquisition programs. By informed policy, sound understanding of scientific and technological advancements, and deliberate management of the AWF across all career fields, we demonstrate our commitment to our men and women in uniform who have dedicated their lives to the protection of our freedom.

The AWF must have professional and technical excellence to deliver the DoN’s complex and highly technical warfighting capability. Highly educated and highly skilled professionals in engineering, cost estimating, financial management, and contracting, as well as in program management, science and technology, life-cycle logistics, information technology, facilities engineering, test and evaluation, small business, industrial/contract property management, and production quality and manufacturing, are required for the acquisition of these capabilities.

The Origin of the Defense Acquisition Workforce Improvement Act

DAWIA was initially enacted by Public Law 101-510, dated November 5, 1990. This law, entitled Defense Acquisition Workforce, was subsequently incorporated into Title 10, U.S. Code, Chapter 87. The primary objective of DAWIA is to enhance the professional knowledge and capabilities of DoD personnel involved in the development, acquisition, and sustainment of warfighting capabilities, systems, and services. The law requires the DoD to establish a process through which persons in the AWF would be recognized as having achieved professional status. It also requires the Secretary of Defense (SECDEF) to establish policies and procedures for the effective management (including accession, education, training, and career development) of DoD military and civilian personnel occupying acquisition positions, referred to as members of the defense AWF. The law also requires the SECDEF to ensure, to the maximum extent practicable, that established AWF policies and procedures are uniformly implemented throughout the DoD.

The Office of the Assistant Secretary of the Navy (Research, Development, and Acquisition)

The ASN(RD&A) serves as the service acquisition executive for the DoN. There are several key positions that support the ASN(RD&A), including the following:

- ASN(RD&A) Principal Military Deputy (PMD): The ASN(RD&A) PMD serves as the ASN(RD&A) principal deputy for executive oversight and leadership of the military AWF and as the co-chair of the Acquisition Career Council (ACC); This person plays a key role in the guidance and monitoring of the implementation of ASN(RD&A) acquisition career management efforts to ensure a qualified and capable AWF.
- ASN(RD&A) Principal Civilian Deputy (PCD): The ASN(RD&A) PCD serves as the ASN(RD&A) principal deputy for executive oversight and leadership of the civilian AWF and as the co-chair of the ACC, which plays a key role in the guidance and monitoring of the implementation of ASN(RD&A) acquisition career management efforts to ensure a qualified and capable AWF.
- Director, Acquisition Career Management (DACM): The DACM serves as the enterprise lead for the professional development and management of the Navy and Marine Corps AWF and is the chief advisor to the ASN(RD&A). This person serves as the focal point within the department for all matters related to the AWF, including strategic planning, policy development and guidance on all matters associated with DAWIA implementation.
- Director, Naval Acquisition Career Center (NACC): The NACC manages and executes the Navy Acquisition Development Program (NADP), the Acquisition Workforce Tuition Assistance Program (AWTAP) under the direction of the DACM, eDACM Support team, DAU course registration, DAU Travel, and on-site DAU training.
- DoN National Leads (NLs): DoN NLs are senior executives assigned to provide oversight of each DoN acquisition career field and to monitor the health of the acquisition career fields and promote acquisition excellence, validate the adequacy of education, training, and other developmental opportunities for respective career fields, and work with the acquisition commands to implement career paths and talent management. Note: Oversight for the PQM career field is the responsibility of DoN SYSCOMs.
- Acquisition Career Council (ACC): The ACC is a cross-competency group that consists of DoN NLs and the DACM and is co-chaired by the ASN(RD&A) principal civilian and principal military deputies. The ACC functions as an advisory council to the ASN(RD&A; see Figure 1).

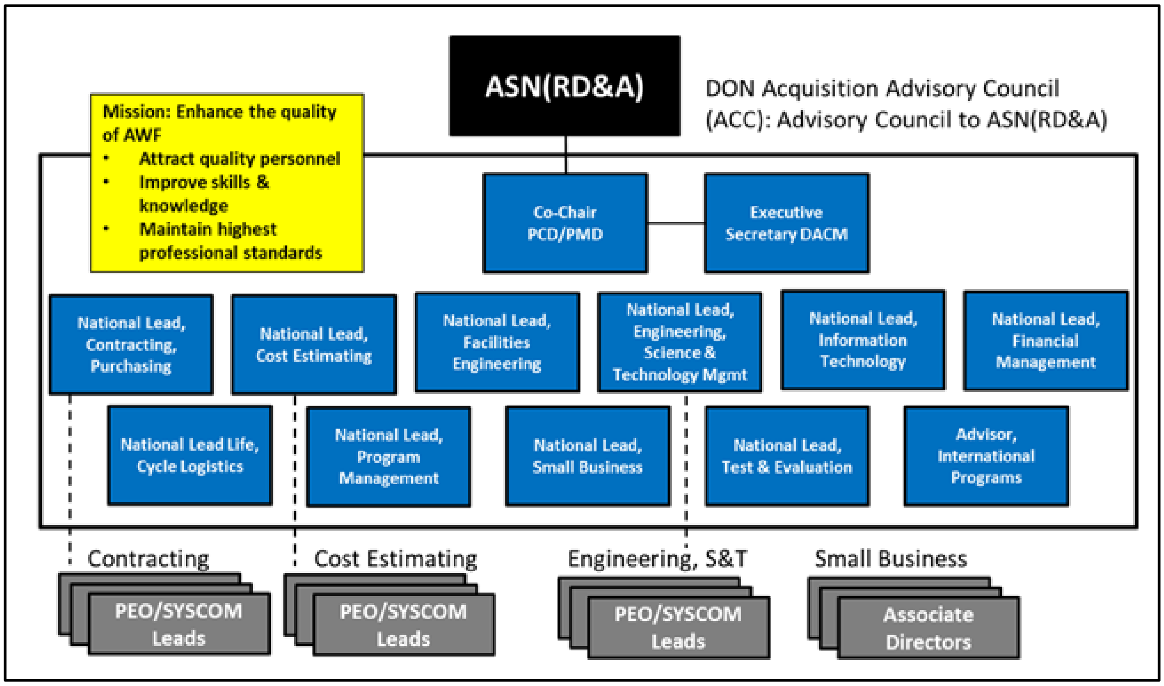


Figure 1. DoN Acquisition Career Council (ACC): Advisory Council to the ASN(RD&A)

Workforce Management Structure

Part of meeting the goal of ensuring technical and professional excellence, major acquisition commands and other organizations with significant acquisition functions must have a clearly defined management structure to support the implementation of DAWIA within the organization. The DAWIA management structure can be incorporated into an actual or virtual organizational structure. Roles and responsibilities at different levels may be consolidated if appropriate as long as the overarching goals are met.

The Governance Heartbeat

The ACC Charter

The rhythm of meetings, conversations, workshops and reports forming the foundation of AWF governance is anchored in the charter for the ACC. The group's name was chosen carefully to denote an emphasis on people and acquisition career development across all career fields over a 30+-year career life cycle. In the charter, the PCD and PMD declare that national leads shall

1. Establish a career field team
2. Meet quarterly
3. Make quarterly agendas, read-aheads, and minutes available to the ACC
4. Develop strategic implementation plans to deploy three goals from the AWF Strategic Plan
5. Define and promote productivity, innovation, and professional and technical excellence
6. Advise regarding talent management
7. Implement viable career path(s)

8. Collaborate at the ACC
9. Understand and articulate workload drivers
10. Understand and articulate links to other career fields
11. Understand and articulate sufficiency/risks of funded requirements with respect to demand signals and acquisition excellence
12. Provide progress report to ACC co-chairs (PCD and PMD)
13. Assess career field health using data, trends, and metrics to identify issues and make recommendations
14. Recommend continuous learning opportunities
15. Represent the DoN at the Office of the Secretary of Defense (OSD) functional integrated product teams (FIPTs) and coordinate FIPT tasking with the DoN DACM
16. Nominate AWF members for acquisition awards
17. Plan for rotation of NLs (three-year term)
18. Establish annual plans and advocacy for DAWIA and other resources to support AWF career development
19. Validate the adequacy of acquisition training and development programs

Career Field Quarterly Reports

To promote effective dialog and problem-solving during meetings, the PCD directed NLs to provide a quarterly written report of progress and status. Elements of the report include

- State of the Career Field
    - Most significant accomplishment this quarter
    - Most troubling problem or issue encountered this quarter
    - Any substantial changes in CF AWF health or performance since last quarter
- Progress Implementing AWF (DoD, DoN, CF) Priorities (i.e., focus areas)
    - Progress on CF objectives and supporting initiatives. Reference progress against baseline POAM.
    - Report changes in measures that indicate progress or impact of implementing initiatives or objectives.
    - Report trend data rather than snapshot data.
    - Describe any new objectives or initiatives started since last quarter, if applicable.
    - Major actions planned (pulled from CF Action Tracker) and accomplishments expected for next quarter
- Major Problems and Issues Impacting the Career Field
    - Description of problem/issue and what has changed since last quarter
    - Steps being taken to address problem/issue
- Help Needed from RDA/PCD/PMD
    - Specific action requested
    - Next step(s) NL will take to support the requested action (e.g., draft correspondence for PCD)

Alternating ACC Meetings and AWF Summit Meetings

Participants in quarterly ACC meetings are typically the people named in the charter and a few invited guests. In alternate quarters, the PCD and PMD convene a larger meeting that includes SYSCOM and PEO flag officers and executive directors. Recent discussions suggest that soon there will be quarterly AWF summit meetings, convening the larger group on a more regular basis. Introduction of the written quarterly report precludes the need for status briefs during the meeting, and recent agendas have been constructed so that breakout groups of executive leaders (sprinkled with a mix of employees with one to three years' experience for a different point of view) can explore strategic focal problems and develop courses of action to address them.

Career Field Workspaces in the AWF War Rooms

A separate ARS 2018 paper by Clark discusses paper-based and digital war rooms developed by DACM to promote ready, relevant lateral learning across programs and career fields so that SYSCOMs, PEOs, national career field leaders, and their teams can move smartly and systematically toward managing the 60,000-member AWF like a major acquisition program. The career field workspaces in the AWF war room display content related to the ACC charter elements discussed previously. NLs often convene working group meetings in the digital AWF war room space to surround themselves with policy, models, measures, and data to guide career field development discussions.

Conclusion

Development of the AWF is a continuous and dynamic process that requires metrics and periodic re-evaluation to ensure that the AWF is meeting the demand signal dictated by the warfighter's need. Recognizing the unique skillset required by the DoD AWF, the focus continues to be on vertical and horizontal integration, increased military and civilian interaction, and an increase in diversity of thought.

Adding Agility to Department of Navy Acquisition Workforce Management in Digital Collaboration Centers

Altyn Clark—is a troubleshooter whom leaders call to diagnose ambiguity and craft a way ahead to address wicked DoN problems. He uses the engineering design process, coupled with industrial psychology and methods of large-scale change, to assess and advance organizational systems. As Chief Solutions Officer of TSI, Dr. Clark invents practical and useful models, methods, approaches, and tools. Dr. Clark brings over 30 years of subject matter expertise in engineering, quality control and assurance, program and personnel management, research, logistics, human capital and strategic management. Dr. Clark received his PhD, MS, and BS all from Virginia Tech. [ac@transformationsystems.com]

Introduction

The Department of Navy (DoN) spends $7 billion annually on salaries and benefits for the acquisition workforce (AWF) that manages about $60 billion of expenditures across a Future Years Defense Program. To help develop smart buyers who understand that every person matters, every day matters, and every dollar matters, the DoN Director, Acquisition Career Management (DACM), has created paper-based and digital war rooms to promote ready, relevant lateral learning across programs and career fields. In aid of managing the acquisition workforce like a major acquisition program, DACM uses the war rooms to promote case study discussion of successful and failed programs; to enable Naval leaders to explore feedback loops and unintended consequences of workforce policy decisions; to develop approaches to better understand workload planning, affordability, and willingness to pay; to investigate organizational system problems in the context of larger systems; to develop a more comprehensive view of workforce health; and to develop cross-career field and cross-SYSCOM career navigators, career guides, and career paths.

The ASN (RD&A) Program Managers' War Rooms

The ASN (RDA) Program Managers Workshop was created in October 2014 to instruct the Navy's program managers and senior acquisition personnel in the history of Navy acquisition and examples of successful Navy program management. Since its creation, the workshop has instructed hundreds of program managers and senior acquisition personnel from over 70 different program offices. The workshop is an intensive five-day course of instruction wherein students are exposed to information relevant to their acquisition duties in the areas of U.S. Naval organization and history, U.S. military material procurement, major program management, and shipbuilding-specific roles and responsibilities. The course material is presented in seven paper-based War Rooms.

1. Evolution of the Navy War Room: Chronicles the evolution of our Navy over its history, with analyses at critical junctures in time.
2. Organization of the Navy War Room: Presents the organizational evolution of the Navy and how it is set up to operate and fight, mobilize, and maintain the Navy.
3. Material & Acquisition War Room: Views the evolution of the Navy's and the nation's material establishment and acquisition history since our founding. Students get an intimate feel for the events of the day as they actually occurred (in many cases different from the conventional history). Provides insights into what might be needed in the next 30 years.

4. Shipbuilding and Modernization War Room: Addresses the unique requirements and challenges of naval shipbuilding and naval systems development.
5. Program Management War Room: Examines how to meet the challenges of a major acquisition program through its life cycle, using the history, lessons, and tenets of three of the Navy's most successful acquisition programs in the modern era: POLARIS, AEGIS, and F/A-18. Also includes a short vignette for program managers on technology and program protection challenges due to cybersecurity threats, supply chain malfeasance, and increasing DoD program protection requirements.
6. Main War Room/CG(X) Case Study: Explores the national and international forces that shape Navy thinking. Provides a comprehensive view of the numerous dimensions and conditions in which a program manager must operate. Includes a postmortem analysis of the U.S. Navy's CG(X) Program.
7. Naval Aviation War Room: Ships are different, so we created a Naval Aviation war room to teach programmatic differences for aircraft that takeoff from and land on ships.

The instruction in each war room also emphasizes four strategic levels of effort:

- Deliver Lethal Capacity
- Increase Agility
- Drive Affordability
- Build a Workforce to Compete and Win

The War Rooms are replete with examples that illustrate the tension that exists between individual acquisition themes, and the trade-offs among them. This course has been characterized as being, "more about experience than academics." The instructors actively encourage the students to share their experiences and points of view in extended discussions. Daily wrap up and course critique sessions are conducted at the end of each day to focus on the lessons gleaned from the instruction, and to share additional lessons based on the experiences of the students in their own programs.

Development of Digital War Room Capability

The ASN (RD&A) COMPASS Room in Crystal City, VA, has an electronic display system called Mezzanine, which consists of 18 vertical large screen panels and three horizontal panels allowing users to immerse themselves in data with layered data displays and multimedia capability. The system is designed to be collaborative, allowing multiple users to connect with their devices via WiFi or HDMI cable to add content to a workspace or to make virtual "sticky notes" to content during a working session. At the end of a working session, the workspace data can be downloaded in an Adobe portable data file (PDF) for archiving and sharing electronically. The bottom line is that the COMPASS Room display system allows us to see and share information simultaneously, so it can be compared and analyzed side-by-side.

Figure 1. ASN (RD&A) COMPASS Room

- Each display panel can hold up to 50 images.
- Images can be resized and moved across multiple display surfaces.
- Video streams can be shared from your own devices wirelessly or via HDMI cables and shared across multiple display screens.
- Mark up workspace content from your smart phone or tablet.
- Save workspace content for later discussion.
- Upload graphics and PDFs to workspaces.
- Interact with any screen from anywhere in the room with the spatial wands.
- Focus attention using the wand as a pointer.
- Real-time interaction.

The Naval Aviation war room, developed in 2017, was designed with the COMPASS Room in mind rather than paper-based as the previous war rooms. DACM has a plan to convert all the existing paper-based war rooms into the digital format.

The Acquisition Workforce War Room

The DoN is strong at teaching technical excellence and capturing technical lessons learned, as evidenced by the traditional paper-based (and soon to be digital) acquisition program war rooms developed by ASN (RD&A) and used to train future program managers. The DoN has been less strong at systematic civilian leadership development and workforce management lessons learned, which drove ASN (RD&A) through DACM to create an Acquisition Workforce (AWF) war room to complement the program rooms. Design of the AWF war room coincided with installation of a digital display system in the COMPASS Room to replace paper-based presentations.

The content of the AWF war room includes an historical AWF timeline, beginning in 1794 when Joshua Humphreys was the only acquisition workforce member (now there are about 60,000!) responsible for acquiring the Navy's first six frigates. The timeline traces AWF evolution and lessons from Six Frigates through the Board of Navy Commissioners, the Bureaus, the First Expansion, the advent of Naval Aviation, the time Between the Wars 1920–1940, the Second Expansion, the emphasis on Research & Development, an AWF Transition, the Fall of the Berlin Wall, the 21st Century AWF, and into today with Naval Aviation—the F-35.

The AWF war room content operationalizes the Acquisition Career Council (ACC) charter elements into semi-standard templates for each career field to promote information-sharing and collaboration across career field national leaders and their action teams. DACM has leveraged the COMPASS Room capabilities to build career field and SYSCOM-specific workspaces detailing the linkages to other career fields and to the DoN AWF Strategic Plan objectives. DACM also identified leadership characteristics needed by every AWF member in all career fields: personal mastery, interpersonal mastery, organizational mastery, and motivational mastery. Career field national leads are using the AWF war room to develop career planning guides; measures of productivity, innovation, technical excellence, and professional excellence; prioritized implementation and action plans; and AWF models and measures—all with the aim of understanding and improving the long-term health of the AWF.

Visibility for Feedback Loops and Unintended Consequences

A separate ARS 2018 paper by Ford and Clark uses system dynamics models to characterize decisions made about the AWF, incorporating feedback loops and unintended consequences.

Visibility for Governance

A separate ARS 2018 paper by Clark and Rosa discusses the ASN (RD&A) governance structure that guides strategic shaping and development of the AWF through National Career Field Leaders.

Visibility for SYSCOM Manpower Planning Models

Warfighting system demand signals articulated in U.S. National Strategy documents, such as the DoN 30-year shipbuilding plan, generate a certain amount of acquisition work that must be accomplished to get capability into the hands of warfighters. The DoN has allocated various domains of acquisition work among several acquisition System Commands (SYSCOMs).

Each SYSCOM uses a tailored approach to translate warfighting system demand signals in their domain into a sequenced volume of work to be accomplished. Each SYSCOM's tailored approach estimates how many people (with appropriate knowledge and experience levels) are needed to perform that volume of work. DoN DACM has expressed a need for ASN(RD&A) to have a more integrated view of these SYSCOM manpower estimates to promote the ability to defend AWF size and conduct trade-offs when required.

Each of the major SYSCOMs has its own approach to translate demand signals into defendable workforce requirements. We do not suggest that all commands should adopt a single standard approach; however, we believe there are some criteria that each approach should adhere to in order to permit integration and decision-making at the enterprise level. We describe these criteria below. We recommend that each SYSCOM self-assess its unique approach against these criteria and if warranted, adopt a plan to address any opportunities for improvement they discover.

Manpower Data/Requirements Output Criteria

1. Explicit Assumptions: The manpower model/process uses operational and organizational assumptions fully revealed without vagueness, implication, or ambiguity.
2. Integrate-able Across the DoN: The manpower model/process produces requirements fully synchronized and integrated with requirements of other program or project resource providers external to SYSCOM.
3. Repeatable and Traceable: The manpower model/process is based on methodology that is not dependent on a single knowledgeable individual for execution; it can be repeated by different individuals and uses data that is traceable back to the original data source.
4. Longitudinal Data Using Common Data Elements: The manpower model/process uses common data elements that include category (CIV,

MIL, KTR), Career Field, total number of FTEs, pricing per FTE, career level, location of work, and capability being supported; using historical data across prior requirements cycles instead of estimates from prior POMs.

Workload Drivers Criteria

5. Defined Workload Drivers: The manpower model/process uses defined workload activity drivers and indicates the type of workforce and skills needed to complete tasks. Aspects to consider include each task's priority level; the required skill mix for task completion; demand patterns; whether there is a shortage in the skillset; and any uncertainty and variation associated with timing.
6. Aligned Priorities (CNO, DoD, WH, etc.): The manpower model/process–defined organizational priorities align to higher echelon priorities used in trade space analysis.
7. Flexible Process: The manpower model/process is flexible and can update mid-POM Cycle based on changing priorities without restarting the whole process.
8. Scalable Process: The manpower model/process demonstrates how FTE requirements adjust up and down based on change in priorities and impact on demand signal.
9. Accounts for Impact to Demand Signal: The manpower model/process has feedback loops linking capabilities demand signal requirements with workforce requirements.
10. Address Operations Impact: The manpower model/process addresses mission impact if less than 100% manning is funded and permits trade space decision analysis.
11. Based on Standard Work Packages: The manpower model/process uses defined and verified measures (work packages) of required FTE's and man-years to complete a micro-product (task below intermediate product).

Risk Management and Trade-Off Analysis Criteria

12. Addresses Risk: The manpower model/process defines and analyzes alternative manpower distributions with their operational and resource implications and the evaluations of various trade-off options.
13. Considers Mission Breaking Point: The manpower model/process identifies a manning level at which the program/project cannot meet mission requirements.
14. Minimum Sustainment: The manpower model/process expresses minimum FTEs to keep the unit running with sustainment operations only—identifies a manning level below which the program/project will not sustain the current level of operations.
15. Options and Sensitivity Analysis: The manpower model/process explains trade-offs and how one manning decision may be affected by changes in another manning decision.

Integrated Total Workforce Criteria

16. Funded/Unfunded Billets: Shows all billets and distinguishes between new and old, funded and unfunded.
17. Work breakout (current, future, navy after next): The manpower model/process can break out work by current Navy operations, future Next Navy operations, and Navy After Next operations.
18. Total Billet Count: The manpower model/process accounts for the total workforce (CIV, MIL, CTR) and can trace all billets, by type, to the appropriate end item and intermediate level products.
19. Funding Source: The manpower model/process can distinguish working capital funded billets from mission funded billets.

Leadership Engagement Criterion

20. Leadership investment, influence, and encouragement to individuals to use the systems.

Work accomplished by the DoN Acquisition Workforce (AWF) in the SYSCOMs enables industry partners in the value stream to construct and deliver warfighting systems. Delivering acquisition, modernization, and maintenance of warfighting capability to warfighters is the aim of this civilian-military-industrial enterprise. The outcome is readiness to fight and win.

Visibility for AWF Affordability and Willingness to Pay

Changes in the National Strategy over time have caused cyclical shrinkage and swelling among the ranks of the AWF. Perceptions of AWF affordability fluctuate among stakeholders across time. Robustly managing the AWF and telling a defendable story about its affordability demands the use of strategic thinking, systems thinking, industrial and organizational psychology, management science, engineering, and principles of major program management.

ASN (RD&A) has a responsibility (SECNAVINST 5300.38, 22 July 2009) to ensure Acquisition Workforce (AWF) capabilities and capacity requirements are balanced with workload. Meeting this responsibility occurs in the context of an ever-evolving national conversation about affordability and willingness to pay (the relative balance among risk, need, value, health, and cost). Articulating a more defendable story about AWF affordability is an ongoing aim of the DACM and the ASN (RD&A) directorate as they seek to manage the AWF as a major acquisition program.

Cogent questions include the following: At what point in time are we making an evaluation or a decision about affordability? What life cycle or time horizon are we considering as we make an evaluation or a decision about affordability? What past evaluations and decisions, made with what life cycle or time horizon in mind, caused the current state of perceived affordability? What is our organizational learning approach to document, manage, and learn from the knowledge, assumptions, constraints, and history that led to perceived affordability in the past, now, and in the future?

Some basic economic principles also apply to our discussion of affordability:

- sunk cost (abandoning a previous investment strategy when new information emerges);
- opportunity cost (investments we don’t make because we are fully committed to other investments);

- design cost (80%–90% of life-cycle cost locks in during initial design assumptions);
- do-nothing cost ("kicking the ball down the road" or not investing in favor of awaiting more information is a decision with implications and perhaps unintended consequences);
- irrecoverable cost (a point at which we have missed the opportunity and no amount of money spent can recover what we lost).

AWF affordability is an inextricably interwoven sub-factor of warfighting capability affordability. Using a similar framework, we define terminology for both. These definitions are offered to provoke critical thinking; they are not given as definitively correct.

Affordability (in the case of both warfighting capability affordability and AWF affordability) is defined simply as "congressional willingness to pay." If we as a nation through our elected representatives collectively decide that we are willing to incur the costs of any National Defense Strategy, then that strategy is acceptably affordable.

Warfighting Capability Willingness to Pay is a function of

- perceived ***risk*** levels based on current and future credible threats, prompting a need for national defense.
- the ***need*** for specific numbers of national defense systems, now and in the future.
- the ***value*** placed on national defense, perceptions of which change over time.
- assessed ***health*** of national defense systems, current and predicted future.
- the ***cost*** of national defense, in the current budget, future budgets, and for a 50-year life cycle.

AWF Willingness to Pay is a function of

- perceived ***risk*** levels we incur based on the current and future composition and expertise of the AWF.
- the ***need*** for specific numbers of AWF members, now and over time.
- the ***value*** placed on AWF work getting done, perceptions of which change over time.
- assessed ***health*** of the AWF, current and predicted future.
- the ***cost*** of the AWF, in the current budget and future budgets, and for a 50-year life cycle.

Warfighting Capability Risk is a function of

- the perceived threat levels, the slope of the threat pace curve, and the gap between U.S. capability and threat capability. To continue with North Korea as the example, we thought they were on one trajectory with ballistic nuclear missile capability and it appears now they may be on a steeper curve than we thought. Our willingness to pay may change with that change in calculus.
- our perceived vulnerability, which is in some ways tied to threat pace, threat slope, and threat gap. If we didn't feel vulnerable based on what North Korea was doing, we would behave differently. If we feel threatened, we perceive greater risk and therefore greater willingness to pay.
- whether we have declared war or not. A formal declaration of war is an obvious expression of the degree of nationalu risk we feel.

AWF Risk is a function of

- the likelihood that unfunded acquisition work will delay or diminish warfighting capability.
- the impact of delayed or diminished warfighting capability on readiness to fight and win.
- the likelihood of a critical expertise diminishing beyond the point of no return.
- the impact on future warfighting capability of losing that expertise.

Warfighting Capability Need is a function of

- stated requirements for warfighting capability; the number of ships, aircraft, and vehicles with associated equipment, combat systems, weapons, and ordnance needed to support the DoN's military missions now and in the future.

AWF Need is a function of

- the quantity of AWF work necessary to meet warfighting capability production requirements.
- the AWF manpower needed to perform that quantity of work.
- AWF productivity.

Warfighting Capability Value is a function of

- prevailing risk tolerance levels.
- the political environment.
- decision-maker connection to warfighting. For example, the number of congresspersons who have been in the service and have experienced what it means to be a warfighter influences how Congress deliberates about the value of national defense.
- the degree of decision-making centralization or decentralization that exists in an administration or in Congress.

AWF Value is a function of

- current readiness and capability delivered by previous AWF.
- future readiness and capability achieved by AWF investment today.
- important work being performed today.

Warfighting Capability Health is a function of

- the assessed health of the Naval Enterprise.
- the health of the acquisition workforce.
- the health of the tools.
- the health of the facilities.
- the health of policies.

AWF Health is a function of

- the assessed current health and projected future health of AWF, including dimensions such as capability, capacity, diversity, experience, certifications, and training levels.

Warfighting Capability Cost is a function of

- workforce costs.
- facility costs.
- tool costs.
- policy costs.
- acquired system costs.
- contracting strategies.

AWF Cost is a function of

- AWF size and the associated salary, benefit, and retirement costs.
- training and development costs.

Expressed as conceptual functions,

$\underline{Warfighting\ Capability\ Affordability\ \equiv willingness\ to\ pay}$

and willingness to pay $= f(risk, need, value, health, cost)$

$where$

$risk = f(perceived\ threat\ levels, threat\ pace\ slope, threat\ gap, perceived\ vulnerability$
$war\ declaration\ state, \dots)$

$need = f(stated\ warfighting\ requirements, \dots)$

$value = f(risk\ tolerance, political\ environment, decisionmaker\ connection$
$to\ warfighting, degree\ of\ decisionmaking\ decentralization, \dots)$

$health = f\ (assessed\ current\ health\ and\ projected\ future\ health\ of\ AWF,$
$industrial\ base, sailor\ training\ systems, energy\ supply, etc\ \dots)$

$cost = f\ (workforce\ cost, facility\ costs, tool\ costs, policy\ costs,$
$acquired\ system\ costs, contracting\ strategies, opportunity\ costs, \dots)$

$\underline{AWF\ Affordability\ \equiv willingness\ to\ pay}$

and willingness to pay $= f(risk, need, value, health, cost)$

$where$

$risk = f(likelihood\ that\ unfunded\ acquisition\ work\ will\ delay\ or\ diminish$
$warfighting\ capability, impact\ of\ delayed\ or\ diminished\ warfighting\ capability$
$on\ readiness\ to\ fight\ and\ win, likelihood\ of\ a\ critical\ expertise\ diminishing$
$beyond\ the\ point\ of\ no\ return, impact\ of\ losing\ that\ expertise\ \dots)$

$need = f(quantity\ of\ AWF\ work\ necessary\ to\ meet\ warfighting\ requirements,$
$AWF\ manpower\ needed\ to\ perform\ that\ quantity\ of\ work,$
$AWF\ productivity, \dots)$

$value = f(current\ readiness\ and\ capability\ delivered\ by\ previous\ AWF,$

$$future\ readiness\ and\ capability\ achieved\ by\ AWF\ investment\ today,$$
$$important\ work\ being\ performed\ today\ \dots)$$
$$health = f\ (assessed\ current\ health\ and\ projected\ future\ health\ of\ AWF, \dots)$$
$$cost = f\ (workforce\ size, salary\ and\ benefit\ levels, retirement\ obligations,$$
$$training\ and\ development\ cost, opportunity\ costs, \dots)$$

There are multiple levels of organizational system to which we can apply conceptual thinking about willingness to pay = f (risk, need, value, health, cost).

Visibility for Different Organizational Systems of Interest

Each level of indenture below implies a smaller subset of the national defense domain than the previous level. Zooming in and out from one level to another elicits different views of willingness to pay = f (risk, need, value, health, cost).

I. Joint Staff and the Office of the Secretary of Defense: Joint warfighting capability, readiness and sustainment levels across all services combined, including soldiers, airmen, sailors, and Marines.

  A. The Department of the Navy, the Chief of Naval Operations, and the Navy Secretariat: Naval warfighting capability, readiness and sustainment levels across Navy and Marine Corps combined, including sailors, aviators, and Marines.

    1. Navy warfighting capability, readiness, and sustainment levels, including sailors and aviators.

    2. Marine Corps warfighting capability, readiness and sustainment levels, including Marines.

      a) Assistant Secretary of the Navy, Research, Development, and Acquisition: Research, development, transition, acquisition, sustainment and modernization of warfighting systems (not warfighters).

      b) National Career Field Leaders. Shepherds of the current effectiveness and long term health of the entire acquisition workforce across Program Executive Offices (PEOs) and SYSCOMs.

      c) Chief of Naval Operations: Effectiveness and safety of fleets, type commanders, sailors, and aviators.

      d) Commandant of the Marine Corps: Effectiveness and safety of Marines.

      e) Program Executive Officers and SYSCOM Commanders: Jointly responsible for the set of AWF members under their command and required to use manpower planning models/approaches to convert warfighting demand signals into expected workload for AWF members and therefore project the required size of AWF Career Fields and projected demand signal on each core equity in Naval Warfare Centers.

      f) Naval Warfare Center Commanders: Responsible to maintain minimum threshold capability for each of their assigned core equities so that the Navy maintains the ability to surge and scale any technology expertise as required.

Leaders and members at lower levels in this hierarchy have different definitions and perceptions of the elements comprising willingness to pay than those at higher levels, and the objective functions and constraints perceived up and down levels may vary greatly.

For example,

- A SYSCOM may have a natural tendency toward an objective function that maximizes the performance of their organization rather than the naval enterprise. Every SYSCOM fights for their mission and warfighting capabilities, maximum fill of their manpower needs, as large an increase in systems and hardware as they can get, and maximum funding for their perceived priorities. SYSCOMs look through a different lens so they have so different objective functions and constraints than the Naval triad. What maximizing performance looks like from a SYSCOM chair differs from an OPNAV chair.
- For a program or person responsible for a mission area, the more they can do in that mission area the more effective they are perceived to be. Maximizing that mission area, however, may not make sense when you look across all the mission portfolios having limited resources that you can invest.
- National career field leaders are charged with improving career field productivity, innovation, and professional and technical excellence. It is at least conceivable that attempts to optimize the performance of one career field will negatively impact the performance of another, and the unintended consequence may be invisible to both national leaders.
- A similar line of thinking applies around service components depending geographically where the conflict is believed to occur, and the type of units required to confront it. We have spent the last 15 years in the desert and the Navy has been the bill recipient or payer for people on land. If North Korea kicks off, it is more a maritime role you are going to see that comes up. So, there is a trade space between Army, Navy, and Air Force, and the next level down you are going to have what level assets and platforms you are going to need to impact trade space options.
- A similar situation applies to Warfare Centers, which have very challenging objective functions. They strive to maintain minimum threshold core equity capability without dedicated mission funding, that is, market and sell core equity products and services to program managers in the hopes there will be enough buyers to maintain the core equity. WFC objective functions and constraints leave much room for well-intended suboptimization.

This line of reasoning leads strongly to a recommendation that future ACC meetings and AWF Summits be conducted in the AWF war room so that Navy leaders may have these trade-off discussions surrounded by more data from the larger system with appropriate context. The model makes for a good acquisition workforce summit discussion about how complex AWF affordability is and how one size does not fit all and it is about having the conversation to understand those impacts.

Visibility for Strategic Options When Exploring the Trade Space

Recall that the aim of developing and exploring this conceptual model is to architect and evaluate possible options for action and for future states and select a preferred way ahead. The options approach considers multiple possible decision pathways in an uncertain environment and allows for making mid-course corrections when new information emerges. Traditional decision models assume a single static decision, while real option analysis assumes a multidimensional series of options where leaders have the flexibility to adapt given feedback loop impacts from previous decisions or a new change in the enterprise ecosystem. If we are in the face of investment or cost decisions, there are at least three sets of options below that we could pursue.

People Levers. Here are some people levers that we as leaders could turn or crank.

1. DAWIA certifications
2. Professional certifications
3. Retention allowance
4. SLRPs
5. Training
6. Hiring
7. Education
8. Government rotations
9. Industry rotations
10. Transition & retrain
11. Partner with other government/industry
12. Job swaps
13. LDPs
14. Succession planning
15. Mentoring
16. Recognition
17. Rewards Shaping through early outs: VSIP/VERA

Workload Levers. Here are some things we could do with workload.

1. Better tools
2. Accelerated acquisition
3. Change ACAT level
4. Reduce documentation requirements
5. Better Buying Power guidance
6. Commonality
7. Redundancy

Tool Levers. Here are some tool levers that we could invest in.

1. Talent management
2. DCPDS
3. eDACM 2.0
4. DAWIA Operating Guide revision
5. AWTAP

6. DAWDF
7. ACQ Demo
8. AWQI
9. Career navigators
10. USC Title X Chapter 87 1701-18xx

Visibility for Career Navigators, Career Guides, and Career Paths

Equipping AWF members and leaders with useful tools to envision several possible alternative career progressions provides a tremendous benefit to the individual and to the Navy. DACM has developed a Career Navigator tool that for that purpose. Career Navigator provides at-a-glance guidance regarding eight key dimensions of a career (see Table 1) across time, spanning from Entry and Journeyman to the Expert and Senior Leader phases.

Table 1. Career Navigator Dimensions

Career Navigator Dimensions	
1. Life Events	Chart the major changes in status or circumstances (e.g., marriage, divorce, death of a spouse) and understand how these changes may affect your career.
2. Results & Awards	Document major professional accomplishments and honors to see if you are tracking toward your goals.
3. Experience/Roles	Visualize how you have grown professionally and if you have been/are/will be in the right roles at the right time.
• RDA Level & Above	Plan your long-term goals.
• Broadening	See how you can expand outside of your main skills domain.
• Within Domain/SYSCOM	Chart your course within your domain/SYSCOM
• Military or Industry	Both military and industry experience can play key roles in career development.
4. Mentoring	Outline your mentor/mentee experience and intentions
5. Certifications	Plan and track certifications required to keep you on track to your career goals
6. Training	Plan and track training that will keep you at the fore of your career field.
• Technical	Manage training for technical aspects of your career
• Professional	Plot training such as: management, team building, soft skills, communications.
• Personal Development	Track training designed to improve personal development such as interpersonal skills, organizational skills, and motivational skills.
7. Education	Track how degrees from DAU and other institutions of higher learning can augment your career success.
8. Character	Document milestones in your career that built and attest to your character.

The Career Navigator tool can help guide people in journaling their respective career paths; envisioning positions they might want to hold in the future; developing actionable plans to become competitive for those positions; and tracking progress. It is a planning tool for career steering regardless of career field, current position or years of experience. Anyone can use it, both inside and outside the AWF, since the principles upon which it stands are universally applicable. Additionally, each national career field leader is chartered to provide career path guidance within their career field.

Visibility for Measures of AWF Health

There are thousands of measures (metrics) that have been used or can be used to assess an organization's workforce "health." Industry and government leaders often find it challenging to select the best portfolio of workforce measures that provide the information that they need to make good data-based strategic and operational workforce management decisions without overburdening the organization with research, data mining, and reporting requirements. There is no magic list of the best workforce health measures; however, there are frameworks that can help.

The Human Capital Management Cycle Framework

The Human Capital Management Cycle Framework defines a healthy workforce as one that is productive, innovative, and excellent as a result of the program/organization/career field successfully executing the critical human capital management processes associated with each phase of the human capital management cycle and achieving the end goals associated with each phase (see Figure 2). “Employing Measures in Managing Acquisition Workforce Health” begins by defining the phases of the human capital management cycle, the specified end goal statements for each phase, and key human capital management processes that are performed in each phase. The end goal statements are a starting point for program/organization/career field leader discussions; leaders may choose to tailor the current statements or adopt new end goals to best fit their organizations if desired. In the Human Capital Management Cycle Framework, measures of the quality and timely execution of the key human capital management processes (or progress status of processes that are being developed but have not been fully implemented) associated with each phase are predictive of the likelihood that the phase end goal will be achieved. The measures associated with the end goals taken at the end of the established goal time-period are the resulting lagging measures.

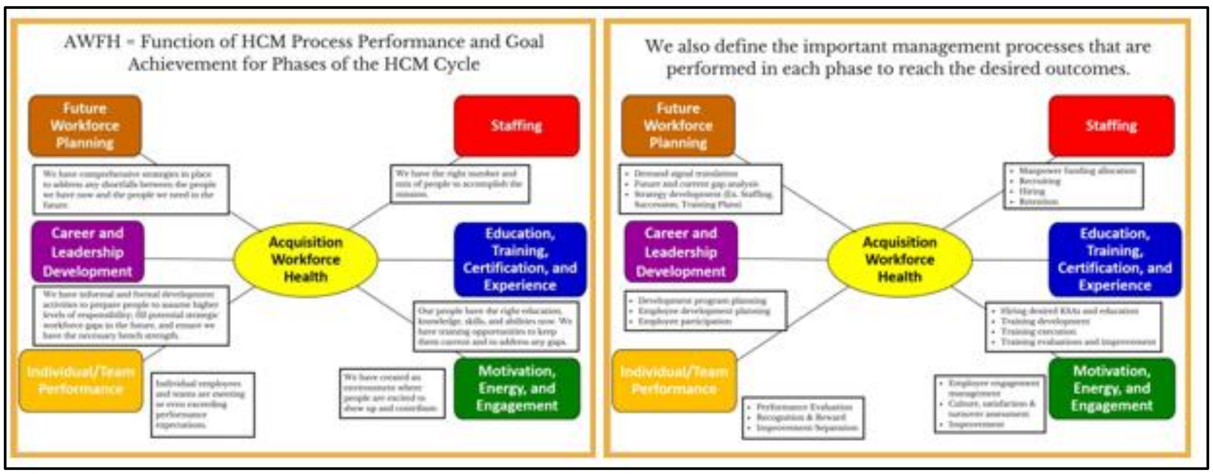


Figure 2. Human Capital Management Cycle Framework

The Enterprise Performance Measurement Framework

Leaders typically limit assessment of their organization’s workforce health to evaluating how well the program/organization/career field performed internal human capital processes to achieve internal human capital goals; the Human Capital Management Cycle Framework. The Human Capital Management Cycle Framework alone is an adequate approach; however, a measurement system based on the Enterprise Performance Measurement Framework accounts for various factors outside of internal human capital process performance that may influence an organization’s human capital goal achievement and workforce health. The Enterprise Performance Measurement Framework also attempts to provide assurance that the workforce is delivering desired results to support the organization’s external customers/stakeholders and the enterprise mission; the driving purpose for the workforce and an organization to exist. Assessing workforce health from the larger Enterprise Performance view instead of from the narrow internal perspective of the Human Capital Management Cycle Framework only results in leaders developing internal human capital goals and metrics that are aligned with both higher level plans/goals and customer/stakeholder needs. In addition, the enterprise perspective helps leaders focus the organization's attention and energy on changing internal processes and external factors that are within their span of control to improve organization workforce health. An enterprise view also provides leaders with a measure of assurance to help determine whether current and future levels of workforce capacity and capability are adequate to meet customer product and service requirements. Finally, an Enterprise Performance Model–based metrics portfolio enables individual organizations to better articulate how their human capital process performance, trade space decisions, and ultimately their workforce's performance, innovation, technical excellence, and professional excellence contributed to overarching enterprise mission accomplishment.

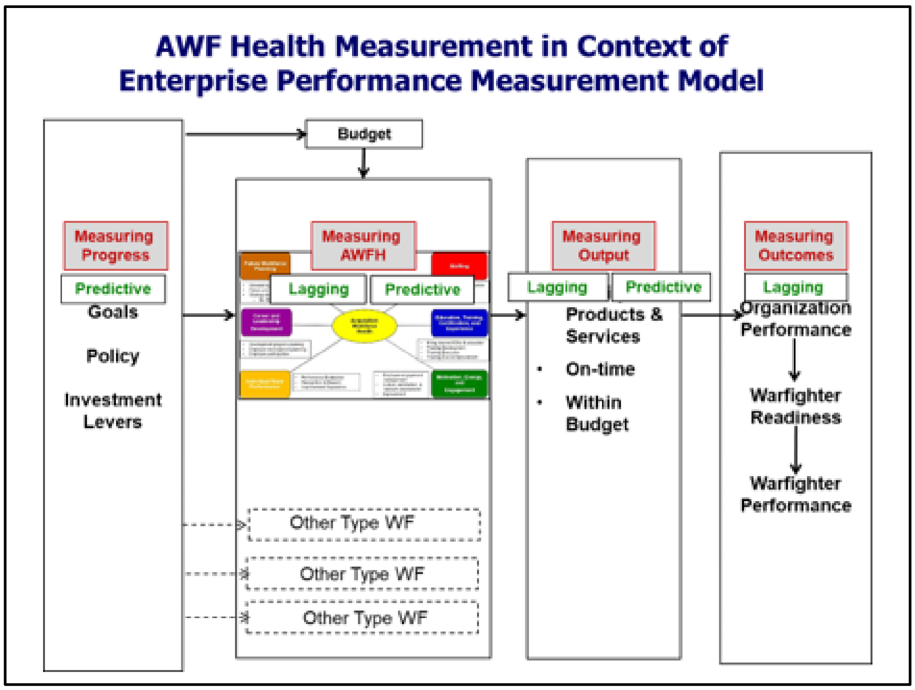


Figure 3. Enterprise Performance Measurement Model

The DoN Acquisition Workforce Health Measure Framework (in Context of National Strategy and Warfighting Performance)

Finally, the presentation concludes with the description of a proposed Department of the Navy Acquisition Workforce Health Measurement framework based on the Enterprise Performance Measurement Framework discussed in the previous section. The framework depicts "real world" elements that impact the management of Navy Acquisition organizations and the Navy Acquisition workforce, actual products and services that acquisition organizations produce and deliver, and a representation of the various customers that use acquisition products and services to ultimately ensure warfighter readiness and support warfighters on the battlefield, in the air, and at sea. The framework begins with the U.S. national strategy as the linchpin that informs DoD and DoN strategic direction, plans, goals, and funding allocation decisions. Strategic-level acquisition organizations develop policy, budget plans, execution plans, and various initiatives to guide lower-level organizations in accomplishing their mission to help the Navy enterprise achieve DoN, DoD, and ultimately

U.S. strategic objectives. The framework acknowledges that most organizations that perform acquisition work for the Navy not only have to manage the challenges associated with an acquisition professional workforce, but also other related professional workforces (research and development, maintenance, support, and others) that may have similar or different goals, objectives, and needs. The Navy acquisition organization's workforce produces specific products and provides specific services that contribute to the successful development, purchase, deployment, and sustainment of ships, submarines, vehicles, weapons, information technology, and other equipment used to ensure the U.S. Navy's readiness to fight. These resources are employed by combatant commanders, sailors, marines, soldiers and airmen to win the fight.

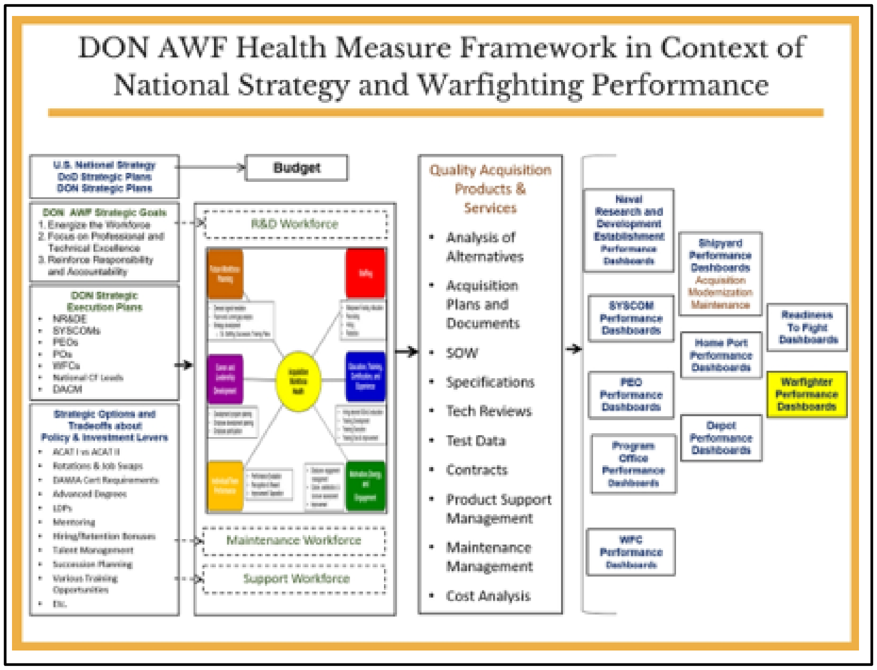


Figure 4. DoN AWF Health Measurement Framework

As depicted in the Enterprise Performance Measurement and associated DoN Acquisition Workforce Health Measurement frameworks, the same workforce health measures may be both predictive and lagging depending on the context in which they are used. A measures portfolio based on these frameworks assumes the following:

1. Measurement results of an organization's progress in implementing external human capital plans, policies, and initiatives levied by a higher headquarters or outside agencies (e.g., OPM, DoD, DoN, Congress, etc.) and the amount of funding allocated are predictive of the organization's level of internal human capital cycle process performance and end goal achievement;
2. Measurement results of an organization's internal human capital cycle process performance and end goal achievement are lagging measures of an organization's progress in implementing external human capital plans, policies, and initiatives levied by a higher headquarters or outside agencies, yet at the same predictive of the organization's capability to consistently deliver quality products/services within the time required by customers/stakeholders and within budget

3. Measurement results of an organization's product/service quality, timeliness, and cost are lagging results of human capital cycle end goal achievement, but also a predictive measure of the organization's external customer/stakeholder's performance satisfaction and success and enterprise mission accomplishment.

A lagging measure of total AWF population can be predictive of the organization's likelihood to achieve adequate levels of personnel in the Critical Acquisition Position (CAP) workforce. In turn, total CAP population measures may allow leaders to predict if the Key Leadership Position (KLP) workforce will be adequately staffed to meet customer demands. The total population of each category of workforce is driven by hiring and retention. Therefore, measures of the success of hiring processes, staff addition rates, departure rates, promotion rates, and retention rates would all be examples of predictive measures used to assess the organization's progress toward meeting desired total population goals in each category. In the context of the Enterprise Performance Measurement and associated DoN Acquisition Workforce Health Measurement frameworks, measures of an organization's implementation of strategic initiatives aimed at improving hiring efficiency and increasing retention would be predictive of the likelihood that the organization would achieve total population goals in each category. The lagging total population results in each category would then be predictive of the likelihood that the acquisition products delivered by the organization would meet customer quality and timeliness expectations. The lagging product quality and timeliness measures (e.g., voice of the customer survey feedback, frequency of meeting product quality and delivery timelines specified by law, policy, or customer demand, etc.) then become predictive measures used to proactively assess the organization's impact on overarching mission and warfighter success.

Closure

The DoN Director, Acquisition Career Management (DACM), has created paper-based and digital war rooms to promote ready, relevant lateral learning across programs and career fields so that SYSCOMs, PEOs, national career field leaders, and their teams can move smartly and systematically toward managing the 60,000-member acquisition workforce like a major acquisition program.

For Further Study

DACM is rapidly exploring innovative ways to promote

- More emphasis on defining and measuring Program and AWF outputs and outcomes
- Judicious application of system dynamics models to appropriate problem sets in all career fields
- Continued exploration of workload forecasting models tied to changing demand signals from the 30-year shipbuilding plan
- Development of talent management systems and toolsets to shape and manage AWF composition
- Transition to Virtual War Rooms with shared data display across multiple geographies
- Understanding the half-life of knowledge and the refresh rate required to maintain currency
- Defining the characteristics of a fully developed professional in all career fields

- Better defining the pool of candidates available for succession planning purposes
- Better understanding the qualities needed in key people, beyond technical training.

Panel 3. Factors Affecting Reported Costs in DoD Acquisition Programs

Wednesday, May 9, 2018	
9:30 a.m. – 11:00 a.m.	**Chair: Richard P. Burke,** Deputy Director for Cost Assessment, Cost Assessment and Program Evaluation ***Further Evidence on Program Duration and Unit Cost Growth*** David McNicol, Institute for Defense Analyses ***Comparing Ship Versus Aircraft Development Costs*** Larrie D. Ferreiro, Defense Acquisition University ***Complexity in an Unexpected Place: Quantities in Selected Acquisition Reports*** Gregory A. Davis, Institute for Defense Analyses Margaret L. Giles, Institute for Defense Analyses David M. Tate, Institute for Defense Analyses

Richard P. Burke—is the Deputy Director for Cost Assessment in the Office of the Secretary of Defense, Cost Assessment Program Evaluation (CAPE). Prior to that, he served in the Office of the Secretary of Defense (PA&E) as the Deputy Director, Resource Analysis; Director, Operations Analysis and Procurement Planning Division; and Operations Research Analyst, Strategic Defensive, and Space Systems Division. Dr. Burke was an International Affairs Fellow at the Council on Foreign Relations, as well as a Visiting Scholar at Stanford University's Center for International Security and Arms Control. He received a PhD, an MS, and a BS in nuclear engineering, all from MIT.

Further Evidence on Program Duration and Unit Cost Growth

David McNicol—joined the Department of Defense (DoD) in 1982. From 1988 until 2002, he was a Deputy Director of Program Analysis and Evaluation (PA&E). Earlier, Dr. McNicol taught at the University of Pennsylvania and the California Institute of Technology. He holds a BA in Economics from Harvard and an MS in Management and PhD in Economics and Finance from MIT. Employed at the Institute for Defense Analyses (IDA) since his retirement from the DoD, he became Director of the Cost Analysis and Research Division in 2006. Still at IDA, Dr. McNicol stepped down in 2012 to return to his previous role as a Research Staff Member. [dmcnicol@ida.org]

Abstract

David McNicol, in "Post-Milestone B Funding Climate and Cost Growth in Major Defense Acquisition Programs," in *Proceedings of the 14th Annual Acquisition Research Symposium, Vol. 1*, explored the association between Program Acquisition Unit Cost (PAUC) growth of Major Defense Acquisition Programs (MDAPs) and funding climates post-Milestone (MS) B. A strong positive association was found for MDAPs that passed MS B in a bust funding climate; the association was weak for programs that passed MS B in boom climates. This paper uses four alternative regression equations to extend these results. In each case, the same pattern of results appears—MDAPs that passed MS B in a bust climate had significantly higher growth than those that passed MS B in a boom climate, the 1969 Packard reforms reduced average PAUC growth, and the reduction persisted through the end of the study (FY 2009)—but changes to the acquisition process after the Packard reforms through 2009 did not further reduce average PAUC growth. The lower PAUC growth after the Packard reforms probably was due mainly to more realistic MS B baselines. This pattern does not depend on the inclusion of post-MS B funding climate and program duration, although those factors have significant effects.

Introduction[1]

McNicol (2017b) explored the association between growth in the unit costs of Major Defense Acquisition Programs (MDAPs) and the funding climates programs experienced after passing Milestone (MS) B.[2] While this topic arose by serendipity, a little reflection establishes that it is plausible to expect average cost growth to be higher for MDAPs that entered a boom climate sometime after passing MS B than it is for those that did not. MDAPs that passed MS B in bust climates probably are especially influenced by a post-MS

---

[1] This paper draws on Chapters 2 and, especially, 3 of McNicol (2018), which is a synthesis of a series of papers on the association of funding climate, acquisition policy, and other factors on cost growth of major acquisition programs.

[2] While the label MS B is used here for all time periods, through the years there have been changes in the labels used for milestones and, to some extent, in their definitions. During FY 1966–1969, there were two milestones in the OSD-level acquisition process, neither of which had a name. Reforms instituted early in FY 1970 provided for three milestones, labeled MS I, MS II, and MS III. In 1987, MS IV and MS V were added. By 1991, MS IV had been eliminated and what had been MS V became MS IV. MS IV had been eliminated by 1996. In 2000, the milestones were changed to MS A, MS B, and MS C, and the definition of MS B modestly changed. See McNicol (2018), Chapter 5.

B boom. Some of these programs presumably had unrealistic baselines and would find a post-MS B boom climate a good time to "get well." Even programs that established realistic baselines at MS B might tend to be less capable than the service wanted and good candidates for adding capability when funding constraints were relaxed.

A strong positive association was found between MDAPs that passed MS B in a bust funding climate and subsequently entered a boom funding climate. The association was much weaker for programs that passed MS B in boom climates. This paper uses a series of four models to extend those results by incorporating acquisition policy variables and program duration. The section titled Cost Growth Due to Program Changes also briefly examines the extent to which PAUC growth post-MS B reflects costs due to subsequent decisions to acquire capabilities beyond those of the MS B baseline.

The Models

The first of the models (Model 1) relates cost growth only to funding climate and acquisition policy configuration.[3] Next (Model 2), post-MS B funding climate is introduced into the model. Two models (Models 3 and 4) that in different ways include both post-MS B funding climate and program duration are then presented. We begin by pointing to important features common to the four models.

Framework of the Models

The topic requires distinguishing between bust and boom funding climates. During the 45 years (fiscal year [FY] 1965–FY 2009) covered by this study, there were two complete bust-boom cycles in Department of Defense (DoD) procurement funding: (1) the bust climate for modernization of weapon systems that began in the mid-1960s and lasted until the Carter-Reagan buildup of the early to mid-1980s, and (2) the long post-Cold War bust climate followed by the post-9/11 boom.

Where a bust funding climate may provide an upward pull on cost growth, acquisition policy and process can be expected to provide a restraining push. For that reason, it is necessary also to recognize changes over time in acquisition policy and process configurations. Five policy and process configurations are distinguished:

1. McNamara-Clifford (FY 1964–FY 1969)
2. Defense Systems Acquisition Review Council (DSARC; FY 1970–FY 1982)
3. Post-Carlucci DSARC (P-C DSARC; FY 1983–FY 1989)
4. Defense Acquisition Board (DAB; FY 1990–FY 1993 and FY 2001–FY 2009)
5. Acquisition Reform (AR; FY 1994–FY 2000)

Policy and *process* tend to be intertwined; process typically is required to implement policy, and the most successful and durable policies are those embedded in process. For this reason, and to avoid constant repetition of "process and policy," the term *acquisition policy* is used here in a broad sense to encompass both policy on particular topics (for

[3] The most developed explanations of funding climate and acquisition policy configuration are provided in McNicol (2018), Chapter 1.

example, contract types) and the Office of the Secretary of Defense (OSD)-level oversight process (for example, definition of the milestones).

Finally, a measure of cost growth is required. The measure used is based on Program Acquisition Unit Cost (PAUC). PAUC is the sum of Research, Development, Test and Evaluation cost and procurement cost, divided by the number of units acquired. For this paper, PAUC growth is computed by comparing the MS B baseline value of PAUC in program base-year dollars—which can be thought of as a goal or a prediction—to the actual PAUC reported in the program's last Selected Acquisition Report (SAR) in program base-year dollars and adjusted to the MS B baseline quantity. Appendix B of McNicol (2017a) describes the conventions used in assembling the database, the sources of the data used, and the quantity adjustment computations. The unit cost growth estimates were updated to the December 2015 SARs. Only completed programs (defined as programs with an end date of FY 2016 or earlier) are used in this analysis because some costs associated with a program may not be fully reflected in its SAR until the program is completed.

To be clear, in what follows, the term *PAUC growth* means PAUC growth as defined previously, that is, growth from MS B through the end of procurement, adjusted to the MS B quantity.

Model 1—The Baseline Model

The "baseline model" is the following assumed relationship:

$$PAUC_i = a_0 + a_1Climate_i + a_2DSARC_i + a_3P\text{-}CDSARC_i + a_4DAB_i + a_5AR_i + e_i \qquad (1)$$

The subscript i denotes the ith MDAP in the sample. This model provides a baseline in that it includes as independent variables only funding climate and acquisition policy configuration.

Climate is a categorical variable[4]; it takes on a value of zero for MDAPs that passed MS B in bust climates and 1 for those that passed in boom climates. The intercept term a_0 is assumed to measure primarily the climate effect.[5] For programs that passed MS B in a bust climate, a_0 is the intercept; for those that passed in a boom climate, the intercept is $a_0 + a_1$. The expectation is that the estimate of a_1 is negative; that is, that MDAPs that passed MS B in a boom climate on average have lower PAUC growth.

The model includes a categorical variable for each of the four acquisition policy configurations. These variables have a value of 1 for the years of the period in question (e.g., FY 1994–FY 2000 for AR), and zero for other years. For technical reasons, one of a set of categorical variables always must be omitted (or the constant term constrained to zero). The selection of the omitted variable is arbitrary insofar as the statistics are concerned; the McNamara-Clifford period was chosen because that is convenient for the exposition. The estimated coefficient of each of the acquisition policy categorical variables is the difference between average PAUC growth in that bin and average PAUC growth in

[4] These are often referred to as "dummy variables" but are more descriptively called categorical variables or indicator variables.

[5] The estimated coefficient of Climate also includes the average net effect of any relevant variables not included in this model and the effect on the estimated intercept of any non-linearity in the response of PAUC growth to the model's explanatory variables.

McNamara-Clifford. That difference is statistically significant if the estimated coefficient of the acquisition policy period categorical variable is statistically significant.[6]

Finally, the error term e_i represents myriad unpredictable factors that influence PAUC growth; it is assumed to be a normally distributed random variable with a mean of zero and constant variance. The coefficients of the model are estimated using ordinary least squares (OLS; also known as multiple regression, linear regression, and least squares regression).[7] The results are presented in Table 1. We use the p-value to characterize statistical significance and refer to any estimate with a p-value of no more than 0.10 as "statistically significant." A p-value of 0.10 means that there is an (estimated) one chance in 10 that the observed estimate would occur by chance even if the true value of the coefficient were zero.

---

[6] Note that for all of the observations of the McNamara-Clifford period, $PAUCj = a_0 + e_j$, and since it is assumed that $E(e_j) = 0$, $E(PAUC_j) = a_0$. If the underlying model is correct and the assumptions of OLS are satisfied, the estimated value of the intercept (denoted $\hat{a}_0$))" is an unbiased estimate of a_0 and of the sample value of the average PAUC growth of the McNamara-Clifford period. Similarly, the expected value of the intercept and the average PAUC growth for the ith acquisition policy bin is "$a_0 + a_i$," and the difference between that and the average for the reference group is $a_0 - (a_0 + a_i) = -a_i$. Hence, if $\hat{a}_i$ is statistically significantly different from zero, the average PAUC growth for acquisition policy configuration i is significantly different from average PAUC growth for McNamara-Clifford. The burden of the assumptions is lightened by the fact that, in this context, "just about" counts. For example, no great harm is done if $E(e_j)$ is small rather than zero.

[7] Readers unfamiliar with this technique can find an explanation in any introductory econometrics text, in many introductory statistics texts, or on the internet. For example, *TIBC Statistica*, http://www.statsoft.com/Textbook/Multiple-Regression; Penn State Eberly College of Science, STAT 501: Regression Methods, https://onlinecourses.science.psu.edu/stat501/node/283; John H. McDonald, *Handbook of Biological Statistics*, http://www.biostathandbook.com/multipleregression.html; and David M. Lane, "Introduction to Multiple Regression," Chapter 14, in *Online Statistics Education: An Interactive Multimedia Course of Study*, http://onlinestatbook.com/2/regression/multiple_regression.html.

Table 1. Estimated Parameters of the Basic Model of PAUC Growth

	Coefficient	p-value
Intercept	86.7%***	< 0.001
Funding Climate		
Climate	-37.3%***	< 0.001
Acquisition Policy		
DSARC	-47.1%***	< 0.001
P-C DSARC	-42.3%***	0.003
DAB	-47.0%***	< 0.001
AR	-65.4%***	< 0.001

* Statistically significant at less than the 10% level.

** Statistically significant at less than the 5% level.

*** Statistically significant at less than the 1% level.

R-Square = 0.25 F = 9.720 (P < 0.001) N= 149. Estimated by OLS. With the Bonferroni correction, Wald's test for the equality of the estimated coefficients the categorical variables for acquisition policy periods yields F = 1.02, p = 0.999.

The dataset used to estimate the model in Table 1 omits four extremely long duration programs. Each of these "programs" is actually a series of modifications and upgrades of an initial program reported on the SAR of the original program. Also excluded are three programs from the early 1980s boom period that were acquired using variants of Total Package Procurement (TPP). These observations were excluded for reasons stated below.

The criteria typically used to judge regression equations readily accept the results in Table 1:

- The intercept and the estimated coefficient of each of the independent variables have the expected signs.
- Their magnitudes are reasonable.[8]
- The intercept and the estimated coefficients of the independent variables are highly significant.
- The estimated equation as a whole is highly significant.
- The proportion of the variation in sample PAUC growth captured by the estimated equation is towards the upper end of what can be expected for panel data.

[8] Evaluations of the reasonableness of the estimated coefficients of the acquisition policy periods must weigh the Climate effect by the proportion of the acquisition policy period spent in a boom climate. This was, for example, zero for AR and 0.154 (=2/13) for DSARC.

In addition, the overall features of the results are consistent with what would be expected from the history of OSD-level oversight of MDAPs over the relevant period (FY 1965–FY 2009).[9]

Four important conclusions are implied by the estimates in Table 1:

- The highly significant negative coefficient of Climate implies that the average PAUC growth of programs that passed MS B in a boom climate was significantly less than that of programs that passed in a bust climate.
- The 1969 Packard reforms of the acquisition process (which define the DSARC bin) resulted in a significant reduction in average PAUC growth compared to that of the preceding McNamara-Clifford period.
- The other three acquisition policy configurations (P-C DSARC, DAB, and AR) also had average PAUC growth significantly lower than that of McNamara-Clifford.
- Average PAUC growth in the four post-McNamara-Clifford acquisition policy bins did not differ significantly from one another.[10]

In brief—funding climate has the expected association with PAUC growth, the 1969 Packard reforms reduced average PAUC growth, and the reduction persisted through the end of the study period (FY 2009), but changes to the acquisition process after the Packard reforms through FY 2009 were not associated with further reductions in average PAUC growth.

The regression in Table 1 contains a remarkable feature. Ordinarily, when outliers (of the dependent variable) are removed from the dataset, the test statistics of the regression improve. This is not the case for the baseline model. If PAUC and duration outliers and programs procured with TPP are removed from the dataset, three of the four estimated coefficients of acquisition policy bins (including that for DSARC) are not statistically different from zero. The point is that the results are driven by the extreme values of PAUC growth. That is to say, the 1969 Packard reforms were effective because they reduced the frequency of MDAPs with extremely high PAUC growth.

Table 2 provides data that can be used to directly test this interpretation. The striking feature of these data is the paucity of outliers after the introduction of the Packard reforms in 1969. The PAUC growth of three of the 16 programs of the McNamara-Clifford years was large enough (at least 134%) to qualify as an outlier[11]; of the 94 MDAPs that passed MS B during the other four periods, only two had PAUC growth of at least 134%. This difference is

---

[9] See McNicol (2018), Chapter 5.

[10] This statement rests on the results of Wald's test with the Bonferroni correction. Wald's test, as used here, tests whether, considered jointly, any of $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$, and $\hat{a}_4$ are significantly different from the others. The Bonferroni correction effectively increases the critical value used to judge statistical significance to recognize that in multiple comparisons there is a considerable probability of a significant difference arising by chance even if the underlying population values are identical.

[11] We use the word *outlier* here as defined by John Tukey: observations 1.5 times the Inter Quartile Range above the third quartile or below the first quartile. None of the outliers had exceptionally low PAUC growth.

statistically significant.[12] Similar differences were not found for PAUC growth of at least 50% and at least 100%.[13] It appears then that the Packard reforms worked mainly by reducing the frequency of very high cost growth programs rather than by reducing cost growth on programs generally.

Table 2. Average PAUC Growth by Acquisition Policy Configuration and the Number of High Cost Growth MDAPs in Each Cohort, Bust Funding Climates for Completed Programs

Acquisition Policy Configuration	Period (FY)	Average PAUC Growth*	≥ 50%	≥ 100%	≥ 134%
McNamara-Clifford	1964–1969	74% (16)	9	4	3
DSARC	1970–1980	37% (49)	18	4	0
Post-Carlucci DSARC	1987–1989	34% (11)	2	2	1
DAB	1990–1993 2001–2002	40% (15)	5	1	0
Acquisition Reform (AR)	1994–2000	31% (19)	5	1	1

Model 2—The Basic Model Plus Boom Effects

Model 2 is prompted by the conjecture that boom climates facilitate PAUC growth of ongoing programs that enter them. If this is so, average PAUC growth presumably will be higher for MDAPs that entered a boom climate sometime after passing MS B than it will for those that did not. This would in particular be expected of programs that passed MS B in a bust climate, but it might also be true of programs that passed in a boom climate.

A two-part naming convention is used to label programs that encountered a boom climate post-MS B and those that did not. The first part of the label gives the funding climate prevailing when the program passed MS B—bust or boom. The second part—0, 1, or 2—denotes the number of boom climates the program entered post-MS B. Programs that passed MS B in a bust climate and were completed entirely within that bust climate will be referred to as Bust0—Bust because they passed MS B in a bust funding climate and zero because they were completed without entering a boom climate. Programs that passed MS B in a bust period and continued into a subsequent boom period make up Bust1 or, for the few programs that extended into two boom periods, Bust2. Programs that passed MS B in boom climates are, similarly, denoted Boom0 or Boom1. (There are no programs in Boom2 as of this writing because programs that passed MS B during the Carter-Reagan boom climate had only one subsequent boom climate they could enter, the post 9/11 boom.)

These definitions capture just one feature of the post-MS B funding climates experienced by programs. They exclude other features that possibly are relevant. For example, they do not take into account the time spent in different funding climates or

---

[12] Fisher's Exact Test (FET); p = 0.021. Application of FET to the five bins of the 134% column of Table 2 yields p = 0.016.
[13] FET; p = 0.297 and p = 0.271 for 50% and 100%, respectively.

transitions from boom to bust of programs that passed MS B in boom periods. These simple definitions do nonetheless provide a way to examine whether boom effects are visible in the data.

We start with average PAUC growth for Bust0, Bust1, and Bust2 presented in Table 3 for the post-McNamara-Clifford portion of the first bust-boom cycle period[14] and the entire bust portion of the second cycle. Recall that only data for completed programs are used. In both periods, average PAUC growth for the treatment group (Bust1) is higher than it is for the control group (Bust0)—42% compared to 18% for the first period and 44% compared to 12% for the second. These differences are statistically significant.[15] For programs that passed MS B in a bust period, subsequent entry into a boom period is then associated with higher PAUC growth. PAUC growth for Bust2 is higher than that of Bust0 but less than that of Bust1.[16] Average PAUC growth for Bust2, however, is not significantly different from that for either Bust0 or Bust1.

Table 3. Average PAUC Growth by the Number of Boom Periods Experienced for Completed MDAPs That Passed MS B in Post–McNamara-Clifford Bust Climates

Bin	1st Bust Period FY 1970–FY 1980	2nd Bust Period FY 1987–FY 2002
Bust0	18% (7)	12% (10)
Bust1	42% (38)	44% (35)
Bust2	28% (4)	none

The next step is to include boom effects in the baseline model. We define two variables, C_{boom}Bust and C_{boom}Boom. C_{boom}Bust is 1 for all programs in Bust1 and Bust2; these programs passed MS B in a bust and then experienced one or two boom periods post-MS B. For all other programs, C_{boom}Bust is zero. Similarly, C_{boom}Boom is 1 for programs in Boom1 and zero for all other programs; these programs passed MS B in the Carter-Reagan boom and then experienced the 9/11 boom period post-MS B. The results are presented in Table 4.

[14] Average PAUC growth for Bust0 programs is 87% and that for Bust1 is 34%. The anomaly here is not the average PAUC growth for Bust1—which is in line with the averages for the other bust periods—but the exceptionally high cost growth of Bust0.
[15] Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) find the PAUC growth data in each of the three bins of the first bust period to be consistent with a normal distribution. An F-test found the variances for Bust0 and Bust1to be significantly different. A two-tailed t-test assuming unequal sample variances found the means of Bust1 and Bust0 for the first period to be significantly different ($p = 0.003$). K-S finds the distribution of PAUC growth for Bust1 of the second bust period to be non-normal. The means of Bust0 and Bust 1 for the second bust period are significantly different by the Man-Whitney U test: $p = 0.018$, $U = 97.5$, $n_1 = 35$, $n_2 = 10$.
[16] The programs in Bust2 are the CVN 68, with a PAUC growth of 7%; the NAVSTAR GPS (85%); ATCCS-MCS (-34%); and the UH-60A (54%). A two-tailed t-test, with unequal variances as appropriate, found the mean of Bust2 not to be significantly different from that of Bust0 ($p = 0.732$) or Bust1 ($p = 0.440$).

Table 4. Estimated Coefficients and p-Values for a Model That Includes the Effects of Climate at MS B and Post-MS B

	Coefficients	p-value
Intercept	81.5%***	< 0.001
Funding Climate		
Climate	-26.6%**	0.027
C_{boom}Bust	18.0%*	0.099
C_{boom}Boom	8.8%%	0.702
Acquisition Policy		
DSARC	-56.5%***	< 0.001
P-C DSARC	-47.1%***	0.001
DAB	-55.9% **	< 0.001
AR	-77.1%***	< 0.001

* Statistically significant at less than the 10% level.

** Statistically significant at less than the 5% level.

***Statistically significant at less than the 1% level.

R-Square = 0.27 F = 7.420 (P < 0.001) N= 149. Estimated using OLS. Boom2 programs and the three mid-1980s MDAPs acquired using TPP-like contracts are omitted. With the Bonferroni correction, Wald's test for the equality of the estimated coefficients of the categorical variables for acquisition policy periods yields F = 1.47 (p =0.90).

Three MDAPs in Boom1 were acquired using a TPP contract. These three programs have much higher PAUC growth than Boom0 programs—because they were acquired using a TPP contract, not because they passed into a boom period post-MS B. For that reason, they are dropped from the sample.[17]

The estimated coefficient of C_{boom}Bust (18.0%) is marginally significant. We do, then, see evidence of a boom effect. The estimated coefficient of C_{boom}Boom (8.8%) is smaller and not significant. As with the previous model, estimates imply that the Packard reforms of 1969 resulted in a decrease in PAUC growth; that decrease persisted, but subsequent changes in acquisition policy apparently did not result in further significant decreases in PAUC growth.

Model 3—Program Duration

The longer a program's duration, the greater its chance of moving into a boom funding climate. For that reason alone, longer duration presumably is associated with higher PAUC growth.

Figure 1 provides some evidence on the premise of the discussion. It plots the average PAUC growth for the bust and boom bins along with the corresponding average program duration (defined as the number of years from MS B through the end of the

---

[17] See McNicol (2018) Chapter 3, Section B.

acquisition phase).[18] The prefixes 1st and 2nd indicate the bust-boom cycle—FY 1965–FY 1986 (1st)[19] and FY 1987–FY 2009 (2nd). Programs that passed MS B in bust climates added more PAUC per year of duration than did programs that passed MS B in a boom climate.

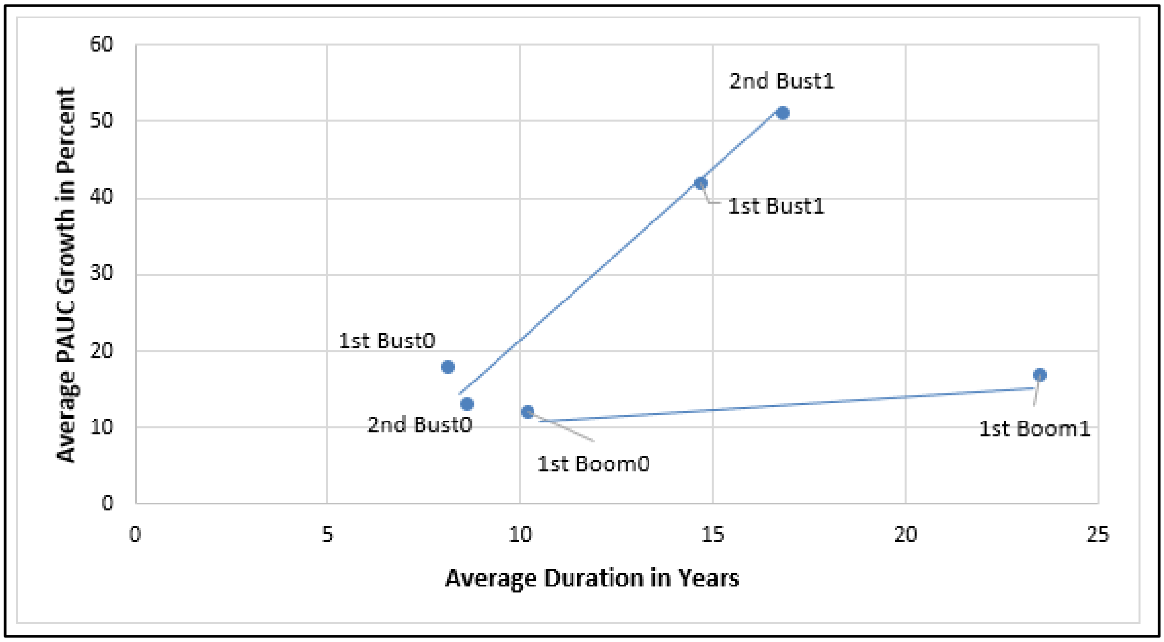


Figure 1. Average PAUC Growth and Average Program Duration for Boom and Bust Periods

A plausible approach to separating the boom effect from a duration effect is to enter into the model a variable defined as the number of years spent in boom climates (T_{boom}) and another variable that is the number of years spent in bust climates (T_{bust}). Very simple definitions of T_{boom} and T_{bust} were adopted:

- T_{boom} = number of years in boom climates post-MS B
- T_{bust} = number of years in bust climates post-MS B

Note that this definition counts a year during which the program was in Engineering and Manufacturing Development the same as a year in which the program was in Low Rate Initial Production (LRIP) or full rate production. There are several alternatives to this definition. For example, the duration variables might be defined as the years in boom and bust climates, respectively, after the program enters LRIP.

Setting aside for the moment the categorical variables for the acquisition policy configurations, the core model considered is shown in Equation 2:

$$PAUC_i = a + bClimate + cT_{boom,i} + dT_{bust,i} + v_i. \quad (2)$$

[18] The end of the acquisition phase was defined as the final year in which substantial procurement funding was obligated, as reported in the program's final SAR.
[19] The McNamara-Clifford period was excluded from the first cycle. See footnote 14

where v_i is the error term. Note that c and d are measured in units of percentage points per year; they are the rates at which programs' PAUC growth increases per year in boom and bust climates, respectively.

We expect the estimated coefficient of Climate to be negative, implying that programs that passed MS B in boom climates have lower PAUC growth than those that passed in bust climates. This specification also allows for climate effects in that the estimates of c and d may be different. In particular, we would expect the estimate of c to be larger than that of d—that is, that PAUC growth accumulates more rapidly in boom than in bust years. Estimates of the parameters of this model expanded to include the categorical variables for the acquisition policy configurations are presented in Table 5.

Table 5. Estimated Coefficients and p-Values for a Model That Includes the Effects of Post-MS B Funding Climate and Duration

	Coefficients	p-value
Intercept	76.5%***	< 0.001
Funding Climate		
Climate	-40.2%***	< 0.001
T_{boom}	4.2%/yr***	0.008
T_{bust}	0.22%/yr	0.804
Acquisition Policy		
DSARC	-57.5%***	< 0.001
P-C DSARC	-44.5%***	0.002
DAB	-57.5%***	< 0.001
AR	-83.1%***	< 0.001

* Statistically significant at less than the 10% level.

** Statistically significant at less than the 5% level.

***Statistically significant at less than the 1% level.

R-Square = 0.30 F = 8.471 (P < 0.001) N= 149. Estimated using OLS. Boom2 programs and the three mid-1980s MDAPs acquired using TPP-like contracts are omitted. With the Bonferroni correction, Wald's test for the equality of the estimated coefficients of the categorical variables for acquisition policy periods yields F= 2.18, p = 0.3724.

All of the estimated coefficients have the expected signs, and all except that for T_{bust} are statistically significant. Like the estimates for Models 1 and 2, the estimates for Model 3 imply that the Packard reforms of 1969 resulted in a decrease in PAUC growth; that decrease persisted, but subsequent changes in acquisition policy apparently did not result in further significant decreases in PAUC growth. The new result provided by Model 3 is that PAUC growth on average increases by 4.2 percentage points (the estimated coefficient of T_{boom}) for each year spent in a boom climate. Note that Model 3 assumes that the effect on PAUC growth of a boom year is the same for programs that passed MS B in boom climates as it is for those that passed in bust climates, which probably is not the case.

Model 4—Alternative Representation of Climate Effects

There is a way to overcome this limitation of Model 3. The alternative uses what are called slope categorical variables, one for boom years (T_{boom}*Climate) and one for bust years (T_{bust}*Climate). In this approach, climate effects are captured in the estimated coefficients of T_{boom}, T_{bust}, and the slope categorical variables. As is illustrated later, introduction of these variables allows the regression to pick different rates of cost accumulation for MDAPs that passed MS B in boom climates than for those that passed in bust climates. We expect that MDAPs that passed MS B in boom years accumulate less PAUC growth in both bust and boom years than MDAPs that passed MS B in bust years. The estimated coefficients for T_{boom}*Climate and T_{bust}*Climate are then expected to be negative.

Table 6 presents the estimated coefficients and p-values for this alternative. Once again, the estimates imply that the Packard reforms of 1969 resulted in a decrease in PAUC growth; that decrease persisted, but subsequent changes in acquisition policy apparently did not result in further significant decreases in PAUC growth. These estimates, however, shed considerable light on why programs that passed MS B in bust climates on average had higher cost growth than those that passed in boom climates. To see how requires being clear about the estimated rates at which programs accumulated cost growth over time.

The estimated coefficient of T_{boom} (4.8%/yr) is the rate at which a program that passed MS B in a bust climate accumulates cost in boom years. The rate for programs that passed MS B in boom climates is much lower, and in fact, negative: -0.1%/yr. This is the sum of the estimated coefficient for T_{boom} and the coefficient for T_{boom}*Climate (4.8%/yr – 4.9%/yr = -0.1%/yr). The uncertainties in the estimates are such, however, that the estimated rate could about as easily be 0.1%/yr as 0.1%/yr.[20] The point here is that programs that passed MS B in bust climates evidentially accumulate PAUC growth much more rapidly when they encounter a boom period than do programs that passed MS B in a boom climate.

---

[20] The negative estimates are not unreasonable *a priori.* Some programs that pass MS B in a boom climate may be used as "banks"—that is, relatively safe places to hold out-year claims on funding. A program used as a bank will show negative quantity normalized PAUC growth when "withdrawals" are made. It may be relevant in this regard that about one-third of the programs that passed MS B in boom climates show negative PAUC growth.

Table 6. Estimated Coefficients and p-Values for a Model That Includes the Effects of Post-MS B Funding Climate and Duration

	Coefficients	p-value
Intercept	64.2%***	< 0.001
Funding Climate		
T_{boom}	4.8%/yr***	0.006
T_{boom}* Climate	-4.9%/yr*	0.095
T_{bust}	1.2%/yr	0.232
T_{bust}* Climate	-1.8%/yr	0.230
Acquisition Policy		
DSARC	-57.0%***	< 0.001
P-C DSARC	-48.1%***	< 0.001
DAB	-60.4%***	< 0.001
AR	-82.9%***	< 0.001

* Statistically significant at less than the 10% level.

** Statistically significant at less than the 5% level.

*** Statistically significant at less than the 1% level.

R-Square = 0.29 F = 5.690 (P < 0.001) N= 149. Estimated using OLS. Boom2 programs and the three mid-1980s MDAPs acquired using TPP-like contracts omitted. With the Bonferroni correction, Wald's test for the equality of the estimated coefficients of the categorical variables for acquisition policy periods yields F = 1.95 p = 0.4948.

Programs that passed MS B in a bust climate accumulate cost in subsequent bust years at an estimated rate of 1.2 percentage points per year. This estimate is not statistically significant. The corresponding rate for programs that passed MS B in a boom climate is -0.6%/yr (= 1.2%/yr – 1.8%/yr). The estimated coefficients for T_{bust} and T_{bust}*Climate also are not statistically significant. The estimates, then, do not say much about the rate at which PAUC growth accumulates in bust years.

Cost Growth Due to Program Changes

The duration variables of Model 4 direct attention to the question of the extent to which cost growth of programs that passed MS B in bust climates is due mainly to unrealistic MS B baselines. At one extreme, most programs that pass MS B may have unrealistic MS B baselines and use entry into a boom period as a chance to "get well." At the other extreme, the tendency in bust periods may be to approve austere programs. When these programs enter a boom climate, they are expanded to acquire capabilities beyond those in their MS B baseline, which is to say that PAUC growth may be largely a matter of program changes.

Selected Acquisition Report: Global Broadcast System (DoD, 2003) provides an example of a program whose content was increased early in the post-9/11 boom. (It passed MS B early in FY 1998 and is accordingly a Bust1 program. GBS was ongoing at the end of FY 2016.) According to the report,

> The current GBS architecture is based on Asynchronous Transfer Mode (ATM) technology. … In December 2002, DoD directed GBS's migration to a more sustainable commercial and standards-based open architecture, based

upon the Internet Protocol (IP). Also, the GBS program received FY03 Iraqi Freedom Funds (IFF) supplemental funding for IP Acceleration of production units to replace deployed ATM units. Based upon extensive warfighter inputs, the accelerated IP production effort included design and development of a new, single case version of the Receive Suite (88XR) for the Army, Navy, and Marine Corps.

Space Based Infrared Satellite-High (SBIRS-High), like GBS, is a Bust1 program. It passed MS B early in FY 1997. As of the December 2015 SARs, funding for the Baseline SBIRS-High program was expected to end in FY 2018. SBIRS-High is a useful contrast to GBS. A large portion of the growth in SBIRS-High unit procurement cost for the baseline program—roughly one-third—occurred before FY 2003, while most of the other two-thirds occurred during FY 2003–FY 2009. This increase was not driven by increased capability, however, but by the unrealistic cost estimate in the MS B SBIRS-High baseline (See Porter et al., 2009; Younossi et al., 2008; Kim et al., 2015).

In the GBS example, it seems clear that capabilities beyond those in the MS B baseline were added to the program. While unit cost did increase, that was a matter of paying more for more. For SBIRS-High, in contrast, it appears that the advent of a boom funding climate provided a program experiencing severe problems an opportunity to "get well." In effect, what otherwise would have been capability shortfalls were converted into cost growth and, relative to MS B, the DoD eventually paid more for the MS B SBIRS-High capability than had been anticipated.

As these examples indicate, the boom effect in general results from acquisition of capability beyond that in the MS B baseline and unrealistic assumptions in the MS B baseline. In examples, the PAUC growth associated with the boom climate mainly appeared in the SARs for the boom years. While we have no examples to offer, PAUC growth for Bust1 and Boom1 programs also occurs between MS B and the subsequent boom. Again, this growth can reflect either acquisition of capability beyond the MS B level or recognition that the cost of acquiring the MS B capability is higher than anticipated.

During a period of nearly 20 years starting in 1989, the Office of Program Analysis and Evaluation (PA&E), predecessor of the Office of Cost Assessment and Program Evaluation (CAPE), funded development of a database that separated cost growth due to program changes[21] from cost growth due to what PA&E called "mistakes."[22] The data in Table 7 are drawn from the version of the PA&E database updated through the December 2002 SARs.[23]

---

[21] A major difficulty in separating program changes from Errors of Inception is ambiguity in statements of capabilities to be acquired. Those responsible for compiling the PA&E database were well aware of this problem.

[22] In about 2010, the Office of Program Assessments and Root Cause Analyses (PARCA) defined top-level proximate causes of cost growth. These included both Errors of Inception and Errors of Execution. As defined earlier by PA&E, the "mistakes" category is the sum of Errors of Inception and Errors of Execution. See McCrillis (2003).

[23] This is the database used in McNicol (2004).

Table 7. PAUC Growth Due to Errors and Program Changes

Cycle	Period (Fiscal Years)	Number of MDAPs that Passed MS B	Errors†	Program Changes‡	Total	Program Changes as a Percent of Total
Boom	1981–1986	35	4%	11%	14%#	79%
Bust	1970–1980	42	24%	14%	38%	37%
	1987–1997	46	21%	10%	31%	32%
	Combined bust	88	22%	12%	34%	35%

† Errors of Inception plus Errors of Execution.

‡ Changes made as a result of decisions to alter from the MS B baseline the capabilities the program is to acquire.

Components do not add to the total because of rounding error.

In the boom climate FY 1981–FY 1986, program changes were almost 80% of the total PAUC growth. In the bust periods, however, PAUC growth due to program changes was about one-third of the total. These data imply that the higher PAUC growth of programs that passed MS B in bust climates is primarily due to errors.

This analysis can be carried forward another step. The most interesting number in Table 7 for this purpose is the 4% for errors in the boom period FY 1981–FY 1986. This number is the sum of Errors of Inception and Errors of Execution. It is reasonable to assume, however, that Errors of Inception are on average small for programs that passed MS B in a boom period. Pushing that assumption to its limit, we have an estimate for Errors of Execution for the programs for the first boom period of 4%. Unfortunately, comparable data for the second boom period (FY 2003–FY 2009) are not available, so we have no check on how representative this estimate is; it is the only estimate we have of the average Errors of Execution for a substantial number of programs. If it is accepted as representative, the data in Table 7 imply that the average PAUC growth of MDAPs that passed MS B in bust climates due to unrealistic MS B baselines is about 17% to 20%.

Conclusion

Each of the models yielded the same pattern of results:

- MDAPs that passed MS B in a bust climate on average had significantly higher PAUC growth than those that passed MS B in a boom climate;
- The 1969 Packard reforms reduced average PAUC growth;
- The reduction persisted through the 1980s, 1990s, and 2000s;
- Changes to the acquisition process after the 1969 Packard reforms are not associated with further reductions in average PAUC growth.

Incorporation of the boom effect and program duration in the models does not provide further policy insights. These factors were significant, however, and must be considered in future analyses of MDAP cost growth.

The PA&E data on PAUC growth due to program changes suggest that the lower PAUC growth after the Packard reforms probably was due mainly to adoption of more realistic MS B baseline lines. We also find in those data an indication that cost growth baked into the MS B baselines—that is, Errors of Inception—are several times larger than Errors of Execution. That conclusion, however, amounts to less than it might seem to at first glance. The classic Error of Inception occurs when the DoD contracts for a Lincoln and budgets for a Ford. Eventually, additional funding must be added to the budget to buy the Lincoln. The DoD must make the necessary budgetary adjustments within a given top line—usually within funding for acquisitions. These adjustments include such measures as stretches, delays, cancellations, and descoping of programs. It is the cost increase imposed by these adjustments, rather than the difference between the cost of a Lincoln and a Ford, that is the relevant cost of Errors of Inception.[24]

References

DoD. (2003). *Selected acquisition report: Global broadcast system.* Defense Acquisition Management Information Retrieval (DAMIR) System.

Kim, Y., Axelband, E., Doll, A., Eisman, M., Hura, M., Keating, E. G., Libicki, M. C., & Shelton, W. (2015). Acqu*isition of space systems, Volume 7: Past problems and future challenges* (RAND MG-1171/7-OSD). Santa Monica, CA: RAND. Retrieved from http://www.rand.org /pubs/monographs/MG1171z7.html

McCrillis, J. (2003, January). *Cost growth of major defense programs*. Briefing presented at the Department of Defense Cost Analysis Symposium, Williamsburg, VA.

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (2nd ed.). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017a). *Post-Milestone B funding climate and cost growth in major defense acquisition programs* (IDA Paper P-8091). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017b). Post-Milestone B funding climate and cost growth in major defense acquisition programs. In *Proceedings of the 14th Annual Acquisition Research Symposium* (Vol. 1, pp. 86–97). Retrieved from http://acqnotes.com/wp-content/uploads/2017/08/SYM-AM-17-034_Wednesday-Vol-1_5-1-2017.pdf

McNicol, D. L. (2018, forthcoming). *Acquisition policy, funding climate, and cost growth of major defense acquisition programs* (IDA Report R-8396). Alexandria, VA: Institute for Defense Analyses.

O'Neil, W. D., & Porter, G. H. (2011). *What to buy? The role of director of defense research and engineering (DDR&E)—Lessons from the 1970s* (IDA Paper P 4675). Alexandria, VA: Institute for Defense Analyses.

Porter, G., Gladstone, B., Gordon, C. V., Karvonides, N., Kneece, Jr., R. R., Mandelbaum, J., & O'Neil, W. D. (2009). *The major causes of cost growth in defense acquisition: Volume I—Executive summary* (IDA Paper P 4531). Alexandria, VA: Institute for Defense Analyses.

Younossi, O., Lorell, M. A., Brancato, K., Cook, C. R., Eisman, M., Fox, B., … Sollinger, J. M. (2008). *Improving the cost estimation of space systems: Past lessons and future recommendations* (MG-690-AF). Santa Monica, CA: RAND. Retrieved from http://www.rand.org/pubs/monographsMG690.html

[24] The magnitude of Errors of Inception is extremely difficult to measure. The only published attempt to do so seems to be McNicol (2004), pp. 9–10 and Appendix B. This computation recognizes some considerations in addition to stretches, which adds to the complexity.

Acknowledgments

David M. Tate (the dean of the reviewers and co-author of one of the papers in the cost growth series), David A. Sparrow, Daniel Cuda, Prashant Patel, and Philip Lurie, all of IDA, provided insightful comments on successive drafts. I am similarly grateful to Gregory Davis for his support as critic and task leader for the past two years. Brian Gladstone and Sarah Burns (co-author of one of the papers), of IDA, and Mark Husband, of Defense Acquisition University, provided valuable comments on the first three papers of the series. Linda Wu, formerly of IDA, managed data acquisition and the database. More recently, J. M. Breuer provided helpful assistance with the statistical work. The research presented in this paper was sponsored by the Director, PARCA.

Comparing Ship Versus Aircraft Development Costs

Larrie D. Ferreiro—is the Director of Research at the Defense Acquisition University at Fort Belvoir, VA. He also teaches systems engineering at DAU, at the Naval Postgraduate School, and, as an adjunct professor, at Georgetown University. Dr. Ferreiro has almost 40 years of experience in ship engineering for the Navy, Coast Guard, private industry, and on assignment as a naval engineer inside the British and French navies. He is a Pulitzer Prize finalist in History for his book *Brothers at Arms: American Independence and the Men of France and Spain Who Saved It.* [larrie.ferreiro@dau.mil]

Abstract

Both warships and military aircraft are highly complex, engineered products that can cost hundreds of millions of dollars each. But the development cost for an aircraft is frequently many times the cost for a ship, in some cases one to two orders-of-magnitude greater (DDG 51 development cost $3 billion, F22 development cost $30 billion). This paper first examines and compares the top-line development costs for a broad range of ships and aircraft, from commercial (e.g., passenger ships and aircraft) to military (destroyers versus fighters), using publicly available cost numbers. It then takes a deep dive into two cargo platforms, T-AKE *Lewis and Clark* and C-17 Cargolifter, using cost data from primary sources. It then compares the development expenditures for the two platforms as a function of time and products (e.g., the use or lack of full-scale models as part of the respective development processes). It finally provides a broad historical perspective to explain how these differences between ships and aircraft actually began in their original development communities during the 19th and early 20th centuries.

Introduction and Research Methodology

Both warships and military aircraft are highly complex, engineered products that can cost hundreds of millions of dollars each. But the development cost for an aircraft is frequently many times the development cost for a ship, in some cases one to two orders-of-magnitude greater. The literature on why this is the case is almost non-existent. The only published study that examines this disparity was recently carried out by RAND, appropriately titled *Are Ships Different?* (Drezner et al., 2011). It focused on the acquisition process of ships compared with that of missiles, aircraft, and tanks. The study highlighted the fact that ships are typically built in low numbers of units compared with other programs. It showed that "ship programs do not typically design and build prototype units designated solely for test," which is almost always the case for other program types, in order to de-risk the final production run. Finally, in part because the lead operational ship acts as the de facto prototype for the rest of the class, full-scale production for ships begins at Milestone B, whereas other programs include extensive prototyping in the engineering development phase after Milestone B, before committing to full scale production at Milestone C (Figure 1).

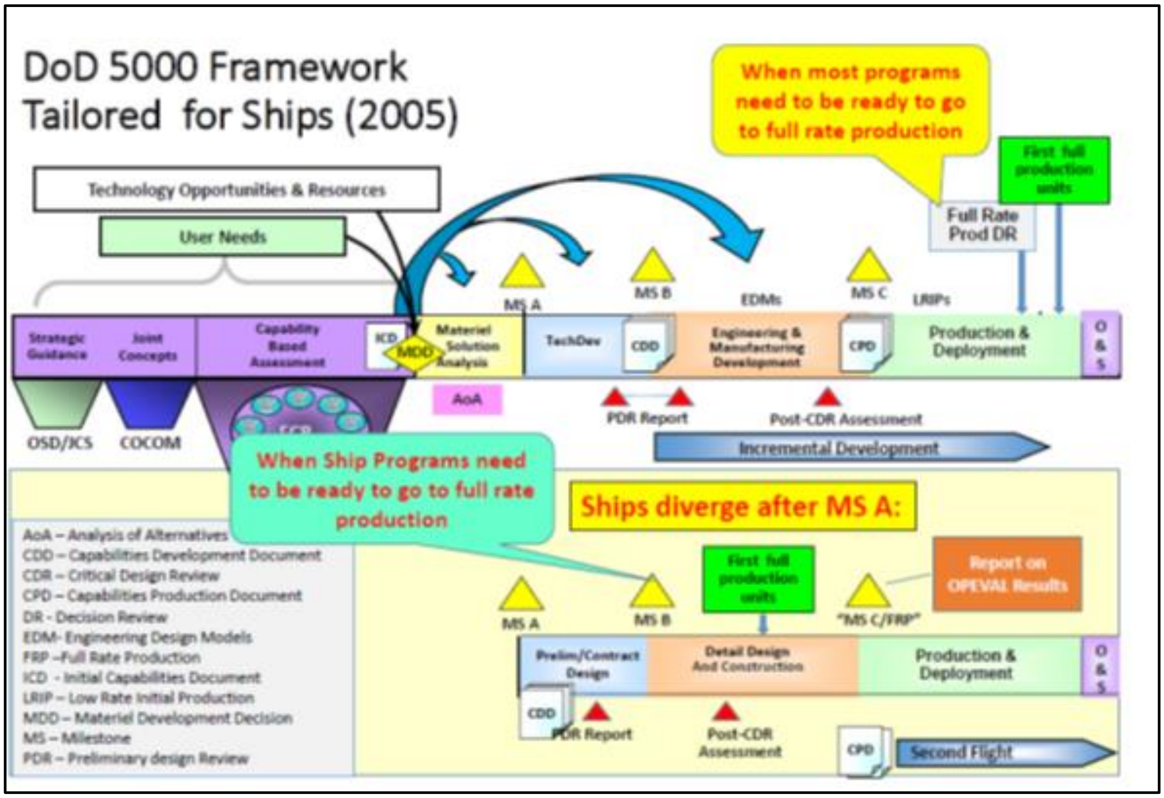


Figure 1. Ship Acquisition Timeline Compared With Other DoD Program Timelines

This RAND study highlights the need for a deeper examination of the cost disparities between the development of ships and aircraft, not only military but also commercial ones. For this reason, I first examined the development cost disparities at a high level between several different ship and aircraft programs, then took a deep-dive comparison between two cargo platforms, T-AKE *Lewis and Clark* and C-17 Cargolifter. These candidates were selected for the following reasons:

- The two platforms are broadly similar in mission: to carry cargo. This largely removes disparities between, say, multi-mission destroyers and single-mission fighters.
- The two platforms have very few weapons systems and combat systems, which can complicate the costing structures for both system development and platform integration.
- The detailed development cost data for the two military platforms are relatively straightforward to obtain via public domain sources; by contrast, detailed development cost data for both commercial aircraft and commercial vessels are proprietary and closely held by companies.

Ship Versus Aircraft Development Costs in Context

My first task was to compare a variety of ship and aircraft programs to determine if there was indeed a general trend of higher development costs for aircraft compared with ships, and to get a rough order-of-magnitude assessment of the difference between them. Table 1 shows these comparisons across both military and commercial platforms, in the United States and in the United Kingdom.

Table 1. Ship Versus Aircraft Costs as of 2005
(Ferreiro, 2016)

	UK Type 23 frigate		UK Typhoon fighter	US DDG 51 destroyer		US F22 fighter
Units	16		620	62		187
Development	**$0.7B**	34x	**$24B**	**$ 3B**	10x	**$28B**
Procurement	$4.3B		$23B	$60B		$34B
Total	$5.0B		$47B	$63B		$62B

	T-AKE Cargo ship		C-17 Cargo plane	Cruise Passenger Ship		Airbus A380 Passenger plane
Units	12		190	10		65+
Development	**$0.1B**	50x	**$ 7B**	**$0.06B**	200x	**$13B**
Procurement	$4.6B		$59B	$ 6B		$22B+
Total	$4.7B		$66B	$ 6B		$35B+

The trends show an order-of-magnitude difference between military platform (Type 23 vs. Typhoon, DDG 51 vs. F22, T-AKE vs. C-17) and a two orders-of-magnitude difference between commercial platforms (passenger ship vs. A380). This confirms that the disparity between development costs is not limited to warships and combat aircraft, but instead is a systematic trend across platform types, whether military or commercial.

Figure 2. C-17 and T-AKE

C-17 Versus T-AKE Development Costs

The next step in this study was to take a more in-depth look at the development costs between the T-AKE *Lewis and Clark* and the C-17 Cargolifter (Figure 2). The cost data was obtained from public domain sources (Defense Acquisition Management Information Retrieval [DAMIR], 1997; GAO, 1991; Naval Sea Systems Command [NAVSEA], 2017; Naval Surface Warfare Center Carderock Division [NSWC-CD], 2018; DAMIR, 2011) and is shown in Table 2. Of specific note is that for the T-AKE, the detailed design costs for production is accounted for in a separate line item that is part of Ship Construction, Navy (SCN) and not part of the Research, Development, Test and Evaluation (RDT&E) budget. By contrast, the detailed design costs for production of the C-17 is spread among the various elements included in the development costs, and cannot be readily broken out as a separate cost. Rather than follow the specific Work Breakdown Structure (WBS) for each platform, I have attempted to correlate major cost categories between the platforms where possible, and break out unique cost categories for each platform where needed.

Table 2. C-17 Versus T-AKE Research, Development, Design and Test (RDDT) Costs in $ Millions, Rounded to the Nearest $1 Million

(DAMIR, 1997; GAO, 1991; NAVSEA, 2017; NSWC-CD, 2018; DAMIR, 2011)

C-17 (1991)		T-AKE (2001)	
		Early stage designs	1
		Baseline designs	3
		Model basin testing (hull)	1
Structures (fuselage, wing, tail)	221		
Structural analysis	115	Survivability analysis	1
Power system (engines)	119		
Electrical system	26		
Avionics and flight control systems	203		
Mechanical systems (environmental, landing, control surfaces)	95	Environmental, safety and health	1
Mission equipment	11	Mission systems (cargo)	3
Other	11	Other studies	5
Test vehicle manufacturing (1 flyable test aircraft, 2 ground test airframes)	211		
Other unallocated	40		
Systems engineering, design, and integration	114	Systems integration design	6
Project management, test & evaluation, and support equipment	900	Program management and support	6
		Detailed design	120
Other unspecified, including full-scale testing of 1 flyable test aircraft and 2 ground test airframes	2,130		
TOTAL RDDT (1991)	**4,200**	**TOTAL RDDT (2001)**	**147**
Actual RDDT (2004)	6, 687		

Analysis of Development Expenditures

Major cost items for the C-17 were as follows:

- Structures, which includes development of the fuselage, wing, and tail section (Each of these was adapted to the unique short-field landing requirement of the aircraft.)
- Structural analyses of the above
- Power and electrical systems, including development of high-capacity thrust reversers for the four main engines
- Avionics (cockpit) and flight control (fly-by-wire) systems, which included the development of full-scale cockpit mockups
- Test vehicle manufacturing and full-scale testing of one flyable aircraft (i.e., a prototype) and two ground test airframes, including static and dynamic structural loading tests
- Systems integration, including mating surfaces and equipment for subsystems and major systems

- Project management, test & evaluation, and support equipment

Small-scale model testing (e.g., in wind tunnels) was not broken out directly, but is presumably included in the above items.

The actual development cost for C-17 escalated from $4.2 billion in 1991, when the detailed data for this study was generated, to $6.7 billion as of the Selected Acquisition Report (SAR) that formed the 1999 President's Budget, which showed costs out to fiscal year (FY) 2004. Therefore, the final detailed numbers for the items in Table 2 are likely to be, on average, 50% greater than shown.

For the T-AKE, the major cost items were as follows:

- Early stage design work, which included feasibility studies, point designs using computer-aided design tools, and hydrodynamics testing of small-scale models up to 10 meters long, at facilities such as the Naval Surface Warfare Center Carderock Division (NSWC-CD)
- Mission systems, including computer-aided cargo flow modeling
- Systems integration design; program management and support, which includes all of the documentation necessary to pass Milestone decision authorities; and support from the classification society American Bureau of Shipping
- Detailed design costs, including direct shipyard and subcontracted engineering to develop detailed plans for production

These cost items are current, as the ship entered service in 2006, and these numbers align with the 2011 SAR.

The most remarkable difference between the C-17 development program and that of the T-AKE is the testing. For the T-AKE, the small-scale model testing for hydrodynamics (e.g., speed-power) is on the order of $1 million. For the C-17, the full-scale construction and testing of one flyable, prototype aircraft plus two ground test airframes is about $2.3 billion, about half the total development cost for the aircraft, and also 2,300 times (three orders of magnitude) greater than for the T-AKE. Other full-scale testing included the cockpit mockups. Although other examples abound (e.g., structural analysis for the C-17 is one hundred times greater than for the T-AKE survivability analysis), it is the use of full-scale testing versus small-scale testing that accounts for the lion's share of the difference in development costs between the two platforms.

Explanations for Differences in Development Expenditures

The differences between the development costs for aircraft and ships are seen in their overall program approaches. Figure 3 highlights these differences, while Figures 4, 5, and 6 compare the activities for each at the different phases of their respective development programs. Note again (as shown in Figure 1) that ship production begins at MS B, while aircraft production begins at MS C.

Figure 3. Overall Differences Between Ship and Aircraft Development Programs

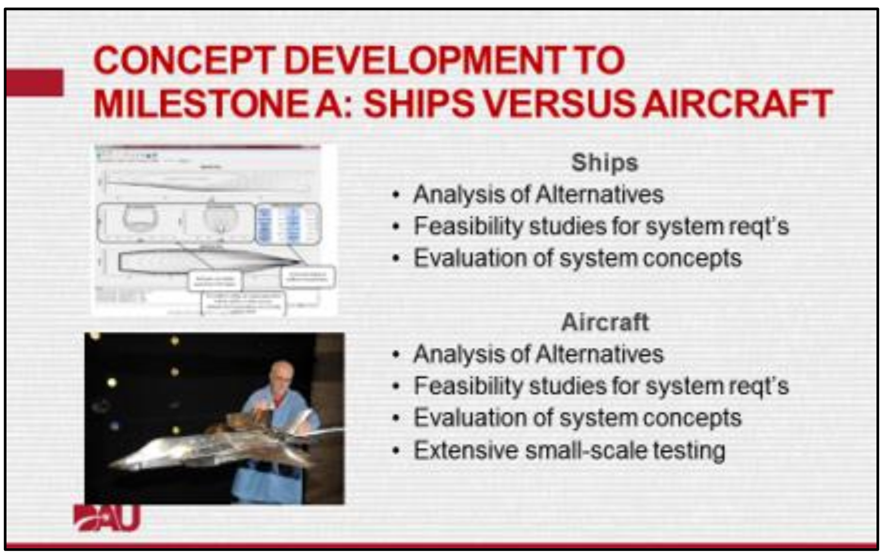


Figure 4. Differences Between Ships and Aircraft at the Concept Development Phase

Figure 5. Differences Between Ships and Aircraft at the Technology Development Phase

Figure 6. Differences Between Ships and Aircraft at the Systems Development Phase

Specific to T-AKE versus C-17, the most noticeable difference between the two platforms lies in the verification and validation processes for the designs and production models. Verification and validation of the T-AKE involves having the ship classed by the classification society American Bureau of Shipping (ABS). Classification is a process where a society like ABS develops internationally-recognized rules for design and construction of ships, which can also include national and international safety regulations (e.g., for stability). In this circumstance, ABS reviews plans and calculations done by the shipbuilder to verify compliance to design code, and regularly inspects the vessel while under construction and in trials to ensure adherence to code. In many cases, the vessel remains "in class" (i.e., the owner contracts with ABS to carry out regular surveys in service in order to ensure continuing compliance with the standards). Many warships today are also classed by classification societies (e.g., Lloyd's Register Naval Ship Rules, Bureau Veritas Rules for the Classification of Naval Ships). In other cases, like for the U.S. Navy, shipbuilding standards and specifications are developed by the Navy itself, which carries out its own inspections. The first in-service ship is also the first one to go through the classification process, in a sense serving as a "test vessel" for the entire class.

By contrast, the verification and validation program for C-17, as explained, involves the construction and testing of many full-scale models, mockups, and prototypes for the full platform, as well as subsystems. Unlike the T-AKE program, the C-17 program developed and extensively tested full-scale cargo hold mockups, full-scale engine mockups, full-scale cockpit mockups, and full-scale wing sections, which were tested to destruction. It also had, as noted, one flyable, prototype aircraft plus two ground test airframes.

Rationale Behind Differences in Development Expenditures

We have identified full-scale prototyping for aircraft verification and validation, versus the rules-and-standards-based system for ships, as the primary driver of the difference between the costs for aircraft and ship development. The next question is, "Why should this be the case?" There are a number of myths that have been proposed to explain this, and they all fall apart upon close inspection. These myths fall under three general categories:

1. Criticality and Safety. Aircraft accidents are seen to be particularly horrific events, especially when the accident causes the plane to literally fall from the sky. A case in point is the catastrophic explosion (due to faulty wiring and poor design) of TWA 800 off Long Island, NY in 1996, killing all 230 people aboard. Thus, the need for extremely high levels of safety afforded by rigorous, full-scale testing of critical systems. By contrast, ships floating on water certainly *appear* safer than aircraft. Yet this is patently not true. In 1994, MV *Estonia* foundered in the Baltic Sea with the loss of 852 lives, about four times the number killed on TWA 800. The blame was ultimately placed on faulty design and operation of a safety-critical system, the bow doors, in part because the wave loads were underestimated—a problem that might have been avoided with rigorous full-scale testing. (Note that both C-17 and T-AKE each carry about 140 military personnel, so, in theory, they should employ equivalent means of achieving appropriate levels of safety. They do not.)
2. Number of units built. Some of the interviewees in the RAND study *Are Ships Different?* claimed that "because of the relatively high unit cost and low total production quantities, ship programs do not typically design and build prototype units designated solely for test" (Drezner et al., 2011). This is a red herring. The previous two classes of U.S. Navy destroyers were built in quantities comparable to, or greater than, those of military aircraft. The DDG 51 class has 62 units and is projected to have 77 units; the DD 963 class had 62 units, including the follow-on series DDG 993 and CG 47/52. By contrast, the F-22 fighter has 187 operational units, while the B-2 bomber has just 21 units.
3. Complexity. Aircraft are perceived to be more complex than ships, thus require more rigorous testing to iron out the bugs. Again, this is false. Using parts count as a straightforward if unsatisfactory proxy for

complexity, the *Ohio*-class submarine, with 350,000 parts (and which is verified and validated via the same type of rules-and-standards method as surface warships) is more complex than the F-16 fighter with just 175,000 parts (Drezner et al., 2011).

There are many valid reasons why shipbuilding programs could and should incorporate full-scale prototyping as part of the verification and validation process. This will not happen, of course, so the question remains, “Why are ships and aircraft different?”

The answer lies in the origins of the modern shipbuilding and aircraft industries. In the 19th century, the same men who built iron and steel ships also constructed bridges, buildings and railroads, and both used rule-of-thumb methods and visual inspections as their means of verification and validation of designs. In the 1860s, the British engineer William Fairbairn used the same methods for calculating bridge girder loads and stresses in his foundries, as he used for building newfangled iron ships in his shipyard. He even used the same factors of safety for structural loading. Those methods were carried on by the many steelyards in the 20th century that also built ships, such as the Missouri Valley Bridge & Iron Co., which constructed more LSTs (Landing Ship, Tank) than any other yard in World War II. This civil engineering inheritance is especially noteworthy when comparing the aforementioned ABS rules with civil building codes (Figure 7). For this reason, RAND was correct when it noted that “Ships are more like a major military construction project than weapon-system procurement” (Drezner et al., 2011).

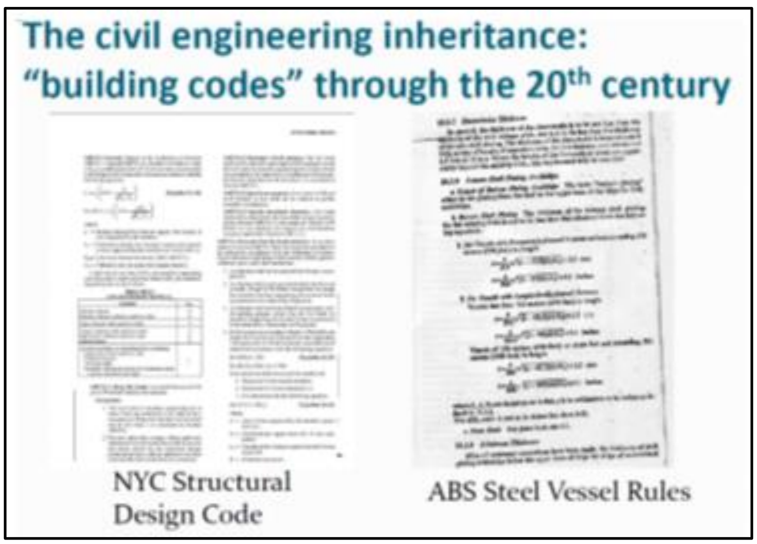


Figure 7. Civil Versus Maritime Building Codes

Aircraft, by contrast, were born in the 20th century, just when physics-based engineering was coming of age. Right from the start, aircraft design was dominated by the likes of German physicist Ludwig Prandtl, who developed advanced theories to explain the aerodynamic performance of lifting surfaces. This was reflected in the amount of research funding poured into aircraft development. In the early 20th century, the U.S. Navy had led the shipbuilding industry in scientific experimentation by funding the construction of two ship model test basins—the Experimental Model Basin (EMB) at the Washington Navy Yard, and another at the University of Michigan. At the same time, almost a dozen wind tunnels sprang up around the nation, including six run by the newly-created National Advisory Committee for Aeronautics (NACA). In the 1902s, the EMB received less than $100,000 annually in appropriations, whereas NACA was being funded to the tune of $1.3 million per year (Ferreiro, 2014).

For some time, in fact, ship classification societies attempted to extend their rules-and-standards methods to aircraft. In 1929, the Aircraft International Register (AIR) was established "to be for commercial aircraft what Lloyd's Register is to shipping" (i.e., intended to provide an internationally accepted set of classification rules for flying machines). For several years, ABS, Lloyd's, and Bureau Veritas established independent aeronautical branches to help the fledgling aircraft industry develop and codify these new procedures and practices. Within a few years, however, national governments took on the role of issuing airworthiness certificates for aircraft, making the role of classification societies redundant. With that, most classification societies shuttered their aeronautical branches, and by 1939, the AIR was disestablished (Ferreiro, 2014).

Conclusions and Recommendations for Further Research

In summary, the reason that development cost for an aircraft is one to two orders of magnitude greater than for ships is primarily due of the extensive use of full-scale prototyping in the aircraft industry for verification and validation. This does not reflect any inherent differences in the two platforms in terms of safety, production numbers, or complexity, but rather it reflects the fact that, even in the 21st century, shipbuilding remains a product of 19th century rule-of-thumb engineering, while aircraft development is the product of 20th century physics-based engineering. Engineering culture, more than the technology itself, is very difficult to change.

Although these cultures are entrenched throughout both industries, it does not mean that change is impossible. Full-scale prototyping, as part of the verification and validation toolkit employed by shipbuilders, can and should be investigated as a through-life-cost benefit (GAO, 2017).

Figure 8. (left) Collision Damage to USS *John S. McCain*, August 2017; (center) Damen Shipyard Full-Scale Test (1998) of Collision-Resistant Ship Structure; (right) Structure Intact After Collision

Such an approach should be looked at in terms of payoff of the initial investment compared with life-cycle improvements to performance and safety. In addition to reducing the teething problems inherent in first-of-class ships, it would also permit the development and validation of systems to protect the vessel and its crew. For example, a recent spate of ship-to-ship collisions, such as the ramming of the destroyer USS *John S. McCain* by a bulbous-bow-fitted tanker in August 2017, has resulted in the loss of lives, property, and combat availability. Such losses might be avoided in future by carrying out full-scale prototyping of collision-resistant systems as were carried out by the Dutch shipyard Damen in 1998 (Figure 8), which demonstrated that a novel structural configuration could absorb the impact of a colliding bulbous-bow tanker with no hull penetration (Ferreiro, 2002).

The U.S. Navy took the lead in scientific experimentation in the early 20th century by funding and constructing model test basins, at a time when the shipbuilding industry was firmly against it. The Navy can once again take the lead by investigating development practices more like those of the aircraft industry, especially in terms of full-scale prototyping of ships to verify and validate the performance of safety-critical systems.

References

Battershell, A. L. The DoD C-17 versus the Boeing 777: *A comparison of acquisition and development.* Washington, DC: National Defense University.

Defense Acquisition Management Information Retrieval (DAMIR). (1997). *C-17A Globemaster III Selected Acquisition Report.*

Defense Acquisition Management Information Retrieval (DAMIR). (2011*). LEWIS and CLARK class dry cargo/ammunition ship T-AKE 1 Selected Acquisition Report.*

Drezner, J. A., Arena, M. V., McKernan, M., Murphy, R., & Riposo, J. (2011). *Are ships different? Policies and procedures for the acquisition of ship programs*. Arlington, VA: RAND.

Ferreiro, L. (2002). ONR Global Newsletter Number 2: Damage-tolerant structural research in the Netherlands. Arlington, VA: Office of Naval Research Global.

Ferreiro, L. (2014). The mutual influence of aircraft aerodynamics and ship hydrodynamics in theory and experiment. *Archive for History of Exact Sciences, 68*(2), 241–263.

Ferreiro, L. (2016). The warship since the end of the Cold War. In D. Moran & J. A. Russell (Eds.), *Maritime strategy and global order: In markets, resources, security.* Washington, DC: Georgetown University Press.

GAO. (1991). *Cost and complexity of the C-17 Aircraft Research and Development Program* (GAO/NSIAD-91-5). Washington, DC: Author.

GAO. (2017). *Weapon systems: Prototyping has benefited acquisition programs, but more can be done to support innovation initiatives* (GAO-17-309). Washington, DC: Author.

Johnson, C. L., II. (1986). *Acquisition of the C-17 aircraft: An historical account* (Thesis). Montgomery, AL: Maxwell Air Force Base, Air Command and Staff College.

Kennedy, B. R. (2004). *Globemaster III: Acquiring the C-17.* Scott Air Force Base, IL:

Air Mobility Command Office of History.

Naval Sea Systems Command (NAVSEA). (2017). Research, development and design costs for the T-AKE 1 [Information prepared for the author].

Naval Surface Warfare Center Carderock Division (NSWC-CD). (2018). Basin estimates [Information prepared for the author].

Complexity in an Unexpected Place: Quantities in Selected Acquisition Reports

Gregory A. Davis—has worked at the Institute for Defense Analyses (IDA) since 2006, conducting research on as broad a range of topics as he can find. Before coming to IDA, he was an AAAS Science and Technology Policy Fellow in OSD(PA&E). Most of his recent work has focused on major defense acquisition programs that have experienced cost growth. Except for GPS-OCX, all of them had complicated definitions of a unit. [gdavis@ida.org]

Margaret L. Giles—joined IDA's Cost Analysis and Research Division in 2016 after spending five years as a civilian with Headquarters, Department of the Army. As an Army civilian, she participated in the development of multiple Army programs through several budget and POM cycles, focusing in particular on finding greater efficiencies within the Army budget. Giles holds a master's degree in international affairs from the George Washington University and a bachelor's degree in political science from Southwestern University. [mgiles@ida.org]

David M. Tate—joined the research staff of IDA's Cost Analysis and Research Division in 2000. Since then, he has worked on a wide variety of resource analysis and quantitative modeling projects related to national security. These include an independent cost estimate of Future Combat Systems' development costs, investigation of apparent inequities in Veterans' Disability Benefit adjudications, and modeling and optimization of resource-constrained acquisition portfolios. Tate holds bachelor's degrees in philosophy and mathematical sciences from the Johns Hopkins University, and MS and PhD degrees in operations research from Cornell University. [dtate@ida.org]

Abstract

We have looked at the definition of units in numerous acquisition programs and discovered that the units reported are almost never simple; in some programs, no two units are the same, and almost invariably the units produced at the end of a long production run are substantially different from the early ones. We have identified three reasons why the units may differ. The first reason is changes over time, generally as system capabilities are improved. The second is due to mixed types, where units that are inherently dissimilar—such as CH-47F and MH-47G helicopters—are produced by the same program and each is called one unit. The final reason why units can differ is reporting accidents. We give examples of all three and discuss possible methods of improving the reporting requirement

Introduction

Acquisition data are primarily about a few questions: "How much funding?," "How much are we getting?," "When are we obligating the funds?," and "When are we getting what we paid for?" All of these questions are interesting, and none are straightforward. Most have been addressed elsewhere and continue to get attention. The question of "What are we getting?," however, is generally treated as though it were simple. Our experience tells us that counting quantities is often not straightforward. This report describes the findings of research that has taken us deeper into this question, showing that quantities are almost always complicated.

The Director of Performance Assessments and Root Cause Analyses (D,PARCA),[1] asked the Institute for Defense Analyses (IDA) to review the quality and utility of data used for acquisition oversight; we started with the question of quantities.

Selected Acquisition Reports

Section 2432 of Title 10 U.S.C. requires the Secretary of Defense to submit to the Congress a yearly status report for each Major Defense Acquisition Program (MDAP), known as the Selected Acquisition Report (SAR), which provides performance, schedule, and cost data. Each SAR includes separate cost estimates for several categories. Both past actual costs and future anticipated costs are reported, as well as quantity of units, for the expected life of the program (DoD, 2016).

Within the Defense Acquisition Management Information Retrieval (DAMIR) system—the repository for SAR data—the *Track to Budget* section identifies the budget program elements (PEs) and procurement line item numbers (LINs) for each appropriation associated with a program in a particular fiscal year, allowing the user to find the equivalent cost and quantity data in the President's Budget (PB) Submission prepared in the same year.[2] Reconciling SAR data with the equivalent PB Submission proves difficult, however, as cost estimates can vary between the two sources, and some PEs and LINs are shared among multiple programs in a non-transparent way. In some cases, the SAR and PB define quantities differently.

Neither the PB nor the SAR is perfect. In general, the Justification Books that the Services produce annually to support the PB contain more detail, which is good for analysis, but if it extends beyond the Future Years Defense Program (FYDP), it is as a single column labeled *To Complete*. The PB also does not include much history, with most of it in a single column labeled *Prior Years*. The SAR reports costs in both Then Year (TY) and Base Year (BY) dollars, while the PB reports exclusively TY dollars. The SARs are the Office of the Secretary of Defense (OSD)'s primary data source for analyzing MDAPs. This dataset is what analysts from many different organizations typically use, per the recommendation of the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]) staff, who describe SAR data as "the official numbers."

Why Selected Acquisition Reports Matter

The SARs are not the dataset used most often for decision-making inside the DoD. When senior leaders make large resource decisions, analysts most often assemble datasets to suit the needs of the decision-maker by pulling data from non-public systems or conducting data calls. Why then do we care about the quality of data in the SARs?

The SARs matter for two reasons: *triggering* and research. What we call *triggering* is why the SARs were created. The Services trigger investigations when they seek milestone

---

[1] PARCA is an office that was under the aegis of the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]).

[2] The President's Budget and annual SAR submissions both generally come out in the second quarter of each fiscal year. The years on a matching budget and SAR set differ by two. The budget is named for the year ahead, and the SAR is a snapshot of the program in the recent past. For example, in the second quarter of FY 2016, the FY 2017 budget was released, quickly followed by the December 2015 SAR.

authorities from the OSD. The OSD can also trigger analyses for program reviews based on the Service's annual submissions, such as the Program Objective Memorandum. Only the SARs provide regular information at the program level. For example, no other annual submission can tell OSD or the Congress about the projected procurement costs for a program that is expected to leave the development stage in five years.

Research on defense acquisition is continuously occurring in government agencies, think tanks, universities, and other organizations. In the past, researchers looking across programs have considered amount of cost growth (McNicol, 2004), setting of production rates (Rogerson, 1991), comparisons among different commodity types (Drezner, 2011), and many other subjects. This research helps the government, and SARs are the best source for comparisons across programs. While it is the nature of research that we cannot predict which research projects will yield fruitful results, we know that better quality data will yield better research results.

Nunn-McCurdy Breaches

Critical Nunn-McCurdy (N-M) breaches are established by statute. If an MDAP sustains too much cost growth, a review takes place that generally leads to either changes in the program or, occasionally, termination. PMs generally want to avoid N M breaches. Too much cost growth is defined in terms of Average Procurement Unit Cost (APUC) or Program Acquisition Unit Cost (PAUC):

- APUC = Procurement Costs/Procurement Quantities
- PAUC = Total Program Costs/(Procurement + Research, Development, Test, and Evaluation [RDT&E] Quantities)

There are four possible critical N–M breaches, two for APUC and two for PAUC. The breach calculation is performed by measuring the percentage growth in APUC or PAUC. A critical breach occurs when the variable has increased by at least 25% against the current Acquisition Program Baseline (APB) or 50% against the original APB. The original APB is the APB that was established during the Milestone (MS) B decision (formerly Milestone II).

Each SAR contains a unit cost report that compares the current APUC and PAUC estimates to the original APB and a second unit cost report comparing the estimates to the current APB if the current APB is not the same as the original one.

Subprograms

An MDAP's baseline may indicate that it has multiple subprograms to increase visibility into the program's activities. If so, each unit produced and each dollar spent is assigned to one of the subprograms. Subprograms have been used to distinguish variants of a system, such as two similar but not identical missiles, or to look at different parts of a system, such as engines and airframes. Each year, each subprogram has its APUC and PAUC calculated and compared to the baseline. According to the N-M Act (10 U.S.C. § 2433), if any subprogram exceeds its thresholds, an N-M breach is declared for the entire program, not just the subprogram that exceeded its baseline.

The popularity of subprograms has changed through the years, as can be seen in Figure 1. The total number of programs each year did not change much, but declaring subprograms became less common from 1998 to 2009, when a rebound started. It is not clear what has caused these changes.

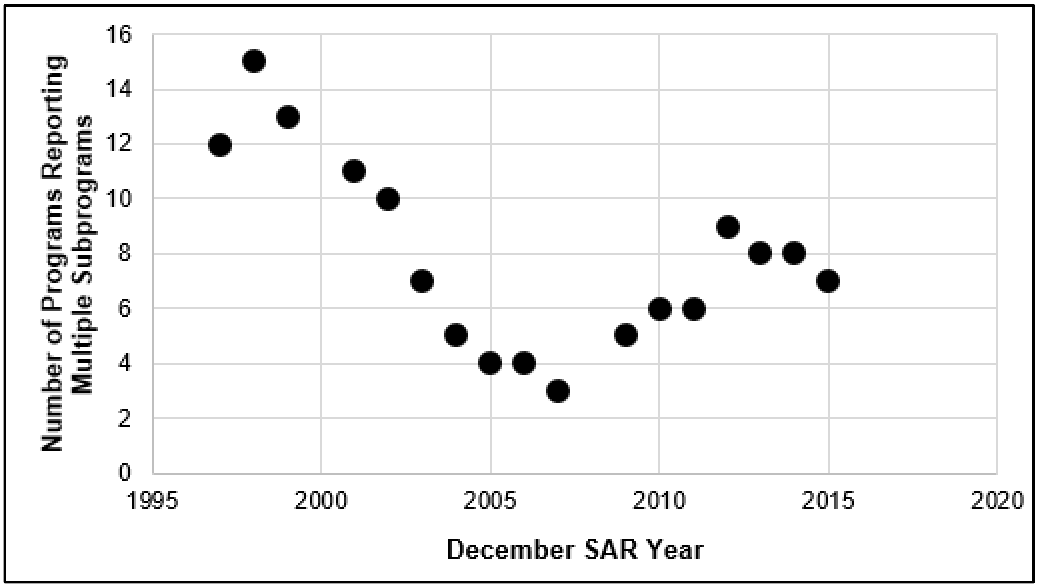


Figure 1. Subprograms in MDAPs 1997–2015

An Example of Budget and SAR Discrepancy: Gray Eagle

Quantity reporting in the SAR is the focus of this report. We begin with a few illustrative examples. The Army's MQ-1C Gray Eagle program acquires unmanned aerial systems (UASs). In the *Track to Budget* section of its 2015 SAR, the program identifies the following LINs within the *Aircraft Procurement*, *Army* appropriation:

- A00005 (MQ-1 UAS);
- A01001 (MQ-1 Payload, which includes funding for other programs); and
- A01005 (Common Sensor Payload Full Motion Video (CSP FMV), a sub-Line Item Number to A01001).

Both A01001 and A01005 are listed as *shared*. The quantities and costs found in these LINs in PB 2017, however, differ from those in Gray Eagle's 2015 SAR, as shown in Table 1. Note that both sources project the program to finish in FY 2018.

Table 1. Gray Eagle Program SAR and Budget Comparison

	FY 2015		FY 2016		FY 2017		FY 2018	
Data Source	Q	Cost (TY $)	Q	Cost (TY $)	Q	Cost (TY $)	Q	Cost (TY $)
PB 2017	19	$246,490 K	17	$355,445 K	0	$60,117 K	0	$10,806 K
Dec 2015 SAR	2	$246,400 K	3	$322,200 K	0	$60,200 K	0	$15,200 K
Difference	**17**	**$90 K**	**14**	**$33,245 K**	**0**	**-$83 K**	**0**	**-$4,394 K**

Note: K – thousand; Q – quantity.

The cost differences in FY 2015 and FY 2017 are minimal, but there is no obvious explanation for the more significant differences in FY 2016 and FY 2018 costs. In PB 2017, the unit of accounting for this program is one unmanned airplane. However, the capability is also dependent on how many ground assets for operating the systems are acquired and on the differences between aircraft, as they are not all the same. In the SAR, the quantity is measured in companies, each of which contains several aircraft with different configurations and some amount of ground equipment. There is a standard measure for what a company is, but not all companies fit the standard description. While the SAR does include a great deal of detail in various written sections, this makes it difficult to use the quantities in the data for quantitative analysis.

A Complex Example: The CH-47F Chinook Program

The Army's CH-47F Improved Cargo Helicopter program demonstrates challenges that can occur when counting quantities across years in both the PB and SAR. This program builds Chinook helicopters, which are easy to count, yet there are serious questions when looking at the data.

First, the CH-47F program's definition of one *unit* has changed over time. In the early days of the program (as reflected in the original June 1998 SAR), the plan was to SLEP[3] 300 existing CH-47D helicopters to an updated configuration, which would be called the CH-47F. In PB 2005, the plan was to SLEP 287 CH-47D helicopters to the CH 47F configuration and 50 MH-47E Special Operations helicopters to a new MH-47G configuration. The definition of a unit had changed to include both CH-47D/F conversions and MH-47E/G conversions, which produce distinct end items and have different expected costs.

The Army's February 2007 budget justification forms expanded the set of planned activities to include all of the following:

- SLEPs of CH-47D to CH-47F
- SLEPs of MH-47E to MH-47G
- New builds of CH-47F from scratch for active duty Army units
- New builds of CH-47F in a different configuration for National Guard units

The reported and projected unit costs for these activities were all different. More to the point, the definition of a unit now included not only a remanufactured existing helicopter, but also a newly built helicopter of the same design. While these may be functionally identical from an operational point of view, there are reasons why an analyst would want to know how many of each were to be built—and at what cost. To further complicate matters, the helicopters produced (both SLEP and new build) employ a mix of mission subsystems, some of which could be repurposed from a remanufactured helicopter or other existing decommissioned helicopter, and some of which had to be built (and purchased) new. The type and number of repurposed subsystems continued to vary from year to year, so that the production inputs (and price) even for new build active component CH-47Fs were different from year to year.

The end result of these changes is that any given unit produced by the CH-47F program might have any one of the MH-47G, CH-47F Army, or CH-47F National Guard configurations. A CH-47F unit might be remanufactured or built new. Whether remanufactured or new, it might include some unspecified mix of government-furnished (free) and contractor-furnished (at a price) mission subsystems. For example, as of the 2013 PB submission (February 2012), 43 new build units had been produced at an average cost of $15 million, of which $1.1 million per unit was for government-furnished equipment (GFE).

---

[3] SLEP is the acronym for "Service Life Extension Program" and is often used as a verb in defense circles. A SLEP can be funded with either procurement or Operations and Maintenance dollars.

The estimated cost to complete the new build program was $2.19 billion for 112 units, or $19.5 million per unit, of which $2.4 million per unit was expected to be GFE. This reflects the expectation that units authorized through FY 2013 would use recovered avionics suites from existing aircraft, but that half of the new build units after that would require new-build (contractor-furnished) avionics. There were clearly anticipated differences in components and cost between units produced up to that point and units expected to be produced in the future.

Furthermore, there are inconsistencies between the SAR and the PB submissions regarding which units comprise the CH-47F program. How new builds versus SLEPs are counted in different years is unusual and is described in detail in the Accidents chapter.

Organization of This Paper

We have divided the common differences among unit definitions into three buckets: changes over time, mixed types of units, and reporting accidents. It is not uncommon for more than one category to apply to a given program; the Chinook has all three. The next three chapters describe what each of these categories means, how confusions arise, and what analysts should do when trying to use cost reporting data. In Suggested Adjustments to Reporting, we make some modest recommendations for modifications to acquisition data reporting that could help make the data more useful for many sorts of analyses. As part of those recommendations, in A Thought Experiment: JLTV, we consider the Joint Light Tactical Vehicle (JLTV) program—how its reporting might have been done differently and what the ramifications of those differences might have been.

Changes Over Time

Implicit in the concept of a *unit* is that every instance of the unit should be identical. Every inch should be the same length, every second should have the same duration, and every run scored in a baseball game should count equally. As noted above, this is often not true of procurement units in MDAPs. One reason that non-identical units might arise is that the product may evolve over time. Even when counting quantities is simple, such as when counting helicopters or ships, the units procured at different times are usually different in both cost and capability. In our full report (Davis, Giles, & Tate, 2017), we detail changes over time in ships, tactical aircraft, and tactical land vehicles. In this excerpted report, we look only at one program, the Air Force's AIM-120 Advanced Medium Range Air-to-Air Missile (AMRAAM) program.

The AMRAAM program was established at a Defense Systems Acquisition Review Council MS I Review in November 1978. After an extended development period, an acquisition baseline of 24,320 units was set in December 1988. The first production units were authorized under the FY 1987 budget and fielded in 1991. The acquisition target was reduced to 16,427 missiles in a 1992 re-baselining that also doubled the expected per-unit cost.

The AIM-120 is still in production. The Air Force now intends to buy a total of 12,851 missiles, and the Navy an additional 4,461 missiles, for a total of 17,312. The final unit is projected to be authorized in FY 2025, almost 40 years after the first unit.

The explanation for the continued utility of the AIM-120 is that the missiles being produced today are nothing like the missiles that were produced in the early 1990s. Figure 2 shows the history of average unit cost by annual production lot for AMRAAM missiles, with filled shapes showing historical data and open shapes, projections. After a typical initial learning curve, it is clear that there have been major changes to the program over its history. In fact, many upgrades, modifications, and wholesale redesigns of the missile have occurred over time; the Teal Group reports seven (Teal Group Corporation, 2014, p. 133). Some were simply improvements, while others had new functions, such as the AIM-120C3, designed with smaller control surfaces to fit inside the weapons bay of an F-22 Raptor, and the AIM 120D, which includes many new features such as Global Positioning System navigation and a two-way datalink.

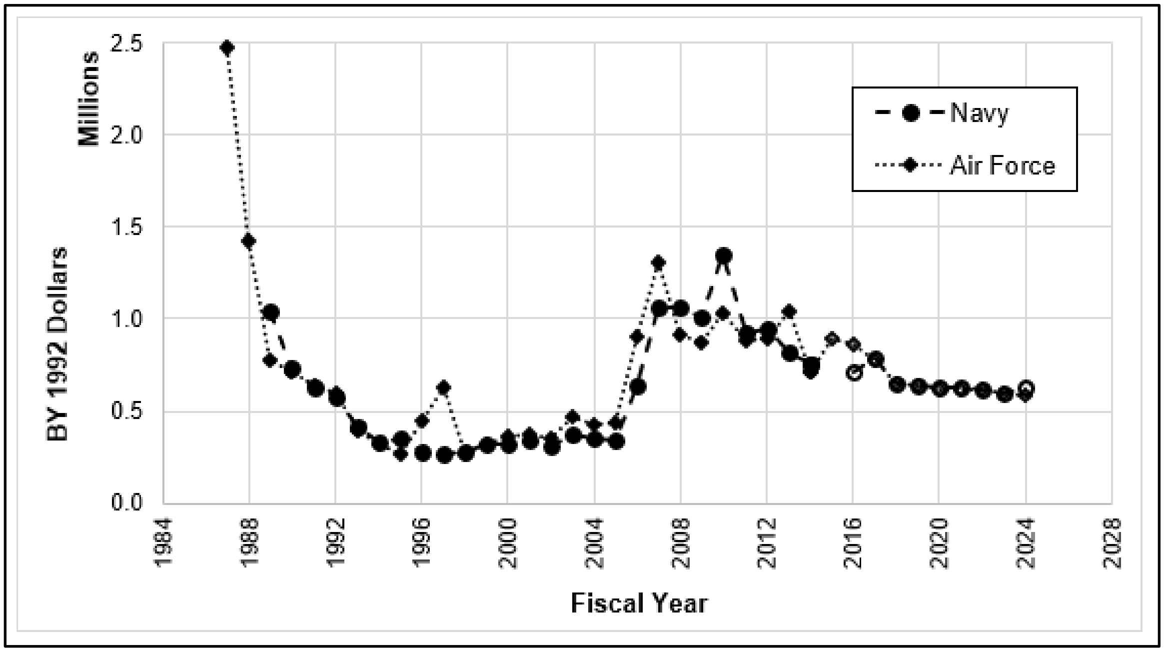


Figure 2. Annual Average Unit Cost for AMRAAM Missiles

There is no sense in which an AIM-120D is "the same thing" as an AIM-120A, or even an AIM-120C7. This is a clear instance in which the implicit assumption that units are interchangeable has been violated.

Of course, within the AMRAAM program, there is no confusion about the kinds of missiles that are currently being produced, their capabilities, or plans for future improvements. The question, then, is how the program might adjust its data reporting to enhance transparency for planners, analysts, and oversight bodies.

Mixed Types

Program offices often procure different end items at the same time. These items are usually similar to each other but substantially different; yet, for quantity reporting purposes, each is considered one *unit*. This often comes about because of different missions or end users. Sometimes, the types are completely different. To illustrate this concept, in this paper we look at an electronics suite.

The Navy's Integrated Defensive Electronic Countermeasures (IDECM) program acquires electronics suites to protect the various F/A-18 aircraft from radio frequency guided missiles. IDECM achieved MS II approval in October 1995, although it was too small at the time to be an MDAP. Because of changes, it became an MDAP in March 2008 and issued its first SAR in June 2008. The *Mission & Description* section of the December 2015 SAR describes the blocks as follows:

- IDECM Block 1: A federated suite, consisting of the ALQ-165 On-Board Jammer (OBJ) and ALE-50 expendable decoy
- IDECM Block 2: An integrated suite, consisting of the ALQ-214 OBJ and ALE-50 expendable decoy
- IDECM Block 3: An integrated suite, consisting of the ALQ-214 OBJ and ALE-55 Fiber Optic Towed Decoy
- IDECM Block 4: A Hardware Engineering Change Proposal to the ALQ-214 OBJ to render it suitable for operation on F/A-18C/D aircraft, while retaining all functionality, when installed on F/A-18E/F

The SAR contains two subprograms: IDECM Blocks 2/3 and IDECM Block 4. The December 2015 SAR reports an APUC of $2.502 million for Block 4 and a far lower APUC of $0.090 million for Block 2/3. This is because the quantities are so different. Block 4 has a quantity of 324, roughly the number of airplanes they will be protecting. Block 2/3 has a quantity of 12,805, although the Navy bought fewer than 600 F/A-18E/Fs in total. Eighty-five of the 12,805 were purchased with *1506 Navy Aircraft Procurement* funds and the balance were or will be bought with *1508 Procurement of Ammunition, Navy and Marine Corps* funds. We presume that those purchased with ammunition funding are only the disposable decoys. The unit costs based on the *End Item Recurring Flyaway* column in each year are presented in Figure 3.

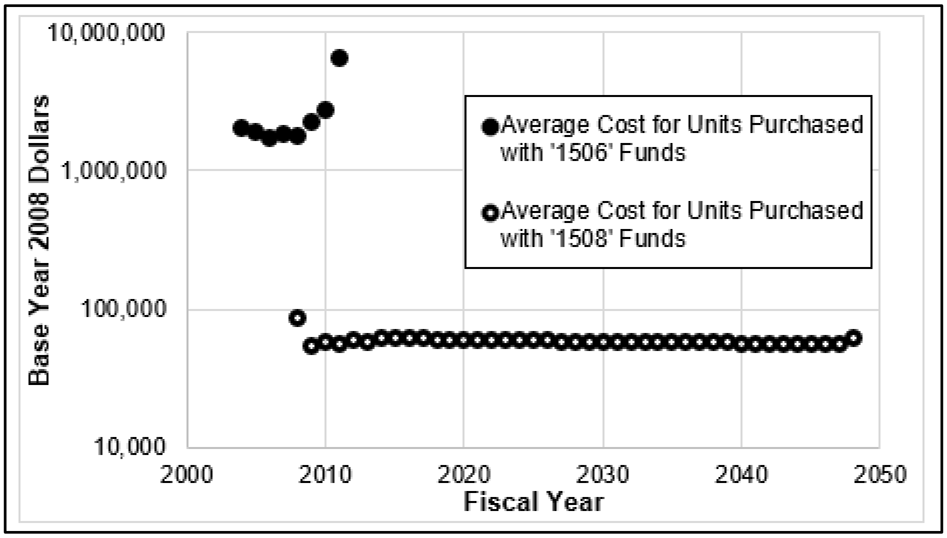


Figure 3. IDECM Block 2/3 Annual Unit Cost by Appropriation Type From the December 2015 SAR

Even though showing the two unit costs on the same chart requires plotting them on a logarithmic scale, the two are both considered *units* for the official unit cost calculation. Just within the more expensive *1506* units, it is clear that there have been significant changes, as the cost there does not follow a typical learning shape, which would be expected to slope down.

While the details have changed with time, the IDECM program has used this reporting system since it issued its first SAR in June 2008.

Accidents

The confusions described above generally come about because of some decision by leadership about how the data should be presented;[4] this category, in contrast, is about cases in which it seems there were also outright errors in how the quantity numbers were put together. We do not know how frequently this happens, but we know that it does happen and can sometimes persist for several years. We do not suggest that any of the cases described below involve intent to confuse analysts, but they did have that effect.

We used the term *accidents* (as opposed to *mistakes* or *errors*) because it was the term a government official in AT&L applied to reporting anomalies for programs like Chinook. We identified three in the December 2015 SARs: Chinook helicopters, the Evolved Expendable Launch Vehicle, and the ICBM Fuze modification programs. It is quite possible there are more. We only present the Chinook situation in detail, as the others were about how dollars were assigned within the program.

[4] Or, more precisely, leadership makes a decision about how the program should be managed and what systems it should produce, possibly without considering the impact this will have on the coherence and consistency of quantity or unit cost reporting.

As described in the section titled A Complex Example: The CH-47F Chinook Program, the CH-47F Chinook Improved Cargo Helicopter program made a number of changes to its definition of *unit* over the course of the program. In the December 2015 SAR, however, the program apparently lost track of how it had been defining a unit and submitted quantity and cost forecasts that did not include all of the units identified in the simultaneous PB submission.

Figure 4 shows the discrepancy between predicted future quantities in the December 2015 SAR and the corresponding 2017 PB. Through FY 2017, the total quantities match perfectly. Beginning in 2018, units described as SLEP units in the PB are missing from the SAR forecast. As a result, the projected cost of these units is not included in the SAR calculations of APUC, PAUC, APUC growth, or PAUC growth.

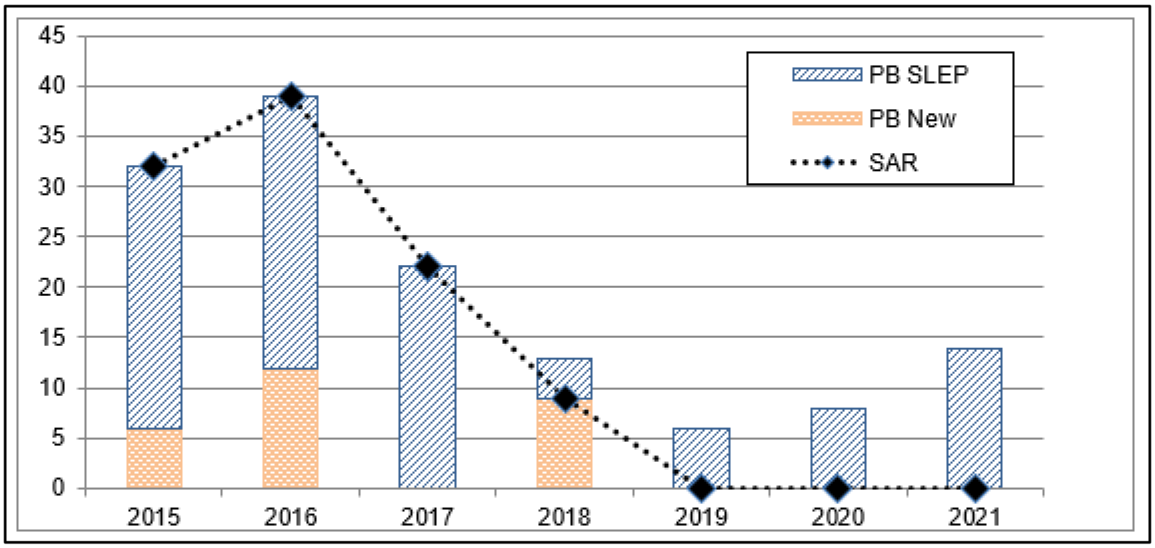


Figure 4. SAR Versus PB Production Quantities for CH-47F

Since the SAR and the concurrent PB are required by law to agree on costs and quantities within the FYDP, this is clearly an accident.

There is also an ongoing mismatch between the SAR and PB with regard to the past quantity produced. In the SAR, every past unit produced is counted, regardless of whether it was a SLEP unit or a new build. In the PB, in the early years of the program, there were no top-level quantities reported, presumably on the basis that upgrades to CH 47F configuration were just one of many ongoing upgrades in the Army's helicopter fleet. Typically, programs that perform multiple types of upgrades, but are not applying all of them to every legacy platform, report the number of each type of upgrade performed separately. They do not typically roll these up into a total quantity for the program's LIN because the individual upgrades are not comparable and the number of platforms modified does not match the total number of any one type of modification.

When the decision was made to build CH-47F helicopters as new builds, the program began reporting a total quantity of units at the line item level and chose to include both SLEP and new build units in this total. However, they never looked back to include previously produced SLEP units in the *Prior Quantity* total. As a result, each new SAR and PB submission disagree both on how many helicopters have been procured and on how many will have been procured in total when the program is finished.

Suggested Adjustments to Reporting

One possible response to the issues described previously is to tell programs never to change what they are buying; once the baseline is set and the program is approved, the plan should be followed and the systems should not change. This assumption is implicit in the data reporting process. And yet, this has never been government practice, and we do *not* recommend that it be adopted. Our military goes to great lengths to provide our warfighters with the best possible equipment, and we should not forbid that just to make bookkeeping easier. We do offer some modest proposals that could make the reported data more useful, but first we need to be careful about incentives.

Data and Incentives

Data recording systems provide incentives. “You get what you pay for” is an old expression. In 2007, Dr. H. Thomas Johnson wrote, “Perhaps what you measure is what you get. More likely, what you measure is *all* you get” (2007). If the acquisition system’s data requirements are not aligned with the system’s goals, suboptimal performance may follow. This is exacerbated when penalties are associated with data reporting. Generally, people would prefer to report accurate data, but when the data will be used to justify punishment, the reporters are incentivized to either change the facts that lead to the data—possibly in creative and unproductive ways—or to provide incorrect data.

The statute that defines the N-M breach specifies PAUC and APUC thresholds that influence program behavior. Since N-M reviews impose costs on programs and can trigger cancellation of a program, many people in defense acquisition, including program managers (PMs), try to avoid them. This likely accounts for some of what we see in data reporting today.

Any changes made to the system need to be considered in this light. If people’s careers will depend on what data they report, at times those data are more likely to reflect what is needed to satisfy the checker rather than reality. Furthermore, people will bend reality to make the data look “right” even if that will not yield the best actual result for national security.

Monitoring Changes Over Time

If we accept that the units produced during the course of a program will change over time, PMs should be given useful and standardized ways to describe (and ideally quantify) those changes, both for past units produced and planned future production.

The current taxonomy of SAR Variance Categories recognizes seven possible reasons for cost growth. Cost growth due to design changes must always be categorized as “Engineering,” lumping together planned and unplanned changes, as well as optional versus necessary changes. For oversight and analysis, it would be useful to be able to distinguish at least three sub-categories of “changes over time”:

- Pre-planned product improvements (P3I),
- Unplanned changes (necessary and unnecessary), and
- Block upgrades or evolutionary acquisition.

Pre-Planned Product Improvements (P3I)

P3I is a form of *spiral* acquisition, in which the first units produced do not include all of the capabilities that the procuring Service has identified as being required. The reasons for delaying might be budgetary, technical, operational, or some combination. The key is that the program has a plan from the beginning to add specific known improvements and has developed cost and schedule estimates for those improvements. This allows P3I costs to be included in the SAR and other program submissions.

In the current SAR, or even the more detailed PB, it is difficult to report current or anticipated P3I costs in a transparent fashion. The additional costs beyond what the program would cost if the improvements were not made will be a mix of RDT&E costs (for developing and testing the new design), nonrecurring costs (for things like new documentation and tooling), end-item recurring flyaway (EIRF) costs (for actual production of the improved units), possible non end-item recurring flyaway (NEIRF) costs (if improvements are made to non-end-item systems), and support costs (if the cost of support and/or spares for the new design is not exactly the same as for the original design).

For the marginal cost of improvements to be visible in the SAR, reporting would need to explicitly include P3I costs. One way to do this would be as follows:

- If the planned improvements are small in number and to be done at a few discrete times during the production run, treat them like Block Upgrades (see Block Upgrades, Evolutionary Acquisition, and Agile Development section).
- If the planned upgrades are more numerous and continuous, establish a Planned Upgrades subprogram and report the RDT&E and Procurement costs associated with planned changes to the original design under that subprogram. For each year in the SAR Annual Funding report, the program should report the following:
    - Under the main end item subprogram, report the quantity produced or planned and the estimated cost if those units had been made to the original design.
    - Under the Planned Upgrades subprogram, report zero quantity and the additional marginal procurement cost for the lot due to design changes. This additional cost should be split among EIRF, nonrecurring, and support costs in the usual way.
    - Report RDT&E costs for the original design under the primary end item subprogram.
    - Report RDT&E costs associated with planned design changes in the Planned Upgrades subprogram.

This system would allow analysts to clearly understand how much of the price change over time was driven by planned improvements and how much was unexpected. It would support meaningful learning curve modeling and also provide some progress tracking of new capability insertions. The narrative portions of the SAR would describe the capability enhancements obtained to date, the plan for future insertion of new capabilities, and the unexpected changes made to the base program.

On the other hand, this system introduces a potentially onerous new type of reporting—namely, the hypothetical cost of the units if they had all been made to the original design. This is not information programs currently possess, and there are potential pitfalls and perverse incentives in how programs might choose to compute and report these counterfactual costs. In particular, cost growth due to design changes that might have been necessary in the base program (e.g., for safety reasons, to meet threshold requirements, or due to diminishing manufacturing sources) could be allocated either to the base subprogram or to the P3I subprogram, whichever seemed least likely to risk an N-M breach.

For N-M purposes, several regulatory changes might be beneficial. First, the primary end item and the Planned Upgrades should be treated as separate triggers. The primary end item would use the usual PAUC and APUC thresholds. The Planned Upgrades subprogram might only have limits based on total cost growth, or perhaps time-phased cost growth (e.g., average cost per year, rather than average cost per unit). Ideally, a breach on the Planned Upgrades subprogram would *not* imply a breach on the base subprogram (although the reverse would not be true).

Under this system, the main temptation for struggling programs would be to mischaracterize some of their core program cost growth as P3I, so as to avoid an N-M breach on the primary end item. By shedding planned improvements, the program could avoid having an N-M breach on either subprogram. This is not necessarily a bad thing. The oversight challenge would be to align operational test criteria with the phased capabilities to be produced.

Unplanned Changes

It is not uncommon for systems already in production to incorporate significant design changes that were not foreseen by the program. Reasons for this can include urgent operational needs from the field, correction of defects discovered post-fielding, implementation of Value Engineering proposals, or response to changes in the adversary/threat environment.

It is clearly unreasonable to require programs to report things they are not yet planning to do. For unplanned changes, the challenge is how to report them as they are discovered and after the fact, in ways that transparently describe the reasons for any corresponding cost and schedule changes.

It would be ideal if SAR reporting of unplanned changes distinguished clearly between design changes driven by new performance requirements and changes required to meet the original program requirements. One possible way to accomplish this would be to add a new category, “Requirements,” to the list of SAR variance categories. Cost changes due to design changes required to meet original program requirements (as of the current APB) would be classified as engineering variances. Cost changes due to new or modified performance requirements would be classified as requirements changes. For a program with a P3I subprogram, the base program and P3I subprogram would report separate cost variances, using the new category where appropriate.

Unfortunately, it is unlikely that programs would report these categories accurately. Not only are there strong incentives to categorize all cost growth as being due to new requirements, but there is often genuine confusion within the program office about which requirements are part of the baseline and which have been added during the course of development and production. In theory, the Cost Analysis Requirements Document and other mandatory acquisition documents establish the baseline requirements assumed by the baseline cost estimate. In practice, this is not as clear, especially for programs that have been re-baselined at some point.

Block Upgrades, Evolutionary Acquisition, and Agile Development

Some programs know in advance that they intend to upgrade or replace the initial design with an improved future design, but do not yet know what those changes will be or what they will cost. They may not know which attributes will be enhanced, since that decision will be based on developments in the future. If multiple changes are made to the weapon system design at a few discrete points in time, these are often termed *block upgrades*. If many changes are made on an ongoing basis as their usefulness becomes known, this is sometimes referred to as *evolutionary acquisition*. In the special case of software programs doing repeated rapid insertion of new features in close collaboration with the users of the software, it is called *agile development*.

In each of these cases, the reporting challenge is that the program knows that they intend to spend money in the future, but they do not know what they will be spending it on, what it will cost, or when it will happen. The challenges for oversight and management are obvious—especially when a program being managed in this way is shoehorned into a reporting system designed for unchanging units. This is part of what happened to the RQ 4B Global Hawk program, which was intended from the beginning as an evolutionary acquisition but was required to guess both content and schedule of future upgrades as part of its original acquisition baseline. Those guesses were then treated as firm requirements by the acquisition system, even after Air Force leaders had changed their minds about both priorities and threshold performance.

In the case of block upgrades, one possibility is to simply declare a new program for each block. This is the approach taken by the AIM-9, AIM-9X, AIM-9X Block II missile programs; the F/A-18C/D and F/A-18E/F fighter aircraft programs; and the UH-60L and UH-60M Blackhawk helicopter programs (among many others).

Other programs have treated successive blocks as distinct official subprograms. This approach was taken by the Joint Air-to-Surface Standoff Missile (JASSM) program. The original program had no subprograms and developed the AGM-158 missile. During that development, the Air Force studied possible improvements to the missile and decided to develop a second variant with longer range. The original AGM-158 was redesignated AGM-158A, and the new "JASSM-ER" (Extended Range) was designated AGM-158B. The program was split into two subprograms for reporting purposes, with JASSM-ER schedule, development costs, and production costs (and cost variances) reported separately. The Navy went even further with the new AGM-158C (LRASM) variant, deciding to make it a distinct program[5] rather than creating a new subprogram within the JASSM program. This may be because the new program is Navy-only, while JASSM is an Air Force program.[6]

An advantage of these approaches is that they isolate the unit cost of the new block from the past, rather than computing an average over all past blocks. It would defeat the purpose of the N-M legislation if 50% APUC growth in what is essentially a new weapon system became invisible because it was being averaged together with thousands of past units of completely different design.[7] A second advantage is that the block upgrade is clearly identifiable as design changes to meet new requirements, as opposed to design changes to overcome technical difficulties in achieving the original requirement.

One disadvantage of the subprogram approach, as currently implemented, is that an N-M breach by any block triggers a mandatory review of every subprogram, as is discussed later.

---

[5] PNO 449, "Offensive Anti-Surface Warfare Increment 1 (Long Range Anti-Ship Missile)," abbreviated as "OASuW Inc 1 (LRASM)"

[6] The "J" in JASSM stands for "joint," and at one point there was consideration of mounting this weapon on Navy aircraft. However, that has not happened, and all of the funds in the SAR are reported from Air Force appropriations.

[7] This is what has happened with the AIM-120 AMRAAM program, as described in the Unplanned Changes section.

A disadvantage of both subprograms and separate programs is the difficulty of accounting for shared RDT&E, nonrecurring, and support costs, such as for testing equipment or software that is used by multiple blocks. For example, the RQ-4B Global Hawk family all uses a common ground station. If this program had used separate subprograms for each distinct aircraft design, it would be inappropriate for the original RQ 4A subprogram to bear the cost of all upgrades to the ground station systems and software, given that all blocks benefit from those upgrades.

A logical response to this problem would be for the Global Hawk program to make the ground station systems a separate subprogram. The difficulty with this is that it would create the possibility of an N-M breach due to cost growth in a subprogram that accounts for only a small fraction of total program cost. A more reasonable approach would be for programs to be able to declare a single subprogram responsible for procurement of items other than end items. This subprogram would only be liable for an N-M breach if its estimated total cost (RDT&E + Procurement) grew to exceed a threshold percentage of the estimated PAUC for the overall program, which would require new legislation from the Congress.

Possible Methods for Handling Mixed Types

As the examples in Mixed Types show, many solutions have been found to the mixed type problem, but all of them have drawbacks.

Subprograms

For some programs, subprograms have provided an elegant solution. For example, the Army's original Multiple Launch Rocket System (MLRS) program distinguished two subprograms: the mobile rocket launcher and the tactical rocket it would fire. This allowed the program to accurately track unit cost growth for both of the fully configured end items being developed and produced. The launcher was produced within its original cost estimate; the rocket experienced a critical N-M breach.[8] Similarly, the Army's PAC-3 suite of upgrades to the Patriot missile system was (after several schedule breaches in the first few years of development) divided into subprograms for the Missile Segment and the Fire Unit.

The fact that a unit cost breach in any subprogram triggers a breach in the whole program discourages their use, even where it seems like an obvious solution. A program without subprograms often has more leeway to do things that will make the cost growth look smaller. For example, if the MLRS program had not defined subprograms, but had treated the rockets as the end-item units, they would have shown a lower percentage cost growth for the combined program than was seen for just the rocket subprogram. In addition, the program could have decided to produce fewer launchers than originally planned, reducing both PAUC and APUC without changing the official number of *units* being produced. Doing so might have avoided the N-M breach, at the cost of greatly reduced transparency regarding cost growth and reduced capability.

Making subprograms more appealing would require congressional action, possibly in an annual authorization bill, which seems possible if some way to maintain program cost accountability could be devised. The Congress might be willing to allow the Milestone

[8] Unfortunately, the program did not similarly distinguish the variant rockets being produced or the later conversion of the entire system from an unguided rocket launcher to a guided missile launcher.

Decision Authority to designate alternative triggers for programs with subprograms, especially if some of the subprograms involve far fewer dollars than others.

In theory, SAR reporting could be expanded so that each program could report simultaneously on multiple distinct end items without declaring subprograms. The principal distinction between this approach and subprograms would (presumably) be the mechanisms for deciding cost and schedule breaches. As with subprograms, it would be important in implementing this change to avoid creating perverse incentives to PMs. In particular, accurately defining multiple end items should not increase a program's chances of experiencing an N-M breach.

Multiple Programs

If a Service is planning to buy a mix of different end items in response to a given set of mission needs, they have some flexibility in deciding how to group those efforts into programs. It is not always obvious which grouping would best serve the needs of both the Service and the oversight community.

At one (unfortunate) extreme, the Army decided to make Future Combat Systems a single program with literally hundreds of different physical products ranging in size and complexity from light tanks down to man-portable UASs, along with many tens of millions of lines of software implementing communications, mission command, and networked fires. The official units for that program were Brigade Sets, of which 15 were to be produced. A prime "lead systems integrator" contract was awarded, with authority to reconsider the mix and capabilities of systems to be developed and procured in each Brigade Set. This offered no useful insight into the program's activities or progress.

At the other extreme, the Army decided to split procurement of their new AH-64E Apache helicopters into two separate programs, one for remanufactured aircraft and the other for new builds. A 2008 acquisition decision memorandum signed by the Army Acquisition Executive contains the following language.

> As a recently delegated Acquisition Category IC program, the AH-64E Apache program is comprised of two separate programs, the Remanufacture program and the New Build program. Each of these programs are separate and distinct with respect to the Acquisition Program Baselines (APB), and their funding lines; however, they have identical configurations and are produced on the same production line. (Shyu, 2013)

The choice to create two MDAPs creates challenges for both the Army and OSD because it adds extra reviews and recordkeeping. Having multiple programs, as with subprograms, creates two triggers for an N-M breach, but it also means that any breach would affect only one of the two programs, whereas creating two subprograms would expose the entire program. It also splits what naturally feels like one program—indeed, the previous language refers to it both as one program and as two in the same paragraph. Since both programs produce identical new AH-64E helicopters, why should they be separated? Although distinct for reporting purposes, they have common goals and management. They share a PM and a production contract,[9] but only the remanufacture program reports any RDT&E costs. Even within Apache, both programs list "Other Support" funds in their SARs, and since the two programs are producing identical helicopters, it is unclear how the Army decides whether a given support purchase will be credited to one program or the other. One cannot understand what is going on in either program without considering the other, which would seem to violate the notion of what constitutes a program. Where there is only one distinct end item, having multiple programs is questionable.

Multiple programs should only be considered as an option in the case of block upgrades to an existing program (as discussed in the section titled Block Upgrades, Evolutionary Acquisition, and Agile Development), or when the set of things to be procured by a proposed new program involves the following:

- Significantly different product types with different acquisition risks,
- Multiple independent contracts with no real synergies, and
- Few significant interoperability requirements among systems.

In general, it is rarely appropriate to split a new proposal into multiple programs.

An example of a program that perhaps should have been split into multiple programs is the Stryker (originally "Interim Armored Vehicle") program. This program procured eight specialized variants of an existing non-developmental armored vehicle. Of these eight variants, six were relatively minor modifications of the existing design, while two[10] required extensive engineering changes to the original. An appropriate program management strategy would have been to make the six "minor modification" variants a single program (with six subprograms), and the two major redesigns either a second program with two subprograms, or two additional separate programs. That would have isolated the development risks of the two most risky projects from the more straightforward projects and would have given the OSD and the Congress better visibility of how the various projects were progressing. As it happened, the Stryker program experienced a significant (but not critical) N-M breach, driven entirely by problems in the two major redesign vehicles.

Different Cost Categories

Using the different cost categories in current SAR reporting can give some visibility into what is happening in a program, but generally does not allow better identification of different unit types. The distinction between *end items* and *non-end items* was not designed to capture differences among multiple distinct end items.

The Air Force's UAS MQ-9 Reaper program plans to procure 347 units, where each unit is an aircraft. The total procurement cost for the program is $9.2 billion in BY 2008 dollars, but only 52% of that is EIRF. Another 22% is categorized as NEIRF, and the remaining 26% is Total Support. This information is useful for cost analysts, although this

---

[9] The December 2015 SAR for the remanufacture program lists four procurement contracts and two RDT&E contracts. The new build SAR only shows one contract, which is one of the four procurement contracts in the remanufacture program.

[10] The two were the M1128 Mobile Gun System (MGS), which attempted to mount a tank-like 105 mm direct fire cannon on a relatively light wheeled vehicle; and the M1135 Nuclear, Biological, Chemical, Reconnaissance Vehicle (NBC RV), which required a suite of sophisticated environmental sensors and a positive-overpressure internal environment.

distribution has no impact on N-M reporting.[11] The aircraft quantity can be compared to the EIRF to understand those units, but there are no quantities reported for ground stations, so an analyst can only know what has been spent on them in total, not what each costs. In this case, NEIRF is something like a subprogram for the ground stations, but it is less transparent than actual subprograms would be.

Reducing Accidents

When humans carry out activities, accidents are inevitable. Reducing accidents requires good processes. We have not analyzed the process for generating SARs or PB submissions. In principle, that could (and perhaps should) be done from a quality assurance point of view.

We were also told that AT&L/Acquisition Resource and Analysis (ARA) performs the OSD's checks on Service-submitted data, and they do not have enough time to do it thoroughly. All of the draft SARs arrive at the OSD in the same season. About a week after the data arrive, ARA meets with each program for about one hour, at which time ARA can ask questions. They feel that this process is insufficient and clearly there are changes that could reduce the accident rate.

The best way to improve ARA's review is probably not only to add more time. While more time might help, ARA would probably also benefit from specialized tools to help them analyze the draft SAR data and quickly compare them to budget submissions, prior year SARs, and general rules about how acquisition programs typically behave. Proposing improvements to that process is beyond the scope of this report.

A Thought Experiment: JLTV

To illustrate the kind of reporting that would be necessary to improve both oversight and data utility for cost analysts, we looked at the JLTV program. The full analysis is in our full report (Davis et al., 2017). We determined that at the beginning of the program, seven subprograms might be appropriate, as indicated in Table 2.

Table 2. Suggested Initial JLTV Subprograms

Number	Subprogram	Description
1	Utility	Base design Utility Vehicle
2	General Purpose (GPV)	Base design GPV
3	Heavy Guns Carrier (HGC)	Base design HGC
4	Close Combat Weapons Carrier (CCWC)	Base design CCWC
5	P3I – Common	P3I common to all variants
6	P3I – HGC and CCWC	P3I specific to HGC and CCWC
7	Support equipment	Trailers, armor kits, etc.

This would not be practical if an N-M breach could be triggered by any one of them.

[11] One could imagine the Air Force lowering the ratio of ground stations to aircraft, not for operational reasons but rather because they want to control APUC.

Conclusions

The default assumption for any acquisition program is that all of the units it produces are identical and interchangeable. This is seldom true—consider asking an F 35A to land on a ship. Any analysis that assumes interchangeable units is making an unwarranted assumption that can lead to mistaken conclusions. The importance of these mistakes will vary, both with the details of the program and the nature of the analysis. We hope that this work can lead to two kinds of changes: one for analysts using acquisition data, and a second for policy makers defining reporting requirements for programs.

For analysts, the primary message is “Beware.” It is not uncommon for invisible differences between units to be important to an analysis, as we saw with previous IDA studies of hedonic price indices for aircraft and tactical vehicles discussed in the Changes Over Time chapter. Without additional data from non-SAR (and sometimes non-PB) sources, it is often impossible to understand the relationships between price, cost, and quantity in many programs. Such additional data are, unfortunately, not always available. Analysts need to know the limits of what can be inferred from the existing data.

For policy makers, there are many opportunities to improve data reporting requirements and guidance, and these come in three varieties. First, there ought to be explicit acknowledgment that not all units are identical, and some effort should be made to quantify unit-by-unit or lot-by-lot differences for analysis and oversight. Second, the rules need to encourage the desired behaviors. The current N-M rules are an excellent example of how rules incentivize behavior in ways that may be counterproductive. For example, IDECM’s unit costs could be reduced by purchasing more towed decoys. When designing new reporting requirements, policy makers need to keep this in mind. Finally, the quality assurance processes applied to official data ought to be studied and improved. While some accidents are inevitable, the system today probably lets through more than it should. SARs are much like custom manufactured parts. Each one is unique, but good processes could still make them more uniform and useful.

References

10 U.S.C. § 2432. Selected Acquisition Reports. Retrieved March 28, 2017, from http://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title10-section2432&num=0&edition=prelim

10 U.S.C. § 2433. Nunn-McCurdy Act. Amendment to the FY 1982 National Defense Authorization Act. Retrieved from https://www.congress.gov/bill/97th-congress/senate-bill/815

Davis, G. A., Giles, M. L., & Tate, D. M. (2017). *Complexity in an unexpected place: Quantities in selected acquisition reports* (IDA Paper P-8490). Alexandria, VA: Institute for Defense Analyses.

Defense Acquisition University. (n.d.) *Glossary of defense acquisition acronyms and terms*. Defense Acquisition Management Information Retrieval (DAMIR) System. Retrieved from https://dap.dau.mil/glossary/pages/1742.aspx

DoD. (1992). *Cost analysis guidance and procedures* (DoD 5000.4 M). Washington, DC: Assistant Secretary of Defense (Program Analysis and Evaluation). Retrieved from https://biotech.law.lsu.edu/blaw/dodd/corres/pdf/50004m_1292/p50004m.pdf

DoD. (2016). *Department of Defense selected acquisition reports* (Release No. NR-106-16). Retrieved from https://www.defense.gov/News/News-Releases/News-Release-View/Article/703529/department-of-defense-selected-acquisition-reports-sars

Drezner, J. A., Arena, M. V., McKernan, M., Murphy, R., & Riposo, J. (2011). *Are ships different? Policies and procedures for the acquisition of ship programs.* Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/monographs/MG991.html

Feickert, A. (2017). *Joint Light Tactical Vehicle (JLTV): Background and issues for Congress* (CRS Report RS22942). Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/weapons/RS22942.pdf

Johnson, H. T. (2007). Lean dilemma: Choose system principles or management accounting controls—not both. In J. Stenzel (Ed.), *Lean accounting: Best practices for sustainable integration*. New York, NY: John Wiley & Sons.

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (IDA Paper P-3832). Alexandria, VA: Institute for Defense Analyses.

Rogerson, W. P. (1991). Excess capacity in weapons production: An empirical analysis. *Defence Economics, 2*(3), 235–249. doi:10.1080/10430719108404695

Shyu, H. (2013). *Acquisition decision memorandum for the AH-64H Apache new build program* [Memorandum for Program Executive Officer, Aviation]. Washington, DC: Office of the Assistant Secretary of the Army.

Teal Group Corporation. (2014). *World missile briefing*.

Acknowledgments

Thank you first to Dr. Peter Eggan of PARCA for supporting us throughout this project, even as we shifted gears so significantly. We also want to express our gratitude to Kathy Conley, Strider McGregor-Dorsey, and Brian Rieksts, all of IDA, whose technical reviews kept us moving (and complaining), leading to a better final product.

Panel 4. Innovations in Test and Evaluation

Wednesday, May 9, 2018	
9:30 a.m. – 11:00 a.m.	**Chair: James Cooke,** Director, U.S. Army Evaluation Center ***Acquisition Challenges of Autonomous Systems*** David M. Tate, Institute for Defense Analyses David A. Sparrow, Institute for Defense Analyses ***Using Developmental T&E to Inform Operational T&E Decision-Based Analysis*** Dashi Singham, Naval Postgraduate School ***An Empirical Study on Content Analysis Use in Test and Evaluation Deficiency Report Analysis*** Karen Holness, Naval Postgraduate School Rabia H. Khan, Naval Postgraduate School Gary Parker, Naval Postgraduate School

James Cooke—is responsible for the synchronization, integration, and coordination of evaluations of joint weapon systems, where the Army is the DoD executive agent, and of all Army systems. In addition, he is responsible for test design, strategic planning, and the development of methodologies critical to the evaluation of Army and Joint weapon systems.

Cooke is a West Point graduate, a distinguished honor graduate from the Infantry Officers Advanced Course, an honor graduate from the Command and General Staff College, a graduate of the U.S. Army Comptrollership School, a Distinguished Honor Graduate of the Army Logistics Management Program, a graduate of the Army Program Manager course at the Defense Systems Management College, and a graduate of both the Air War College (Correspondence) and the Army War College. He has held military secondary skill identifiers in comptroller, intelligence, engineering, and operations research/systems analysis, and has academic degrees in engineering, general science, and mathematics. He spent nearly 25 years in the military as an airborne and ranger qualified infantry officer and held command and staff positions from platoon through brigade and on COCOM (CINCLANT), Army, Joint, and OSD staffs.

Prior to his current position, Cooke served as the Assistant Deputy Under Secretary of the Army for Test and Evaluation (T&E), where he provided oversight of all T&E capabilities for Major Army Acquisition Category I and II programs and for all DoD Chemical and Biological Defense Program activities. He was the integrator of strategic change and primary agent for the Secretary of the Army to coordinate T&E positions on issues and reports with the other Military departments, the Office of the Secretary of Defense, the Joint Staff, and Congress. Cooke was also responsible for oversight of the T&E Enterprise and Test Ranges; development of the Army T&E Strategic Plan; all Army T&E regulations and policies; and the resources and funding for labor, infrastructure, instrumentation, and test-related models and simulations throughout the Army T&E Enterprise. Prior to this assignment, Cooke was assigned temporary duty as Director, United States Forces–Iraq (USF-I), Office of Security Cooperation (OSCI), Transition Team from November 2010–August 2011. He and his joint team planned the establishment of the largest Office of Security Cooperation in the world.

Cooke was the Army's Director for Models and Simulations under the Deputy Chief of Staff for Operations, G-3/5/7 from 2007–2009. He was responsible for strategic oversight of all Army M&S

programs and developments in order to create unity of effort and purpose. He was also the functional proponent and Branch Head for Modeling and Simulation, which encompassed over 400 officers and 1,000 civilians. From 2006–2007, Cooke was the Director of Army Battle Command/LandWarNet. He served as the Special Assistant for Systems and Director for Analysis on the Army Secretariat Staff from 2002–2006. From 1995–2002, Cooke served in the Program Analysis and Evaluation Directorate of the Office of the Secretary of Defense. He was responsible for leading analysis of major Army programs and portfolio investment analysis across all Army program and budget lines with direct responsibility for weapons and tracked combat vehicles, missiles, and the Army's testing and analysis funding. Cooke served as the Director of the Joint Staff J8 Resources and Joint Planning Division and the Chief of Analysis for the J8 from 1993–1995.

Acquisition Challenges of Autonomous Systems

David M. Tate—joined the research staff of the Institute for Defense Analyses' (IDA) Cost Analysis and Research Division in 2000. Prior to that, he was an Assistant Professor of Industrial Engineering at the University of Pittsburgh, and the Senior Operations Research Analyst—Telecom for Decision-Science Applications, Inc. At IDA, he has worked on a wide variety of resource analysis and quantitative modeling projects related to national security. These include an independent cost estimate of Future Combat Systems development costs, investigation of apparent inequities in Veterans' Disability Benefit adjudications, and modeling and optimization of resource-constrained acquisition portfolios. Tate holds bachelor's degrees in philosophy and mathematical sciences from the Johns Hopkins University, and MS and PhD degrees in operations research from Cornell University. [dtate@ida.org]

David A. Sparrow—received a PhD in physics in 1974 and spent 12 years as an academic physicist. He joined IDA in 1986 and has been a Research Staff Member ever since, with brief forays into management and government service. He was the first Director of the IDA Simulation Center from 1989 to 1990, and Assistant Director of the Science and Technology Division from 1993 to 1997. He then joined the government for a two-year stint as Science Advisor on Modeling and Simulation to the Director, Operational Test and Evaluation. Since returning to IDA, he has focused on technical issues in system development, especially ground combat systems—expansively defined to include unexploded ordinance (UXO), counter mine, and, occasionally, missile defense. He has authored ~100 refereed papers and invited talks on various academic and national security topics. [dsparrow@ida.org]

Abstract

The Department of Defense has stated publicly that future defense capabilities will depend strongly on autonomous systems—systems that make sophisticated judgments about the world and choose appropriate courses of action, and perhaps even adapt and learn over time. Developing and deploying such systems poses more than just a technical challenge in robotics and artificial intelligence—it also poses many challenges to the acquisition process and workforce. From cost estimation to sustainment planning, every aspect of acquisition will be affected. Test and evaluation, in particular, may require not only novel methodologies and resources, but organizational and process changes as well.

Acquiring Autonomy—Bottom Line Up Front

We consider the life cycle of a typical major acquisition program, and identify the processes and activities that are complicated by the presence of autonomy. We argue that every aspect of acquisition planning, management, and execution will be more difficult and less certain for systems with autonomous capabilities—and significantly more so for some of the ambitious autonomous capabilities currently envisioned by the Department of Defense (DoD) and emphasized in the National Defense Strategy (DoD, 2018). Designing and implementing the autonomous capabilities will force a different approach to system development and program management than is customary—simultaneously requiring more rigor and more flexibility.

Because there will be a lack of historical precedent to guide planning, execution, and oversight activities, a number of key "control loops" within the development effort will be new, different, and/or more difficult:

- Diagnosis of performance issues will be harder and will thus take longer than for non-autonomous systems.
- Determining and implementing corrective design changes will be harder and may require simultaneous changes to hardware, software, and concepts of operations.
- The division of responsibilities between humans and machines, and the protocols and concepts of operations (CONOPS) that enable effective teaming, will necessarily be part of the system design, rather than something to be figured out and perfected after the system has already been designed and built.
- Achieving acceptable performance will almost certainly involve *iterative experimentation* of a kind usually found only in Science and Technology (S&T) or Advanced Concept Technology Demonstration (ACTD) projects.
- Developmental Test and Evaluation (DT&E) and Operational Test and Evaluation (OT&E) will both require more frequent tests and new kinds of test instrumentation. Testing will also need to be more closely coupled with contractor and Program Office design processes than is typical.
- Achieving acceptable performance will require representative human users/operators/teammates far earlier and more frequently in the development cycle than is common under current practice.
- Developmental and regression testing will continue throughout the acquisition life cycle, often including occasional post-fielding regression testing.

The features of autonomous capability that will drive these changes include the following:

- The complexity and general lack of transparency of the core artificial intelligence (AI) modules enabling autonomy: *perception, reasoning, course of action, selection,* and *adaptation*;
- Substantive Human-Machine Teaming, involving shared situational awareness and understanding of mission objectives between human and machine agents; and
- The potential for undesired emergent behaviors of the complex system.

The literature on autonomous capabilities tends to focus on technical challenges of how to implement autonomy. Very little literature exists on the practical aspects of turning promising new AI technologies into commercial products or effective and suitable government systems that authorities will be willing to see fielded. We consider the technical aspects of implementing autonomy only to the extent that they can be expected to affect acquisition success—that is, timely and affordable delivery of effective and suitable systems. The main body of the paper discusses how the interactions among these AI modules and between the AI modules and the human team members, coupled with certain anachronistic features of the DoD acquisition system, create challenges throughout the acquisition process.

Why Autonomy Breaks Acquisition

How Autonomy Works

What do we mean by “autonomous systems”? One way to think about this is in terms of the “OODA loop” first described by Col. John Boyd (1995). Human beings performing complex tasks in complex environments repeatedly:

- **Observe**—take in data about the environment and themselves
- **Orient**—use that information to create a mental model of what is going on
- **Decide**—identify possible courses of action and choose one
- **Act**—implement the decision

Autonomous systems are those that implement a nontrivial OODA loop of their own, especially in the Orient and Decide modes. They collect sensor data for their own use, they process the data they collect to maintain a complex world model describing their environment and current state, and they develop possible courses of action and select which action to implement without direct human instruction.

To distinguish the machine version of OODA from the human version (and to emphasize the ways in which it is different), we will use a slightly different terminology. The corresponding capabilities that enable autonomy are

- **Perception**—collecting data about the environment and making sense of it; building a world model
- **Reasoning**—extrapolating from the world model to interpret events, intentions, unobserved entities, etc.
- **Selection**—identifying available courses of action and choosing one

Some of these capabilities are more sophisticated versions of familiar system features. Perception, for example, can be thought of as a natural extension of computer vision and sensor fusion capabilities. Selection allows more nuanced and complex behaviors than past state-action lookup tables.

These defining capabilities are in turn enabled by a wide array of specific AI methods and algorithms. These include various forms of Machine Learning (ML)—supervised learning, unsupervised learning, reinforcement learning—combined with complex constraint processing, theorem-proving, and other Reasoning techniques. The ML subsystems are implemented using specific learning architectures, such as Deep Learning or Generative Adversarial Networks (GAN), and their training is accomplished using specialized optimization techniques such as backpropagation.

The result, particularly for neural network-based forms of ML, is a system that does not so much process information algorithmically, but instead reaches a snap judgment when presented with an input. The system has “hunches”; the purpose of the training is to make those hunches accurate. In general, these hunch-making systems will be nested and combined, with feedback loops that make it essentially impossible to trace the “logic” of how the output relates to the input. This lack of transparency turns out to have important consequences for acquisition.

Engineering Design vs. Experimentation and Discovery

If you want to build a bridge, you start by deciding how much traffic of what type the bridge will be required to carry. You consider the geology of the planned site, the force of the current of the river you are bridging, the prevailing winds, and many other factors. You then consider possible bridge designs that could carry that much load under those conditions, and decide which design best meets your needs, taking into consideration cost, useful lifetime, maintenance required, time to build, and other measures. The key here is that you have a very good idea before you build the bridge just how much load it will be able to carry. You don't have to experiment; there is no trial-and-error involved. The only testing you need to do is to confirm that you built the bridge you designed—proper materials and processes, correct measurements, and so forth.

If you want to build an autonomous system to perform a given set of missions, there is (at present) no corresponding engineering science you can rely on. For any but the simplest missions, you can't look at the performance requirements and know that if you use Algorithm A, trained on training data set B, with decision logic C, the system will perform at overall level X. If the autonomous system is going to interact significantly with humans in a relationship that could be characterized as "teaming," even very simple missions can require considerable experimentation and fine-tuning before achieving the desired overall level of mission performance.

Consider the case of early aircraft autopilot systems. It is fairly straightforward to design a system that can maintain a given bearing and altitude without human intervention. It proved to be much harder to design a CONOPS and protocol for human-autonomy interaction that allowed humans and autopilots to cooperate smoothly without occasionally crashing an airplane.

The value of engineering design is that it completely characterizes the behavior of a proposed system. This means not only that you can know what a system can do without actually having to build it, but that you can know what that system will *not* do. In the absence of a well-established body of knowledge that supports predictive engineering design, all system development will need to rely on a process of experimentation and discovery—not only to figure out a design that works, but also to establish the dependability of that design.

For the program manager (PM), the presence of significant experimentation during development eliminates an important breakpoint that is built into the acquisition system. There are critical differences between activities before and after Milestone (MS) B—the Technology Maturation and Risk Reduction phase prior to MS B is about technology development, whereas the Engineering and Manufacturing Development (EMD) phase after MS B is about product and manufacturing process development. Technology development is inherently event-driven, but the theory has been that by requiring mature technologies at MS B, it is possible to make EMD predictable in cost and schedule. The need for experimentation as part of the design of autonomous systems pushes the inherently event-driven elements of the program well into EMD, where the supporting activities of Systems Engineering (SE), DT&E, and Project Management are ill-equipped to deal with it.

Proving a Negative Is Hard

For most current systems, safety and cybersecurity requirements are the only requirements of the form "the system must NOT ___." These proscriptions are typically not treated as "requirements" (in the Capability Development Document Key Performance Parameter sense) at all, but rather as "certifications" to be granted by stovepiped authorities separate from both the developers and the Test and Evaluation (T&E) processes. In addition, the scope of what is considered a "safety issue" is quite limited. A rifle program's safety certification will be focused on ensuring that the rifle will not explode in the operator's face, will not cause burns during normal operations, and will fire bullets only in the direction it is being aimed. The certification will *not* be concerned with the possibility that the rifle could be fired at friendly forces, stolen and used against noncombatants, or deliberately wedged into the hinges of a troop transport—those are risks to be mitigated by personnel screening and training, not by design of the rifle.

For autonomous systems, many unwanted outcomes that are typically avoided through proper screening and training of human operators will have to instead be avoided through the design of the system. Compliance with these "negative requirements" is inherently more difficult to demonstrate, for the familiar reason that you cannot prove a negative. It is easy to demonstrate that a skilled marksman can hit a six-inch target at 400 meters, by having a marksman do that. Reliability is more difficult, but can be demonstrated statistically. However, demonstrating that something will *never* happen is much more challenging. We don't even attempt to demonstrate that an infantryman will never shoot a civilian—accidentally or deliberately—but an autonomous system may well require us to provide convincing evidence of that.

In practice, we do not require proof that unacceptable outcomes will not occur; we require reasonable confidence. This makes T&E of autonomous systems an exercise in risk management; however, this is very different from the typical 5-by-5 "risk cube" approach typically used in defense program management. The familiar approach focuses on *identified* risks that are both serious and reasonably likely (~5% probability) to occur. In the case of complex autonomous capabilities, the program will need to produce evidence that the *unidentified* risks that may be catastrophic are extremely rare (multiple zeros in the probability). Further, in general, the argument supporting this assertion will need to be persuasive outside the program office.

State Space Explosion, Design of Experiments, and Autonomy

If you were given the task of testing a new passenger automobile tire, it would be a mistake to only test the tire on dry, straight, smooth, asphalt roads. Anyone who has done much driving knows that tires also need to be able to cope with curves, water (or ice), a variety of paving materials and surface conditions, and perhaps a range of temperatures. All of those things occur naturally in typical driving, and all of them matter.

For some systems, the set of parameters that matter, and the set of values they can take, is small enough that you could actually test every combination, to verify that the system performs acceptably in all of them. Similarly, for sufficiently simple software, it is possible to explicitly test every possible execution path of the software, to verify that the application behaves as intended. This is *exhaustive testing*. It is not very efficient, but it conveys a very high degree of confidence in the dependability of the system that was tested. The set of all possible combinations of relevant parameters is called the *state space* of the system in question.

For our automobile tire example, there is no chance that you could test every single possible combination of surface, surface condition, temperature, moisture, curviness, and so forth. This is *state space explosion*—when the number of configurations of interest grows faster than our testing needs can handle. However, we know the range of possible values each of those parameters can take, and we have a very strong physics-based understanding of which parameters are important and how performance varies with these factors and their interactions. It is possible to infer what a test of every possible combination would reveal by testing a much smaller number of well-chosen combinations of the parameters, using statistical inference to draw conclusions about how the tire would perform in situations *between* the ones that were explicitly tested. Design of Experiments (DOE) is the branch of statistics that studies how to do this efficiently and effectively, choosing a minimum set of design points to characterize system performance everywhere in the state space. DOE has been incredibly important in establishing the effectiveness and suitability of military systems with increasingly large state spaces.

DOE only solves the problem of state space explosion when

- We know which parameters are important.
- There aren't too many important parameters.
- We can reasonably expect that changes in system performance will be smooth in the regions of the state space that are between design points.
- We know in advance which regions of the state space are likely to exhibit rapid changes in performance for small changes in the parameters.

Unfortunately, those will generally not all be true for autonomous systems.

As noted previously, we have no engineering design theory for autonomy the way we do for tires, or for ships or aircraft or missiles. We don't know which inputs to the autonomy software will be important, we can't exhaustively test all of the possible execution paths of the software, and we can't statistically characterize performance over the entire state space using DOE. We are going to have to do something else—something that the Defense Acquisition System assumes you are never, ever going to do. We are going to have to experiment—a lot—at every stage of the acquisition process, probably including post-fielding sustainment.

The PM is then confronted with the following situation—the state space is too large for comprehensive exploration. It is not well enough understood for purely mathematical or statistical approaches such as DOE to ensure adequate coverage. The program office will need to use exploratory tools that are designed to preferentially find areas of potential concern, dynamically guiding the state space exploration in both simulation and real testing. There are a number of such approaches in development, but none that have become part of the standard development or T&E toolkit.

Emergent Behavior and Transparency

Emergent behavior is the general term for high-level behavior exhibited by a system that is hard to predict from the characteristics of the individual elements of the system. In physics, a classic example is trying to predict whether a given molecule will behave as a solid, liquid, or gas at a given temperature and pressure. Similarly, the tendency of snowflakes to exhibit hexagonal symmetry is an emergent behavior of water.

Human systems also exhibit emergent behavior. True gridlock—that state of urban traffic in which nobody can move—is an emergent behavior of individual driver behavior. In satellite communication systems, the tendency of speakers to speak and pause simultaneously is an emergent behavior of the combination of normal human speech patterns and the delay inherent in the communications system.

Autonomous systems, and autonomous systems interacting with humans, are even more prone to undesired emergent behavior than human systems are. Avoiding unwanted emergent behavior will be an important part of system development—which brings us back to the problem of proving a negative. Furthermore, when emergent behavior does arise, the problem of diagnosing the causes will be made much more complex by the inherent lack of transparency of the underlying algorithms in the autonomy.

For traditional systems, a common approach to avoiding unintended and undesired behaviors is to consider the worst-case scenarios of system behaviors, identify what states the system would have to be in for those behaviors to occur, then work backwards to determine how those states could be reached during actual operation of the system. For autonomous systems, there are severe limits on the effectiveness of this approach. In particular, the backwards causal trace of what conditions could lead to the undesired state may not be feasible if the undesired state is in part the output of a "black box" ML algorithm. Characterizing which inputs to, for example, a Deep Learning network would lead to specified outputs is not generally possible.

This lack of transparency is a big problem for verification and validation (V&V) of ML-based capabilities, even without the added issue of emergent behavior. For example, suppose that a target identification and cueing system is sometimes failing to warn its human teammates of a certain type of threat. Is the problem that the sensors are not seeing that threat? That the *perception* is not correctly identifying it? That the *reasoning* is concluding that it isn't important? That the *selection* is incorrectly choosing not to report that target? Some combination of the above? If the *perception* is the problem, is the failure due to the algorithm being used, the input data used to train the ML, or bugs in the code?

We will need to find novel ways to instrument what (and how) the autonomy is thinking, if we are to be able to develop and certify these systems on useful timelines. This is a new requirement. We will also need to open up the black box of the software to DT&E. This is also new, and both technically and organizationally challenging. It is technically challenging because this is fundamentally a new type of instrumentation with few precedents. It is organizationally challenging in part because of intellectual property and trade secrets issues: The vendor may regard this "instrumentation" as an effort to acquire intellectual property that was not contracted for, or as putting intellectual property at risk of unauthorized disclosure.

Testing Autonomy Will Have to Be Different

Developmental Test and Evaluation (DT&E)

DT&E has many purposes, including

- To help produce the information necessary for efficient and successful development of the desired system capabilities
- To verify the adequacy of the system design
- To quantify reduction of technical risk
- To verify contract technical performance
- To certify readiness for OT&E

These purposes can be roughly lumped into three overarching categories:

- **Characterization**—how exactly is the system (or some subsystem) behaving?
- **Diagnosis**—why isn't the (sub)system behaving properly?
- **Certification**—can we conclude that an interim performance goal for the system has been achieved?

What makes these categories distinct is that each of them requires *a different kind of testing*, collecting different measurements under different ranges of conditions. A common error in test planning is to assume that the same test event (or kind of test event) can support all of these disparate goals simultaneously. Just as a physician orders different kinds of blood tests, depending on whether the purpose is a general physical exam, diagnosis of a specific set of symptoms, or confirmation that a particular result has been achieved, so too DT&E must tailor its test designs to the particular purpose at hand. Early in the DT&E process, characterization will be primary. Late in the DT&E process, with luck, certification will be primary.

Autonomous capabilities complicate all three of these categories of DT&E. Autonomy is fundamentally a software-enabled capability, which means that autonomous systems inherit all of the T&E problems associated with complex software. In addition, the particular challenges posed by autonomous capabilities are not well addressed by traditional T&E practices, organizations, and resources. Although the kind of testing will vary with the characterization, diagnosis, or certification goals, there are new tools and new uses of existing approaches that will apply to all three. Although exhaustive testing is presumed infeasible, virtual testbeds allowing for extensive testing in a Live/Virtual/Constructive environment will be essential. The testing in this world will require the following:

- **Novel Modeling and Simulation (M&S) techniques**, enabling exploration of the decision space, rather than high fidelity representation of the physical space. Without this capability, characterization is impossible.
- **Novel instrumentation techniques**, enabling visibility into the perception, reasoning, and selecting functions. Without this capability, the data to support diagnosis (and ultimately certification) cannot be obtained.
- **New approaches to test design**, probably including adversarial test design, to allow efficient exploration of the extensive decision space.

The existing resources are not sufficient to support these activities, and in some cases include techniques that do not yet exist.

To be effective, DT&E will need to be more of a continuous engagement than a sequence of episodic events, with much more feedback into the architecture and design parts of development—the “D” in DT&E. This will require shorter test-redesign-test cycles that will probably continue much longer into the development cycle before morphing into the more familiar test-fix-test cycles. The “E” part of DT&E might require maintaining a detailed log of system and subsystem performance throughout the development cycle, to be used by downstream T&E and certification processes. This paradigm of continuous accumulation of evidence, rather than passing a convincing test event at the end of development, would be new—and is not supported by current organizational structures, division of responsibilities, or test resources.

When substantial human-machine teaming is intended, there will be additional challenges. The teaming CONOPS must be engineered into the machine. This will require active participation by users and user surrogates during the early stages of development. However, CONOPS development is not usually a program management office (PMO) responsibility, and the PMO may not have the authority to influence it.

With these challenges in mind, we turn back to the categories.

Characterization

For software systems, the question "What is the system doing?" is traditionally answered at the code execution level, by tracing the execution path and verifying that it is implementing the desired logic. If the complexity of the code is such that tracing the execution is impractical, or the logic of the algorithm is not understandable by humans, then some other way of interpreting what the system is doing will be required.

For systems with significant autonomous capabilities, much of the important system behavior will be implemented by algorithms that cannot be interpreted by humans as sequential logic. For example, any subsystem that relies on neural network models trained by supervised learning will not be executing sequential logic in the usual sense. Instead, when the neural network is presented with input data, it generates a "hunch" about the appropriate output, based on its training. Tracing the execution of the code that generates this output is unhelpful; it is merely a large number of weighted sums and transfer functions, uninterpretable at that level. Making sense of the hunch would require reverse-engineering the functioning of the neural network, in order to assign human-understandable meanings to patterns of weights in the network. This is a new requirement; we have not historically needed to instrument the internal states of mission software in order to decide whether it is performing well for the right reasons.

Diagnosis

A vital role of DT&E is to provide the measurements that enable testers to understand *why* the system is not behaving as intended. For software systems, this has traditionally meant finding bugs in the software that are causing it to implement incorrect logic.

For autonomous systems, because we have no engineering theory that would enable predicting system behavior from the software design, we cannot assume that undesired system behavior is being caused by bugs. The problem might just as easily be due to incorrect or inappropriate training data, a poor choice of algorithm, or unanticipated emergent behavior. If, for example, the system is failing to react appropriately to the actions of certain entities, it is difficult to tell from the outside whether the system is failing to see those entities (*perception*), failing to identify them correctly (*reasoning*), or failing to consider or choose the appropriate response (*selection*). If this is a system that continues to learn and self-modify after fielding, the system might have begun working correctly, but has learned a "bad habit" that is impairing performance (*adaptation*).

To diagnose the source of the problem, it will be necessary not only to instrument the internal states of the software, but also to have a normative model of what correct function looks like, in all modules and at multiple levels of descriptions, to compare the resulting measurements against. The instrumentation will be especially challenging on platforms that are highly constrained in available space, weight, or power—it will be difficult to *observe* system function without *distorting* system function. The development of normative models may be equally challenging, especially for enabling capabilities like ML, where proper function depends as much on how the model was trained as it does on correct implementation of the algorithm.

It is also important to mention that correct diagnosis does not always lead to a unique potential corrective action. If the *reasoning* functions are drawing incorrect conclusions from the world model built by the *perception* functions, it may not be obvious which module should be changed. Both may be functioning correctly according to the original design specification. Again, it will require experimentation to find the best way to achieve the intended functionality. Sometimes, the solution might involve adding a new run-time monitoring process that watches for problematic cases and intervenes when appropriate. This not only adds to the design complexity; it also introduces new modules that themselves will need to be verified and validated.

All of these challenges are exacerbated by an extreme version of the traditional chicken-and-egg problem posed by simultaneous hardware and software development. In general, software developers would prefer representative working hardware to test on, in order to verify that the software is working as intended. At the same time, hardware developers need representative working software to execute, to verify that the hardware is working as intended. Avoiding gridlock is not easy, especially if you can't be sure whether a given observed problem is due to hardware problems, software problems, or integration problems. Autonomous systems have all of these issues, but add the further complication of needing to develop the human-machine teaming CONOPS in parallel with both hardware and software. Correct diagnosis will require determining whether the hardware, the software, or the human team members are not behaving as intended, or whether the problem is in the algorithms chosen, the CONOPS design, or some combination of those things.

Certification

For autonomous or semi-autonomous weapon systems subject to DoD Directive (DoDD) 3000.09, *Autonomy in Weapon Systems* (DoD, 2017), there are certification requirements even before formal program initiation:

> Before a decision to enter into formal development, the USD(P), USD(AT&L), and CJCS shall ensure:
>
> *(1)* The system design incorporates the necessary capabilities to allow commanders and operators to exercise appropriate levels of human judgment in the use of force.

In other words, if DoDD 3000.09 applies, there are certifications about the design that must be made before development begins. The basis for these certifications is not yet well-defined.

In general, the T&E to support the various certification activities will involve open air testing of full up systems or major components. However, even these activities will require range safety releases, which are themselves a kind of certification. The *Joint Software Systems Safety Engineering Handbook* (DoD, 2010) addresses the relationship between autonomy embedded in the software and safety-critical functions.

For autonomous systems, safety release will generally depend upon a combination of a safety argument based on the entire development history and some specific "kill switch" function. A "kill switch" feature might itself depend on internal monitoring by an autonomous process, which could pose additional challenges, since internal monitoring may affect mission performance by competing with mission systems for power and computational resources.

Finally, the DT&E results are key inputs into Operational Test (OT) Readiness Reviews. While not a formal certification, readiness for OT typically requires the system under test to be "production representative." The current working definitions of "production representative" are all highly hardware-centric—they refer to tooling and production lines. For continually evolving complex software systems in general, the question of whether the system is yet production representative is difficult to answer. This is particularly true for ML systems. Many approaches to ML require extensive data for training. If the system were retrained using different data, it would exhibit different performance. Thus, the training data set itself needs to be production representative in some sense.

This vision for a persistent, intrusive instrumentation and experimentation process that includes CONOPS and training as design attributes is very different from what DT&E has looked like in the recent past. Even M&S, which is already familiar, would need to be used in novel ways for which little support currently exists. However, these are at least activities well within the traditional scope of a PMO's authority. Human teaming, human training, and CONOPS development are emphatically not part of a PM's traditional authority, and yet they will be essential elements of the system design. Safety issues introduced by the possibility of an incompetent or insane operator also demand novel T&E data collection, test designs, and information sharing across traditional organizational boundaries. This will introduce many of the same development issues as are seen when a program's design depends on decisions made in an external program—but in an unfamiliar context for which no administrative processes yet exist.

Operational Test and Evaluation (OT&E)

The purpose of operational testing is to determine whether or not a particular system is effective and suitable when used by warfighters in execution of their missions. The presumption is that DT&E has already established system performance over the intended range of operating conditions; all that needs to be confirmed is mission effectiveness and suitability when in the hands of intended users.

The scoping of OT under these conditions is straightforward—in some cases, a single scenario will suffice. For example, Air Superiority aircraft are usually tested in 1v1, 2v2, and 2v4 scenarios, but the number is manageable and predictable. For a ground vehicle, one would conduct both Major Combat Operations and Stability Operations scenarios, but again the scope is manageable and predictable.

For systems with autonomous capabilities, we have already seen that it may not be feasible to cover the state space during DT&E. For operational systems with substantial autonomous capabilities, there is no current understanding of how to scope the set of test missions by analogy to the air combat and ground combat examples. If suitability for autonomous systems requires the absence of unwanted emergent behavior, we do not know in general what would constitute sufficient evidence of that. In a previous paper (Tate et al., 2016), the authors argue that "sufficient evidence" might need to include the entire test history of the system, with the time series of improving performance and elimination of failure modes providing a degree of assurance not obtainable through pass/fail testing at the end of the development cycle.

Surprising uses of weapon systems by their operators have been observed during operational tests, even for systems with no autonomous capabilities. The probability of surprising emergent behavior is much higher when humans are teaming with autonomous systems rather than merely operating them, or when the systems are operating by themselves. This increases the probability of a test that cannot be conducted safely, or a test for which the wrong instrumentation or supporting data was available. As with DT&E, autonomous systems will generally require more test events in OT&E than traditional systems.

For systems subject to DoDD 3000.09, there are additional requirements that apply after IOT&E. In particular, Enclosure 2 of the Directive states,

> (b) After initial operational test and evaluation (IOT&E),any further changes to the system will undergo V&V and T&E in order to ensure that critical safety features have not been degraded.

(1) A regression test of the software shall be applied to validate critical safety features have not been degraded. Automated regression testing tools will be used whenever feasible. The regression testing shall identify any new operating states and changes in the state transition matrix of the autonomous or semi-autonomous weapon system.

(2) Each new or revised operating state shall undergo integrated T&E to characterize the system behavior in that new operating state. Changes to the state transition matrix may require whole system follow-on operational T&E, as directed by the Director of Operational Test and Evaluation (DOT&E).

In general, we do not yet know how to determine how much testing will be enough—or even whether the testing that has been done so far is sufficient. Even for systems for which DoDD 3000.09 does not apply, ongoing regression testing of software for safety and performance determination is certainly a best practice, where the same questions apply.

The earliest challenges for the PMO will be planning for the IOT&E. Scoping the tests will be more challenging than for systems without autonomous capabilities. In the case of robust teaming, scoping an OT that adequately explores the teaming arrangement remains an unsolved problem. The prospect of emergent behavior during the test may render the test unexecutable or uninformative. More time and resources will typically be required to execute the test events needed, and IOT&E will probably need to instrument more completely and archive more data in order to support post-fielding regression testing.

Cost and Schedule Estimation

Before you can do a meaningful Analysis of Alternatives (AoA), you need to have at least a rough guess at the cost, schedule, and operational effectiveness associated with each of the alternatives. (Ideally, you should also have a good idea of the risks and uncertainties associated with those attributes. The DoD does not have a stellar record in that regard.)

Standard cost and schedule estimating techniques were developed for sequential processes, in which the number, nature, and precedence relationships among tasks is known in advance. They work very well for routine construction projects. They work quite well for new system development when the new system is similar to past systems. They can even be useful for unprecedented new system development. They do not work well at all for projects with branching or looping logic, where the set of tasks and/or the number of times you will have to do each of them is not known with certainty.

In practice, cost estimators assume that experimentation is over—that the proposed design is (essentially) the design that will be built. They also tend to assume that the total cost of the program will be the sum of the costs of the components, failing to account adequately for critical path dependencies and costs of integration. The process of refining the design during development might involve test-fix-retest cycles to confirm that the design was implemented correctly, but it will not involve test-diagnose-redesign cycles at the higher levels of system architecture, design, or CONOPS.

This is already a bad assumption for software-intensive systems, first-of-a-kind science facilities, or complex system of systems integration (not to mention systems whose requirements keep changing over time). As noted above, autonomous systems partake of all three of those problematic categories, with the additional problem of necessary experimentation due to a lack of underlying engineering theory. Standard cost and schedule estimates for autonomous systems will always underestimate development time and resources required—and thus will underestimate cost as well. Until a substantial body of historical autonomy program data can be amassed, a new kind of estimating methodology will be needed, specially crafted to account for the uncertainty in how many diagnosis, redesign, and integration cycles will be required, to produce accurate forecasts of cost, schedule, and development risk.

To make matters worse, the cost and schedule estimates depend on the details of the CONOPS—how autonomy and human-autonomy teaming are to be used in performing the mission. The details of this CONOPS will not be known early in the program; they will necessarily be the result of substantial experimentation. This means that even this hypothetical novel estimating methodology, specially adapted for autonomous systems, will be subject to additional uncertainty due to reliance on educated guesses about key design features.

This suggests that it will be especially important for the PM to get updated cost and schedule estimates on an ongoing basis, keeping the cost estimators up to date on any changes in the autonomy or teaming design. In practice, although DoD guidance strongly encourages this kind of “living cost estimate,” there are strong political and organizational incentives that drive programs to avoid doing that.

Conclusions

Current U.S. military strategy has placed great weight on developing and fielding advanced AI-enabled systems with autonomous capabilities that will allow these systems to operate themselves to a significant degree. They will also need to team with human warfighters as collaborating agents rather than as tools to be operated. We have shown that even if the technical challenges of implementing these new autonomous capabilities can be solved, they pose significant unprecedented challenges to the defense acquisition system, its organizations, and its processes.

At root, these challenges all arise from the absence of a mature scientific theory or engineering practice for autonomous systems that would enable designers to predict the macro-level behavior of an autonomous system from a description of its enabling algorithms and data. Until we can accurately predict how a given design will behave without building and testing it, we must instead develop systems through iterative experimentation and adjustment. At the same time, we must invent test, evaluation, and certification techniques that will enable reasonable assurance that a given autonomous system will perform dependably—safely, securely, effectively, reliably, etc.—within a specified range of mission contexts.

Establishing dependability with certainty would require proving a series of negatives—that the system will *not* behave unsafely, will *not* exhibit undesired emergent behavior, does *not* have exploitable cyber vulnerabilities, does *not* have exploitable training biases, does *not* have exploitable reasoning or selection algorithm foibles, and so forth. Since proving a negative is impossible, we recommend an approach inspired by Karl Popper’s (2002) notion of falsification: namely, that a system can be considered dependable to the extent that we have tried as hard as possible to prove that it is *not* dependable and have failed. This approach would be a significant departure from current test practice, which is focused on doing the minimum possible amount of testing that can support a conclusion that a performance threshold has been met.

Implementing this approach will require novel T&E methodologies, such as intelligent adversarial testing in highly virtualized environments. These methodologies will in turn depend on T&E resources and infrastructure that do not yet exist—live/virtual/constructive simulation testbeds, instrumentation of AI internal states, sophisticated human/machine teaming support, copious data collection and archiving to establish dependability over time, etc. Such testing also implies developmental and operational test plans that are fundamentally unpredictable in duration and scope.

The defense acquisition system is predicated in large measure on the assumption that such unpredictability has been ironed out of a program by the time it passes Milestone B. That is generally not true even today; it will be far less true for the kind of experimental discovery processes needed to field effective autonomy. The DoD needs to be aware of that if they are serious about making autonomy a cornerstone of future military capability. Without significant changes to how we develop and test systems, the DoD will either not field autonomous systems, or will have very little reason to believe that its fielded systems will be safe, secure, and effective.

References

Boyd, J. R. (1995). *The essence of winning and losing* (Briefing).

DoD. (2010). *Joint software systems safety engineering handbook* [Ver. 1.0]. Retrieved from http://www.acqnotes.com/Attachments/Joint-SW-Systems-Safety-Engineering-Handbook.pdf

DoD. (2017). *Autonomy in weapon systems, incorporating Change 1* (DoD Directive 3000.09). Retrieved from http://www.dtic.mil/

DoD. (2018). *Summary of the 2018 national defense strategy of the United States of America: Sharpening the American military's competitive edge*. Retrieved from https://www.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Popper, K. (2002). *The logic of scientific discovery*. New York, NY: Routledge. Retrieved from https://archive.org/details/PopperLogicScientificDiscovery (Original work published 1935)

Tate, D. M., Grier, R. A., Martin, C. A., Moses, F. L., & Sparrow, D. A. (2016). *A framework for evidence-based licensure of adaptive autonomous systems* (IDA Paper P-5325). Alexandria, VA: Institute for Defense Analyses.

Acknowledgments

Valuable comments reflected in the paper were provided by Phil Sarnecki. John Biddle, Nick Kaminsky, and Poornima Madhavan contributed valuable discussions.

Using Developmental T&E to Inform Operational T&E Decision-Based Analysis

Dashi Singham—is a Research Associate Professor of Operations Research at the Naval Postgraduate School, where she researches, teaches, and advises student theses. Dr. Singham's primary areas of focus include simulation modeling, simulation analysis, and applied statistics, with most of her work on developing new methods and metrics for analyzing simulation output. Her areas of application include energy and intelligence systems. She received her PhD in Industrial Engineering and Operations Research from the University of California Berkeley in 2010. [dsingham@nps.edu]

Abstract

We describe two-stage sequential experiments that are used in building and testing valid simulation models. In the first stage, preliminary samples are taken to estimate performance and inform the parameters for the experiments in the second stage. These two-stage experiments can be mapped to test and evaluation (T&E) by having the first stage applied to developmental test and evaluation (DT&E) and the second stage applied to operational test and evaluation (OT&E). By considering DT&E and OT&E as part of a combined two-stage experiment, we can better leverage the results of DT&E to inform OT&E.

Introduction

Statistical experimentation in test and evaluation is critical to obtaining clear, valid results and recommendations regarding the quality of a system being tested. We refer to test and evaluation (T&E) of a system, where a system can be a weapon, computer program, piece of machinery, and so forth. While much of the methodology for T&E has been developed, there is still much room for improvement in terms of ensuring widespread knowledge and implementation of statistical methods. Hill (2017) states, "The current T&E workforce, while very competent in the engineering domain and mechanics of test, will benefit by improving their baseline level of statistics, their statistical fluency, thus firming up their overall knowledge base" (p. 123).

This research develops two-stage statistical procedures that use developmental test and evaluation (DT&E) data to design and conduct operational test and evaluation (OT&E) plans. Two-stage procedures rely on data collected in a first stage to estimate key parameters that are needed to determine what types of future tests should be run to answer a research question. In a T&E setting, these estimated parameters have some uncertainty given that testing conditions may be limited in the first stage or approximated using simulation. This uncertainty can be used to determine what tests and statistical parameters to use in the second stage. For example, if DT&E reveals strong performance in some areas and weaker performance in others, we can design OT&E tests that allocate more effort to quantifying the effect of the weaker performance areas on overall system sustainability.

Two-stage statistical procedures are commonly used in analyzing simulation models. The first stage runs some preliminary experiments to estimate key parameters, like the variance and distribution of the output. Then, second-stage experimental parameters are chosen and the results from the second experiment contribute to the final assessment of the system. This research draws on two-stage procedures by mapping first-stage methods to DT&E, where simulation or less-costly experimental methods are available. The second-stage method is then mapped to OT&E with an emphasis on the fact that these experiments may be much more costly.

Examples of highly cited two-stage procedures include Chick and Inoue (2001) and those reviewed in Goldsman and Nelson (1998). These methods use the first-stage samples to estimate the variance, among other parameters, of multiple systems. Estimation of system variance is critical to determining the details of an OT&E experiment. Giadrosich (1995) describes how an estimate of the standard deviation can be used to choose the sample size, and sequential sampling methods that rely on this variance estimation are presented in Singham (2014). We note that much of the simulation literature now focuses on fully sequential sampling rather than two-stage sampling, but these fully sequential methods may not always be appropriate for T&E because of high sampling costs and potential for bias.

This paper exploits two-stage statistical procedures to provide a better link between statistical methods used in DT&E and OT&E testing. The case study presented addresses the unique challenges present within a T&E environment, such as specific capabilities requirements, limited budgets, and risk associated with an incorrect evaluation. OT&E often requires a much higher budget due to the operational nature of the testing. Thus, the information from the first-stage is critical in determining where effort should be focused in the second stage. However, in some cases, sophisticated simulation models can be employed for integrating testing, combining aspects of developmental and operational testing. For example, Allen (2010) describes the Boeing Engineering Development Simulator in its ability to replicate many operational settings while testing the enhanced capabilities of the aircraft, saving costs by using a simulated environment.

The next sections summarize the background in T&E and two-stage procedures, present a proposed two-stage algorithm, and apply the algorithm to a case study.

Background

DT&E and OT&E each pose their own set of unique challenges. DT&E is often performed under highly controlled or even simulated environments, so there are limitations on how much this data can be extrapolated to estimate performance under operational conditions. Modeling and simulation (M&S) can help quickly obtain initial data sets, perform sensitivity analyses, and drive additional testing questions. M&S can be a cost-effective method when there are limits on physical experimentation, though it should not replace operational testing (Marine Corps Operational Test & Evaluation Activity [MCOTEA], 2013). Simulation methods can be integrated with a test process, especially in developmental phases before a final assessment is made, and can be especially important in DT&E (*T&E Management Guide*, 2005).

DT&E can usually inform the types of experiments run in OT&E. DT&E plays a major role in evaluating a potential system and its ability to meet the capabilities requirements. In order to ensure that a proposed system meets the requirements, a detailed DT&E process is needed to test system capabilities, limitations, costs, and safety. The data carefully collected in these experiments provides a wealth of information that can be used to inform efficient OT&E exercises. Because the questions used to design an operational test plan are motived by the results of DT&E, there is a unique opportunity to leverage two-stage statistical methods to efficiently answer questions about whether the capability requirements have been met.

For example, DT&E can be used to screen potential tests that may be unnecessary in OT&E because it is deemed that certain configurations of a system are likely to have poor operational performance and no further effort should be wasted on these settings. While Design of Experiments (DOE) is often considered a critical part of OT&E, using it in DT&E can only enhance the types of experiments that could be run in OT&E. Ortiz and Harman

(2016) argue for the use of DOE in DT&E in addition to OT&E because randomization, replications, and blocking can be more easily implemented. Such experiments in DT&E can narrow the space of possible feasible configurations to test in OT&E. This is part of the "shift-left" mentality to do more analysis in earlier stages of development to save costs and improve results throughout the entire acquisitions process.

Because OT&E assesses the performance of a system under more realistic conditions, testing can be much more expensive and constrained. Thus, it is even more important to design a test plan that is able to obtain the best information possible given constraints on the overall testing budget across the two stages. Additionally, the research questions and decisions that need to be made may have changed as a result of DT&E. Understanding integrated testing and evaluation is critical to efficient implementation of modeling and simulation results (United States Marine Corps, 2010).

Confidence Intervals

Confidence intervals are commonly used to assess the risk associated with the system by evaluating mean performance. Here we give a brief summary of confidence intervals to define notation and introduce key parameters. A confidence interval is collected from n samples of system performance results to estimate the mean of the system μ using $\bar{X}$ as the centerpoint. The half-width on either side of the centerpoint defines the confidence interval

$$\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \quad (1)$$

where σ^2 is the variance estimate for the data. If the data is normally distributed and the variance is known, then the confidence interval can be estimated exactly using standard z-tables. If the variance of normal data is estimated, then t-tables are used. The two key parameters we study are the variance estimate, which is critical to understanding the risk associated with an estimate, and the sample size n, which is often controllable by the user. A larger variance estimate leads to a larger confidence interval. If the variance is underestimated, the confidence interval will be too narrow and there will be more certainty (than there should be) in the result. The sample size n is critical for estimating the variance, and it also determines the width of the confidence interval. More samples are better for reducing uncertainty in estimates, but often come at high cost in a T&E setting.

Choosing the Sample Size

Sequential methods for generating confidence intervals have been studied most recently in Singham and Schruben (2012) and Singham (2014). These methods increase the sample size until a confidence interval with a half-width smaller than some pre-specified level can be generated. They have traditionally been studied in the context of simulation models where large numbers of samples can be collected.

Suppose the estimate of the standard deviation is s, and we have some desired precision in our confidence interval δ, which is the half-width of the interval. Then, the sample size that guarantees (for independent and normally distributed data) that the confidence interval for μ has a half-width smaller than δ is

$$n \geq \frac{t_{n-1,\alpha}^2 s^2}{\delta^2} \quad (2)$$

and this can be used to choose the sample size. Johnson, Freeman, Hester, and Bell (2014) study sequential methods for estimating ballistic resistance of armor, and note that the methods used by the Department of Defense (DoD) have not changed recently. The methods can be simple to implement and do not require much statistical analysis, and the authors conduct simulation experiments to determine which tests are most effective at estimating different percentiles for the probability that the armor is perforated. Such tests are often used as part of Lot Acceptance Testing to determine whether a production item is acceptable.

Two-Stage Procedures

Two-stage procedures are often used instead of single-stage procedures because initial data collected in the first stage can be used to enhance the efficiency and quality of results in the second stage. A main example of this is using the first stage of an experiment to estimate the variance of the system. The variance is usually unknown ahead of time, yet it is a crucial part of estimating confidence intervals or other measures of performance. A poor variance estimate can lead to low validity of statistical results. Results from DT&E can be used to estimate the variance of the system, which in turn helps decide how many runs are needed in OT&E. For example, if the variance of the system is high, then more runs will be needed in OT&E to assess the feasibility of the system. If the variance of the system appears low, perhaps fewer runs will suffice.

Given a set of *n* independent and identically distributed (i.i.d.) samples of system performance estimates, then

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad (3)$$

is used as the variance estimate. When the data is dependent and is normally distributed, we can quantify the dependence using autocorrelation with lag *h,* which is a measure of dependence between sample X_i and X_{i+h}. If the output data of a series has positive dependence, we hope that this dependence decreases over time as *h* increases, so that observations far apart are relatively independent. If the dependence between samples is positive, the variance estimate will be smaller than it really is. This means that the risk in the system will be underestimated, and we would proceed to OT&E with more certainty in performance than what actually exists.

Positive dependence between samples can exist for many reasons. For example, if a machine is not completely reset and recalibrated between samples, then the state left by the previous run can affect future runs. If the same operator tests the machine or weapon for each run, there may be correlation between outputs based on the habits or practices of the operator. In reality, there may be more variance in an operational setting because there will be many different people using the equipment. Thus, it is important to ensure independence between samples in the first stage. It may be useful to employ a confidence interval for the variance:

$$\left[\frac{(n-1)\hat{\sigma}^2}{\chi^2_{\alpha/2}}, \frac{(n-1)\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] \qquad (4)$$

where the chi-squared term is the relevant quantile of the chi-squared distribution with *n-1* degrees of freedom. This means that we can assess the uncertainty in the variance estimate based on the number of samples taken in the first stage, and inflate our estimate of the variance in the second stage using the upper confidence level of the variance estimate. Inflating the value of the variance estimate will encourage more samples to be taken in OT&E and will protect against the potential underestimation of risk resulting from a too-low variance estimate.

Ranking and Selection

Ranking and selection procedures attempt to determine the best system inputs when the system configurations are discrete options that can be listed. There is uncertainty ahead of time about the actual performance of the system, and the feasibility of the system to meet some constraints. Figure 1 shows the potential layout from the first stage of a ranking and selection experiment. The *x*-axis measures the feasibility of the system, while the *y*-axis measures the performance along the main objective or measure of effectiveness (MOE). The goal is to select the system with the best objective that is feasible. Based on the figure, it makes sense to invest more time in the second stage on the "Feasible, good objective" system and the "Infeasible, best objective" systems. It is possible the latter system may actually be feasible if we tested more, or it's possible the former system may actually be the best system. In any case, it probably does not make sense to spend resources in the second stage on the "Feasible, poor objective" and the "Infeasible, poor objective" systems.

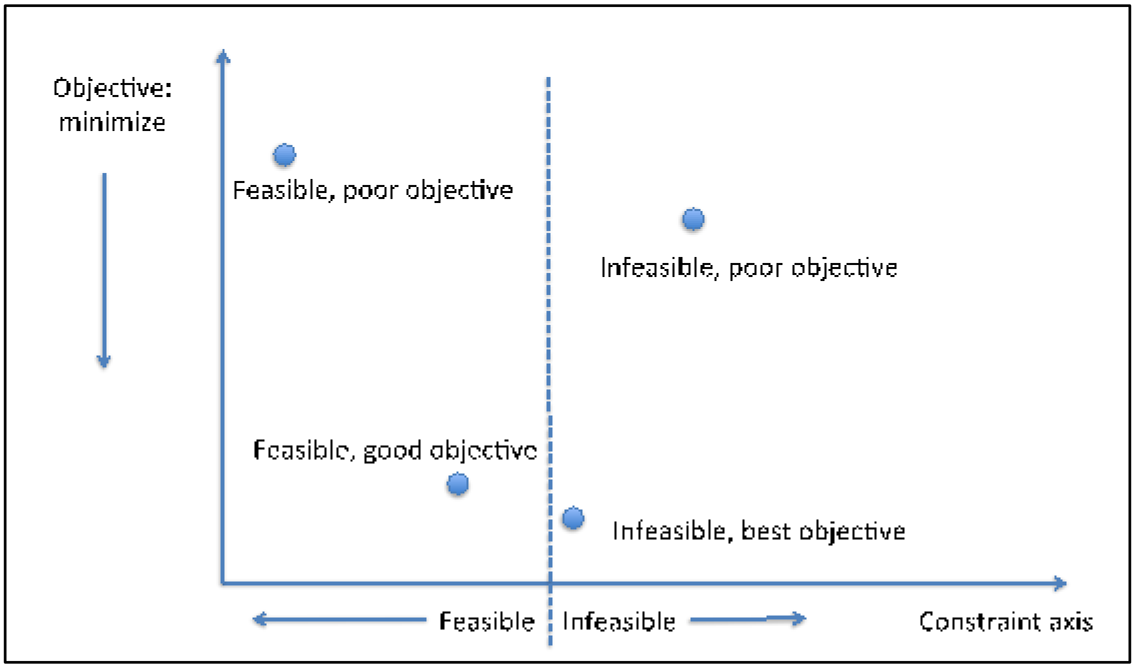


Figure 1. Comparison of System Configurations by Feasibility and Objective Function

Subset Selection/Screening

A number of subset selection procedures exist that screen out potential system configurations that are deemed suboptimal or infeasible. The first stage takes some initial number of samples from each system in the hopes of obtaining information that can be used for a more efficient second stage. In some cases, many system configurations can be eliminated from consideration in the second stage. This is something that occurs naturally in the transition between DT&E and OT&E; we do not usually bother to test options in OT&E that clearly did not work in DT&E.

One such subset selection procedure is Singham and Szechtman (2016), which uses information in the first stage to estimate the variance of the system and then allocate effort to the second stage accordingly. Systems with higher variance obtain a higher allocation of effort because they have more uncertainty. Similar methods can be used, as in Figure 1, to allocate sampling effort to systems close to the feasibility boundary, or close to optimality.

Then, in the second stage, a subset of the systems is chosen which is likely to contain the best systems with high probability.

A Two-Stage DT&E/OT&E Integrated Procedure

We now describe a two-stage statistical procedure that can be mapped to the stages of DT&E and OT&E. There are many different contexts to consider, but here we study the case where DT&E experiments allow for an arbitrary number of trials. For example, computer simulation experiments can often be used to test the potential readiness of a system, and it can be easy to run many replications.

The goal of the experiment is to determine which systems meet the requirements for performance, and, if more than one system meets the requirements, to determine which one is the best, or most cost-efficient, option. There are two main objectives of the first stage. The first is to screen out any system configurations that are highly likely to fail in OT&E, thus saving valuable experimentation resources. The second objective is to allocate resources to the remaining systems so that in OT&E the best system determination can be made. As in Figure 1, more resources would go to systems that are close to the feasibility boundary for meeting performance. Additionally, systems that display a high variance in the first stage would receive more samples in order to reduce their confidence intervals to make an operational suitability determination.

Next, we present the details of the two-stage statistical experiment. We run the first-stage experiments to estimate the mean and variance. These are used to calculate *p*-values, which are used to determine which systems can be eliminated from contention as worse than the threshold. Then, an inflated variance estimate is used to assign sample sizes to each system. This inflated estimate is used to account for potential model error resulting from the simulation setting being different from an operational setting.

1. The objective is to select the best alternative system that performs at least as well as the benchmark system, which determines the feasibility/capability requirements.
2. Develop DT&E experimentation parameters to answer objectives.
    a. For example, when analyzing performance of a sensor, two factors are (1) the coverage area of the sensor and (2) the location and number of sensors.
    b. Given the first stage is a simulation stage, we can run a large fixed number of replications of each system configuration to estimate the variance. However, to illustrate the effect of variance estimation in a limited budget, we run 30 replications of each configuration.
3. Run first-stage DT&E and analyze results.
    a. Estimate the mean $\bar{X}_i$ and variance $\hat{\sigma}_i^2$ for each system configuration *i*, including the benchmark system. Call the estimated mean for the benchmark $\bar{X}_0$ and, if the capabilities threshold for the benchmark is known, then its mean is fixed at *μ*.
    b. Reassess critical issues and specific objectives for the system, screen out factors and configurations if possible.
        i. Calculate *p*-values for each system for comparison to the system mean. Let *n* be the number of samples, and $F_{t_{n-1}}$ be the cumulative distribution function of the *t* distribution

with *n-1* degrees of freedom. If the benchmark is estimated then replace *μ* with $\bar{X}_0$ (see Singham and Szechtman, 2016, for an example of this type of calculation).

$$p_i = F_{t_{n-1}}\left(\frac{\overline{X}_i - \mu}{\hat{\sigma}_i/\sqrt{n}}\right) \quad (5)$$

ii. Use *p*-values to determine which systems to eliminate. These systems have a low probability of having performance that is better than the benchmark. For example, if

$$p_i \leq \alpha \quad (6)$$

then typically for $0 \leq \alpha \leq 0.1$,, eliminate the system from contention for having a mean performance level that is so small to be unlikely to be better than the benchmark *μ*. This will remove systems that have a small mean relative to *μ* while also having a relatively a small variance because we are fairly certain these systems will perform poorly.

c. Using confidence intervals for the sample variance, we can choose the upper confidence limit to deal with uncertainty associated with future OT&E experiments giving a conservative performance estimate.

$$\tilde{\sigma}_i^2 = \frac{(n-1)\hat{\sigma}_i^2}{\chi^2_{1-\alpha/2}} \quad (7)$$

d. Determine the budget allocation for the second stage based on first-stage results by comparing outcomes to the threshold objectives.

i. Calculate the sample size needed for each system to compare it to the threshold using properties of absolute and relative precision sampling as determined in Singham (2017).

$$n_i \geq \frac{t^2_{n-1,\alpha}\tilde{\sigma}_i^2}{|\overline{X}_i - \mu|^2} \quad (8)$$

ii. We need to do a similar calculation for the benchmark system if its true performance *μ* is not known. We decide a precision *δ >0*, which is the allowed deviation from *μ* that would be acceptable in a confidence interval estimate of the benchmark. Then, the second stage number of samples for the benchmark is

$$n_i \geq \frac{t^2_{n-1,\alpha}\tilde{\sigma}_i^2}{\delta^2} \quad (9)$$

iii. Rescale the sample sizes to be proportions for the second stage given a total budget *N*, and *S* total systems under testing.

$$\hat{n}_i = N\left(\frac{n_i}{\sum_{i=1}^{S} n_i}\right) \quad (10)$$

iv. If the unscaled n_i values are much too large for OT&E, then run n_i samples for system *i* in DT&E to obtain further information and repeat the screening, as in Step 3.b.ii, to remove additional systems that appear unlikely to beat the benchmark.

4. Run second-stage OT&E and analyze results.
   a. Run experiments on the potential subset using $\hat{n}_i$ sample sizes for each system *i*.
   b. Determine whether the requirements and objectives have been met by comparing the final results to the threshold. A similar *p*-value calculation to the one above can be used to determine if a system is significantly better or worse than the threshold.

What will most likely occur is that the first-stage experiment will determine a large number of samples n_i that will be needed to test each system. If these sample sizes are too large for OT&E, then we recommend running these experiments in DT&E to obtain as much information as possible and repeating Steps 3 and 4. The idea is that with enough samples, the difference $|\bar{X}_i - \mu|$ becomes large relative to $\hat{\sigma}_i/\sqrt{n_i}$ so that a clear determination can be made whether system *i* is better or worse than the benchmark *μ*. This can be used to screen out systems that are worse than the benchmark, and determine the allocation of effort toward systems better than the benchmark. Afterwards, if the number of samples is still too high for OT&E and there is a total budget *N* for samples, the rescaling can be done to allocate the budget towards systems that require more samples to make a determination.

In some cases, the T&E analyst may want to further reduce the subset from those that appear better than the benchmark for OT&E. For example, if seven out of 10 configurations are in the selected subset, the analyst may only choose the top three for consideration in OT&E to determine the best one.

Case Study—Unmanned Sensors for Intelligence Collection

To illustrate the procedure, we use a simulation experiment designed to test the performance of sensors for tracking targets such as pirates or smugglers. These sensors are designed to report information on potential targets of interest in large unpatrolled areas of water. Different sensors have different properties. For example, some have larger areas of coverage, while others may be more accurate and have a higher probability of detecting a target. The goal is to determine whether a particular sensor configuration can achieve the performance needed to be successful in finding targets, while balancing the cost and number of sensors to be purchased.

The simulation model has been built by the author and colleagues and is part of ongoing research being conducted at the Naval Postgraduate School. The full theoretical model details are available in Nunez, Singham, and Atkinson (2018). The model simulates numerous target paths given intelligence about the target's trajectory. Sensors can then be placed, and the number of target paths that are successfully observed can be recorded. Experiments can be run to determine a number of objectives, for example, which configuration is the best, or how often a particular setup successfully observes the target. We note that in this study, we do not consider whether physical specification requirements are met, but rather focus on whether the particular system can meet operational requirements.

Cheng (2016) studied different sensor configurations using this model to compare their performance. The benchmark given was Lynx multi-mode radar, which has a range of 80km (about 0.72 degrees) and an endurance of 48 hours. The Lynx radar system delivers high quality results but can be quite expensive (close to $7 million). Thus, we want to determine if we can obtain similar performance results using two cheaper unmanned sensors that may have smaller coverage areas. We use the sensor simulation model as the model to test the two-stage procedure. The model is flexible and allows for infinite input possibilities, and, as it is a computer simulation model, it is relatively inexpensive to run multiple replications to collect data.

Experimental Results

The experiment runs by simulating multiple potential target paths based on intelligence. Sensors are placed at the beginning of the run to attempt to locate the target as it passes through the area, and the simulation records the proportion of paths that intersect the sensor coverage areas. There will be variation each time the experiment is run due to randomness in the simulated paths. Thus, it is important to run multiple replications to estimate the potential error in the estimated probability of success.

We place sensors along the central expected path of the simulated target to obtain the maximum probability of success. Figure 2 shows the benchmark sensor placement for a target that is predicted to depart off the coast of South America towards the western coast of Mexico (red box). The blue heatmap shows the relative likelihood of the target's location given the intelligence at hour 25, with a higher probability in the middle. The Lynx sensor is positioned to anticipate observing the target at hour 50, but there is a high probability the target will not pass through the sensor and will remain undetected.

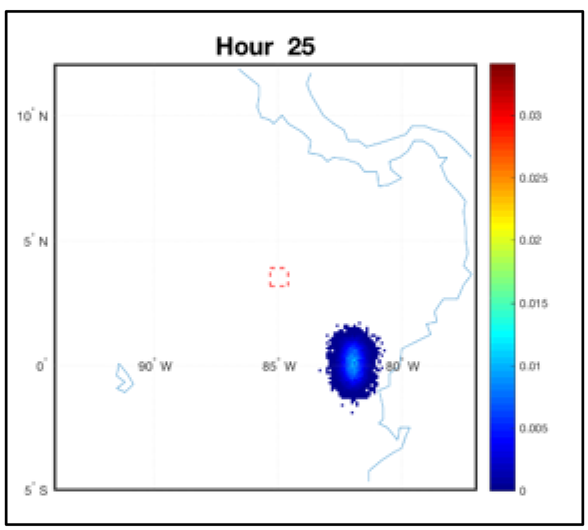


Figure 2. Benchmark Sensor Placement (Red Box) and Target Distribution (Blue Heatmap)

The alternative systems to the benchmark include those with two sensors with smaller coverage areas. We place the sensors to anticipate where the target will be at hours 35 and 70. While these sensors are smaller, there are two of them, so the second sensor may capture targets that remained undetected by the first sensor.

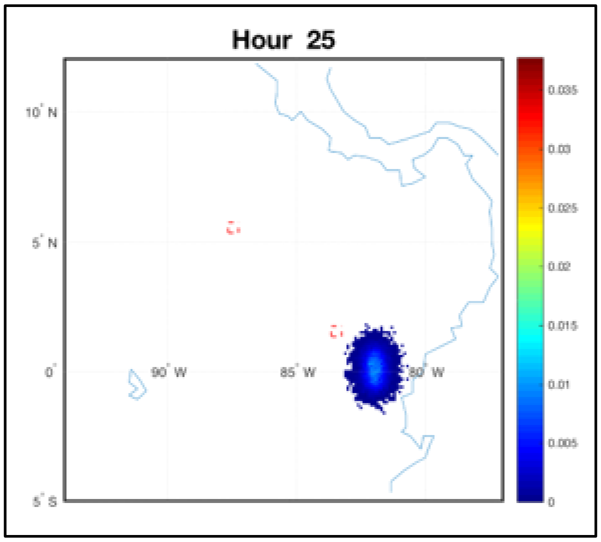


Figure 3. Dual Configuration: Alternative Sensor Placement With Two Smaller Sensors

We conduct first-stage experiments to compare different alternative configurations against the benchmark. The results of these experiments decide which systems have the potential to be better than the benchmark, and how to allocate second-stage experiments in OT&E. We note that sensors with smaller coverage areas are assumed to be cheaper and are preferred. A second-stage experiment could consist of more comprehensive simulation runs, or operational testing of the sensor in practice to see how it performs. Table 1 summarizes names of the system configurations, with the benchmark, dual sensor configurations, and their coverage widths.

Table 1. Names and Coverage of System Configurations

System Configuration	Coverage Width (each sensor, degrees)
Lynx (single) benchmark	0.72
Dual20	0.20
Dual30	0.30
Dual35	0.35
Dual37	0.37
Dual38	0.38
Dual39	0.39
Dual40	0.40
Dual50	0.50

All of the sensors in the dual configuration have much smaller coverage widths than the Lynx. We apply the algorithm to a series of first-stage experiments, as described previously, by running 30 replications of the experiment for each configuration and saving the mean and variance of the proportion of targets detected. Each replication simulates 200 target paths based on intelligence. We use these values to calculate p-values relative to the benchmark, and then eliminate systems who have *p*-values smaller than α=0.05, as these are unlikely to be better than the benchmark. For the remaining systems still in contention, we calculate the upper bound on σ^2 to determine the number of replications needed to distinguish the system from the benchmark mean. Table 2 summarizes the results of the experiment.

Table 2. First Stage Experiment Performance

System Configuration	First Stage Mean	*p*-value	Number of Samples	Proportion of Samples for Second Stage
Lynx (single) benchmark	0.1457	---	35	9%
Dual20	0.0408	0	---	---
Dual30	0.0892	0	---	---
Dual35	0.1155	0	---	---
Dual37	0.1273	0	---	---
Dual38	0.1407	.17	220	58%
Dual39	0.1543	.96	65	17%
Dual40	0.1542	.97	57	15%
Dual50	0.2235	1	2	1%

We require a precision of 1% on the estimate of the benchmark, so the allowable deviation in the estimated performance of the benchmark is 1%. The systems with two sensors with small coverage areas (Dual20, Dual30, Dual35, Dual37) all have estimated performance significantly below that of the benchmark, so the *p*-value is 0. We can eliminate these systems from consideration in the second stage. It is apparent that Dual50 has the best performance by far, with Dual38, Dual39, and Dual40 having performance close to that of the Lynx single sensor system. Depending on the requirements, we may want to choose the sensors with the smallest coverage width if they are cheaper.

We use the algorithm to calculate the number of samples needed in the second stage for the remaining systems and the benchmark. The Lynx system requires 35 samples to estimate the mean performance down to 1% absolute error. The Dual50 system only requires 2 samples, mainly because its performance is much higher than the benchmark, so little additional testing is needed to distinguish it as an improvement. The Dual38 system requires 220 samples because its performance is closest to that of the benchmark, so many more samples are required to distinguish whether or not it is better. Dual39 and Dual40 require 65 and 57 samples, respectively, to ensure they are better than the benchmark.

The last column shows the percentage of effort needed for each system. If the second stage cannot complete the recommended sampling effort because of cost or operational constraints, the last column shows the relative effort that should be expended on each system, with 58% of the effort going to Dual38. At the end of the second stage, we hypothesize that Dual39 is the “cheapest” system that has performance at least as good as the benchmark, where the smaller coverage area sensors are cheaper. However, we must still expend significant effort on Dual38 because it could be better or indistinguishable from the benchmark.

We conduct a second-stage experiment, which is meant to represent a more expensive operational setting but still involves a simulated model. Each replication now simulates 20,000 independent target paths (instead of 200 in the first stage), resulting in a more accurate estimate. In reality, the second-stage experiments would be in an operational setting where real information could be obtained.

Table 3. Second Stage Experiment Performance

System Configuration	First Stage Mean	*p*-value
Lynx (single) benchmark	0.1437	---
Dual38	0.1404	0
Dual39	0.1473	1
Dual40	0.1542	1
Dual50	0.2279	1

The second-stage results in Table 3 show clearly that Dual38 does not perform as well as the benchmark, while Dual39, Dual40, Dual50 are superior to the benchmark. Thus, the conclusion is that Dual39 is the cheapest system that performs at least as well as the benchmark, meaning two sensors with a coverage width of 0.39 would perform at least as well as one sensor with a coverage width of 0.72. However, the analyst could still choose Dual38 if she or he felt it was close enough to meeting the requirements. We note that the first stage required 270=9x30 total replications, while the second stage required 379 total replications. By eliminating some systems after the first stage and reallocating effort, we are able to focus effort on obtaining the best system. This saves effort over continuing to employ equal allocation over all systems in the second stage.

Conclusion

We present a two-stage statistical method that can be used to link experimental parameters in DT&E and OT&E experiments. The first-stage experiments can be used in DT&E to estimate the performance of different systems. These results can be analyzed to determine which system configurations to test in OT&E and how to allocate effort in the second stage. Typically, more effort should be allocated towards systems with high variance or those close to the feasibility boundary or capabilities requirement, which can be determined by a benchmark or other metric. We apply the algorithm to a model designed to compare different sensor configurations.

References

Allen, C. L., Jr. (2010). The role of simulation in test and evaluation. *ITEA Journal, 31*, 378–383.

Cheng, C. C. (2016). *A Brownian bridge movement model to track mobile targets* (Master's thesis). Monterey, CA: Naval Postgraduate School.

Chick, S. E., & Inoue, K. (2001). New two-stage and sequential procedures for selecting the best simulated system. *Operations Research, 49*(5), 732–743.

Giadrosich, D. L. (1995). *Operations research analysis in test and evaluation*. Reston, VA: American Institute of Aeronautics and Astronautics.

Goldsman, D., & Nelson, B. L. (1998). Statistical screening, selection, and multiple comparison procedures in computer simulation. In *Proceedings of the Winter Simulation Conference, 1*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Hill, R. (2017). The test and evaluation workforce and a base of sand issue. *ITEA Journal of Test and Evaluation, 38*(2).

Johnson, T. H., Freeman, L., Hester, J., & Bell, J. L. (2014). A comparison of ballistic resistance testing techniques in the Department of Defense. *IEEE Access, 2*, 1442–1455.

Marine Corps Operational Test & Evaluation Activity (MCOTEA). (2013). *MCOTEA operational test and evaluation manual*. Quantico, VA: United States Marine Corps.

Nunez, J. A., Singham, D. I., & Atkinson, M. P. (2018). *A particle filter approach to estimating target location using Brownian bridges*. Manuscript submitted for publication.

Ortiz, F., & Harman, M. (2016). DOE in DT: The place to be*! ITEA Journal of Test and Evaluation, 37,* 241–245.

Singham, D. I., & Schruben, L. W. (2012). Finite-sample performance of absolute precision stopping rules. *INFORMS Journal on Computing, 24*(4), 624–635.

Singham, D. I. (2014). Selecting stopping rules for confidence interval procedures*. ACM Transactions on Modeling and Computer Simulation (TOMACS), 24*(3), 18.

Singham, D. I. (2017). Decision-based metrics for test and evaluation experiments. In *Proceedings of the 14th Annual Acquisition Research Symposium*. Monterey, CA: Naval Postgraduate School.

Singham, D. I., & Szechtman, R. (2016). Multiple comparisons with a standard using false discovery rates. In *Proceedings of the 2016 Winter Simulation Conference*. Piscataway, NJ: Institute of Electrical and Electronics Engineers.

Test and evaluation management guide (5th ed.). (2005). Fort Belvoir, VA: Defense Acquisition University Press.

United States Marine Corps. (2010, May 6). *U.S. Marine Corps integrated test and evaluation handbook* [Memorandum]. Washington, DC: Author.

An Empirical Study on Content Analysis Use in Test and Evaluation Deficiency Report Analysis

Karen Holness—is an Assistant Professor in the Department of Systems Engineering at the Naval Postgraduate School (NPS) in Monterey, CA. She holds a BS, MS, and PhD in Industrial Engineering from the University at Buffalo. Prior to NPS, she worked as a Navy civilian in the acquisition workforce for eight and a half years in various industrial engineering, systems engineering and human systems integration roles. She also previously worked for three years as an Industrial Engineer at Corning, Incorporated in Corning, NY. [kholness@nps.edu]

Rabia H. Khan—is a Faculty Research Associate in the Systems Engineering Department at the Naval Postgraduate School (NPS) in Monterey, CA. Khan's research background includes studies in systems engineering competency modeling and development, cognitive processing, content analysis, energy studies, curriculum development and measuring self-efficacy within the field of systems engineering. She earned her BS in Psychobiology from the University of California, Davis and an MS in Engineering Systems from NPS. [rhkhan@nps.edu]

Gary Parker—is a Faculty Research Associate in the Systems Engineering Department at the Naval Postgraduate School (NPS) in Monterey, CA, with 41 years of combined military, industry, and government experience in intelligence, systems engineering, and defense systems acquisition. His research areas include systems behavioral modeling, complex adaptive systems, and system of systems engineering. Parker's education includes a BS in Aerospace Engineering from the University of Colorado, an MS in Business Administration from Boston University, and MS degrees in Systems Engineering and Physics from NPS. He is currently a PhD student in the Systems Engineering Department. [gwparker@nps.edu]

Abstract

This research investigated strategies and heuristics used to prioritize system deficiencies identified during test and evaluation. Five participants were recruited to participate in this laboratory study and were assigned to an experiment condition either with or without content analysis training. Content analysis is a well-known methodology for identifying patterns and themes in qualitative datasets. In either experiment condition, subjects were asked to (1) classify a set of flight simulator deficiencies, (2) develop a deficiency resolution priority order using those classifications, and (3) complete a set of questionnaires regarding the completion of these tasks and demographic information. Across the five subjects, there was fairly high variability in the strategies and methods used. Therefore, the impact of the content analysis training was inconclusive. However, the variety of observed approaches warrants future research, specifically into the use of multiple categorization schemes when deciding upon a deficiency resolution priority order.

Introduction

Like other data analysis efforts within a typical Department of Defense (DoD) acquisition program, test and evaluation (T&E) data analysis efforts are impacted by constraints of program cost, schedule, and resource availability. The choice of analysis methodology also impacts the quality and reliability of the data analysis results. Government and contractor engineers who work with T&E data come from a variety of backgrounds and have their own intuitive approaches to evaluating data. The analysis of T&E data is further impacted by the inherent mental models, heuristics, and biases each government and contractor engineer brings to working on the same dataset based on their individual backgrounds and experience.

Holness (2016) described the potential for research in the use of content analysis in various systems engineering (SE) activities, including the Integration, Verification, and

Validation processes of which T&E is a part. This current empirical study investigated the types of data evaluation strategies, and corresponding decision-making and planning strategies, used when analyzing primarily qualitative T&E data and leveraging a content analysis framework.

This research addressed one primary question: How can technical decision-makers use patterns and themes in T&E data to prioritize the correction of system deficiencies discovered during test events? The following were the research objectives for this study:

a. Investigate the strategies and heuristics used by decision-makers to
    i. identify patterns and themes in T&E datasets
    ii. use those patterns and themes to classify the deficiencies into categories
    iii. use those categories to prioritize deficiencies for resolution
b. Investigate the perceived level of effort and value of classifying data into categories

Throughout this paper, variations on the terms *deficiency*, *discrepancy*, *anomaly*, *issue*, *problem*, *failure*, and *fault* are considered synonymous and are used interchangably.

Literature Review

The standard process for conducting a T&E event involves adherence to a pre-established T&E plan that supports either system verification or validation activities with an approved set of test procedures. After executing the test procedures and recording the results, observed anomalies are analyzed and resolved using some form of quality assurance process to determine compliance with established requirements (International Council on Systems Engineering [INCOSE], 2015).

Kossiakoff et al. (2011) state that the cause of discrepancies is not always obvious since they can result from any number of factors, including issues with "(1) test equipment, (2) test procedures, (3) test execution, (4) test analysis, (5) the system under test, or (6) occasionally, to an excessively stringent performance requirement" (p. 467). Wasson (2006) includes additional issues, like test environment and human error. He also states that when test failures occur, a discrepancy or deficiency report (DR) is written, the significance of the problem on the system under test and the test plan needs to be determined, and the source of the failure must be isolated.

When documenting an observed deficiency, it is important to provide sufficient detail on what happened and provide an assessment of the deficiency's severity and implications. This assessment typically starts with a judgment of the system's ability to meet its operational and/or maintenance requirements in light of this failure. The most common way to do this uses a pre-determined classification scheme. For example, Kenett and Baker (2010) describe six generic severity classes for software, each with a corresponding generic definition: catastrophic, severe, moderate, minor, cosmetic, and comment. For example, *minor* is defined as when "things fail under very unusual circumstances, and recover pretty much by themselves. Users don't need to install any work-arounds, and performance impact is tolerable" (p. 196). Providing a descriptor for each severity class is important to support consistent use across developers and testers.

As shown in Figure 1, the sample DR summary format, originally from the *Memorandum of Agreement (MOA) on Multi-Service Operational Test and Evaluation and Operational Suitability Terminology and Definitions* (2010) and shown in the DoD (2012) *Test and Evaluation Management Guide*, includes a column for deficiency description and an additional column for remarks. The deficiency shown in Figure 1 was classified as minor.

The systems engineering team members also evaluate the qualitative and quantitative data contained in written text and the deficiency codes in discrepancy descriptions to determine the best way to resolve the deficiencies.

Equip Nomen	Report I.D.	Report Date	Type of Deficiency	Deficiency Description	Cog. Agency	Closure Code	Action Ref	Remarks	Status	Date Information		
										Action AC CLO Date	Test for CLO Date	Last Update
	A		B		C	D						
AN/TCY-38 CNCE, ETC.	EPR 101-41.11-23001-YC-20-JFT, ETC.		INFO. MINOR, OPERATIONAL, ETC.	SHORT TITLE, PART NO, SUBASSEMBLY, ETC. PLUS PROGRAM EXAMPLES 1. OX-34 INVERTERS FAILED 2. SOFTWARE FLT-8 (E7R31) (DIAG) TRAINING PROBLEM WHEN TTY ON LINE. 3. YDU 8 CARD FAILURE	GTE, ESO, RCA, ETC.	NEEDHAM, FORT HUACHUCA, ETC.	FM-MS-404, ESD LTR 18 MAR 79	DEPOT REPAIR/REPLACE. TAPE PATCH DUE BY 24 AUG 79. SEE ECP AK-000, ETC.				

A. SERVICE UNIQUE REPORT NUMBER, i.e., EPR KH-41
B. TERMS LIKE "MAJOR," "MINOR," ETC.
C. WHERE THE CORRECTIVE ACTIONS WILL TAKE PLACE
D. PROBLEM REPORT #, DATE OF LETTER SENT TO AGENCY, ETC.

Figure 1. Sample Deficiency Report Summary
(MOA, 2006)

In another DR summary example, the Naval Air Warfare Center Training Systems Division (NAWCTSD) uses a format that includes ample space for both a deficiency description and corrective action recommendation. It also includes a numerical deficiency category scale for any hardware, software, or process issue. As described on the NAWCTSD (2017) website, "A Part I (critical), Part I* (safety/critical), Part II (major), or Part III (minor) DR classification shall be assigned to each deficiency."

Following an investigation into the failure's root cause, there is a subsequent assessment of what it might take to fix it, what should be done to address it, and corresponding impacts to program cost and schedule. The order in which to work on the deficiencies is also determined. As stated in the DoD's (2012) *Test and Evaluation Management Guide*, "A comprehensive and repeatable deficiency reporting process should be used throughout the acquisition process to report, evaluate, and track system deficiencies and to provide the impetus for corrective actions that improve performance to desired levels" (p. 26).

Using the NAWCTSD categories as an example, it is clear that Part I and Part I* deficiencies must be addressed first, since they are critical and impact safety or mission execution. The Part II and Part III DRs must be reviewed for some order of precedence to be resolved and potentially retested by the test engineers. Depending on the size of the system, the number of DRs that need to be prioritized for resolution can vary from a few to many.

There is variability in how best to tailor an approach for a specific work domain. The common approach across a variety of deficiency classification and prioritization tasks is some combination of calculated numerical scores and human judgment. Of particular interest in this research is the creation and use of additional classification categories to complete a prioritization task. This emphasis on embedding classification within prioritization warrants a discussion about the fundamentals of content analysis as a categorization process for qualitative data.

As defined by Patton (2015), content analysis refers to "any qualitative data reduction and sense-making efforts that takes a volume of qualitative material and attempts to identify core consistencies and meanings. … The core meanings found through content analysis are patterns and themes" (p. 541). Under this general definition falls various methods for gathering relevant text segments, searching for occurrences of specific data points, iteratively coding the data, clustering data, then analyzing the results of the clusters and subsequent classifications for meaning and conclusions. This is the fundamental approach for grounded theory, defined by Birks and Mills (2012) as "an approach to research that aims to produce a theory, grounded in the data, through the application of essential methods" (p. 179). Further analyses using descriptive and inferential statistics such as frequency counts, chi-square, percent agreement, and alpha and kappa statistics are used to evaluate classification schemes and gauge their validity when used by multiple coders (Krippendorff, 2013; Miles & Huberman, 1994; Patton, 2015). When determining inter-rater agreement and reliability, the best statistic to use in a specific content analysis study basically depends on the coding scheme, the number of raters, and the number of categories.

The objective of this research study is to investigate different ways that system issues with assigned deficiency classifications are prioritized for resolution. Of particular interest are the strategies individuals use to prioritize a list of deficiencies for resolution, with or without prior knowledge of the content analysis methodology. The next section describes the design of this study.

Methodology

All research design and execution activities were completed at the Naval Postgraduate School (NPS) by the authors of this report. The experimental protocols and materials were approved for use by the NPS Institutional Review Board (IRB) prior to the start of the experiment. The test materials used in the experiment were

- unclassified and non-proprietary,
- understandable by a typical NPS Engineering and/or Graduate School of business and Public Policy student, and
- designed to target a specific deficiency prioritization solution.

Experiment Design

The research study was designed as a laboratory experiment, where study participants sat in front of a computer and performed reading and assessment tasks using files created in standard office software such as Microsoft Word and Excel and Adobe Acrobat.

The primary target population for this research was current NPS resident systems engineering (SE) students. Additional students were recruited from the following curricula: Naval/Mechanical Engineering (Total Ship Systems Engineering track), Systems Acquisition Management, and Modeling, Virtual Environments & Simulation. No previous experience with T&E was required to participate, no incentives were given to recruit subjects, and no compensation was provided to the volunteers at completion of the experiment. An informed consent form was used that explained participation was completely optional and that all data collected would be anonymized.

Study participants were assigned to one of two experiment conditions where they either (a) received a training session about content analysis and how to find patterns and themes using this method or (b) received no training. In both conditions, each participant was asked to categorize a list of deficiencies that were already assigned a technical priority by test personnel using the previously described NAWCTSD deficiency codes. Then, using the categories they created, subjects were asked to prioritize the deficiencies for resolution and explain the thought processes they used to accomplish these tasks. The study was designed to be completed within two hours, regardless of experiment condition.

There were three key hypotheses guiding this study. First, the subjects in the content analysis training condition were expected to produce more well-defined categories than those in the non-training condition. Ideally, the training would assist with their category identification and classification strategy. Second, the perceived difficulty of the categorization and prioritization tasks (i.e., frustration level, mental and temporal demand, etc.) would be higher for those subjects in the non-training condition. Third, participants were expected to leverage the issue prioritization assigned by the test personnel in order to come up with a resolution priority order. In other words, all of the Part II issues labeled by the test personnel would have higher resolution priority numbers than the Part III issues, regardless of the issue categories the subjects created on their own. This was the expected deficiency prioritization solution. This strategy was also expected by all participants, regardless of training condition.

No power analysis was performed to determine the sample size for this study. The expected number of participants was 10–20 SE department students, based on the approximately 45–50 eligible students in the resident systems engineering curricula during the 2017 summer quarter. This number seemed reasonable, based on sample sizes reported in similar studies from the research literature. As described in the previous chapter of this report, Henningsson and Wohlin (2004) had eight participants, while Linkov et al. (2009) had 21 participants. In a policy capturing study reported by Lafond et al. (2015), 60 university students performed a radar contact classification task in a naval air-defense scenario using a simulated combat control system microworld. Finally, in the Cropp, Banks, and Elghali (2011) study, 30 industry professionals reviewed hypothetical case studies and rated potential risks associated with each one.

Data Collection Method

A pilot study was conducted prior to the main experiment. One person volunteered to participate in the timeframe allotted. After evaluating this person’s data, no changes were made to the methodology or data collection process.

For the main experiment, student participants were recruited via email. A copy of the informed consent form was attached to the email so potential participants could read it ahead of time and decide if they wished to participate in this study. In addition to email, some classroom visits were made to advertise the availability of the study and promote responses to the email. Students were asked to contact the research associates listed in the email if interested in participating and indicate a day and time that worked best with their schedule. Recruitment took place in July and August 2017, and data collection took place in the month of August. Only four students volunteered to participate.

At the beginning of each experiment session, subjects were first asked to sign the informed consent form. Then, they were given an overview of what they were expected to do. Those in the training condition were asked to review a PowerPoint file with an 18-minute narrated instructional brief on content analysis methodology before starting the main experiment task. All subjects were asked to complete the following tasks:

- Read the provided T&E deficiency report that described testing performed on a generic aircraft flight simulator system.
- Using an Excel spreadsheet, look for patterns and themes in the provided deficiencies and create categories to help them prioritize the issues for resolution.
- Create a prioritized deficiency list indicating the order they think the deficiencies should be resolved.
- Complete a demographics questionnaire about their backgrounds and T&E experience.
- Complete questionnaires that assessed
    a. the classification strategies they used,
    b. perceived classification task difficulty,
    c. the value they assigned to doing the classification task as part of deficiency prioritization, and
    d. the impact the categories had on the priority order.

The provided T&E deficiency report was both generic and realistic, describing tests conducted on the flight simulator and deficiencies discovered during testing. The deficiencies were defined as issues found in the simulator's hardware and software by test personnel while executing a set of approved simulator test procedures. The deficiency list provided in the T&E report contained 25 issues. A brief description was provided for each issue, along with the deficiency priority assigned by the test personnel and the name of the organization primarily responsible for resolving the issue. All of the deficiencies were either a Part II or Part III deficiency, as defined by the NAWCTSD guidance described previously. The T&E report provided definitions of all of the NAWCTSD classifications for each subject's reference.

Subjects were asked to view themselves as a government systems engineer, read through the list of identified deficiencies, group them into relevant categories, and use those categories to prioritize the deficiencies for resolution. The subjects were specifically instructed via a hardcopy instruction sheet to assign each deficiency a unique priority number (i.e., two or more deficiencies could not be assigned the same priority number). For the purposes of the study, subjects were instructed to assume the following:

- Both funding and personnel are available to work on all identified issues.
- All issues must be resolved within the next 1–2 months.
- A resolution for each issue can be either a fix, a workaround solution, or planned deferral of resolution until something else is obtained.

Subjects did not have to identify a course of action to resolve each issue; they were asked to assume that one would be created for each deficiency after the priority order for resolution was completed. The subjects were asked to complete the categorization and prioritization task within one hour using the provided T&E report as a reference and working with the list of deficiencies in a Microsoft Excel spreadsheet. A pen and paper were provided to each subject during the course of the study, should they have wanted to write notes to assist in completing the tasks.

At the end of the prioritization task, the research associate noted the subject's completion time, then gave each subject an additional 15 minutes to complete a series of questionnaires in a separate Excel spreadsheet. These questionnaires were designed to capture the subject's demographic information, classification strategies, perceptions of task difficulty, and perceptions of the value of doing classifications as part of deficiency prioritization.

On completion of the questionnaires, the research associate provided a short debriefing, then collected any notes the subjects may have taken. Subjects were allowed to read and leave with a copy of the debrief form at the conclusion of the two-hour experiment block.

Data Analysis Method

The research associates uploaded all individual subject data files to a secure NPS file server. All of the subject responses to both the categoriziation/prioritization exercise and the questionnaire were anonymized and aggregated into a master Excel spreadsheet. For analysis purposes, the pilot study results were included in the final dataset, bringing the total number of participants to five.

The initial data analysis approach was to apply a content analysis approach to the qualitative data collected from the subjects and apply descriptive and inferential statistics to the quantitative data. The low number of subjects that responded to the recruitment campaign limited the usefulness of inferential statistics. Instead, only frequency counts, averages, standard deviations, and pairwise comparisons of the numerical data were performed.

Results Summary

The participants included one NPS employee and four NPS students. Two students were from the SE curriculum, and two were from the Systems Acquisition Management curriculum. Two of the students were current active duty, and two were civilians. Across all five subjects, the reported bachelor's degrees included communication studies, mechanical engineering, business management, and oceanography. The reported master's degrees included management, aerospace engineering, and national security and strategic studies. No subjects held a PhD in any field.

Only two subjects had prior experience evaluating T&E data, each reporting five and seven years of experience. Three of the five subjects were assigned to the content analysis training condition; two did not receive the training. Both subjects in the non-training condition took slightly more than an hour to complete the classification and prioritization task, as did one of the subjects who received the training. The other two subjects in the training condition took less than one hour to complete the task. Across the five subjects, the average time to complete these tasks was 58 minutes.

Categorization Results

Table 1 shows a sample of the results of the categorization exercise for the flight simulator Part II issues. The results were grouped by training condition to highlight any substantial similarities and/or differences between the two subject groups. Subjects 1 and 4 created one category scheme, while the remaining subjects created two category schemes. Subject 3 was the only person to incorporate the test personnel prioritizations into their categorization and prioritization scheme. Subjects 1 and 3, who were both in the training condition, had the most similar hardware and software categorizations. Subjects 2 and 5 created categories related to specific types of hardware, software, and other system elements (e.g., instructor, procedure). Of particular interest is the fact that four out of five subjects created a scheme with an inherent or defined hierarchy. Even Subject 3, who used the test personnel issue priority values, assigned an order of precedence to the second category set: (1) Additional information required/Possible Part I, (2) Hardware functionality missing/Testing not completed, (3) Software bug functionality missing/Testing not completed, (4) Software bug, (5) Non-functional hardware deficiency.

Table 1. Sample Part II Deficiency Categorization Results

Issue #	Issue Title	Category Subject 1 (T)	Category Subject 3 (T)	Category Subject 5 (T)	Category Subject 2 (NT)	Category Subject 4 (NT)
6	Missing Battery Indicator	Hardware	Part II. Hardware functionality missing. Testing not completed.	Ancillary Priority D	Physical component, Part III	Minor
7	Headset Mic Problems	Hardware	Part II. Additional information required on availability of workaround and what the contract specified. Potential to be a Part I.	Ancillary, Priority D	Interface Part II	Major
8	Instructor Station—Screen capture software test incomplete	Hardware	Part II. Hardware functionality missing. Testing not completed.	Instructor, Priority B	Data capture, Part III	Minor
9	Digital Map malfunction	Simulation Software	Part II. Software bug.	Cockpit, Priority B	Procedure mismatch, Part II	Critical
13	Flap display not working	Hardware	Part II. Software bug.	Cockpit Priority B	Procedure mismatch, Part I	Minor
15	Visual Scene—Time of Day mismatch	Simulation Software	Part II. Software bug.	Visual, Priority C	Visual system delta, Part III	Critical
22	Trainer automatic power shutdown did not work	Hardware	Part II. Software bug? Functionality missing. Testing not completed.	Ancillary, (safety), Priority A	Physical component, Part I*	Major

Key: (T) – Training condition; (NT) – Non-training condition

It is noteworthy that the two subjects assigned to the non-training condition seemed to leverage the NAWCTSD deficiency code definitions provided in the T&E report to create their categories. Table 2 shows a sample of the results of the categorization exercise for the flight simulator Part III issues.

Table 2. Sample Part III Deficiency Categorization Results

Issue #	Issue Title	Category Subject 1 (T)	Category Subject 3 (T)	Category Subject 5 (T)	Category Subject 2 (NT)	Category Subject 4 (NT)
1	Coldstart media missing	Technical Software	Part III. Hardware functionality missing. Testing not completed.	Data, Priority A	Physical component, Part I	Critical
2	Can't play back recorded mission	Technical Software	Part III. Software bug.	Instructor, Priority A	Data capture, Part I	Major
4	Lighting system mismatch	Hardware	Part III. Hardware functionality missing. Testing not completed.	Cockpit, Priority D	Physical component, Part II	Minor
10	Ice Shedding/ Removal	Simulation Software	Part III. Software bug.	Visual, Priority C	Procedure mismatch, Part III	Major
11	Gross Weight	Simulation Software	Part III. Software bug.	Instructor, Priority B	Procedure mismatch, Part III	Critical
12	Engine Fire Extinguisher malfunction buttons	Hardware	Part III. Software bug.	Cockpit, Priority A	Procedure mismatch, Part I	Safety/critical
23	No audio captured in mission recording	Technical Software	Part III. Additional information required on availability of workaround and what the contract specified. Potential to be a Part I.	Instructor, Priority A	Data capture, Part I	Critical

Key: (T) – Training condition; (NT) – Non-training condition

The results from Tables 1 and 2 highlight the differences in approach to assigning issues to the created categories. Given the aforementioned observations on the categorization strategies used by the test subjects, it appears that subjects used heuristics to focus on high-level attributes of the system, perhaps as a way to manage and consolidate the data in a meaningful way. Each subject made a judgment of circumstance, scope, and criticality using the provided descriptions of each issue and their own interpretations and mental model of each issue. Despite the similarities in some of the category names, each person's working definition of these categories was different enough to preclude the same issues all being assigned to the same categories. It is difficult to tell what their categories would have looked like if they had been specifically instructed to use the test personnel prioritizations. Based on these results, the impact of the content analysis training was inconclusive.

Prioritization Results

Each subject was asked to first categorize the issues and then prioritize the issues for remediation. Table 3 lists the assigned priority numbers for the Part II issues. The results were grouped by subject training condition to highlight any substantial similarities and/or differences between the two subject groups.

As directed by the experiment instructions, subjects were specifically asked to assign a unique priority number to each issue, without duplication of ranking (i.e., two or more deficiencies cannot be assigned the same priority number). Subjects 3, 4, and 5 used a 1–25 scale and assigned a unique resolution priority number to each issue. For the remaining two subjects,

- Subject 2 assigned all issues either a 1, 2, or 3. Even though this person created two category schemes, only the scheme with the inherent hierarchy (Part I, Part I*, Part II, Part III) was used for resolution prioritization. This resulted in multiple #1, #2 and #3 issues that require further prioritization within each of these subsets.
- Subject 1 used a scale dependent upon the number of issues in each category. In other words, the 10 issues assigned to the "hardware" category were assigned resolution priority numbers 1–10. The twelve issues assigned to the "simulation software" category were assigned resolution priority numbers 1–12. The three issues assigned to the "technical software" category were assigned resolution priority numbers 1–3. This strategy also resulted in multiple issues with the same resolution priority ranking that require further prioritization within each of these subsets.

There were 25 issues total: 11 Part II and 14 Part III. It was expected that all of the Part II issues would appear within the top 11 rankings of the prioritization list had the subjects leveraged the priority from the test personnel. As shown in Table 3, this was the case for Subject 3. For Subjects 4 and 5, who also used a 1–25 scale, this was not the case because of their interpretation of the issues and the categories they used. It is noteworthy that Subjects 3, 4, and 5 rated only one issue the same resolution priority number (Part III issue 20).

Table 3. Part II Deficiency Prioritization Results

Issue #	Issue Title	Priority Assigned by Test Personnel	Priority for Subject 1 (T)	Priority for Subject 3 (T)	Priority for Subject 5 (T)	Priority for Subject 2 (NT)	Priority for Subject 4 (NT)
6	Missing Battery Indicator	II	6	2	24	2	24
7	Headset Mic Problem	II	2	1	22	2	9
8	Instructor Station–Screen capture software test incomplete	II	5	3	8	3	22
9	Digital Map malfunction	II	5	8	10	2	5
13	Flap display not working	II	4	5	9	1	21
15	Visual Scene–Time of Day mismatch	II	11	9	11	3	6
17	Incorrect weather depiction	II	4	10	18	3	8
18	Cross winds setup	II	2	6	4	1	17
19	Night FLIR not working	II	3	7	12	1	19
22	Trainer automatic power shutdown did not work	II	1	4	1	1	15
25	Weather visual scene and cockpit display mismatch	II	1	11	19	3	16

Priority Ranking Statistics

Table 4 summarizes the priority rankings assigned by the five subjects to each of the 25 deficiencies. Subjects 1 and 2 did not follow the instructions given to them to assign a unique priority ranking to each deficiency. Their responses are presented for completeness but grayed out to indicate their incompatibility for use in any statistics. The average and standard deviation of the rankings by Subjects 3, 4, and 5 are shown at the right of the table. A low standard deviation (like issues 20, 5, and 9) indicates closer agreement among the subjects than those issues with large standard deviations like issues 6, 14, 7, and 3.

Since the same average ranking could be obtained from different sets of widely differing data, it is instructive (given the small number of subjects) to do a pairwise comparison of rankings between subjects.

Figure 2 shows graphically the spread of priority rankings for the 25 deficiencies between pairs of subjects. Such a graph highlights issues where there was close agreement (e.g., issue 20) and wide disagreement, such as Subjects 3 and 4 on issues 10, 11, 12, 13, and 14.

Table 4. Average and Standard Deviation of Issue Priority Ranking by Subjects 3, 4, and 5

Issue Prioritization by Subject #						Subjects 3-5	
Issue #	Subj 1	Subj 2	Subj 3	Subj 4	Subj 5	Average	Std dev
1	3	1	15	3	2	6.7	7.234
2	1	1	16	10	6	10.7	5.033
3	9	3	23	18	3	14.7	10.408
4	7	2	13	23	25	20.3	6.429
5	8	3	14	14	15	14.3	0.577
6	6	2	2	24	24	16.7	12.702
7	2	2	1	9	22	10.7	10.599
8	5	3	3	22	8	11.0	9.849
9	5	2	8	5	10	7.7	2.517
10	7	3	22	11	16	16.3	5.508
11	6	3	18	4	13	11.7	7.095
12	3	1	17	2	7	8.7	7.638
13	4	1	5	21	9	11.7	8.327
14	12	3	21	1	21	14.3	11.547
15	11	3	9	6	11	8.7	2.517
16	10	3	25	25	17	22.3	4.619
17	4	3	10	8	18	12.0	5.292
18	2	1	6	17	4	9.0	7.000
19	3	1	7	19	12	12.7	6.028
20	8	3	20	20	20	20.0	0.000
21	10	3	24	12	23	19.7	6.658
22	1	1	4	15	1	6.7	7.371
23	2	1	12	7	5	8.0	3.606
24	9	3	19	13	14	15.3	3.215
25	1	3	11	16	19	15.3	4.041

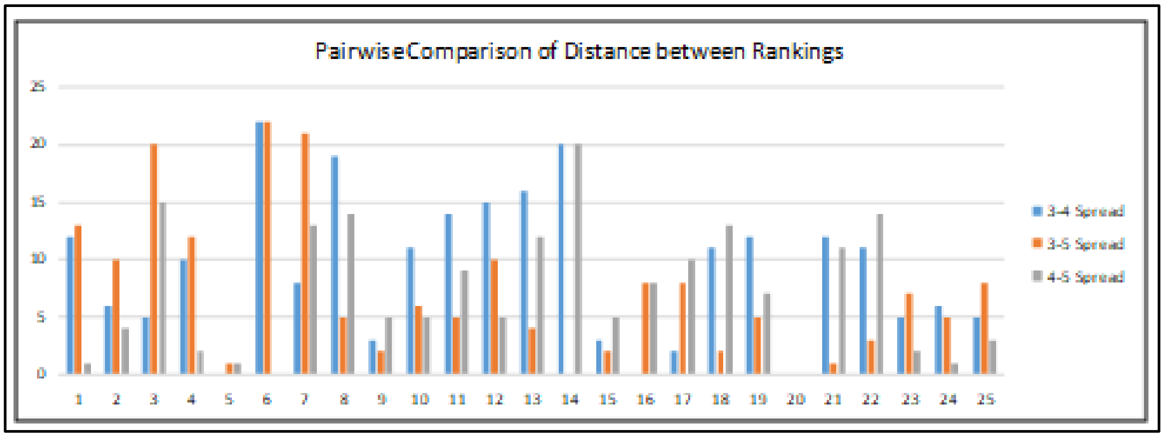


Figure 2. Pairwise Comparison of Distance Between Rankings by Issue

Given these findings, no statistically significant differences were observed in the perceived value, between those in the training condition and those in the control condition. Once again, based on these results, the impact of the content analysis training was inconclusive.

Classification Strategies Questionnaire Results

In the classification questionnaire, subjects were asked to describe the rationale they used to create categories and assign a resolution priority number to each issue. Figure 3 shows the reported answers summarized into seven categories.

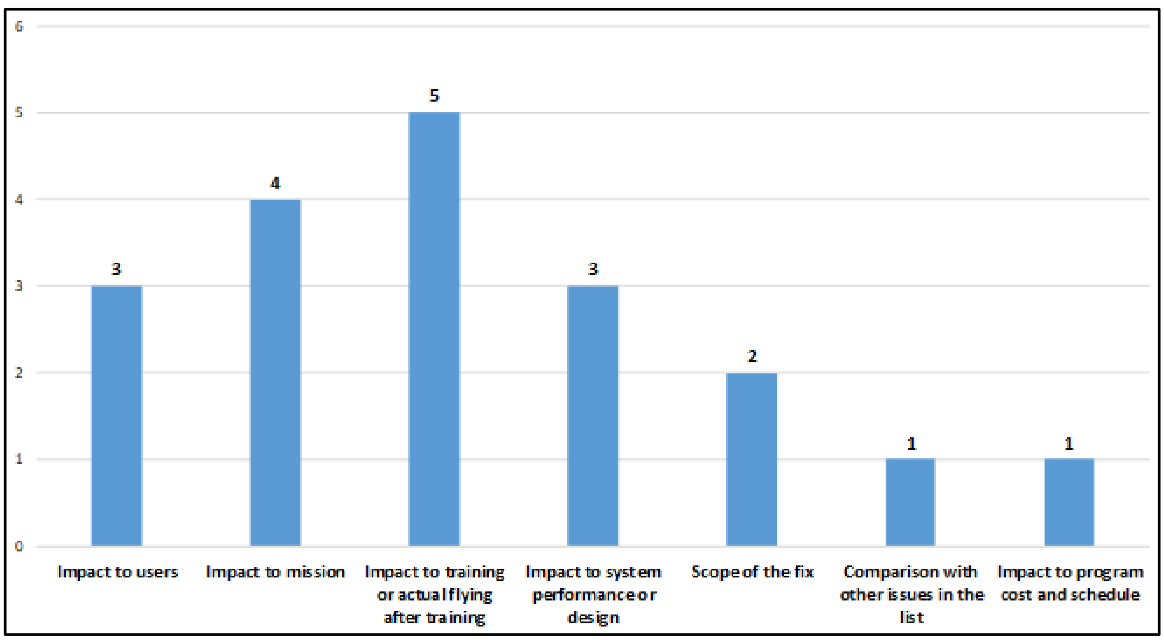


Figure 3. Counts of Reported Rationale

Impact to users, impact to mission, and impact to training or actual flying after training seem similar and could possibly be consolidated. However, more detailed rationale is required to group them together. No noticeable differences between subjects in the training versus non-training condition were found. It is interesting to note that one subject specifically noted looking for "patterns of deficiencies" as a classification strategy. This subject was in the non-training condition but did have a background in the T&E domain.

Prior problem solving was a commonly cited theme across all subjects when asked if they leveraged anything from their previous training or experience.

Workload Assessment Questionnaire Results

Table 5 summarizes the subject responses to the workload assessment questionnaires. Subjects were asked to rate their perceived level of workload on a number of factors, on a scale from 1 to 10, with “1” reflecting a “poor” level and “10” being a “good” level.

Table 5. Workload Assessment Questionnaire Results

	Mental	Temporal Demand	Performance	Effort	Frustration Level
Subject 2 (NT)	4	5	7	6	3
Subject 4 (NT)	9	10	8	10	9
Subject 1 (T)	9	9	4	6	4
Subject 3 (T)	8	2	6	8	3
Subject 5 (T)	6	6	4	5	3
Average Rating:	**7**	**6**	**6**	**7**	**4**

In general, subjects in the training condition rated the mental demand to be high, but the frustration level low. The high scores for Subject 4 in the non-training condition were attributed to the fact that this person was an international, non-native-English-speaking student who had no prior T&E experience. Subjects 2 and 3, who rated the lowest temporal demand, were the ones that took the longest to complete the task. Even though subjects were told they had up to one hour to complete the categorization and prioritization tasks, Subjects 2 and 3 exceeded the allotted hour by 8 minutes and 14 minutes, respectively. Subjects 1 and 4, who reported the highest mental demand and temporal demand scores, both used one category scheme to group similar issues, judge issue severity, and come up with a resolution priority order.

An interesting observation on the performance attribute is that those in the training condition rated their overall level of satisfaction with completing the tasks lower than those in the non-training condition. Additional data is needed to determine an explanation.

Perceived Value Questionnaire Results

The subjects rated two factors: (1) the value of categorizing deficiencies before prioritizing them, and (2) the impact of categorizing on prioritization order. The subjects were asked to use a scale from 1 to 10, with “1” reflecting a “low” perceived value and “10” being a “high” perceived value. Table 6 summarizes these responses. No significant differences were observed between those in the training condition and the control condition on the value scores.

Table 6. Value and Impact of Categories Questionnaire Results

Subject	Value of Categories	Impact of Categories
Subject 2 (NT)	No score provided	No score provided
Subject 4 (NT)	10	10
Subject 1 (T)	10	10
Subject 3 (T)	6	8
Subject 5 (T)	8	1
Average Rating:	8	6

Subject 2 provided comments for both of these questions instead of a numerical score. Subject 2 described categorizing the deficiencies after prioritizing them and stated the belief that mission impact is the most important consideration in prioritization. An interesting observation is that Subjects 2 and 5 used two category schemes: one with an inherent hierarchy and one specific to system characteristics. However, neither of them used the latter during the prioritization task. This also explains the low impact score provided by Subject 5. The remaining Subjects 1, 3, and 4 all used their categories to help them assign resolution priority numbers to the issues.

Discussion

The goal of this study was to evaluate ways that deficiencies are classified into categories using available information and how the correction of the deficiencies is prioritized using those categories.

The first hypothesis for this study was that subjects in the content analysis training condition would produce more well-defined categories than those in the non-training condition. No firm conclusion could be made regarding any training impact on the types of categories created or the number of category schemes used.

The second hypothesis for this study was that the perceived difficulty of the categorization and prioritization tasks (i.e., frustration level, mental and temporal demand, etc.) would be higher for those subjects in the non-training condition. Based on the workload assessment results, no training impact was observed. The determining factors of perceived difficulty were the types of categories created and the number of category schemes used.

The third hypothesis for this study was that participants would leverage the issue prioritization assigned by the test personnel in order to come up with a resolution priority order. This strategy was expected by all participants, regardless of training condition. Only one subject actually used the test personnel categorizations. The remaining four subjects created their own criteria to judge each issue's technical priority in order to sort them for resolution. Only one subject explained why they did not use the issue priority assigned by the test personnel. In this subject's opinion, test personnel often do not have adequate training or operational experience as a system user to judge the criticality of issues identified during test. It should be noted that this bias was stated by a subject that self-reported no prior T&E experience.

All subjects realized a need to judge the severity of each issue using the information provided and their own experience with classification and prioritization to come up with a resolution priority order. However, the strategies they used were very different, with a high degree of subjectivity in methodology used. It was not possible to determine which interpretations and approaches were the most efficient in terms of time to complete and level of effort. There were no apparent correlations between educational background, prior T&E experience, and strategy used. With a greater number of study participants, more repetition in similar strategies might have been observed.

Future Research

Because of the small number of participants recruited in this study, it would be worth repeating, but with incentives provided to increase volunteer enrollment. The results of this study indicate that using both a technology-based and priority-based categorization scheme might produce results that are more consistent across subjects. It would be interesting to revise this research study to investigate how subjects assign issues to pre-defined technology-based and priority-based categories provided to them. Another variation would be to pre-assign issues to such categories and then ask subjects to create a resolution priority order. Finally, it seems worth investigating the preferences people have for resolution prioritization criteria. The results of this study indicate a preference for ordinal versus interval criteria and measurement scales.

The ultimate objective of further research in this topic is to generate a categorization and prioritization scheme that produces consistent results across personnel from a variety of backgrounds. Ideally, with a valid scheme, the only key differentiating factor between personnel would be their level of domain knowledge and T&E experience with a specific type of system. With such a scheme identified, further research to develop software tools and/or training for workforce development would be logical next steps.

References

Birks, M., & Mills, J. (2012). *Grounded theory: A practical guide.* Thousand Oaks, CA: Sage.

Cropp, N., Banks, A., & Elghali, L. (2011). Expert decision making in a complex engineering environment: A comparison of the lens model, explanatory coherence, and matching heuristics. *Journal of Cognitive Engineering and Decision Making, 5*(3), 255–276. doi: 10.1177/1555343411415795

DoD. (2012). *Test and evaluation management guide.* Retrieved from https://acc.dau.mil/docs/temg/Test%20and%20Evaluation%20Management%20Guide,%20December%202012,%206th%20Edition%20-v1.pdf

Henningsson, K. & Wohlin, C. (2004). Assuring fault classification agreement—An empirical evaluation. In *Proceedings of the 2004 International Symposium on Empirical Software Engineering*. doi: 10.1109/ISESE.2004.1334897

Holness, K. S. (2016). Content analysis in systems engineering acquisition activities. In *Proceedings of the 13th Annual Acquisition Research Symposium* (Vol. 1, pp. 57–62). Monterey, CA: Naval Postgraduate School. Retrieved from http://www.acquisitionresearch.net/files/FY2016/SYM-AM-16-025.pdf

International Council on Systems Engineering (INCOSE). (2015). *Systems engineering handbook: A guide for system life cycle processes and activities.* Hoboken, NJ: John Wiley and Sons.

Kenett, R. S., & Baker, E. R. (2010*). Process improvement and CMMI for systems and software* [Online version]. Retrieved from https://doi.org/10.1201/9781420060515-c6

Kossiakoff, A., Sweet, W. N., Seymour, S. J., & Biemer, S. M. (2011). *Systems engineering principles and practice*. Hoboken, NJ: John Wiley and Sons.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage.

Lafond, D., Vallieres, B. R., Vachon, F., St. Louis, M., & Tremblay, S. (2015). Capturing nonlinear judgment policies using decision tree models of classification behavior. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59*(1), 831–835. doi: 10.1177/1541931215591251

Linkov, I., Satterstrom, F. K., & Fenton, G. P. (2009). Prioritization of capability gaps for joint small arms program using multi-criteria decision analysis. *Journal of Multi-Criteria Decision Analysis, 16,* 179–185. doi: 10.1002/mcda.446

Memorandum of agreement (MOA) on multi-service operational test and evaluation (MOT&E) and operational suitability terminology and definitions. (2010). Retrieved from http://www.public.navy.mil/cotf/OTD/OTA%20MOTE%20MOA.pdf

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis.* Thousand Oaks, CA: Sage.

Naval Air Warfare Center Training Systems Division (NAWCTSD). (n.d.). Deficiency reporting for training system testing. Retrieved from http://www.navair.navy.mil/nawctsd/Resources/Library/Acqguide/testingdeficiencyreporting.htm

Patton, M. Q. (2015). *Qualitative research & evaluation methods*. Thousand Oaks, CA: Sage.

Wasson, C. S. (2006). *System analysis, design, and development.* Hoboken, NJ: John Wiley and Sons.

Panel 5. New Dimensions in Managing Systems of Systems

Wednesday, May 9, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Rear Admiral Jon Hill, USN,** Deputy Director, Missile Defense Agency ***Capability Composition and Data Interoperability to Achieve More Effective Results Than DoD System-of-Systems Strategies*** Nickolas H. Guertin, Carnegie Mellon University Douglas C. Schmidt, Vanderbilt University William Scherlis, Carnegie Mellon University ***Managing Complex Systems Engineering and Acquisition Through Lead Systems Integration*** Ronald Carlson, Naval Postgraduate School Warren Vaneman, Naval Postgraduate School ***Inherent Moral Hazards in Acquisition: Improving Contractor Cooperation in Government as the Integrator (GATI) Programs*** William Novak, Carnegie Mellon University Julie Cohen, Carnegie Mellon University Andrew Moore, Carnegie Mellon University William Casey, Carnegie Mellon University Bud Mishra, New York University

Rear Admiral Jon Hill, USN—is a native of Texas, born and raised on Fort Bliss. A surface warfare officer designated as an engineering duty officer, he is a graduate of Saint Mary's University and earned a Master of Science in Applied Physics and Ordnance Engineering from Naval Postgraduate School.

Hill's first flag officer tour was Program Executive Officer for Integrated Warfare Systems (PEO IWS). In this role, he was accountable for developing, delivering, and sustaining all surface ship combat control systems, radars, missiles, launchers, electronic warfare, naval gunnery systems, and surface and subsurface anti-submarine warfare mission capabilities.

Other leadership and acquisition engineering positions include AEGIS Shipbuilding (PMS 400), Naval Surface Warfare Center (NSWC) Dahlgren Division and Port Hueneme Division, PEO Theater Surface Combatants, and on the Assistant Secretary of the Navy staff for Research, Development and Acquisition (ASN RD&A).

He also served on the Joint Staff (J-6), U.S. Army Staff for Missile Systems, and as a senior fellow on the Chief of Naval Operations Strategic Studies Group (CNO SSG XXVII). He served the Missile Defense Agency (MDA) as technical director for Aegis Ballistic Missile Defense then as Aegis Combat Systems major program manager responsible for delivering Naval Integrated Fire Control and Counter Air (NIFC-CA) and Integrated Air and Missile Defense (IAMD) capabilities to forces afloat.

Hill assumed duties as deputy director of the Missile Defense Agency in November 2016. He advises the director, MDA, in fielding the Ballistic Missile Defense System (BMDS) and manages MDA support to BMDS operational programs. Hill supports the development and implementation of BMDS policy, capabilities, priorities, and resources, and serves as the operational interface with the services, combatant commands, Joint Staff, and allies.

Personal awards include the Navy Distinguished Service Medal, Defense Superior Service Medal, Legion of Merit (two awards), Defense Meritorious Service Medal, Meritorious Service Medal (three awards), Joint Service Commendation Medal, U.S. Army Commendation Medal, Navy and Marine Corps Commendation Medal (two awards), and Navy Achievement Medal (two awards).

Capability Composition and Data Interoperability to Achieve More Effective Results Than DoD System-of-Systems Strategies

Nickolas H. Guertin—is a Senior Systems Engineer with the Software Engineering Institute working in the Intelligence and Defense Acquisition Communities. He received a BS in mechanical engineering from the University of Washington and a MBA from Bryant University. He is a registered Professional Engineer and is Defense Acquisition Workforce Improvement Act–certified in Program Management and Engineering. Guertin served 23 years in the Navy in submarines and as a retired Navy Reserve Engineering Duty Officer for nuclear propulsion, ship construction, and repair. He is also retired from the civil service with a career in undersea weapons, sensors, combat systems, and enterprise transformation. [nhguertin@sei.cmu.edu]

Douglas C. Schmidt—is the Cornelius Vanderbilt Professor of Computer Science at Vanderbilt University and a Visiting Scientist at the Software Engineering Institute. His research covers software patterns, optimization techniques, and analyses of middleware frameworks for distributed real-time embedded systems. From 2010 to 2014, he served as a member of the Air Force Scientific Advisory Board, and from 2000 to 2003, Dr. Schmidt served as a Deputy Office Director and a Program Manager at DARPA. Dr. Schmidt received BS and MA degrees in sociology from the College of William and Mary in Williamsburg, VA, and an MS and a PhD in computer science from the University of California, Irvine (UCI). [d.schmidt@vanderbilt.edu]

William Scherlis—is a full Professor and Director of the Institute for Software Research (ISR) in the School of Computer Science at Carnegie Mellon. His research relates to software assurance, software evolution, and technology to support software teams. He joined the Carnegie Mellon faculty after completing a PhD in computer science at Stanford University, a year at the University of Edinburgh (Scotland) as a John Knox Fellow, and an AB magna cum laude at Harvard University in applied mathematics. Dr. Scherlis served at DARPA for more than six years. He is a Fellow of the IEEE and a Lifetime National Associate of the National Academy of Sciences. [scherlis@cs.cmu.edu]

Abstract

This paper investigates how layered business and technical architectures can leverage modular component design practices to establish new approaches for capability acquisition that are more effective than existing “system of systems” (SoS) strategies. We first examine proven methods, approaches, and patterns for crafting large-scale services, real-time capabilities, and military-specific Internet of Things (IoT). We then propose elements of a new approach that applies a coherent set of methods to develop military mission capabilities as sets of composed modules.

Our approach builds on a broad range of prior work related to functional decomposition of requirements into modules of capabilities for deployment in an open environment. We also extend prior work related to using technical reference frameworks as foundations for modules that meet capability needs. We tie this prior work with emerging development practices to describe a new approach for crafting capability. Finally, we assemble these findings into a new overarching model of financial, organizational, programmatic, quality-management, and business patterns needed to deliver payloads onto fighting platforms more effectively. Implementing the recommendations in this paper will establish a DoD acquisition environment shaped to be more efficient and deliver much higher quality—with far greater innovation—in a fraction of the time.

Introduction

The warfighting capability employed by the United States is, for now, the envy of all nations. We have made incremental changes in our acquisition practices for building and deploying military capacity. This capacity can be viewed as “platforms” (e.g., tanks, ships, aircraft, etc.) and the mission system “payloads” (e.g., sensors, command and control, weapons, etc.) that are populated onto those platforms to deliver the desired capability (Greenert, 2012). The requirements to design these systems have historically been defined independently to address specific military gaps. Moreover, upgradability and extensibility were not widely perceived as military requirements at the time they were created. These systems have evolved to become more software-reliant over time, and that trend is increasing (Scherlis et al., 2010).

Performance improvement by upgrading the existing portfolio of systems, using the existing pattern of activities, has been perceived as lower risk, taking less time, and being more affordable than instantiating a new product. Those existing products, however, were not initially designed to support incremental upgrades or even routine ongoing software and hardware sustainment. They were instead purpose-built and are therefore not architecturally structured to scale and address adjacent solution opportunities. As a result, the current capacity for breadth and pace of change is impeding our ability to evolve capability quickly and robustly enough to meet new requirements in emerging technical and warfighting environments.

Technologies that we use to build these cyber-physical/software-intensive systems are widely available to all nations and non-state actors. The practices that were successful in the past for incorporating *commercial-off-the-shelf* (COTS) technologies, on a system-by-system basis, are insufficient by themselves to meet these rapidly evolving challenges. To stay ahead of our adversaries—and continuously increase our pace of change for delivering innovation—the DoD needs new approaches that achieve rapid delivery, flexibility, and capacity to provide continuous improvement to fielded capability.

Military capability provides differentiation between belligerents. In addition, adversaries benefit from our impediments to responsiveness that are self-inflicted from our approaches to acquisition, testing, and evaluation. If the building blocks for crafting military capabilities are all available in COTS form, then all nations could end up on the same playing field for military capability and warfighting advantage. Our nation both needs and deserves unfair advantage wrought by having different and better performing products. Achieving this goal requires new approaches for capability architectures that are intentionally designed to support a military capability requirement for upgradability and responsiveness. In particular, cyber-adversaries are very nimble, so our approach thus enables nimble responses to nimble adversaries.

The remainder of this paper is organized as follows. First we describe several emerging opportunities related to the trend towards modularity and open systems architectures; then we examine key change drivers and technical/organization structures associated with the new model of acquisition we propose for the DoD; next we examine the impacts associated with the implementation and organizational structure of our proposed acquisition model; and finally, we summarize our recommendations and present concluding remarks.

Emerging Trends and Opportunities

Addressing the limitations with conventional acquisition approaches described earlier requires a new set of business and technical practices to achieve different results and more advanced capacities than our adversaries. In particular, new acquisition structure and associated technical architecture are needed to harness the innovation engines of all sectors of the American and global economies. The leading characteristic of applying *modularity* to an *open system architecture* (MOSA) approach is that different components can be created by independent parties and can evolve at different rates.

When the DoD relies on the ecosystem that makes MOSA attractive, it loses some control but gains by "riding the growth curves" of capability and quality. As such, conventional approaches must be rethought at every level, including the ways the DoD (1) funds capabilities, (2) organizes these capabilities to create new products, (3) builds and assesses quality, (4) converts those quality innovations into affordable, broadly usable capacities that are reliable and delivered rapidly, and (5) continues to evolve and modernize products and their components.

Examples of Modular Open System Architectures Adoption in the DoD

Segments of the DoD have aggressively innovated their acquisition practices in the past. In each case, there was a "burning platform" to drive a capability need and/or a financial/programmatic change, including the following:

- The Navy's Program Executive Office (PEO) for Submarines instituted the Advanced Processor Build and Technology Insertion (APB/TI) process. This multifaceted and phased approach provided dramatic performance improvement that was validated through peer-reviewed and independent measurement and analysis. Full commitment to wholesale replacement of submarine combat systems involved new approaches to delivering these systems into both new construction and existing classes. To apply all available resources to the transition, the Navy abandoned support for legacy MIL-SPEC products to concentrate on employing new capabilities and functional performance to a demanding customer (Guertin & Miller, 1998). This submarine-focused federated system-of-system construct improved enterprise value and supported integration of innovation.
- The Navy's PEO for C4I systems performed an enterprise architecture approach to provide common compute-plant and capability integration environment under the *Consolidated Afloat Network Enterprise Services* (CANES). This initiative collected together infrastructure needs and provided a landing pad for the Navy's C4I suite. Though a powerful example, CANES is programmatically applied only to the PEO C4Is family of systems.
- The Army's PEO for Aviation has declared the Future Airborne Capability Environment (FACE^{tm}) open standard as the Common Operating Environment (COE) for their new capability development (Adams, 2014). The strategic vision for the Army's use of FACE is to open up opportunities for multiple offerors of innovation, improve interoperability, and reduce the cost and time for capability indoctrination ("Future Vertical Lift," 2018). Industry supports the FACE^{tm} approach for three primary reasons: (1) to avoid being left behind as others find new opportunities and (2) to take advantage of new methods to improve internal corporate efficiency, as well as to (3) increase market share and increase profits (Nichols, 2017). The government's incentive for creating and continued participation is to enable increased

productivity and effectiveness, especially for integration and interoperability, as well as to reduce programmatic risk.

- The Air Force is developing the *Open Mission Systems* (OMS) specification, which is a non-proprietary architectural standard designed to enable affordable technical refresh and insertion, simplified mission systems integration, service reuse and interoperability, and competition between suppliers across the life cycle. Industry and the government have developed and agree upon a set of open key interfaces and architectural guidelines to achieve the goals of OMS (Unmanned Aerospace Systems, 2014).

Trends and Opportunities Enabled by Advances in Technology and Strategy

The changes in underlying COTS technologies used by the MOSA-enabled DoD programs described in the previous section have continued to evolve due to innovations in software technologies and architectures. It is now feasible to address backward compatibility and to use a variety of hardware implementations in any one system instantiation or data center while new technologies continue to evolve (e.g., using Graphic Processor Units specifically for performing Artificial Intelligence processing). This change in market dynamics enables greater support for backward compatibility of software onto other operating system and hardware environments by invoking widely used standards (Schmidt et al., 2013).

The COTS software building blocks available to develop, deliver, and manage capability have matured in the commercial market. It is now time to take a fresh look at how the acquisition, testing, and resourcing communities are structured to develop and rapidly deliver highly reliable, intuitively operable, innovation in warfighting capability.

One enabling step recently taken by the Navy was the establishment of the Digital Warfare Office (DWO). The DWO is *inter alia* a leader in the area of decomposing the performance attributes of a system into functions (Serbu, 2017). The DWO focuses on methods for decomposing capability into elements that are internally tightly coupled, but loosely coupled externally, which can then be applied to illuminate software modules needed to deliver the required military performance. It would be tempting to stop there and create a specification for a system that would be comprised of these functional elements. To reach greater performance and speed capability deliver, however, the Navy must then extend this logic to structure the technical architecture to facilitate continuous delivery of innovations and avoid current independent system-based delivery epochs, which classically stretch from two to five years (DeLuca et al., 2013).

An Architecture First Approach

A new “architecture-first” strategy is thus needed that addresses enterprise performance equities, conformant quality attributes, and managed variability while sustaining minimally-coupled and inherently interoperable designs. This strategy should establish a framework of support infrastructure that provides an integration environment in which modules of capability can be a hosted. New development methods and architectural constructs facilitate loose-coupling of capabilities and deployment of software onto containerized or virtualized environments, thereby eliminating the need for hardware-dependent deployments.

The primary function of an “architecture-first” strategy is to establish rules of construction. These rules are set to ensure quality attributes are known and followed throughout the life cycle of a warfighting system. Likewise, these rules also ensure that loose coupling (which enhances systematic reuse), low cyclicality (which is a metric that illuminates corruption of the benefits of modularity through overindulgent interplay across modules), and that strategic architectural attributes (often called "non-functional requirements") are addressed.

While components and functions are separated, it is nonetheless the case that mission capability can be manifested to operators in a tightly-integrated manner. This integration is a consequence of effectively "matrixing" component capabilities and the design of both end-user experience and the application program interfaces for traditional SoS.

Trends and Opportunities Enabled by Advances in Hardware

Advances in COTS hardware are enabling new opportunities for a hardware support model that facilitates continuous deployment of warfighting capability. The past practice has been to configure a specific hardware baseline, procure precisely those parts as a block-buy that will last the life of the deployed configuration (anywhere from 10 to 20 years), and then plan for the program to not run out of spare parts. Block-upgrades, however, are not a sustainable business model for the commercial sector.

Innovations in hardware sustainment strategies have fundamentally changed the methods and mechanisms of retaining high-end capability needed by any organization whose business depends on modern data centers and cloud computing environments. These technologies support advanced software-centric technologies, such as virtualization and modularization. Commercial organizations, such as Google, Amazon, LinkedIn, and Facebook, apply these technologies to continuously upgrade their data centers with new hardware in a manner that allows them to deploy new software capabilities rapidly and reliably (Clark, 2004).

Forerunners of Advances in Acquisition Models

In recent years, various DoD efforts have combined several Programs-of-Record (PoRs) to improve efficiency and to "commonly do what is commonly done." For example, the SubLAN architecture from 2004 was a progenitor of the broader naval effort for providing Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) capacity to host nontactical and crew services capabilities under the auspices of the Consolidated Afloat Network Enterprise Services (CANES; Anderson, 2009). As mentioned earlier, CANES is consolidating and modernizing shipboard, submarine, and shore-based command, control, communications, computers and intelligence (C4I) networked systems to increase capability and affordability.

All Transformation Efforts Must Address Culture

The largest challenges faced by the enterprise-focused transformation effort of CANES were programmatic and cultural. The value proposition of integrating common artifacts and components that were not initially designed for common use was relatively straightforward to articulate (Wang et al., 2004). PEO C4I also established other programmatic elements, including a shared and evolving build environment, the capacity to host a wide array of capabilities, and a PEO C4I organizational policy of rewarding creation of common elements. These behaviors are antithetical to the classic acquisition behavior of protecting the PoR and preferring a system-by-system go-it-alone approach. PEOs will have to face portfolio optimization issues directly if they wish to pursue a completely redesigned approach for *continuous delivery* of modularized, advanced, reliable, and innovative capability into a continuously modernized and shared environment.

The Defense Acquisition Executive (formerly USD[AT&L]) has recently been split into the Undersecretaries of Research & Engineering (R&E) and Acquisition & Sustainment (A&S). This decomposition is illuminating for a path on how to organize around the principles of focusing on innovation for the warfighting domain (the R&E portfolio), while devising a highly reliable and flexible integration environment for those innovations (the responsibilities of A&S). One of the most valuable outcomes of splitting these activities is the acknowledgement that each entity works on different activities, with different skillsets and business drivers, yet each must depend on the other if either is to succeed. The organizational construct of the former USD(AT&L) was predicated on a different strategy and orientation of engagement for oversight of acquisitions performed by the military services. The existing staff will have to undergo a deep culture change if the split into R&E and A&S is to succeed.

Towards a New Model of Acquisition for the DoD

It is widely recognized that the DoD needs to have nimble response to nimble adversaries. Incremental improvement to existing capabilities, granular delivery of new “payloads,” and the ability to continuously deliver to the military platform. The current pattern of upgrading ships and aircraft applies a system-by-system, rip-out and installation process. This pattern, however, incurs prolonged periods to upgrade capabilities and reduced operational availability, and makes interoperability more challenging.

Another area that is widely agreed to in principle—but has been even more elusive in practice to achieve—is taking successful prototypes and productionizing the capability with excellent quality, full support, and training. The benefit of rapidly attempting new ideas and quickly declaring success or failure may be lost, however, if those prototypes are fielded in a way that does not match the business needs of the organization. Without good architecture practices, those efforts might provide a near-term salve on an urgent problem, only to be exasperated by the user from the long slow slog usually needed for the transition to be production ready, with no overall improvement in capability.

In both of these cases, the use of an enterprise technical framework, the mission or threat-driven (i.e., market-driven) quality attributes, and data architectures that support interoperability can change the game for delivering the “unfair advantage” to the DoD. A different programmatic and technical alignment is thus needed to deliver smaller capability improvements, along with associated hardware updates, that can be installed quickly, and certified for use automatically when installed. This approach requires new means to leverage commercial investment in data center technologies, as well as products that are built to take full advantage of new development tools, techniques (Schmidt, 2014), and certification approaches so that the DoD only pays for unique military capability that can be delivered quickly and reliably, as shown in Figure 1.

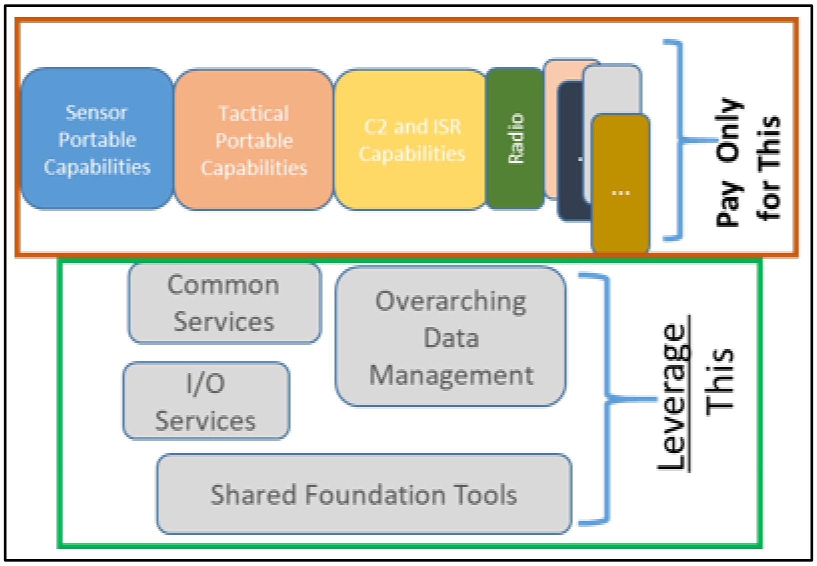


Figure 1. Do in Common What Is Commonly Done and Pay Only for Military Capability

A new procurement and delivery strategy is needed that values shared responsibility, improved warfighter capability, increased operational robustness, outstanding support and continuous improvement. The aspiration is that this strategy is implemented such that the resulting products are defect-free to the warfighter; are tested early and often, certified for operational use when deployed, fully supported, and highly reliable; and can continue to provide the required capability in the face of component failures for protracted periods of time (Guertin, Womble, et al., 2015).

Relevant Technology Trends

Development paradigms are constantly in motion. Service-Oriented Architecture (SOA) was effective for a time, but the development methods and the underlying technologies that made SOA attractive have changed. Emerging design and development practice that are now achieving broad adoption are containerization and micro-services (Amazon, 2018):

- Containerization is an operating system feature where the kernel supports the existence of multiple isolated user-space instances that enable the deployment and running of distributed applications without launching an entire virtual machine (VM) for each module. Containerization can be applied to turn many architecture design elements into fungible commodities that are robustly available to support evolving software development practices.
- Micro-services are a variant of the SOA architectural style that structures an application as a collection of loosely coupled fine-grained services connected via lightweight protocols. Modular capabilities implemented as micro-services more efficiently use the next-generation of computers being produced by the commodity processor markets, including multiple cores, clouds technology, storage evolutions, etc. ("Kubernetes," 2018b).

Containerization and micro-services also help reduce development risk and increase overall product robustness. Likewise, they can be combined with agile and "DevOps" methodologies, where development and quality assurance teams can focus on a capability as a new unit of functionality that works with other containers as a part of a capability architecture (Kubernetes, 2018a).

Requirements for a New Acquisition Model for the DoD

A new acquisition model for the DoD must address how the organization will evolve into a revamped set of business, organizational, contracting, technical, financial, and ultimately cultural behaviors. The core of this model involves transitioning from a structure based on a collection of independently acquired systems into a highly interdependent ecosystem of rapid capability delivery that integrates and interoperates as a foundational principle. This team-centric approach will require constant communication and collaboration across historically partisan divides (McChrystal et al., 2015).

New practices will be needed to align stakeholders to new organizational identities and reward mechanisms. This transformation will only happen by having a clear-eyed future objective structure, matched to a thoughtful progression from the current state toward that objective (Katzenbach et al., 2016). This model would start with an architectural approach that (1) establishes and ensures loose coupling and independent development for components, (2) establishes early and continuous production/evaluation, and (3) has an orchestration of capabilities crafted to present a user experience that appears fully integrated. The organizational implications of this model are as discussed below (Katzenbach et al., 2016).

Overarching Business Model

The DoD should be organized on the principle that has guided dynamic markets. An analogous example of this kind of enterprise approach is the automobile industry. The trend in that market is to limit the number of different organizations that create similar value elements. That market has evolved to use product line architectures (PLAs) to maximize flexibility and reuse of common elements.

A PLA is a design has built-in flexibility to encompass all the different ways a product could be used. This approach is accomplished through configurable design features that are intentionally built to accommodate customizations that support specific customer use-cases. In this way, the PLA design can maximize reuse, while also providing all the same variations that the customers demand. The move to maximizing the utility of a “platform” to serve multiple vehicle product choices has the combined effect of offering greater flexibility in the products being offered and to do so much faster.

Several industries that market complex, safety-critical, large-scale cyber-physical systems have adopted and improved on the product line approach. Figure 2 shows the breakdown of major functional elements of an automobile into product line segments.

Figure 2. Automotive Example of Product Line Engineering and Payload/Platform Management the Renault-Nissan Common Module Framework

Likewise, Figure 3 shows the strategic trajectory of a major U.S. automotive manufacturer to reduce duplicative infrastructure and embrace product lines as a central organizing theme to continue to create flexible and adaptable products while improving efficiency.

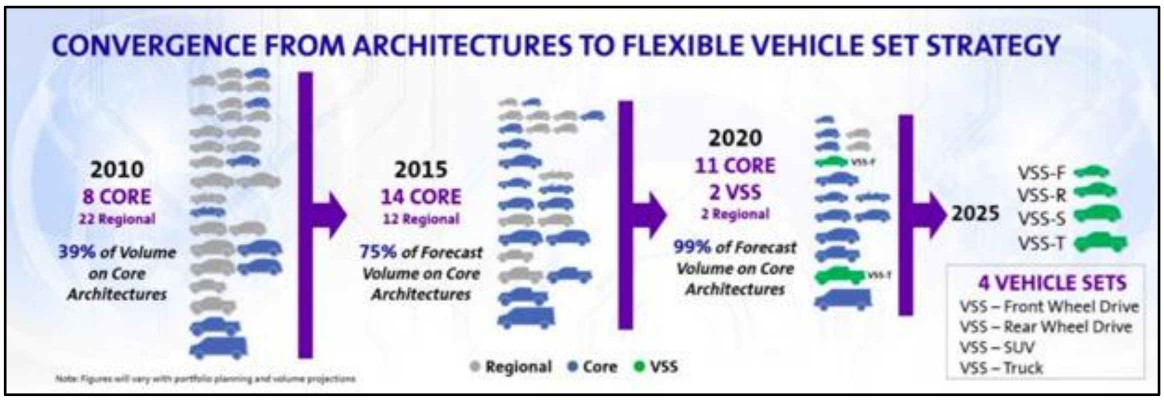


Figure 3. General Motors Platform Reduction Strategy

To achieve the efficiencies experienced in other domains, many aspects of the DoD acquisition structure should be retooled. In particular, the organizational constructs in the Army, Navy, and Air Force services and DoD-affiliated agencies (services/agencies) should be retooled from independent system deliverers into the following three distinct categories:

- Platform acquisition, which provides the outer shell and integration environment of the aircraft, ship, tank, etc.;
- Enterprise architecture product lines, which define a set of software-intensive systems that share a common, managed set of features satisfying the specific needs of a particular market segment or mission and that are developed from a common set of core assets in a prescribed way; and
- Modular capability managers, which provide flexible and adaptable capabilities that get added to or run on the enterprise architectures, and provided to platform integrators. They will need new programmatic approaches, tooling and techniques to manage loose coupling and independence of components, with their ongoing integration. This requires

some discipline with respect to connectors and other internal structural features in the architectural model.

The budget should also be reformed to reflect this strategic approach and embrace a different life-cycle reality of continuous engineering to include early and often validation and verification by the test and evaluation (T&E) community (Guertin & Hunt, 2017). The capabilities developed and deployed into the field can no longer be thought of as produced in their final state (no more fire-and-forget acquisitions). The military environment is constantly changing, and the products the acquisition community provides need to be continuously upgraded and rapidly fielded in quantity, as modularized capability.

Cyber-physical systems can be improved through both software and hardware changes. Although the DoD acquisition framework (DoD, 2017) enables significant tailoring and flexibility, the vast majority of acquisitions still follow a classic spiral development model to achieve a production end-state and a corresponding near-elimination of research and development funding for capability improvement. This approach is particularly problematic in cyber-physical or software-reliant solutions for the following reasons:

- The dynamic cyber threat environment requires constant vigilance for counter penetration and protection measures (even if no capability changes are required).
- The COTS components used to build these systems (hardware, operating systems, tools, etc.) are all in motion responding to market pressures such that the usable in-service lifespan may be much shorter than the longevity of the hardware (e.g., depreciation of software versions or termination of support for obsolete hardware baselines).
- The deployment of new software functions is often an affordable way of improving warfighting performance and addressing evolving mission needs, long after the production run might otherwise be considered complete.

Organizational Impact

The operations of the organization must evolve from a PoR-centric approach to one that values shared resources and focused investment on rapidly deploying military capability (Golden-Biddle, 2013). This change will be hard to manage since three major organizational entities will replace the traditional PoR environment, as shown in Figure 4.

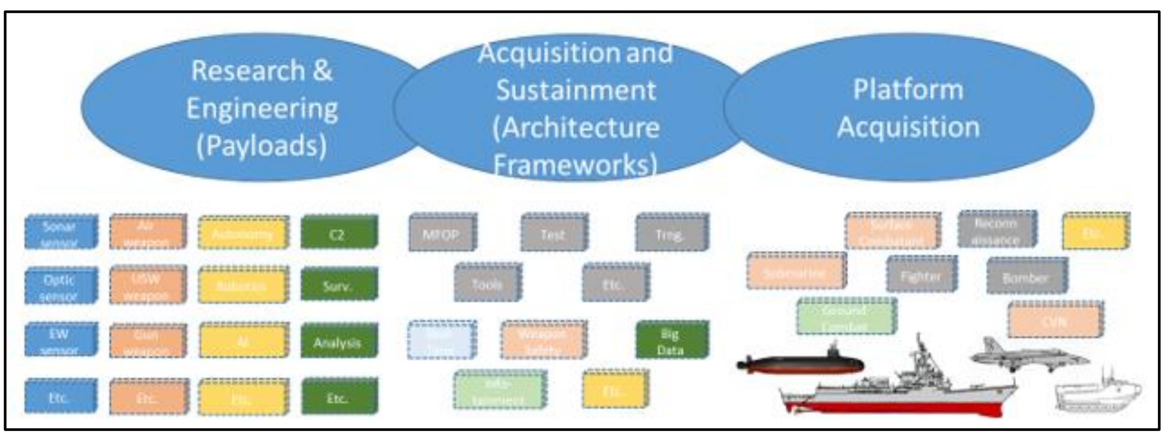


Figure 4. Future PEO Organizing Alignment

Each major entity shown in Figure 4 would shepherd a set of capability managers to ensure interoperability and cooperation pervades the organization, as follows:

- The Research and Engineering (R&E) arm within each service/agency would be populated with capability managers that manage a set of portable modular products (the software of which is decoupled from hardware implementation). A balanced scorecard would be used to adjust modular capability allocation based on a distribution of the particular purposes that address the DoD's strategic needs, such as improving provider diversity or creating new venues for innovation. This scorecard would be shared with the four Defense Committees (in the House and Senate, both Armed Services and Appropriations) so they can identify changes in spending policies and allocation of associated appropriations.
- The Acquisition and Sustainment (A&S) arm will be staffed by architects and systems engineers who manage a technical framework for data interoperability, product development support, module integration, and hardware acquisition. A&S will work with the R&E community to provide artifacts like development kits, integration engines, test harnesses, compliance tools, virtual test bed and hardware definitions, as well as accreditation or certification platforms for delivery of capability modules. A&S will receive architecture investment–related funds and all deployed hardware development/procurement funds. It will also be responsible for delivering certified capability directly to the platform.
- The Platform Acquisition arm will be responsible for acquisition and sustainment of the platforms that host the capability payloads. They also own the platform-unique capability requirements that flow down to both R&E for overall performance and to A&S for platform integration requirements.

Business and Contracting

The acquisition model should be based on having a robust and sustained landing pad for modules of capability that can be risk-prudently and affordably removed, replaced, added, and certified for use, as well as deployed as discrete functional elements that perform integrated functions. At least the following two distinct contracting models should be employed (both of which are already in place and supported by policy or statute):

- The overarching framework should be procured by the Acquisition and Sustainment (A&S) arm described above. All innovators who want to deliver capability to the warfighter need to participate in a continuous evolutionary model (including architectural connectors and data models) for this foundation as a set of living standards. The landing pad for these capabilities could be acquired through a consortium model (e.g., one that is based on Other Transaction Authority [U.S. Air Force, 2015]) and based on industry standards (e.g., the FACE^{tm} Technical Reference Framework [FACE^{tm} Consortium, 2018a]). The intellectual property (IP) strategy for this business environment should be based on collaboration, open standards, and consensus.
- Innovation warfighting functional performance should be acquired through the *Research and Engineering* (R&E) arm described above. Their business relationships should be based on acquiring smaller units of capability (Jones & Womack, 2010) in an agile manner and to sustain a diversity of candidate approaches to cutting edge technologies. The R&E strategy would balance

the need to cultivate organizations that have deep expertise in technical or tactical areas that should be retained for as long as they can competitively deliver warfighting excellence against projecting new capability needs and maturing them through a strategic research and development pipeline. This model is superior to periodic competition because warfighting performance can be removed and replaced by a new competent actor when a capability is ready for (re)use.

A model for establishing a cadre of performers that constantly innovate and compete to deliver new capability will need a different contracting and a remuneration model that awards deployed capability and well-integrated functionality. In one variant of this model, the more software that is delivered, selected as superior, integrated, certified and deployed, the more money the contractor will make. This model will also generate new capability providers through direct industry investment.

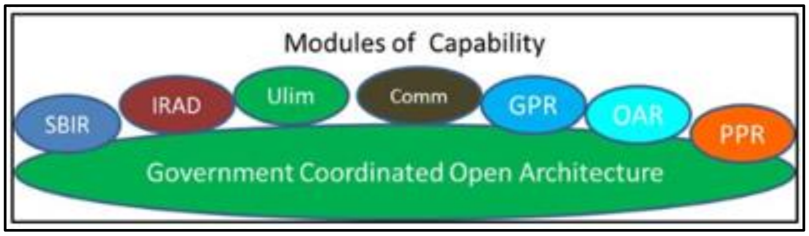


Figure 5. IP Strategy for Capability Module

The intellectual property strategy for this business environment runs the full gamut of data rights, as shown in Figure 5. The government need not attempt to negotiate for greater rights to share IP than the contractor should be bound to offer. The value of a certain capability is based on replicable functional performance, as well as prior investment.

A tension must be managed more artfully than in the past regarding delivery of detailed design data to the government needed to perform test, evaluation, and accreditation activities that ensure elimination of cyber vulnerabilities. This tension has been a divisive issue and the crux of angst associated with IP/data rights issues related to doing business with the government. The government, in turn, must become a trusted steward of industry’s IP that is not destined to be shared with competitors (Limited, Restricted, SBIR, Program Purpose, etc.).

To attract a wide array of potential competitors, the government must also be more nuanced in exercising all of its data rights than it has been in the past. In this way, legitimate use of rights to technical data can be used to gain access to necessary information, while shielding innovators and investors that have independently created designs (i.e., not based on government funding) from unfair practices, corrosive relationships, or counter-productive business threats.

Technical Architecture

The technical architecture described below flows from the business architecture, as shown in Figure 6. This overarching architecture begins with a set of *technical reference frameworks* (TRFs) that support the needs of military systems. Several TRFs have been established through industry collaboration and consortia that represent an excellent starting point development and integration of support loosely-coupled modules of capability. These TRFs are transformed into reference implementations for product lines that support classes of related capabilities. From these reference implementations, product line-specific architectures are crafted that can be deployed as integrated capabilities.

Applying this well-orchestrated cascade of dynamically evolving, industry-supported TRFs can bridge to reference implementations and become specific implementations that support capabilities. In turn, these TRFs are built and verified to support quick integration or removal with few dependencies to other modules (Guertin et al., 2015). The resulting “plug and play” model provides a base capability of the installed infrastructure that is designed for flexibility and growth.

Modular capability elements that are composed into a deployable product should be tested for platform integration in virtualized environments as soon as they are reliable and ready for use. This approach requires new means of continuously updating decoupled hardware and portable software in smaller increments. Cyber-physical capabilities should be expressed as loosely-coupled modules (e.g., containerized micro-services) that can be plugged into the systems architecture with interfaces that are managed through discovery.

Certification of these capabilities are performed as an overall product-line (White et al., 2007) with platform-specific uniqueness certification needs addressed prior to shipment through a virtual test-bed/digital-twin construct (Joshi, 2017). The new capabilities are then delivered to the platform (e.g., a ship, plane, or ground vehicle) as a precertified package, along with targeted hardware changes. The crew (not a civilian installation team) then follows a field procedure to install the changes through simplified—ideally automated—instructions/scripts. The results are then tested automatically on initiation to validate that (1) the certified configuration was accurately completed and (2) the platform is ready for all its assigned missions (Guertin & Hunt, 2017).

Figure 6 also shows how a common data model can be used to support module-level interoperability, such that new functions can be discovered on introduction, complete with full semantic and syntactic descriptions. The Navy has invested in at least two data models that are suitable for this purpose: ASW COI (ASW COI Data Modeling Working Group) and FACE^{tm} (ASW COI, n.d.; FACE^{tm} Consortium, 2018b). This technical architecture also provides a means to decouple software capability into modular units of performance that can be deployed in containers onto an MOSA-enabled platform. Modularization (e.g., containerization and micro-services), is a fundamental tenet to support the overarching business model.

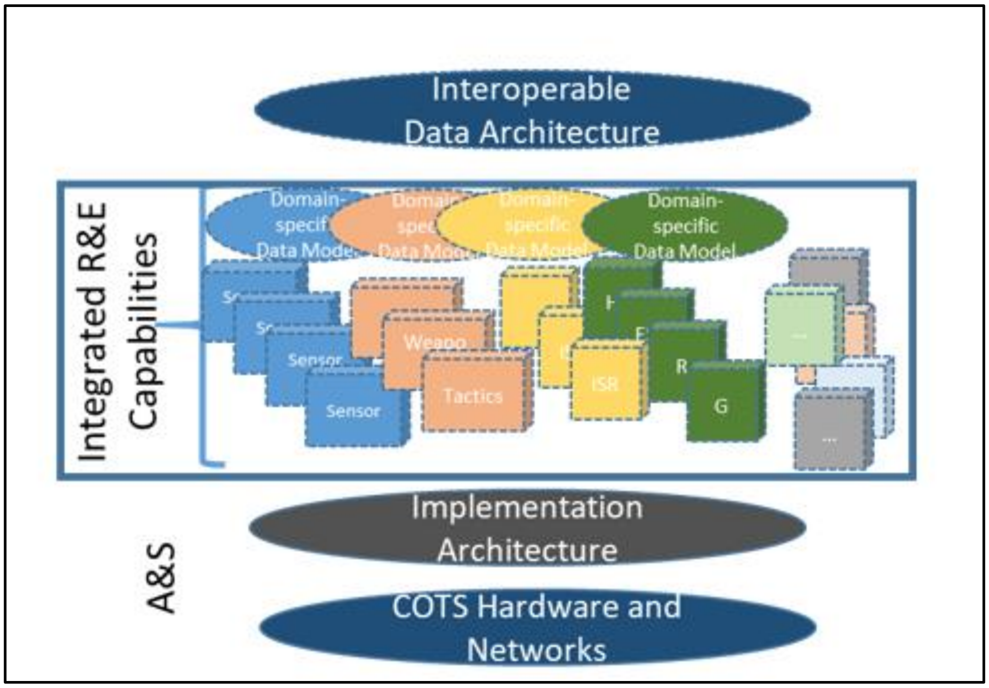


Figure 6. Composition of Severable, Loosely Coupled Capabilities

A virtual testbed (digital twin) will be used to support automated testing and certification of a platform's delivered capability. This testbed can be deployed at as many development sites as are needed by the development community. Capability tests can therefore be performed outside of a single integration laboratory, such that platform differences can be embraced and managed. As a result, the DoD will have the operational flexibility to fit out the capability set that a platform will need for the mission(s) it will perform.

Any hardware kit delivered to the DoD will have gone through an automatic test sequence to ensure it is installed correctly and validated with respect to its digital twin. These kits will be developed by A&S so they can be installed by enlisted technicians to the greatest extent possible. Finally, modern warfighting platform designs are based on standard equipment racks that are already in use on a platform-by-platform basis. These are all predicated on COTS infrastructure, such as electronics-friendly 19-inch rack-mount design. Operational-level and Intermediate or Depot-level actions are thus performed only under the most extreme conditions.

Gradients of Trustworthiness

One of the challenges associated with modular OSA architectures, and the concept of components as payloads, is the presence of "gradients" of reliability, trustworthiness, and security within and across our systems. These gradients are generally unavoidable and require architectural attention to minimize the operational impact they portend. But they can also be beneficial—because they enable nimble approaches to integration of diverse payloads from diverse sources.

The presence of these gradients must be addressed as part of any exercise in architecting and re-architecting. Containers and micro services are an important part of the solution, but there are other aspects as well.

Here are three examples of mechanisms to address the gradients:

1. The design of "connectors" among components in a system, which address issues ranging from governing data flows to enforcement of cross-domain data management policies.
2. Technical methods for isolation and encapsulation, such as sandboxing, which can both protect sensitive components from the broader systems environment and also vice versa, enabling safe use of less trustworthy components.
3. Architectural patterns that enable reduction in those areas where we need the most deep and costly T&E practices, with consequent reduction in cost and delay. Examples of the latter are (a) flight controls vs. other avionics in the DO-178 environment, and (b) doer-checker patterns for advanced heuristic controls, such as might be guided by AI/machine learning components that, in present practice, are relatively opaque to analysis and prediction regarding safety and security attributes.

It highlights the unavoidable deep interplay of architectural technical choices and acquisition strategies.

Financial Architecture

The Financial Architecture will be based initially on the current Program Element structure of the Planning Programming and Budgeting and Execution process, with close coordination between the PEO and the associated Warfare sponsors. The authors assert that if there are sufficient funds to support this number of independent systems, associated infrastructure, and development teams, then there are more than sufficient to support the proposed business model. Eventually, the funding model will need to be changed over the course of several budget cycles to reflect the business model of continuous capability innovation and technology transition.

Impacts Associated With New Implementation and Organizational Structures

Adopting a new acquisition architecture predicated on separating the concerns associated with building new capabilities (R&E) from those associated with a product-line architecture landing-pad, support tools, and shared services delivered to the platforms that would host them (A&S) will yield a number of impacts that are depicted in Figure 7 and described below.

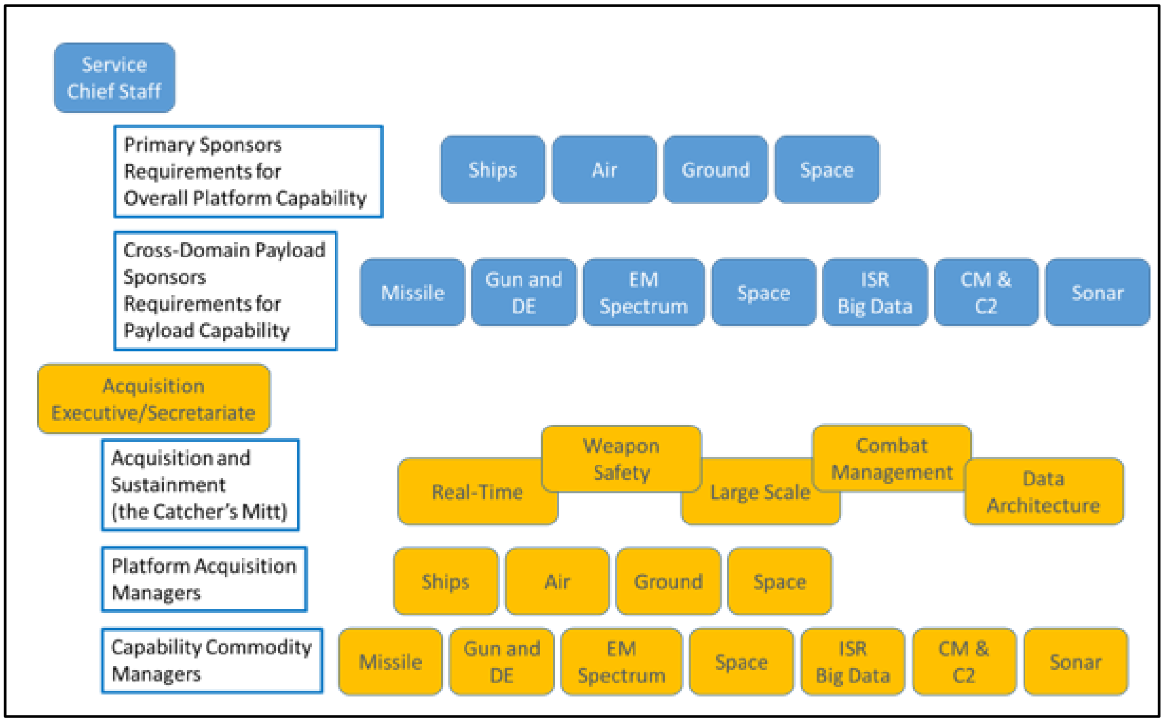


Figure 7. Resourcing and Acquisition Alignment

1. The R&E organizations would focus on delivering cross-platform reusable component capabilities in product-lines that have unique attributes and value, such as Sonar, Imaging, Radar, Communication, Strike, EM/EO/IR, Payload Launch & Control, etc. Those organizations establish requirements for a shared system architecture and work together to integrate products.
2. Likewise, the A&S arm collects the R&E infrastructure requirements and creates a common environment that provides a secure, real-time, safety-critical, and cyber secure environment, including build tools, automated test capability, data architectures, connector models, training environments and integration frameworks. The Platform Acquisition arm would focus on designing, building and sustaining the platform, and specific requirements for installation and non-warfighting system integration (Guertin & Clements, 2010).
3. An important step that some PEOs have begun is to examine modularized capabilities packaged as containers to be deployed as severable, self-healing units of performance such that new products or services can be delivered independently. The system and software architectures need to support loose coupling of those modules so they can be extracted and replaced by new capabilities are well-practiced and

available in the marketplace. The resulting product set could be changed since they are designed as loosely coupled—but highly cohesive—capabilities that are more reliable and self-healing, and that can be integrated quickly with known impacts to existing products and services (Guertin, Womble, et al., 2015).

The supporting elements of the acquisition arena should be refactored to support this model, as summarized in the following:

1. The *Test and Evaluation* (T&E) communities must be a part of this transformation from the inception and be involved in setting the architecture constraints. To ensure that the integration, test, and certification activities validate that the development team has created a highly reliable and critical-bug–free product, the testers should also be a part of establishing a way to check that the deployed product is production-ready. An evolving practice to ensure this alignment is to establish a digital twin environment that would be validated to ensure that all in-lab testing of deployable products represent the installed configuration when the capability is shipped for installation. Only then will the delivery and installation testing be performed in days instead of weeks, with the resulting capability suite certified for full use (Guertin & Hunt, 2017).
2. Product support management takes on a new characteristic. Products that are software-reliant or cyber-physical never encounter a classic sustainment period. Instead they reach a maturity in the productization of the design and enter a continuous engineering and upgrade phase that lasts throughout disposal.
3. PEOs need to perform portfolio management and to decompose functions into modules that can be containerized, apply (not develop) the appropriate a containerization and technical reference framework scheme, establish an infrastructure consolidation plan, to include hardware, networking, storage, and adopt a data architecture that is practiced by a broad community. It is now a good time to consider new standards for architecting this environment that can support the warfighting community for several decades into the future.

The type of change described above will likely imbue classic organizational resistances and text-book rejection responses to strategic change, which are natural human and organizational responses. Fortunately, the mechanisms of resistance to change are better understood now than ever before. To minimize these effects, therefore, a coordinated rollout plan should be developed where members of the organization are welcomed to become a part of how the organization achieves its shared objectives. Likewise, a detailed communication plan should be developed that invites personnel in the existing program offices and subordinate organizations to participate in developing how and where they fit and where the growth opportunities lie. In times of uncertainty, people in these organizations will be primarily interested in how change will affect their lives (Williams, 2017).

Industry will be most interested in the impact to existing tasking and the opportunities that lie ahead. The role of the system integrator would be retooled into an overarching capability integrator, a system architect, and a hardware procurement agent. There is currently an integrator for every system and an overall platform integrator. These duplicative and overlapping roles are ripe for consolidation (Guertin & Womble, 2012).

Successfully implementing the types of change described above will require the full commitment of all members of the acquisition community. Organizations make these kinds of transformations most gracefully when all the members of the organization can see their future in the implementation of the next model. New models for change management have progressed out of the neuroscience and human-centered design communities. These more nuanced approaches draws people in the change strategy such that they feel like they will own the result, which will have the effect of much greater results.

Summary of Recommendations

This section summarizes our recommendations for the DoD along the following dimensions:

- **Organizational/Cultural**—The "burning platform" being addressed is how to reinvent a model of behavior that can achieve a dramatic reduction (at least 80%) in time to flow of capability to the DoD. The resulting environment will shift to a continuous capability delivery engine that is affordable, flexible, adaptable, and reliable. The organization needs to separate the concerns of the payload capabilities, from a supporting enterprise architecture and the host platforms. The resulting managed capability will deliver in smaller increments and will be improved regularly, with higher reliability and in easy-to-install packages that come with training and support.
- **Business**—Conventional federated system-of-systems business relationships currently employed by the services/agencies need to evolve to a model of decoupled capabilities developed by a variety of firms that are experts in their craft. This business model is built on leveraging the commercial marketplace, on valuing private investment, honoring the unique nature of small business, while also maintaining the government's fiduciary responsibility when taxpayers are making investments. Any capability that comes with restrictions on sharing internal design details must come with a certification that the design can be gracefully removed from the system and replaced with equivalent capability derived by a different organization. The overarching architecture on which all this capability will run will become a shared responsibility between industry, the standards community, and government. That open architecture will be co-developed by the stakeholder firms in collaboration with the government who coordinates the effort to ensure that capabilities can be replaced. Other Transaction Authority Consortium models should be investigated as a preferred mechanism for establishing this environment.
- **Technical—**It all begins with a high-level strategic and enterprise approach that is led by the services and supported by the highest levels of the DoD. This transformation is not achievable without the underpinnings of new technical and data architectures. Those underpinnings begin with an approach that is testable and verifiable that the products being developed by industry and accepted by the government comport to the enterprise strategy. Fortunately, we have starting points. Several technical reference frameworks have been established and support a conformance model. These have the support of forward-thinking government and industry teams that used cross-organizational collaboration and standards bodies/consortium models to ensure voices are heard, but not to the exclusion of making progress towards a common goal (consensus-based).

Concluding Remarks

Congressional, DoD, and military leadership of the services have demanded faster and more effective means of achieving the objectives for capability acquisition. Our work reported in this paper describes a new acquisition model that will enable the DoD to plan, buy, field, and certify military capability more effectively by:

- establishing a new budgetary framework based on integrating modular capabilities into open platforms,
- applying containerized and micro-services architecture frameworks that the services/agencies use for integration environments, and
- ensuring resilient and reconstitutable capabilities that can recover from cyber-attacks and combat damage automatically.

Capabilities build in this way will enable services/agencies to update much more frequently to meet warfighting needs and keep the U.S. military ahead of the competition by providing the following benefits to the DoD:

- eliminating classic budgetary overruns and misaligned financial investments for greater life-cycle management and cost of ownership;
- ensuring that software capabilities are durable, self-reporting, and self-healing, as well as enable capabilities to utilize diverse data sources, reducing coupling and increasing reuse;
- allowing the upgrading of products when they are robust and ready, as well as supporting backward compatibility with the other interacting systems on board; and
- enabling software-reliant systems to fallback to a previous version, or even strategically select which software variant is to be loaded next.

This paper has shown how a comprehensive approach—based on current practices and time-proven research—can span the full gamut of the acquisition environment (requirements capture, financial management, programmatic approaches, development methods, and deployment operations) to achieve the national military capability objectives faster, with lower risk and with greater cost performance.

References

Adams, C. (2014, March 7). FACE software effort builds momentum. *Avionics Today*. Retrieved from http://www.aviationtoday.com/2014/03/07/face-software-effort-builds-momentum

Amazon. (2018). Amazon web services. Retrieved from https://aws.amazon.com/microservices/

Anderson, S. (2009). CANES: Consolidated, dynamic and combat-ready. *CHIPS, 27*(3), 6–8. Retrieved from http://www.doncio.navy.mil/chips/ArticleDetails.aspx?ID=2639

ASW COI Data Modeling Working Group. (n.d.). *asw-coi-data-modeling*. Retrieved from https://movesinstitute.org/mailman/listinfo/asw-coi-data-modeling

Clark, D. D. (2004). *An insider's guide to the Internet*. M.I.T. Computer Science and Artificial Intelligence Laboratory.

DeLuca, P., Predd, J. B., Nixon, M., Blickstein, I., Button, R. W., Kallimani, J. G., & Tierney, S. (2013). *Assessing Aegis Program transition to an open-architecture model*. Santa Monica, CA: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR161/RAND_RR161.pdf

DoD. (2017). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: Author.

FACE Consortium. (2018a). *FACE: Future Airborne Capability Environment*. The Open Group. Retrieved from http://www.opengroup.org/face

FACE Consortium. (2018b). *FACE public documents*. The Open Group. Retrieved from https://www.opengroup.us/face/documents.php?action=show&dcat=&gdid=18615

Future Vertical Lift Cross Functional Team. (2018, February 7). Future vertical lift. *STAND-TO!* Retrieved from https://www.army.mil/standto/2018-02-07

Golden-Biddle, K. (2013). How to change an organization without blowing it up. *MIT Sloan Management Review*. Retrieved from http://sloanreview.mit.edu/article/how-to-change-an-organization-without-blowing-it-up/

Greenert, J. (2012, July). Payloads over platforms: Charting a new course. *Proceedings Magazine, 138,* 16–23.

Guertin, N., & Clements, P. (2010). Comparing acquisition strategies: Open architecture versus product lines. In *Proceedings of the Seventh Annual Acquisition Research Symposium* (pp. 78–90). Retrieved from https://calhoun.nps.edu/bitstream/handle/10945/33466/NPS-AM-10-033.pdf?sequence=1

Guertin, N., & Hunt, G. (2017, April). *Transformation of test and evaluation: The natural consequences of model-based engineering and modular open systems architecture.* Paper presented at the Acquisition Research Symposium, Monterey, CA. Retrieved from https://www.researchgate.net/profile/Nickolas_Guertin/publications

Guertin, N., Sweeney, R., & Schmidt, D. C. (2015). How the Navy can use open systems architecture to revolutionize capability acquisition: The Naval OSA Strategy can yield multiple benefits. In *Proceedings of the 12th Annual Acquisition Research Symposium* (pp. 107–116). Retrieved from http://www.acquisitionresearch.net/publications/detail/1498

Guertin, N., Womble, B., Bruhns, P., Schmidt, D. C., Porter, A., & Antypas, B. (2015, May). Management strategies for software infrastructure in large-scale cyber-physical systems for the U.S. Navy. *Cutter IT Journal,* 14–18. Retrieved from https://www.dre.vanderbilt.edu/~schmidt/PDF/itj1505_NG.pdf

Guertin, N. H., & Miller, R. W. (1998). A-RCI—The right way to submarine superiority. *Naval Engineers Journal, 110*(2), 21–33. doi: 10.1111/j.1559-3584.1998.tb03250.x

Guertin, N. H., & Womble, B. (2012, April). Competition and the DoD marketplace. In *Proceedings of the Ninth Annual Acquisition Research Symposium* (pp. 76–82). Retrieved from https://www.researchgate.net/publication/265109556_Competition_and_the_DoD_Marketplace

Jones, D. T., & Womack, J. P. (2010). Lean thinking: *Banish waste and create wealth in your corporation* (2nd ed.). New York, NY: Free Press.

Joshi, N. (2017, June 6). Applications of digital twin. *Allerin*. Retrieved from https://www.allerin.com/blog/applications-of-digital-twin

Katzenbach, J., Oelschlegal, C., & Thomas, J. (2016). 10 principles of organizational culture. *Strategy + Business, 82.* Retrieved from https://www.strategy-business.com/article/10-Principles-of-Organizational-Culture?gko=71d2f

Kubernetes. (2018a). Kubernetes. Retrieved from https://kubernetes.io

Kubernetes. (2018b, March). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Kubernetes

McChrystal, S., Collins, T., Silverman, D., & Fussell, C. (2015). *Team of teams: New rules of engagement for a complex world.* New York, NY: Penguin.

MotorTrend. (2005, April 15). Inside platform sharing: The key to today's increased auto-manufacturing efficiency. Retrieved from http://www.motortrend.com/news/ic-plat/

Nichols, M. R. (2017, May 19). Why is the military interested in the FACE avionics standard? *Avionics Today*. Retrieved from http://www.aviationtoday.com/2017/05/19/military-interested-face-avionics-standard

Scherlis, W., et al. (2010). Critical code. *Software producibility for defense*. National Academy of Sciences.

Schmidt, D. C. (2014, March 3). The importance of automated testing in Open Systems Architecture Initiatives. *SEI Blog*. Retrieved from https://insights.sei.cmu.edu/sei_blog/2014/03/the-importance-of-automated-testing-in-open-systems-architecture-initiatives.html

Schmidt, D. C., Stal, M., Rohnert, H., & Buschmann, F. (2013). *Pattern-oriented software architecture: Patterns for concurrent and networked objects.* Hoboken, NJ: John Wiley & Sons.

Serbu, J. (2017, February 24). Navy opens new 'digital warfare' office, aiming to exploit advances in data science. *Federal News Radio*. Retrieved from https://federalnewsradio.com/navy/2017/02/navy-opens-new-digital-warfare-office-aiming-exploit-advances-data-science

Unmanned Aerospace Systems C2 Standards Initiative. (2014). *Open mission standard (Version 1.0).* U.S. Air Force Virtual Distributed Laboratory. Retrieved from https://www.vdl.afrl.af.mil/programs/uci/oms.php

U.S. Air Force Office of Transformational Innovation. (2015, July). *Other Transaction Authority (OTA) overview*. Retrieved from http://www.transform.af.mil/Portals/18/documents/OSA/OTA_Brief.pdf?ver=2015-09-15-073050-867

Wang, N., Gill, C., Schmidt, D. C., & Subramonian, V. (2004). Configuring real-time aspects in component middleware. In R. Meersman & Z. Tari (Eds.), *Lecture notes in computer science: Vol. 3291. On the move to meaningful internet systems 2004: CoopIS, DOA, and ODBASE* (pp. 1520–1537). Berlin, Germany: Springer-Verlag. doi: 10.1007/978-3-540-30469-2_43

White, J., Benavides, D., Schmidt, D. C., Trinidad, P., Ruiz-Cortes, A., & Doughtery, B. (2010). Automated diagnosis of feature model configurations. *Journal of Systems and Software* [Special Issue on Software Product Lines], *83*, 1094–1107. doi: 10.1016/j.jss.2010.02.017

White, J., Schmidt, D. C., Wuchner, E., & Nechypurenko, A. (2007). Automating product-line variant selection for mobile devices. In *Proceedings of the 11th International Software Product Line Conference* (pp. 129–140). Washington, DC: IEEE Computer Society. doi: 10.1109/SPLINE.2007.19

Williams, R. (2017, February). The psychology of organization change: How neuroscience can help leaders. Retrieved from https://www.business.com/articles/the-psychology-of-organizational-change-how-neuroscience-can-help-leaders

Disclaimer and Distribution Statement

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

Managing Complex Systems Engineering and Acquisition Through Lead Systems Integration

Ron Carlson—served 26 years in naval aviation as a pilot, seven years of which were at NAVAIR where he led NAVAIR systems engineers through several years of systems engineering revitalization. He joined the NPS Systems Engineering department nine years ago. He has a Master of Philosophy degree from Stevens Institute of Technology, master's degrees in Strategic Studies and National Policy from the Naval War College and Business Administration–Aviation from Embry Riddle Aeronautical University, and his Bachelor of Science in nuclear engineering from the University of Michigan. [rrcarlso@nps.edu]

Warren Vaneman—has more than 30 years of leadership and systems engineering experience from various positions within the intelligence community, including as Chief Architect of the Enterprise Ground Architecture at the National Reconnaissance Office. He is also a retired Navy Reserve Captain. He has a BS from the State University of New York Maritime College, an MS in systems engineering and a PhD in industrial and systems engineering from Virginia Tech, and a Joint Professional Military Education Phase 1 Certificate from the Naval War College. The International Council on System Engineering (INCOSE) certifies him as a Certified Systems Engineering Professional (CSEP). [wvaneman@nps.edu]

Both Warren Vaneman and Ron Carlson are faculty members in the Systems Engineering department at NPS, providing research and instruction support to NAVAIR (Patuxent River, MD), Strategic Systems Command, (Washington Navy Yard), NAVSEA (Dahlgren, VA; Carderock, MD), and to NPS SE students in and across the distance learning domain.

Abstract

In the modern operational environment, multiple systems forming System of Systems (SoS) are required to satisfy the spectrum of capabilities needed to satisfy the mission. Accomplishing the mission has always been a SoS endeavor, where integrating multiple systems into a SoS has been left to small communities of "hero engineers," or to the operators responsible for the mission. The acquisition and management of these mission capabilities across the SoS life cycle requires the complex integration of interdependent new and legacy systems from the lowest component level to the highest enterprise level. In 2008, Congress directed government organizations to adopt a Lead System Integration (LSI) process to address the issues with the acquisition, development, and integration of a SoS. The purpose of this paper is twofold: (1) the results of our early exploration of LSI are presented with the definition and development of the LSI Enterprise Framework; (2) it provides an update to our ongoing research that is using a model-based approach to explore the correlation between other frameworks and processes used to engineer and manage SoS employed by Navy Systems Commands.

Introduction

In 2008, Congress enacted Public Law 110-181, directing the Secretary of Defense to properly size and develop the government acquisition workforce to accomplish inherently governmental functions related to the acquisition of major defense systems and to minimize, and eventually eliminate, the use of industry-performed Lead Systems Integration (LSI) functions. Lead Systems Integration is an acquisition strategy that employs a series of methods, practices, and principles to increase the span of both management and engineering acquisition authority and control to acquire system of system (SoS), or highly complex systems (NPS-NAVAIR LSI Cohort #2, 2015). The roles of the LSI are similar to the traditional roles performed by systems engineers and systems integrators. The primary difference is the span of design and integration authority that persists throughout SoS acquisition and life cycle.

The Navy explored the LSI concept and provided a draft implementation plan shortly after the Congressional mandate. Although the Navy did not immediately implement the LSI recommendations, they did pursue processes to engineer and manage SoS. Frameworks such as Information Technology Technical Authority (IT TA) and Integration and Interoperability (I&I) have dominated the SoS process discussion for the better part of a decade. In recent years, the Naval Air Systems Command (NAVAIR) has revitalized LSI and has been striving to better define and implement the concept. While each framework has its strengths, none solely addresses the complete problem.

The purpose of this paper is twofold. First, the results from our early exploration of LSI are presented, with our definition and development of the LSI Enterprise Framework. This framework is a means by which the government can engineer and manage the capabilities and interdependencies of a SoS, across multiple systems, programs, and stakeholder levels.

Second, this paper provides an update to our ongoing research that uses a model-based approach to explore the correlation between IT TA, I&I, LSI, and other frameworks used to engineer and manage SoS employed by Navy Systems Commands. The premise is that by identifying the strengths of each, a revised framework to improve the engineering and management of SoS can be suggested.

The LSI Enterprise Framework Levels

The LSI's purpose is to affordably optimize integrated mission capabilities across the SoS life cycle (NPS-NAVAIR LSI Cohort #1, 2014). To successfully plan, develop, and manage a SoS, a comprehensive development, acquisition, and implementation strategy is required. The LSI Enterprise Framework captures the complex, interdependent, and mission capability areas to characterize the systems from the enterprise to the component levels. This framework establishes the means to engineer and manage the capabilities and interdependencies of a SoS, or complex systems, that can be executed by the government LSI, across multiple systems, programs, and stakeholder levels, where operational and managerial interdependencies exist. The foundation of the LSI Enterprise Framework are the four levels of programs, systems, and stakeholders. The LSI interfaces between the different boundary layers in Figure 1 (NPS-NAVAIR LSI Cohort #2, 2015).

Enterprise Capability Level

The Enterprise Capability Level is the top layer of the LSI framework that consists of a variety of stakeholders, from one or many organizations that represent the complex, sociotechnical systems that comprise interdependent resources of people, information, and systems that must interact with each other and their environment to achieve mission success (Giachetti, 2010). While the majority of the LSI engineering and management activities occur below the enterprise level, this level is important because this is where organizational, policy, and resource decisions are made that provide guidance and governance throughout all levels of the enterprise. It is at this level where the capabilities required to achieve enterprise mission success are defined, decomposed, and allocated to the SoS level to be satisfied as mission capabilities.

Mission Wholeness Level

The Mission Wholeness Level is where a collection of supporting constituent systems and programs are brought together into a SoS to support end-to-end capability effectiveness for the designated mission areas. A SoS is a set or arrangement of systems when independent, and task-oriented systems are integrated into a larger systems construct, that delivers unique capabilities and functions in support of missions that cannot be achieved by individual systems alone (Vaneman & Budka, 2013).

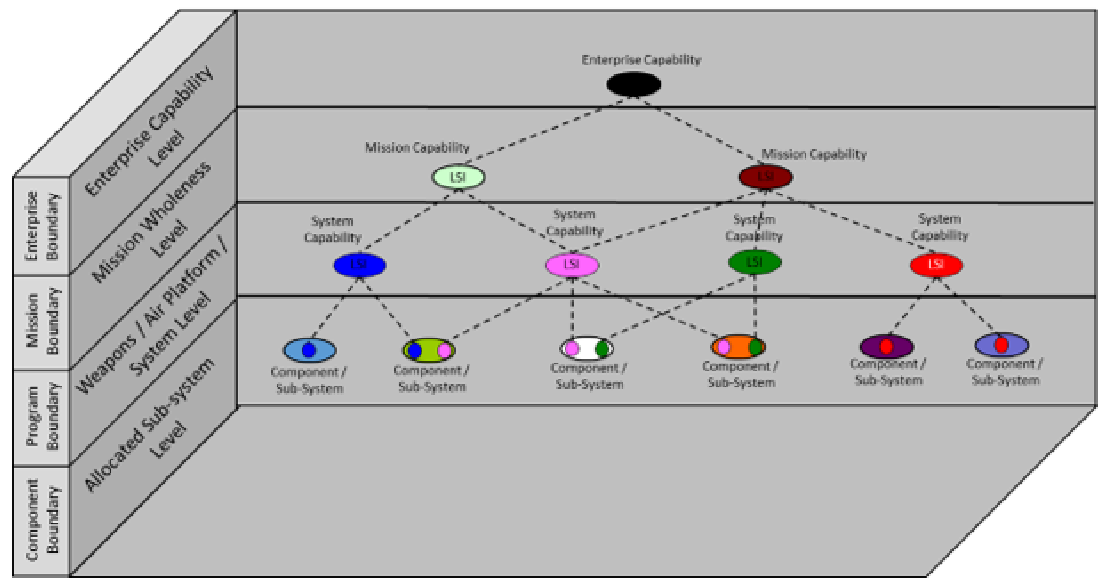


Figure 1. Four Levels of the LSI Enterprise Framework

Accomplishing mission capabilities has always been a SoS endeavor. However, knitting the multiple systems together has frequently been left to small communities of systems or to the operators themselves. The LSI adds the rigor and discipline into combining constituent systems into a SoS, thus reducing the risks that happenstance or chance introduces when integration is left to the operator. Examples of defense mission areas include ballistic missile defense, antisubmarine warfare, counter-air warfare, and surface warfare.

Many of the LSI activities are similar to traditional systems engineering, systems integration, and program management functions. These activities are expanded at the Mission Wholeness Level and are described by those functions encompassed by System of Systems Engineering and Integration (SoSE&I). System of Systems Engineering and Integration is the planning, analyzing, and integrating constituent systems into an SoS capability greater than the sum of those systems (Vaneman & Budka, 2013). The LSI uses the functions defined within SoSE&I to put systems engineering and program management rigor and discipline into development, acquisition, and sustainment decisions at the Mission Wholeness Level. The key elements of SoSE&I are as follows:

1. Managed SoS baseline that directly tracks to mission capabilities;
2. SoS validation, verification, and certification methodology to evaluate delivered capabilities in the context of mission performance;
3. Formal method of governance and change control that puts discipline and rigor into investment decisions at the SoS Level.
(Vaneman, 2016)

Systems Level

The System Level is where a combination of functionally related physical elements are integrated into a usable system to achieve a capability. The systems developed and acquired at this level can operate independently. Examples of systems within the aviation community include different types of aircraft.

Traditional development, sustainment, and management of individual systems is the emphasis of the System Level. The goal of the LSI is to best influence and impact SoS opportunities, and design flexibility into these constituent systems to best adapt to new interfaces, thus extending functionality of the SoS. Three significant roles are important to the LSI in this level. First, the LSI must ensure that the SoS level organization has sufficient insight and understanding of the individual programs and constituent systems within the SoS to understand the functionality, interoperability, and compatibility that will result from the engineering and design effort. This role is important because as decisions are made within program offices to optimize individual systems, they are often made without consideration of the system within a SoS (Vaneman, 2016).

Second, understanding constituent system functionality and programmatic issues is critical since constituent systems in a SoS rely on each other to achieve mission success. Issues such as system schedule delays or technology issues leading to capability shortfalls are critical since other systems that depend on upstream information may not be able to fulfill their missions within a SoS. System retirements are also an area of concern because a premature decommissioning may yield gaps that inhibit the SoS (Vaneman, 2016).

Third, the LSI must ensure a strong governance model is in place that provides the technical authority to govern system baselines so that the system delivered for integration into a SoS meets the requirements that were allocated to it (Vaneman 2016).

Allocated Subsystems Level

The Subsystem/Component Level consists of the allocated sub-systems and components that by themselves may or may not provide a usable standalone end product or capability. These are the lowest level building blocks required for any SoS or complex system that are typically managed by a team in a larger program office or separately by subsystem program offices for large and complex subsystems (Vaneman & Carlson, 2017). Examples of subsystems within an aircraft include avionics, propulsion, and communications.

The LSI Touchpoints

Given the breadth of a SoS acquisition effort and recognizing that a government LSI's resources to manage an effort are limited, an LSI must be able to efficiently focus on the highest payoff "touchpoints" of control or influence to assert and execute trade space—aligned across the enterprise—to enable organizational agility. Although previous efforts have discussed the inherently governmental functions for an LSI, there has been unclear guidance to current program processes (U.S. Navy Chief Systems Engineer, 2010).

The LSI Enterprise Framework defines 12 key touchpoints that apply across all domains as the essential "high payoff" functions and activities. These LSI touchpoints are the functions that implement trade spaces to affordably optimize integrated warfighting capabilities across the SoS life cycle. These touchpoints do not necessarily define new processes but do identify how existing processes can be enhanced and used more efficiently. Figure 2 depicts the traditional organizational and programmatic functions and the 12 touchpoints required for a LSI (NPS-NAVAIR LSI Cohort #2, 2015).

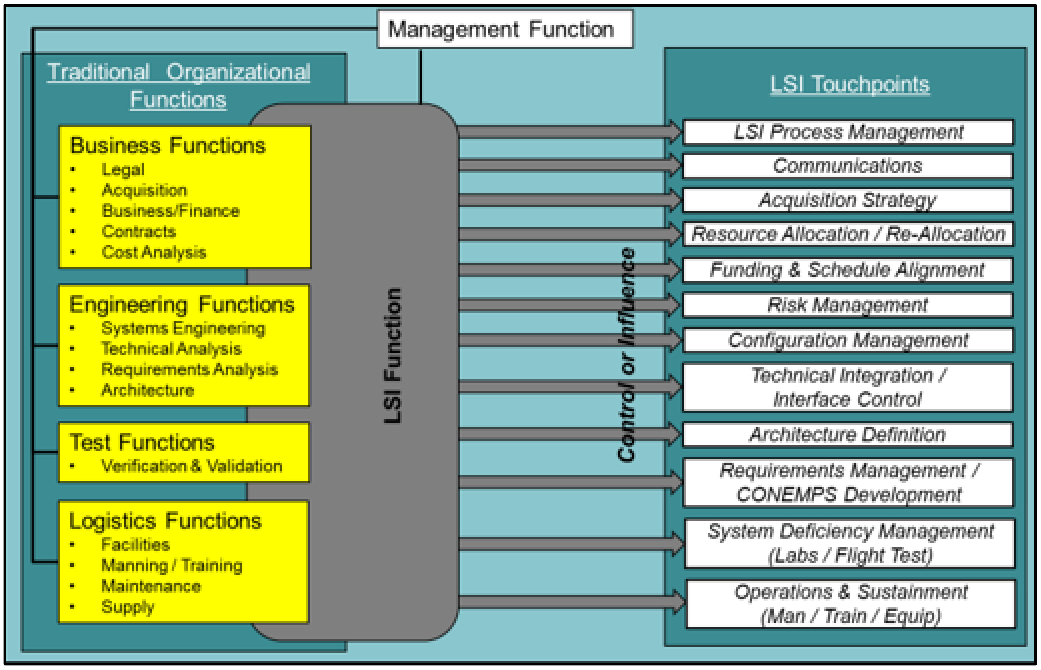


Figure 2. Traditional Organizational Functions and the LSI Touchpoints

LSI Process Management

Responsible for mission wholeness, the LSI defines how their processes interface and interact with legacy processes across multiple stakeholders to meet unique SoS mission capabilities and trade space objectives. These standard work processes document the most efficient known method to produce a system or service, eliminating procedural waste and establishing a baseline for future process improvement initiatives. Standard work packages define process trigger conditions, objectives, enabling factors, inputs, functions, outputs, interfaces, and process time. Furthermore, these standard work processes are the foundation of effects-based staffing, which is critical to defining the skills and resources required to build and maintain an acquisition workforce capable of executing an LSI acquisition strategy (NPS-NAVAIR LSI Cohort #2, 2015).

Communication

The LSI serves as the primary interface and facilitator across a diverse stakeholder constituency. The continuous evolution of SoS capabilities, priorities, mission environments, assumptions, constraints, and threats mandates unprecedented organizational alignment and enterprise agility. Due to the number of typically “stove-piped” teams and program offices, the need to communicate effectively is a key to success. The desired end state of this communication touchpoint is full programmatic, technical, and organizational alignment between the LSI acquisition objectives, and the objectives of the constituent systems (NPS-NAVAIR LSI Cohort #2, 2015).

Acquisition Strategy

The LSI should serve as the principal SoS acquisition strategist. While the U.S. government has been assembling SoS for decades, there is often no overarching acquisition strategy. Given their broad responsibilities, the LSI is often in the best position to develop an overarching acquisition strategy that can be implemented across multiple independent and asynchronous programs and stakeholders to achieve the desired mission capabilities within the resource constraints (NPS-NAVAIR LSI Cohort #2, 2015).

Resource Allocation/Reallocation

The LSI is the primary arbitrator of enterprise resource allocations and reallocations between constituent SoS elements and stakeholders. Requirements and risk mitigation plans should be properly funded across the integrated mission architecture in accordance with an LSI value maximization strategy to achieve the desired capability outcomes. Given the inherent volatility, uncertainty, complexity, and ambiguity of SoS mission environments, allocation of requirements and resources is an iterative process that occurs throughout the mission life cycle (NPS-NAVAIR LSI Cohort #2, 2015).

SoS asynchronous development schedules add a new degree of complexity to LSI resource allocation and re-allocation functions. Given the broad scope of constituent systems encompassed within many SoS mission architectures, it is unlikely that all elements will be in the same acquisition phase. In order to optimize SoS mission value across the SoS trade space, the LSI should also consider the overall mission readiness throughout the SoS life cycle, including existing legacy operations and sustainment activities (Vaneman & Carlson, 2017).

Enterprise Funding and Schedule Alignment

The handling of funding is an inherently governmental function. Enterprise funding and schedule alignment is especially challenging for the LSI since resources are usually budgeted by the resource sponsors to specific programs and systems, and not the SoS to satisfy enterprise or mission-level capabilities. The LSI should be aware of dynamic funding and schedule changes across multiple programs and must align multiple asynchronous schedules of the constituent systems it may control (NPS-NAVAIR LSI Cohort #2, 2015).

System Deficiency Management

SoS deficiency management, supported by laboratory and operational verification and validation activities, is challenging for LSIs in complex mission environments involving multiple programs and stakeholders. The LSI should determine the impact of constituent systems deficiencies at the SoS level. The LSI should also determine the best way to mitigate these deficiencies. The use of simulations and prototypes representing each constituent system that comprises the SoS is a cost-effective method that can be used for early integration risk reduction and may help to refine requirements and identify additional requirements and constraints at the SoS level that may not be apparent at the system level (NPS-NAVAIR LSI Cohort #2, 2015).

Architecture Definition

An architectural definition for a SoS, preferably developed and hosted in a Model-Based Systems Engineering (MBSE) environment, is essential for engineering, analysis, and management of the SoS. The SoS architecture provides a technical blueprint of the SoS, showing the traceability of functional and derived relationships among all constituent systems. The architectural viewpoints enable stakeholders to visualize, define, and bound the component systems, SoS, and identify integration points both inside and outside the systems. From these views, system interoperability issues can be identified. With proper CM and use of compatible databases, new systems entering the SoS family may more easily integrate from an LSI standpoint, and where all disciplines can see integration impacts, dependencies, and interoperability concerns (NPS-NAVAIR LSI Cohort #2, 2015).

Requirements Management and Concepts of Employment

Once a preferred SoS concept is established, the LSI allocates requirements, functions, interfaces, and constraints across constituent systems. This task is especially challenging since the LSI must consider enterprise requirements management and concepts of employment (CONEMPS) across multiple systems and stakeholders. The stakeholders may each hold different assumptions, limitations, or constraints about the expected use of systems, and the mission requirements for the SoS. Constituent system decomposition and integration may also change dynamically or emerge during the evolution of the mission capabilities during SoS life cycle. Requirements management for the LSI is further complicated since the allocation of requirements and resources may be iterative and ongoing across elements that the LSI may not control. The LSI should align requirements, assumptions, limitations, and constraints at the capability level for the overall SoS effort. The CONEMPS may be used as one tool to energize early user and resource sponsor involvement to align stakeholders (NPS-NAVAIR LSI Cohort #2, 2015).

Configuration Management

Configuration management (CM) is the application of appropriate resources, processes, and tools to establish and maintain consistency between the system requirements, the system, and the associated system configuration information. This CM definition must be expanded to address the asynchronous CM across multiple interdependent stakeholders and constituent systems. This asynchronous CM is especially complex for a LSI that must establish and maintain the overall SoS CM baseline throughout the life cycles for all system baselines. Since multiple system program baselines contribute to mission success, the LSI’s CM baseline may change dynamically (NPS-NAVAIR LSI Cohort #2, 2015).

Technical Integration and Interface Control

Technical integration and interface control has a more significant role for the LSI bringing together a SoS, or complex system, than in traditional systems engineering. Since technical trade space management for a SoS occurs at the interfaces between constituent systems, the LSI should focus on enterprise technical integration and interface control. This effort is far more complicated than a traditional acquisition effort, since the technical maturity of the constituent systems within the SoS may be at different levels, and may also be changing at different rates (NPS-NAVAIR LSI Cohort #2, 2015).

Risk Management

Risk management for a SoS is more complex than for traditional systems. Since there are likely many interdependent stakeholders and constituent systems in this effort, the LSI should expand the traditional definition of risk management from the system level and focus on risks at the SoS level. LSI risk management must maintain visibility of risks and opportunities of all systems and critical subsystems, across the SoS trade space. The LSI defines alternative mitigation strategies to combine and normalize these risks across the SoS trade space (NPS-NAVAIR LSI Cohort #2, 2015).

Operations and Sustainment

The LSI’s challenge of affordably optimizing integrated system capabilities across the SoS life cycle is more complex than in a traditional acquisition effort since it may involve multiple independently-developed support strategies or existing legacy system support strategies across the systems in the SoS, which may also be at different levels of maturity. The LSI must understand the support requirements for the entire SoS so that the logistical requirements can be allocated effectively to the constituent systems and supporting stakeholders. The logistics support system should be evaluated across the SoS life cycle to ensure operational supportability with specific attention to minimizing the logistics footprint. Sustainment costs should also be considered during system development and evaluated during testing to ensure that when the SoS capability is fielded, the sustainment costs to support the system are within the constituent systems and/or the LSI's SoS budget (NPS-NAVAIR LSI Cohort #2, 2015).

The Stakeholder Architecture and Governance

The LSI is responsible for defining the enterprise stakeholders, their equities, interest, relationships, and possible impacts to the enterprise trade space to affordably optimize integrated warfighting capabilities across the SoS life cycle. These stakeholder capabilities are represented in an architecture that uses a mission-based, top-down approach to consider the systems from an end-to-end mission perspective (Vaneman, 2016). The LSI responsibilities apply horizontally across the operational, acquisition, and resources sponsor stakeholders within each LSI Enterprise level to ensure coordination and commonality. The LSI responsibilities are applied vertically across the SoS to ensure integration and interoperability for each mission capability area. The Stakeholder Architecture is essential for supporting LSI processes and communication methods to best influence the enterprise trade space.

A cornerstone of an effective SoS is governance. Governance is the structure and relationships among key stakeholders that determine and organization's direction and performance (Hicks, 2008). Governance provides the set of decision-making criteria, policies, processes, and actions that guide the stakeholder architecture to achieve the enterprise goals and objectives (Vaneman & Jaskot, 2013).

The LSI governance challenge is to transition from a program focus, where governance is within the program office, to a mission capability, or SoS, focus, where the governance must occur at the SoS level, where agreements towards achieving a common objective can be agreed to among the various stakeholders, and process for conflict resolution can be defined.

Universal Enabling Resources

Universal enabling resources are those resources that support LSI-unique execution at any of the touchpoints to assert and execute the trade space. The four enabling resources and interrelated enablers apply at all levels in the LSI Enterprise Framework and are outside the responsibilities of the typical program offices. However, the LSI must be aware of these activities and navigate within them. The four enabling resources are shown in Figure 3 (NPS-NAVAIR LSI Cohort #2, 2015).

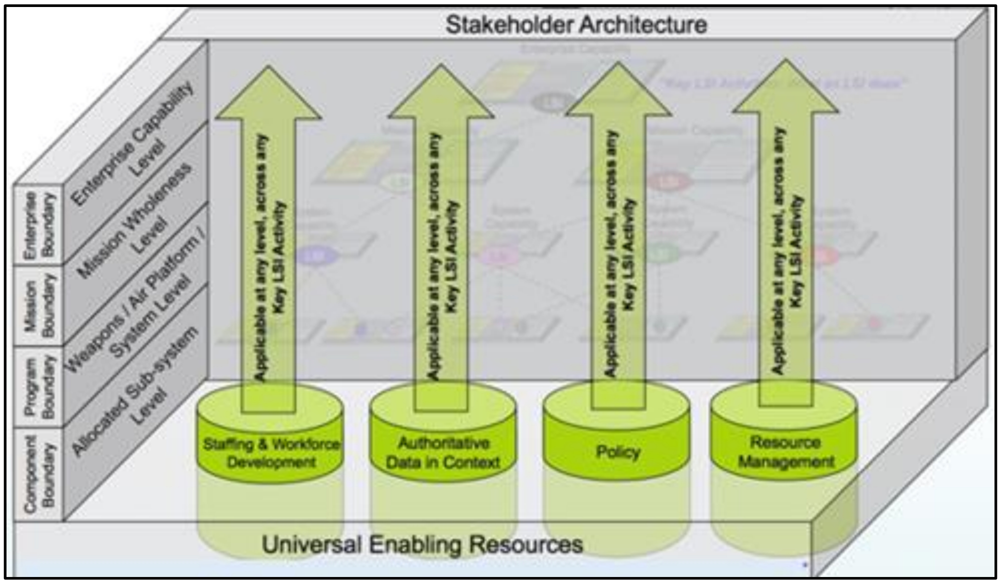


Figure 3. Universal Enabling Resources

Staffing and Workforce Development

Given that Public Law 110-181 specifies that LSI is an inherently governmental function, the key challenge for staffing and workforce development is to recruit and train qualified government engineers to rebalance roles and responsibilities traditionally performed by prime contractors. The government LSI candidates should have a "global" systems perspective and have knowledge across program boundaries. Due to the unique nature of operating in a complex SoS environment, these LSI candidates require additional depth of focus and tailored enhanced knowledge, skills, and experiences beyond that required in traditional acquisition programs (NPS-NAVAIR LSI Cohort #2, 2015; Vaneman & Carlson, 2017).

Authoritative Data in Context

The complex nature of the SoS environment makes asserting and executing the trade space essential and creates the need for sound, authoritative data across systems. In any LSI effort, everyone must have the same data and have a way to validate the authenticity and accuracy of the data to be used for decisions. "Authoritative Data in Context" includes a comprehensive integrated set of programmatic, technical, and stakeholder data that enables a shared common understanding of the trade space (NPS-NAVAIR LSI Cohort #2, 2015; Vaneman & Carlson, 2017).

Policy

Policy consists of the technical, organizational, and legal guidance and constraints of the LSI organization. This may include public law, civil mandates, legal rulings, competency policies, certification requirements, and other overarching guidance that must be accounted for by an LSI when executing any of the touchpoints at any level. These policies provide common guidance across the organizational levels, though the relative impact and flexibility of these policies may vary (NPS-NAVAIR LSI Cohort #2, 2015; Vaneman & Carlson, 2017).

Resource Management

Resource management includes a cost, schedule, and performance resource triad that captures the relationship between the financial, timing, and capability aspects of the total system. When considered against a set of requirements, the resource triad is necessarily constrained by limiting the available resources to a bounded set (NPS-NAVAIR LSI Cohort #2, 2015; Vaneman & Carlson, 2017).

The LSI Enterprise Framework Assembled

Figure 4 (NPS-NAVAIR LSI Cohort #2, 2015) depicts the LSI Enterprise Framework assembled from the four layers, the LSI Touchpoints, stakeholder architecture and governance, and the universal enabling resources. This framework allows for the alignment of key LSI activities across the enterprise to be aligning the appropriate touchpoint to the various LSI levels and tasks. The framework identifies the internal and external organizational dependencies through the stakeholder architecture. Through the universal enabling resource, staffing, and workforce development, policies, resource management, and the authoritative data context can be applied as required throughout the enterprise. Finally, governance empowers decisions across the enterprise by providing a set of decision-making criteria, policies, processes, and actions that guide the stakeholder architecture to achieve the enterprise goals and objectives.

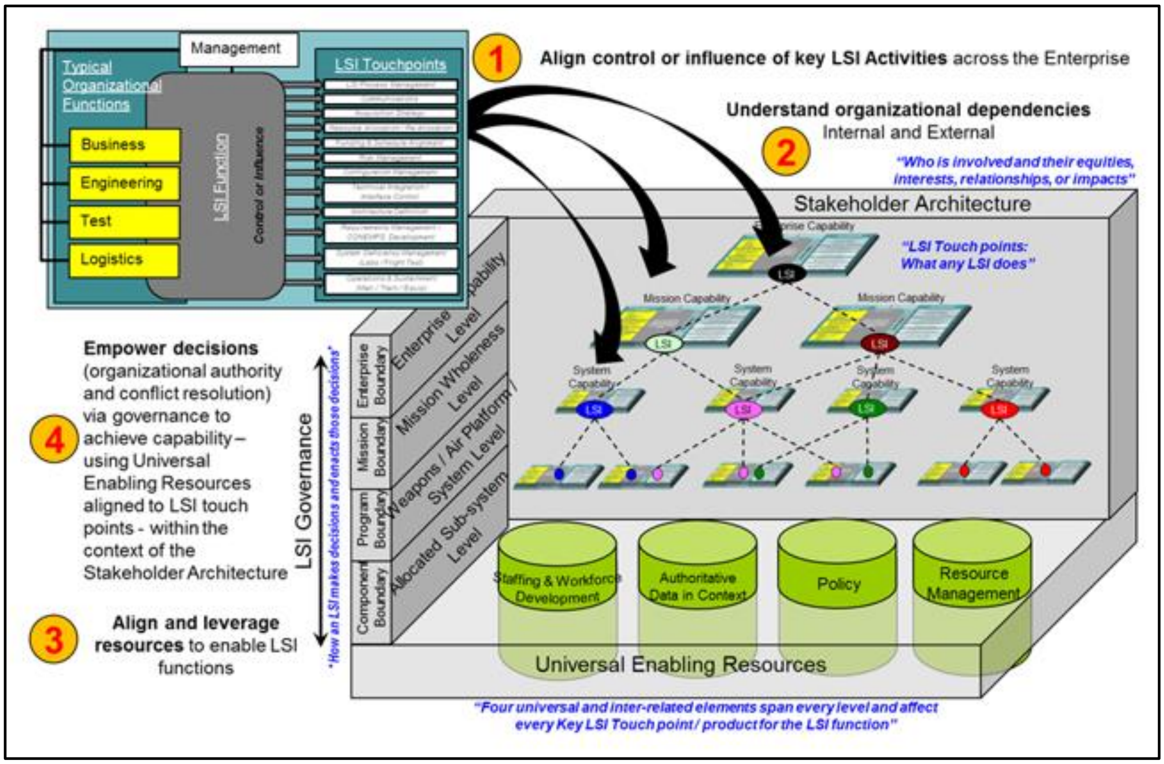


Figure 4. The LSI Enterprise Framework

Ongoing Research

The U.S. Navy has been exploring and developing strategies and approaches to address the engineering and acquisition challenges associated with SoS and complex systems. LSI is the broadest strategy encompassing the widest swath across the SoS life cycle. However, LSI is not the only strategy to address SoSE&I activities. Other strategies to date include Integration and Interoperability (I&I) and Information Technology Technical Authority (IT TA). While each strategy offers insights and partial solutions to the challenges posed by this complex systems development and acquisition environment, none address the complete problem to the depth required.

Our current research is exploring the strengths of each of these concepts and provides a framework that will better define LSI across the SoS life cycle. The following are the research questions, and the proposed methodology:

> **QUESTION 1**: What is the correlation between the System of System Engineering and Integration, Integration and Interoperability, and Lead System Integrator concepts?
>
> **METHODOLOGY 1:** Develop a model that correlates the concepts of SOSE&I, I&I, and LSI. The model will include inputs and outputs of each phase within the SoS life cycle. The model will be generated by a review of existing documentation and collaboration with the SYSCOMS. This model will serve as the baseline for further research tasks and can be tailored to individual organizations.
>
> **QUESTION 2:** How can correlating SOSE&I and I&I with LSI improve the engineering management of SoS and complex systems, and facilitate acquisition strategies that improve the belonging, connectivity, and integration of SoS and complex systems to better satisfy mission objectives?
>
> **METHODOLOGY 2:** Using case studies, derived from SYSCOM interactions, examine how the model will improve the engineering and acquisition of SoS and complex systems. Revise the model as necessary. This analysis will allow the research team to test the generic model against specific cases.
>
> **QUESTION 3**: How does the correlated LSI model apply across non-Navy development and acquisition, and within the Department of Defense?
>
> M**ETHODOLOGY 3**: Apply the LSI model and lessons learned to at least one non-Navy organization within the DoD. Revise and tailor the model as necessary. This analysis will allow the research team to demonstrate that the model is extensible within the DoD.

References

Giachetti, R. (2010). *Design of enterprise systems*. New York, NY: CRC Press.

Hicks, K. H. (2008). *Invigorating defense governance: A beyond Goldwater-Nichols Phase 4 report.* Center for Strategic and International Studies.

National Defense Authorization Act for Fiscal Year 2008, Pub. L. No. 110-181, § 802, 122 Stat. 3 (2008).

NPS-NAVAIR LSI Cohort #1. (2014*). The roles of the government-led Lead System Integrator (LSI).* Monterey, CA: Naval Postgraduate School, LSI Certificate Program.

NPS-NAVAIR LSI Cohort #2. (2015). *An enterprise system integrator framework.* Monterey, CA: Naval Postgraduate School, LSI Certificate Program.

U.S. Navy Chief Systems Engineer. (2010). *Navy Lead Systems Integrator working group final report.* Washington, DC: Author.

Vaneman, W. K. (2016). System of systems engineering and integration. In *Proceedings of the 10th Annual IEEE Systems Conference*. IEEE.

Vaneman, W. K., & Budka, R. (2013). Defining a system of systems engineering and integration approach to address the Navy's information technology technical authority. In *Proceedings of the INCOSE International Symposium*. INCOSE.

Vaneman, W. K., & Carlson, R. (2017). Defining an enterprise lead systems integration (LSI) framework. In *Proceedings of the 12th Annual System of Systems Engineering Conference*. IEEE.

Vaneman, W. K., & Jaskot, R. (2013). A criteria-based framework for establishing system of systems governance. *In Proceedings of the 7th Annual IEEE Systems Conference.* IEEE*.*

Inherent Moral Hazards in Acquisition: Improving Contractor Cooperation in Government as the Integrator (GATI) Programs

William E. Novak—is a Principal Member of the Technical Staff at the Carnegie Mellon University Software Engineering Institute. He is a researcher, consultant, and instructor in the acquisition and development of software-intensive systems. Novak has over 30 years of experience with government acquisition, real-time embedded software and electronics product development, and business management. He has held positions with GE Corporate Research and Development, GE Aerospace, Texas Instruments, Tartan Laboratories, and GTE Automatic Electric Laboratories. Novak received his MS in computer engineering from Rensselaer Polytechnic Institute and BS in computer science from the University of Illinois at Urbana-Champaign. [wen@sei.cmu.edu]

Julie B. Cohen—is a member of the Client Technical Solutions Directorate at the Software Engineering Institute, where she leads diverse work with customers in satellite programs, medical scheduling, and sustainment. Prior to the SEI, Julie was a Program Manager at both Brashear and Marconi. In the Air Force, Julie worked in the F-16 program office, the Flight Training program office, the Air Force Operational Test and Evaluation Center, and the Air Force Research Laboratory. She received her BSEE from Carnegie Mellon University, and an MSEE from the Air Force Institute of Technology. She is a certified Program Manager Professional. [jcohen@sei.cmu.edu]

Andrew P. Moore—is Lead Researcher in the CERT National Insider Threat Center at Carnegie Mellon University's Software Engineering Institute. He has 30 years of experience developing and applying mission-critical system analysis methods and tools. Moore has worked for the Naval Research Laboratory investigating high-assurance system development methods; has edited a journal Special Issue; and has authored a book, two book chapters, and a wide variety of technical journal and conference papers. Moore received an MA in computer science from Duke University, BA in mathematics from the College of Wooster, and Graduate Certificate in System Dynamics from Worcester Polytechnic Institute. [apm@sei.cmu.edu]

William A. Casey—has worked in the areas of threat analysis, code analysis, natural language processing, genomics, bio-informatics, and applied mathematics in academia, industry, and government. He received his PhD in applied mathematics from the Courant Institute at New York University. He also holds an MS in mathematics from Southern Illinois University Carbondale, a master's equivalency in computer science from the Courant Institute at New York University, and an MA in mathematics from the University of Missouri Columbia. Dr. Casey is a member of the American Mathematical Society (AMS) and Society for Industrial and Applied Mathematics (SIAM). [wcasey@sei.cmu.edu]

Bud Mishra—is a Professor of computer science and mathematics at NYU's Courant Institute, NYU. He has degrees in sciences from Utkal University, in electronics and communication engineering from Indian Institute of Technology (IIT), Kharagpur, and MS and PhD degrees in computer science from Carnegie Mellon University. Dr. Mishra is a fellow of the National Academy of Inventors, an IIT Distinguished Alumnus, NYSTAR Distinguished Professor, AAAS Fellow (engineering: robotics, hardware verification and computational biology), IEEE fellow (robotics and automation) and a fellow of the ACM (computational biology and symbolic computation).

Abstract

In the Government as the Integrator (GATI) model of acquisition, prime contractors no longer hand-select the members of the acquisition team or consortium, as they often did in the Lead System Integrator (LSI) model. One drawback of GATI acquisitions, thus, is that independent contractors may have little incentive to cooperate by sharing data and supporting other contractors, potentially resulting in delays, overruns, and poor performance. These problems are considered in this work to be both breakdowns in cooperation and expressions of moral hazards. Since the need for cooperation among contractors is still critical to success, finding ways to motivate that cooperation to improve program performance and outcomes is key to effective GATI acquisition. In this research, potential incentive mechanisms were analyzed for their ability to promote cooperation by applying game theory framing and analysis to this GATI acquisition context, and using system dynamics and agent-based modeling to study the results for their ability to promote cooperation and improve program outcomes.

Introduction

Acquisition programs can suffer from recurring cost, schedule, and quality problems, failing to deliver on time or with full capability, and delaying or leaving the warfighter without needed capabilities (Buettner, 2014). Decision-making impacts acquisition outcomes due to their significant cost and performance implications. Many key decisions are initially made in defining the acquisition strategy (e.g., acquisition approach, solicitation type, monitoring activities, etc.; Ward, Elm & Kushner, 2006) and often must be revisited as program circumstances change. Getting them wrong has consequences: "Acquisition decision-making has not been well managed for these [enterprise IT] systems … and the result has not served today's leaders and soldiers well. … The resulting operational impact is profound" (Dacus & Hagel, 2014). To wit, the DoD portfolio experienced $469 billion in cost growth (over 48%) since programs established their first full estimates (GAO, 2017).

Acquisition program managers often lack the data, evidence, and tools needed to make complex decisions associated with acquisition strategy trade-offs by identifying and evaluating acquisition solution options. They are also in need of better ways to encourage good performance on the part of key program stakeholders. They would be well-served by ways to improve both the quality of acquisition staff decisionmaking, and ways to incentivize desirable acquisition program behaviors. One area that has increasingly come to the fore in terms of its importance to acquisition program outcomes is that of systems integration, which is the focus of the work described here.

The objective of the research described in this paper is to use modeling and simulation to describe and analyze problems with U.S. government system acquisition, and to develop a means to improve decision-making in the acquisition context. We identify different classes of incentive mechanisms—direct financial, future business, and team networking—that can be applied to influence contractor behavior by aligning their interests with those of the program. Specific incentive mechanisms can help to keep contractors focused on program work—but there is no single best incentive. Using a combination of types of mechanisms can achieve greater levels of influence on contractor behavior across a range of organizations with different business models, as will be discussed in detail.

This research uses system dynamics modeling to characterize the acquisition problem and the interrelationship of incentive mechanisms. It also uses game theoretic agent-based modeling in a targeted way to analyze the ability of incentive mechanisms to mitigate system acquisition problems. Incentive mechanisms are analyzed for their ability to promote cooperation and thus improve program outcomes. Upon considering the mathematical models presented, we argue that there is an engineering science involving how best to apply a mix of incentive mechanisms, and in what context they may have the greatest effect. This and other areas of potential future work are discussed at the end of the paper.

Addressing Systems Integration

Largely as a result of the failures of a number of major acquisition programs using the Lead System Integrator (LSI) approach in the 1990s and 2000s, the U.S. Congress enacted the National Defense Authorization Act (NDAA) legislation of 2006, 2008, and 2009 that limited the use of LSI by barring the award of most new LSI contracts after FY 2010 to any contractor who had not already done LSI work previously, and prohibiting the use of an LSI for programs beyond the point of low rate initial production (LRIP).

The poor track record of LSI programs (Young, 2010) and the NDAA legislation led acquisition program management offices (PMOs) to consider one of the most logical alternatives to LSI (DePillis, 2013), which was to have the government act as its own systems integrator, an approach known as Government as the Integrator (GATI). The GATI approach claims several key benefits over the use of an LSI, including government control of the design and architecture of the system and software, better visibility into program status and progress, and the development of higher technical expertise in the government acquisition workforce.

With these seemingly clear advantages, the question arises as to why GATI acquisition doesn't always perform as well as one might expect (Gansler, Lucyshyn, & Spiers, 2009). One key issue is the declining amount of technical expertise in the government acquisition workforce over the past 20 years, producing less than optimal outcomes when the government takes on technical responsibility. In addition, there is a difference in the way GATI teams are formed versus the way in which LSI teams, or "consortia," are formed. In an LSI acquisition, the prime contractor deliberately and carefully chooses a team of select contractors that will bid on, and (if awarded) ultimately perform the work, taking into consideration their areas of expertise, past performance, competitive aspects, and other factors. In such a team, there is an incentive for all members to contribute to the success of the team's work, and the prime helps to enforce this inclination.

In a GATI acquisition, however, no such team is chosen up front, and contracts with various companies are awarded individually and competitively to make up the government's contractor "team." One result of the GATI approach is that unnatural alliances may be inadvertently formed in the process of selecting "best of breed" contractors in each technical area, so that within the final team competing and contentious rivals may be expected to cooperate. Since there may be no pre-existing relationships among the contractors to encourage cooperation, and no prime contractor present to enforce it, mechanisms such as Associate Contractor Agreements (ACAs) may be used to mandate the sharing of information and the cooperation that will be required in a GATI context (AFFARS, 2013).

Unfortunately, but perhaps not surprisingly, contractors generally don't like using ACAs because they may be forced to support, work closely with, and exchange sensitive information with potential or actual competitors. The net effect in such situations can be superficial cooperation, with intense competition, mudslinging, and even backstabbing occurring behind the scenes. Since in acquisition programs the incentives driving the contractors' behaviors are not inherently aligned, rather than mandating cooperation, the most effective way to promote cooperation is to strongly incentivize contractors to behave that way toward one another to advance the program's goals.

The Principal Agent Problem and Moral Hazard

We can broadly characterize a defense acquisition following the GATI model as a government PMO that is coordinating the activities of a set of contractors who are each developing different portions, or *components*, of a system. We can refer to these contractors, or to the government contractor entities responsible for developing these components (e.g., in cases where the major system components are developed by independent acquisition programs), as component performers (CPs). These *component performers* have specialized technical expertise and information that is beyond that of the PMO, which is precisely why they have been engaged to do the work. In the world of microeconomics and game theory, this type of arrangement is referred to as the "Principal Agent" problem, in which expert "agents" (i.e., contractors) apply their expertise to perform work for the "principal" (i.e., the PMO).

> **Illustrative Moral Hazard in Acquisition**
>
> An example of this type of problem occurred on an actual acquisition program, but the names and technologies have been changed in this description. A next-generation cruise missile will be available from its program early next year. However, the program developing the airborne launcher for the missile, which was originally scheduled to be ready at the same time as the missile, has encountered technical problems, and due to the resulting delays, the launcher won't be available until the following year at best. Nevertheless, the cruise missile program plans to proceed with the originally scheduled production and deployment of a large quantity of the new cruise missiles.
>
> To make matters worse, the sensitive electronics and advanced sensors in the missile are known to degrade over time, impacting the missiles' performance, and ultimately rendering the missiles unusable after five years. The sponsoring military service doesn't understand the decision to proceed with pro-duction of the missiles now, but they suspect it may be so that the missile program office will be able to say they met their schedule, even though it will waste a significant portion of the expensive missiles' expected useful life.
>
> This is a moral hazard in that component performers are making decisions either purposefully disregarding or passively ignorant of the needs of the overarching program. The risks to the program associated with those decisions are borne primarily by the program, rather than by the performers making those decisions.

The disparity of expertise (or more generally information) between principals and agents is referred to as *information asymmetry* and creates an imbalance of power between the two parties. When the agents' knowledge exceeds that of the principal, the informational advantage can yield what economists refer to as "utility gains." As an example, consider the schedule to complete a task for which the agent has a high level of experience, while the principal has a low level of experience: the principal finds it inherently difficult to assess the proposed schedule without additional information. The agent, having more knowledge, may propose a schedule that benefits themselves over the principal, choosing an overly optimistic schedule to improve their chances of getting the contract, or a pessimistic one to provide scheduling flexibility should other lucrative side opportunities arise that the agent will want to pursue. Such situations are referred to as a "problem" because while the ideal goal is to have the agent act transparently as the principal needs, the agents' interests are not necessarily aligned with those of the principal, and honesty and high quality work may not be the agents' best strategy.

A concept that is embedded within the "Principal Agent" problem is that of *moral hazard.* Moral hazard refers to situations in which the agent decides how much risk they are willing to accept, but the principal bears the cost if things go poorly—and so the agent may decide to take on more risk than they would have if they bore the full cost of that risk themselves. In the acquisition context, a moral hazard may happen when the contractor and the government have different objectives for the engagement, and the government may not be able to tell whether the contractor is acting in the government's interests (e.g., by providing honest estimates of cost, schedule, and quality), or is acting in its own self interest and forcing the government to accept the risk associated with inaccurate information.

The presence of moral hazard in acquisition can have many different manifestations and consequences. One that is of specific interest to GATI acquisitions deals directly with cooperation. Consider the situation where CPs may decide (without the government's knowledge) to promote the interests and success of their particular component at the possible expense of

- others achieving their own component objectives, thereby adding risks to other team members and indirectly to the program's goals, or
- the global program achieving its objectives, thereby adding risks to program goals.

This constitutes a moral hazard in that these CPs are making decisions that are either deliberately disregarding, or passively ignorant of, the needs of the overarching program. They are thus incurring risks to the program that will ultimately be transferred to and borne by the PMO, rather than by the CPs making those decisions. There could be many different rationales behind such a decision, ranging from optimizing the component's own Earned Value Management (EVM) performance to achieve an incentive fee (despite adversely impacting the performance of other components), to temporarily diverting needed contractor staff to work on a more lucrative "side" opportunity external to the program (while in the process delaying the global program schedule). We will discuss the implications of such side opportunities in greater detail throughout the paper.

Such a situation is made even more acute in a system of systems acquisition context where each CP could be a separate acquisition program in its own right. If each program were a separate financial entity, there would be that much more incentive to pursue self interest, even to the detriment of the overall system.

A Dynamic Characterization of the Problem

Among the critical dilemmas of GATI acquisition is that within a CP with weakly aligned utilities, local objectives can take priority over global objectives either in coordination with the government, or in coordination with other CPs. Figure 1 illustrates a "Growth and Underinvestment" dynamic (Senge, 1990) that can occur during acquisition, yielding increased propensities for CPs to pursue local objectives at the expense of global (i.e., GATI) goals. Since this growth and underinvestment behavior unfolds dynamically based on feedback to the acquisition decision makers, we use a conceptual system dynamics model to represent it.[1]

As shown in the upper left portion of Figure 1, the GATI perspective promotes the desire for government integration and a demand for the government integration function. Government integration fulfills the needs of the CPs in the early stages, as seen in the self balancing feedback loop B1 (dark blue). However, as use of government integration increases naturally as part of an acquisition involving multiple CPs, the CP's satisfaction with the integration decreases, driving the need to grow the government integration capability in the self balancing B2 (purple) feedback loop. The inherent limitations on the speed of this growth relative to the increased need represents the problematic pattern of "Growth and Underinvestment" mentioned previously. The resulting decline in satisfaction with

[1] The notation used in this figure is summarized in the appendix.

government integration promotes the CP to try to take on more of the integration function, leading to a design that is optimized more for that CP's local objectives. This behavior is self reinforcing (see the green feedback loop R1) in that the CP comes to prefer their own local decisions over those of the government integrator. The negative implications of this for the acquisition program achieving program goals is illustrated by the (red) influence arrow, which represents an undermining of the government acquisition program's ability to achieve its global goals.

The stocks and flows in the middle right portion of the figure illustrate staff moving from development to integration functions, development to support functions, or development to other project work. Moving staff to support functions is beneficial to the acquisition program in that it improves the productivity of the other performers. In addition, as shown in the (orange) feedback loop R2, it is self reinforcing in that as a CP supports other performers, those other performers are more likely to support the original CP's own needs. This mutual supportiveness improves the productivity of all performers participating in those support activities. Note that this feedback loop can also run in the opposite direction, such that decreasing the level of support by one performer can then lead to similar decreases by other performers as well.

Finally, CPs' performance on the acquisition program can be undermined if the performer is significantly distracted by other independent business opportunities. As shown in the (light blue) self balancing loop B3, incentive fees are a common means for keeping CPs focused on the acquisition program, and discouraging pursuit of these independent business opportunities. The improved schedule performance that results from keeping developers working on acquisition program–related activities across all performers, combined with appropriate levels of interperformer support, improves the chances that the acquisition program will achieve its global goals. The rest of this paper describes a set of mechanisms that can help incentivize CPs to make decisions promoting the achievement of global goals, thus improving the likelihood of acquisition program success.

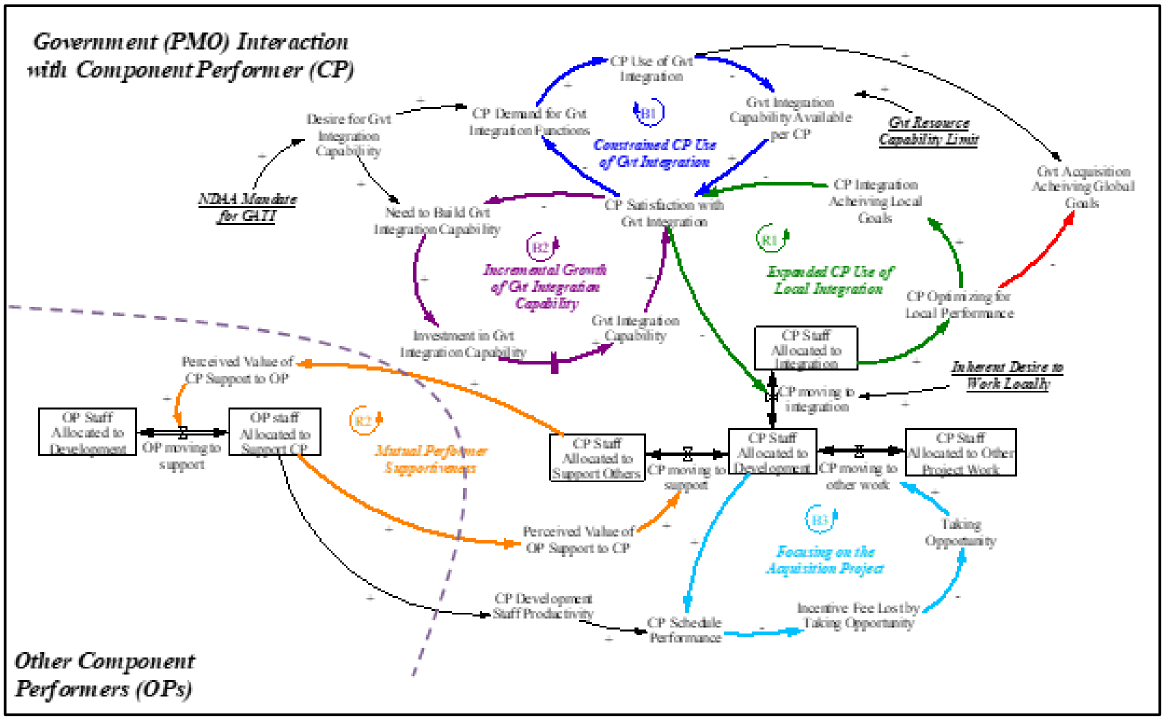


Figure 1. System Dynamics Model Characterizing Aspects of GATI Acquisitions

Component Performer (CP) Cooperation Incentives

Incentives can be used to influence the extent to which a CP focuses on their own local objectives (making themselves look good) or the objectives of the overall program (making the program as a whole look good). In order to understand how CPs are incentivized, we need to understand the drivers of their behavior, which are depicted in Figure 2. CP business is driven both by the revenue obtained from ongoing projects (shown on the right) and new business (shown on the left). Ongoing projects involve the revenue from the acquisition program (AP) as well as other projects the CP is conducting. New business may come as follow-on to the AP (e.g., for good performance) or other opportunities that arise. The CP also needs to maintain its reputation in the community if it is to attract that new business, as shown in the middle of Figure 2.

Understanding the acquisition PMO's objectives from the government's perspective is equally important. After all, it is the misalignment of the objectives of the PMO and the CPs that is the source of problems in acquisition. Figure 3 shows the drivers of PMO behavior oriented around the target AP. AP performance depends on the performance of individual CPs as well as on the performance of their collaboration. The left half of the figure shows that individual performance primarily depends on the CP's schedule and quality performance. The right half of the figure shows that the collaboration among CPs depends on the trust and accountability in the CP relationships. Collaboration can also be greatly enhanced if the CPs do not have to worry about losing key intellectual property (IP) on which their future business depends.

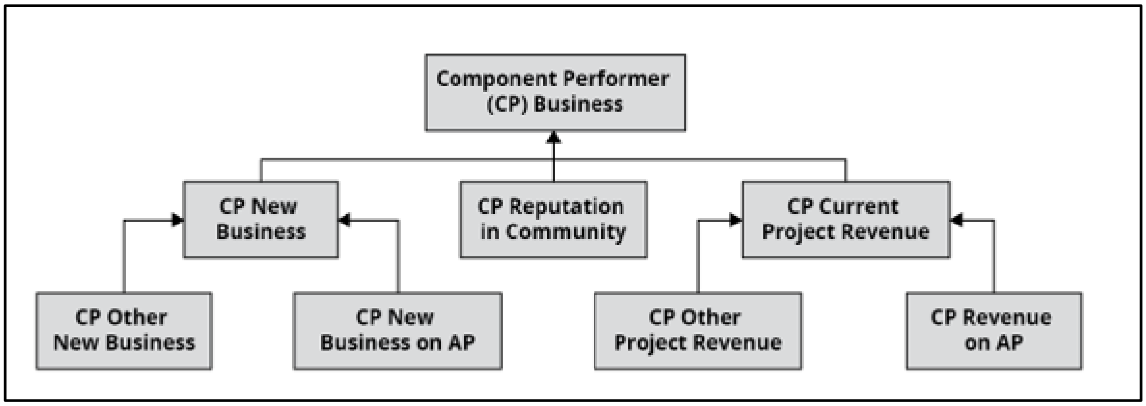


Figure 2. Drivers of Component Performer (CP) Behavior

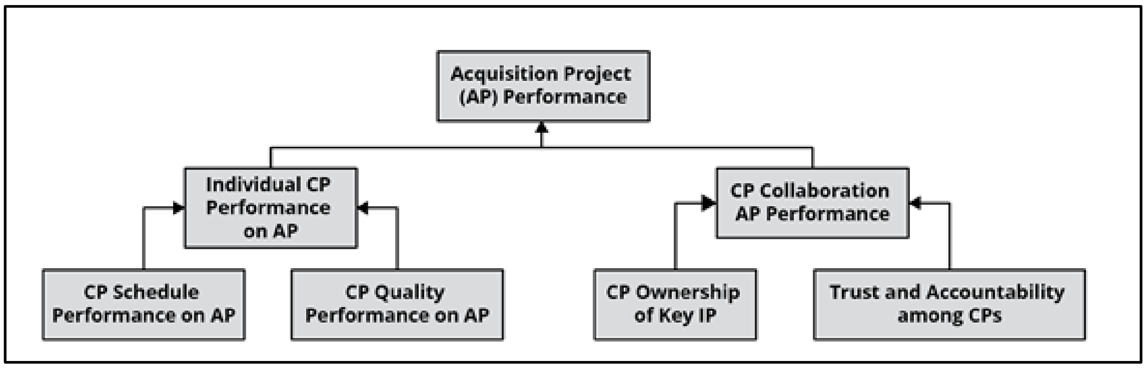


Figure 3. Drivers of Acquisition Program Management Office Behavior

The goal of this section is to show how incentive-based mechanisms can help align the objectives among CPs and between CPs and the PMO. To do that, we need to describe the types of incentive mechanisms we believe are important. Incentives that influence whether CPs take a local or a global perspective in their decision-making have three primary types depending on whether the incentive focuses on the organization's *business*, on the *money*, or on the acquisition program *team* members in the organization. As we discuss each of these categories below, Figure 4 ties example incentives back to their influence on achieving CP and PMO objectives as stated previously.

Business

Future business incentives encourage desired behaviors by increasing the potential for CPs to earn future business. While this ultimately may be financially rewarding for the CP, it is primarily about increasing the performer's competitive edge so as to make money in the future, rather than about paying them money directly for good performance now. Figure 4 shows two future business incentives: *reputation tracking* (Padovan et al., 2001) and *intellectual property/non-disclosure agreements* (IPA/NDA).

- Reputation tracking mechanisms track the performer's performance on the current program and make the results available for consideration in future acquisition program awards. The Contractor Performance Assessment Reporting System (CPARS) is one example of this in current use (CPARS, 2017; (USDoDIG, 2017). Good performance on the current tasking can also be incentivized by allowing options for additional years of contract performance when the CP's performance satisfies certain criteria.

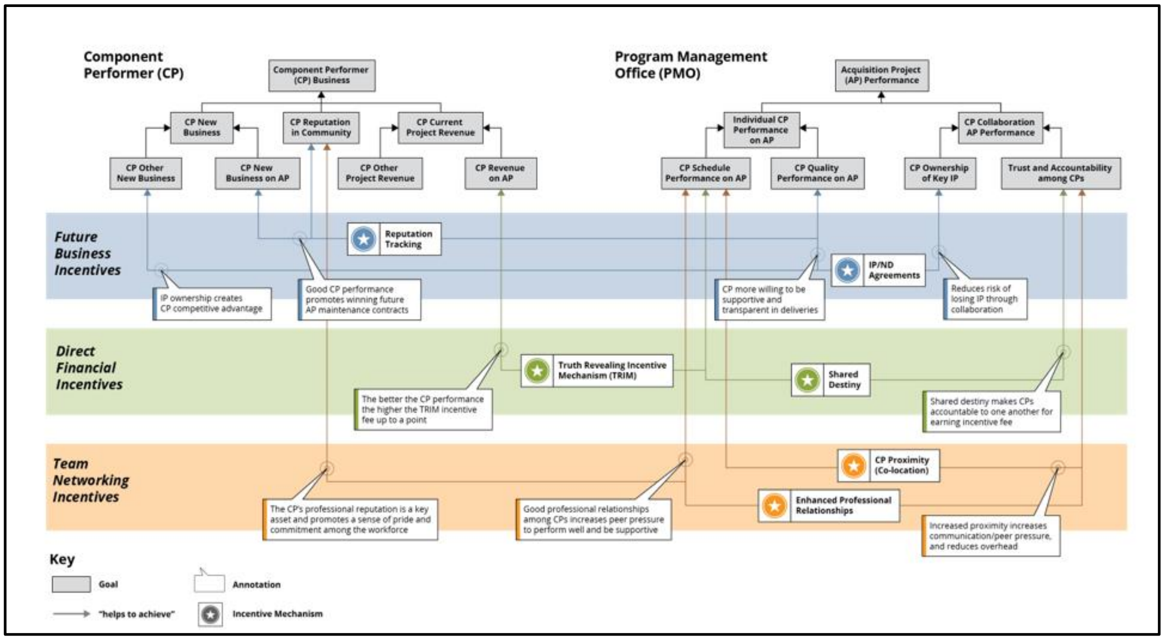


Figure 4. Mechanism-Based Incentive Alignment

IP agreements allow CPs to have certain exclusive rights to IP developed as part of the program activities, enabling them to use this IP in future developments or negotiations. NDAs ensure that collaborators will keep confidential communications that, if divulged, could hamper the CP's business or reputation on current and future ventures. IP agreements and NDAs can help ensure transparent communication and sharing among CPs that is needed for program success.

Money

Direct financial incentives encourage desired CP behaviors by providing financial rewards for those behaviors (Rendon et al., 2012). Figure 4 shows two such direct financial incentives: the *Truth Revealing Incentive Mechanism* (TRIM; Coughlan & Gates, 2009) and *Shared Destiny* (GAO, 2006).

- TRIM promotes the accuracy of performer cost and schedule estimates by making sure that accurate estimates optimize the cost-plus incentive fee awards provided to the CP (Coughlan & Gates, 2009). TRIM consists of a sliding incentive fee based on the promised schedule date, and the fee decreases more rapidly as the schedule moves further out. The smaller the schedule variation, the larger the fee, thus incentivizing on-time delivery. The mechanism encourages the CP to "reveal the truth" regarding their ability to meet the schedule, since by doing so they will earn the largest fee. Proper use of TRIM can thus both ensure accurate estimation that will benefit the PMO, and maximize the incentive fee, to the benefit of the CPs
- Shared Destiny promotes performer collaboration to achieve overall program objectives by financially rewarding performers according to the success of the program. All performers only receive as much fee as the lowest-performing team receives, so all teams are incentivized to improve the capability of the lowest-performing team, who without that help might reduce everyone's fee (GAO, 2006). This motivates both collaborative effort and individual CP accountability for achieving overall program success. CPs must balance the needs of achieving their own local objectives with the support,

communication, and sharing needed to achieve global objectives. Trust among CPs is an essential ingredient of this success, which is promoted by the team networking incentives.

Team

Team networking incentives are social mechanisms to promote positive attitudes among individuals associated with different CPs. Previous research shows that social incentives can be even more influential in shaping behavior than market (financial or business) incentives (Pentland, 2015). Figure 4 shows two team networking incentives: *CP proximity* and *enhanced professional relationships*.

- Professional relationships can be enhanced through team building or other social functions that permit different teams across the CPs to get to know and respect each other both professionally and personally.
- Team networking can be improved by co-locating team members of different CPs. While this can be done virtually using networking and collaboration tools available, physical co-location, if possible, is likely to improve team networking, interaction, and familiarity to the greatest extent (Pentland, 2012; Pentland, 2015).

A Game-Theoretic Basis for Acquisition

Game theory has been characterized as "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers" (Myerson, 1991). While early discussion of two-person games goes back at least as far as the 1700s, game theory did not exist as a unique field until John von Neumann and Oskar Morgenstern's descriptions of the foundations in their book in 1944, *Theory of Games and Economic Behavior* (Leonard, 2010). The 1950s saw the field expand into the logical side of decision science by many scholars in the areas of economics, political science, psychology, computer science, and biology. Cooperation has long been studied in applications of game theory with fundamental insights in repeated games occurring in the 1980s (Axelrod, 1984). Leveraging these insights for business gained significant traction in the 1990s by extending cooperation into competition (so-called "co-opetition"; Axelrod, 1997; Nalebuff & Brandenburger, 1997) and also by generalizing the business ecosystem (Moore, 1993). Interest and publications in these areas have continued through the 2000s into the recent literature (Axelrod, 2006; Brandenburger & Nalebuff, 2011; Daidj, 2017; Dixit & Nalebuff, 2008; Peltoniemi & Vuori, 2004). Another aspect of game theory, evolutionary game theory, as described in *Evolution and the Theory of Games* (Smith, 1982), introduces the idea of the Evolutionarily Stable Strategy (ESS) that can be identified by analyzing evolutionary games using simulation.

Our application of game theory at the most abstract level builds on the earlier work on combining competitive and cooperative business strategies as laid out in Nalebuff and Brandenburger (1997). As in that paper, we ask the question of how we could change the current acquisition "game" to the benefit of acquisition programs more broadly. Figure 5 extends and adapts the value net for the business relationships specified in that paper to the acquisition context. As seen on the left side of that figure, CPs and the PMO are the key strategic players in the game, loosely governed by the acquisition contract. However, each CP is desired to collaborate with other CPs and achieve overall program objectives, usually through vehicles such as an Associate Contractor Agreement (ACA) and possibly additional Memoranda of Understanding (MOU).

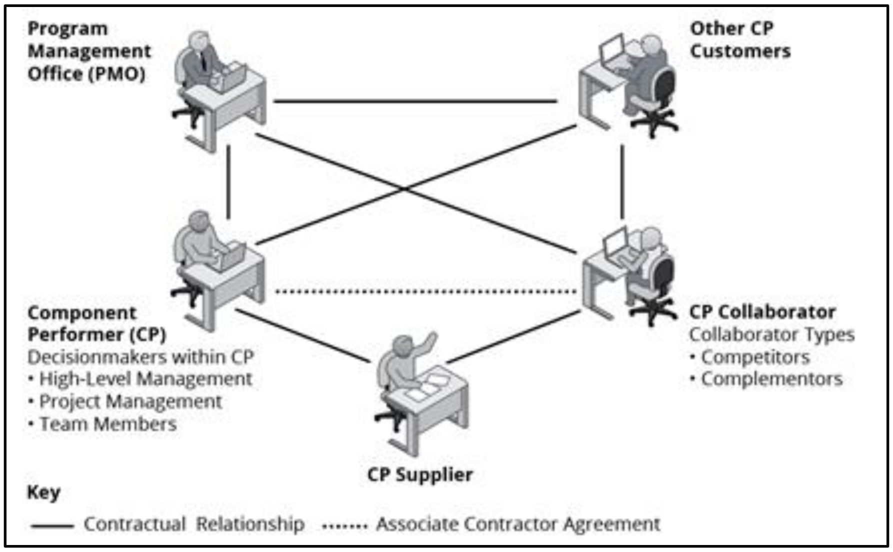


Figure 5. Acquisition Game Players

Both CPs and their collaborators have contractual relationships with the PMO, but will likely have other customers (and suppliers) as well. The figure is *not* meant to imply that all CPs necessarily have the same customers (or suppliers), but rather that all CPs have other customers besides the acquisition PMO. These other customers, both current and potentially in the future, are in some sense in competition with the PMO in terms of adequacy of staffing the acquisition program, since those other customers can draw resources (capabilities and people) away from the program by engaging with the CPs. A focus of our modeling and simulation efforts is on incentive mechanisms that try to avoid this, and can help make CPs concentrate on the PMO's acquisition program.

A key challenge is the extreme diversity of the players: different stakeholders within a CP's organization are likely to be influenced by the incentive types in different ways. High-level management within the CP organization is likely to be most concerned with future business incentives, with direct financial incentives a close second. Project management personnel are likely to be most influenced by the direct financial incentives, since their performance evaluation will presumably depend most on turning a profit. At the lowest level in the organizational hierarchy, project team members of one performer may be most influenced by the professional relationships among project team members of other performers. For example, if management approves of sharing data, but there's little trust of other CPs, a CP may send as little as possible, causing the other CP to continually request more data. Similarly, if team members really trust their colleagues in other CPs, they may share information with them without explicit approval from management (McAllister, 1995). This same dynamic may be true about a CP sharing data with the PMO. The level of influence wielded by these different stakeholders may vary by organization, but all of them likely need to be considered at some level to determine overall organization behavior.

Finally, the quality of the relationship between the CPs on an acquisition program is key to its success, and characterizing it is a central component of our modeling and analysis approach. Keeping the relationship healthy is more challenging if the collaborators are also competitors. A collaborator CP is a *competitor* if they pursue the same business as the other CP. A collaborator CP is a *complementor* if they go after business in a different and complementary domain as the other CP. As characterized in Nalebuff and Brandenburger (1997), a CP's service/product is worth *more* when combined with a complementor's service/product (because of the increased customer value of the combination), whereas it is worth *less* when combined with a competitor's service/product (because of the competition for the same customer space).

While acquisition can be viewed as a game, the very large complexity associated with acquisition programs suggests that a game-theoretical view of acquisition may be constructed from a set of simple, but essential games. Taking this approach, more complex programs are considered as a set of causal connections among essential games. The following section describes a particular game relevant to acquisition. Following that is an overview of the specification of the CP agent as it is affected by incentive mechanisms discussed previously. Simulation results from a detailed system dynamics model of the agent describe some interactions of incentive mechanisms. Later sections describe the vision for possible future extensions of this work.

An Acquisition Game

Since the contractors within an acquisition program are voluntary and work toward shared objectives with common or reciprocal interest, a primary consideration is their coordination and cooperation to improve efficiencies. For this purpose, we consider a class of games known as "coordination" games and note its ability to characterize a large class of social cooperation problems, including those arising within acquisition programs, and specifically GATI acquisitions. In the acquisition context, the CPs may be modeled as rational (utility optimizing) agents. Coordination relies on the actions of others, and naturally presents uncertainty and risk. One example of this is the uncertainty regarding the level of cooperation. Due to uncertainty and risk, CPs may choose a course of action to maximize their individual rewards over that which achieves the best outcome for the acquisition program. To illustrate this dilemma, we use a version of the familiar Stag Hunt game (explained below) to describe how the perception of risks may lead a rational agent (i.e., the CP) to potentially select non-cooperative actions, thus burdening the acquisition program with suboptimal outcomes.

This situation is essential to acquisition strategies and represents the aforementioned moral hazard, as often the government (not the CPs) bears the cost of those risks when program outcomes are poor. Thus, it is important to better understand the

> **Acquisition as a Stag Hunt**
>
> The problem of agents pursuing self-interest despite clear direction from the principal to cooperate in order to achieve more global objectives is well represented in game theory terms by the Stag Hunt game. Stag Hunt is a type of "coordination" game that characterizes a wide range of social cooperation problems. In the game, two players go out hunting together, and each player can individually choose to hunt either a stag or a hare, but they may not be aware of what the other player has chosen to do. A player can catch a hare alone. However, a stag is worth more than twice as much as a hare, but (due to its larger size) both players must choose to hunt the stag if they are to be successful. Thus, the Stag Hunt game has two outcomes (i.e., Nash equilibria) that are most likely: either both players hunt a stag, or each hunts a hare.
>
> The Stag Hunt game clearly describes a situation containing a "moral hazard." It also directly parallels an acquisition scenario where the component performers make decisions more focused on maximizing their individual gain (i.e., pursuing hares in Stag Hunt parlance), and less on achieving the cooperative best outcome for the overall system (i.e., hunting a stag). Even if a hare is worth only half of a stag, hunting a hare represents work that can be done successfully alone, vs. working with a team of others where there's risk and uncertainty around what the collective outcome will be.
>
> However, it should be noted that if the Stag Hunt game is played repeatedly, there may be reputational impacts if one player/contractor repeatedly abandons their partner(s) to pursue another opportunity (i.e., hunting a hare), rather than cooperating to achieve the agreed-upon goal of the program (i.e., hunting the stag). Such "defections" may be met by others being unwilling to engage with them in the future.

nature and degree of the risks, and how those risks are transferred. They are examined here in the context of a specific GATI program, which seeks to obtain the highest satisfaction when all CPs avoid moral hazard and follow through with schedule commitments. As such, the GATI program manager may offer an incentive so that each CP also gains the greatest satisfaction in that case. However, note that the program manager can do little to manage the CP's perception of risk, such as the uncertainty in other CPs' actions.

In summary, we describe a model of cooperation by Stag Hunt games, which involves the essential nature of cooperation and risk. To consider coordination problems more directly on a GATI task, we return to the problem of maintaining a CP's follow-through to schedule commitments, even when they (and their peers, i.e., other CPs supporting the GATI program) are tempted with potentially lucrative side-offers from other customers. Since all CPs on a GATI task are likely to have similar attractive side-offers and are aware of this possibility, the action of following through on task schedule commitments (and declining lucrative side-offers from other customers) will be risky as it can yield losses if other CPs fail at their schedule commitments. The greatest reward achievable for the GATI program is obtained only when *all* CPs follow through on commitments.

The simplest Stag Hunt game involves two players who select how to allocate their effort to either hunt stag (S) or hare (H). If both players hunt hare, the reward is for each h, If both hunt stag, then the reward is $s/2 > h$. However, a stag can only be captured when both players work together, so if one hunts stag while the other hunts hare, the rewards are 0 for the player hunting stag, and h for the one hunting hare (see the payoff matrix in Table 1). The most satisfactory outcome for both players is therefore to cooperate in hunting stag. Unfortunately, it entails risk (as the partner may commit to hunting stag, but not follow through, deciding instead to hunt hare). Its Nash equilibria can be analyzed by computing the expected reward and variance conditioned on selected effort. Letting P_S and P_H represent the probability that a partner will hunt stag and hare (note that $P_H = 1 - P_S$), then by selecting H, the expected reward is h independent of the other's selection (i.e., having variance zero). By selecting S, the expected reward is $s/2 * P_S$ with variance $P_S P_H$.

Uncertain of what a partner player may select, the value P_S (with $0 \leq P_S \leq 1$), may represent a subjective belief of their propensity to play S, and its ability to diminish the expected return can be seen by noting that the expected reward for choosing H is greater than that of selecting S when $P_S < 2(h/s)$. A risk-averse player may also trade off reward to decrease risk, thereby selecting H for higher values of P_S in order to control the downside risk.

With S representing a CP's choice to follow through on commitments, and H otherwise (perhaps delaying the GATI schedule by accepting a lucrative side project), the single-shot Stag Hunt game combines the elements of risk, reward, and cooperation found in the aforementioned GATI scenario. However, the single-shot game lacks the temporal dimension where repeated outcomes may result in dynamic behaviors such as the emergence of reciprocity. We therefore contend that the repeated game form of the Stag Hunt game may be more relevant to capture the dynamic behaviors on a GATI program, and thus suggest that properly monitoring the behavior and reputational metrics of the CPs may form important tools for GATI PM. Additionally, other incentive mechanisms may be considered (e.g., the TRIM and Shared Destiny incentive rewards).

An analysis of how players consider acting in a cooperative game may be extended to consider the following: (1) how various incentive mechanisms alter the dynamics of the game, and (2) how changing mechanism parameters affect the expected behaviors within the game. This type of analysis may be applied to optimize the mechanisms governing how a GATI PMO may best consider various incentives or scheduling actions to yield the best program outcome. To reiterate, in modeling the moral hazard to GATI programs, we only need to consider who assumes the cost arising from risk when outcomes are poor, and also by what means the moral hazard may enter the GATI program. While above we model the *defection* (or the selection H, i.e., to not follow through on a schedule commitment) as facilitated by a lucrative side deal, we note that it could likewise arise from various other causes. For example, the CP may choose to spend more time to improve the quality, or may decrease the quality to free up time to pursue a lucrative side-offer, yet still abide by the schedule commitment (albeit with a lower quality and higher profit). Still, the subject of obtaining a truthful and accurate schedule estimate is presumed as asymmetric information yielding leverage for the CP over the government, as the CP retains the expertise needed to accurately estimate cost and schedule. The estimate of schedule, being asymmetric information similar to other cases listed before, is known to the CP but not to the government, and thus facilitates how moral hazard may enter the scenario. In future, work we wish to consider the entry of moral hazard generally and summarize how risks may be transferred within these or specific GATI scenarios.

Table 1. Stag Hunt Game Normal Form

Stag Hunt (Player 1, Player 2)		Player 2	
		Stag (S)	Hare (H)
Player 1	Stag (S)	(s/2, s/2)	(0, h)
	Hare (H)	(h, 0)	(h, h)

Recognizing that the normal form of the game (above) is ideal, but not necessarily realistic, and that we cannot fully know the CPs' true preferences, we then used a statistical approach to approximate a range of CP preferences regarding the specific incentive mechanisms that the PMO chooses to employ. The image shown in Figure 6 is a single frame from a movie developed by the project to illustrate the outcomes of the Stag Hunt cooperation game. The left-hand side of the figure shows the utilities of Player 1 and Player 2 for various outcomes of a Stag Hunt game with incentives, which in the movie version change depending on how the PMO "weights" or controls parameters of four different incentive mechanisms including TRIM and Shared Destiny. The right-hand side of the figure provides a statistical view considering the range of contractor preferences of the program outcomes (in a color scale), which in the movie is further explored by varying parameters of the different incentive mechanisms. The outcomes for the PMO are shown with better outcomes (i.e., where both players cooperate, or hunt a stag) in lighter colors, and worse outcomes (i.e., both players defect, and hunt hares) in darker colors. The X and Y axes both represent increasing CP satisfaction, so locations in the upper right represent the greatest overall satisfaction among the CPs. A dark area in the lower left is where the CPs are generally incentivized to move, which indicates that poor (uncooperative) outcomes for the program as a whole are most likely. The size of the light or dark areas indicates the variance or risk (uncertainty) of that outcome.

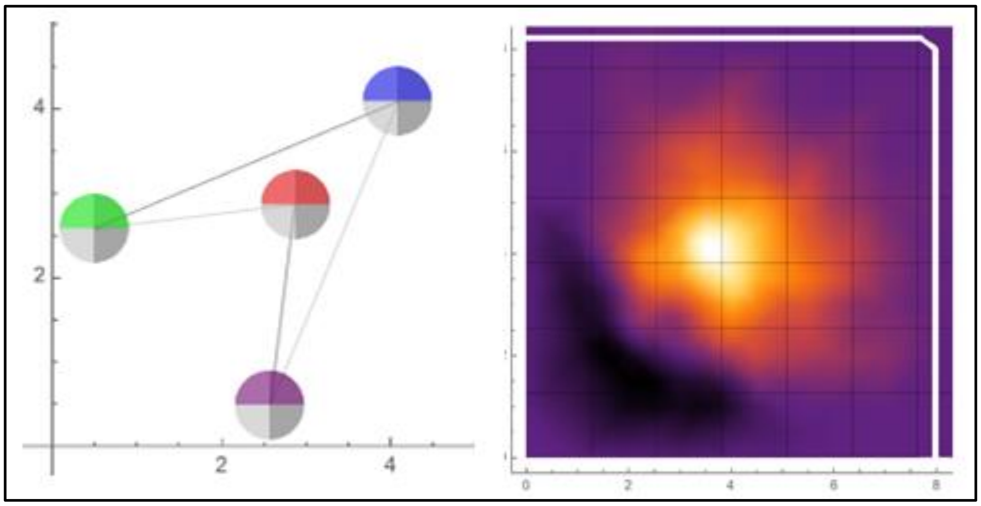


Figure 6. Image From Stag Hunt Analysis Movie

The movie illustrates a key turning point in the outcomes of the Stag Hunt game. In Stag Hunt terms, this is the point at which the value of a hare becomes larger than the value of half of a stag (which is what each hunter would receive for his or her efforts in capturing a stag). This threshold determines where cooperation would break down, as it becomes no longer worthwhile to work together to hunt stag. In acquisition terms, an equivalent scenario would be the simultaneous use of incentive mechanisms such as TRIM (promoting self-interest) and Shared Destiny (promoting cooperation). An overly heavy emphasis on the use of TRIM will undermine the effectiveness of the cooperative effects of the Shared Destiny mechanism. The full movie can be viewed at the SEI website at https://www.sei.cmu.edu/go/moralhazards.

A Systemic View of the Acquisition Game

We now move to a more detailed descriptive model of the dynamic behavior of a CP agent, and its interaction with both the PMO and other CPs. This systemic view of an acquisition program has a number of key dynamics, as depicted in Figure 8. We detail these dynamics below, starting with the basics of getting work done by CPx in dynamics D1 and D2:

- ***D1: CPx Getting Work Done on AP***—D1 is the primary self-balancing loop (shown in dark blue) that drives CPx to do work. CPx allocates FTE to the AP and does work at a level of schedule performance as deemed acceptable by the acquisition PMO.

- ***D2: CPx Getting Work Done on Other Projects***—Of course, CPs are likely to have other development work on which they need to progress. D2 involves a self-balancing loop (also shown in dark blue) very similar to D1 to indicate progress on this other work. Allocating personnel among projects is a key function of CPx management. As other project work grows, personnel may need to be taken from the AP to staff that work.

The dynamics for each of the incentive mechanism categories are specified in dynamics D3 through D5:

- ***D3: CPx Benefit From AP Direct Financial Incentives***—Dynamics associated with Direct Financial incentives are shown (in green) in the figure and include both the TRIM and Shared Destiny mechanisms. The PMO agent controls TRIM through the incentive fee calculation, which is designed to discourage CPx from taking other business opportunities that would take resources away from the AP to a level that would prevent CPx from meeting its schedule estimates for its AP tasking. Likewise, the PMO controls the Shared Destiny incentive fee calculation, but here it motivates CPx to work with its AP collaborators sharing information and providing assistance, so that all CPs on the AP can perform in an effective and efficient manner.
- ***D4: CPx Benefit From AP Future Business Incentives***—Dynamics associated with Future Business incentives are shown (in light blue) in the figure and include both Reputation Tracking and IP/ND Agreement mechanisms. Both of these mechanisms motivate CPx decision-makers by enabling and supporting the development of new business ventures, which can draw personnel and capabilities away from the AP. Therefore, the use of these mechanisms needs to be balanced with other incentives that promote CPx focusing on achieving AP objectives.
- ***D5: CPx Benefit From AP Team Networking Incentives***—Dynamics associated with Team Networking incentives are shown (in orange) in the figure and include mechanisms based on CP proximity and enhancing professional relationships. CP Proximity mechanisms increase interaction frequency and the potential for observing collaborator behaviors, which have been show to reinforce behavior and increase trust (Pentland, 2015). Combined with mechanisms for Enhancing Professional Relationships, good foundations for trust can be enabled, which spur the CP interaction necessary to achieve program goals. Good collaboration can be a self-reinforcing dynamic in the positive direction. However, poor relations between CPs can cause this feedback to go in the negative direction, resulting in a downward spiral that reflects dynamics seen in actual programs.

No matter what incentives are adopted, over time the players will likely find ways to game the resulting system to their own advantage, and to the disadvantage of the overall program. The ways to game a particular system are likely numerous and certainly hard to predict; one possible way is described by dynamic D6. Modeling and simulation can be used to help analyze these schemes by applying concepts from evolutionary game theory.

- ***D6: Gaming the System—Shortcutting Collaboration Support to Enhance Own Business***—The dynamic associated with one potential way of gaming the system of incentive mechanisms introduced so far is shown (in light purple) in the figure. It is based on the notion that CP reputation is relative to the reputation of other CPs, and a reputation tracking system may have the unintended effect of having deficiencies in one CP's performance being blamed on other CPs as a way for the guilty CP to keep their own reputation intact. Even worse, CPs may disrupt their own support to other CPs in subtle ways so as to bolster their claims. Another more obvious potential motivation for shortcutting a CP's support of its collaborators is that this frees up resources to allow a CP to take on other opportunities.

Although the constraints of this paper will not allow a description of the details of the simulation model, we can discuss a few simulation results that show the conceptual value of modeling and simulation in evaluating combinations of incentive mechanisms in the acquisition context. We use a measure called Composite Program Performance, defined as the product of three component measures: two measuring the CP's schedule performance and productivity, and one measuring the extent to which the overall program requirements are satisfied. Each of the component measures are normalized to be between 0 and 1, as is the Composite Program Performance. While a more general model may weight the component measures differently, we've assumed equal weighting in the following analysis.

Figure 7 shows the complementary aspects of team networking and Shared Destiny incentives. Both types of incentives are intended to improve collaboration among CPs, but they do so through different means. Team networking incentives stimulate the social connections between personnel in different CPs (e.g., through proximity, team building, etc.). Shared Destiny stimulates collaboration through direct financial incentives. These incentives are complementary—they target different decision-makers within a CP. Team networking is more likely to affect team members, whereas Shared Destiny will have a bigger impact on management. Since we do not yet know the individual impact of these incentives on CP behavior, we can simulate them over a range of possible impacts, from low to high as shown in the figure. We note that there are diminishing returns with respect to applying more powerful team networking or Shared Destiny incentives, especially as the cost of those incentives grow. As we shall see next, incentive mechanisms are not always complementary in nature—stronger incentives can, in some instances, actually cause worse outcomes in terms of program performance.

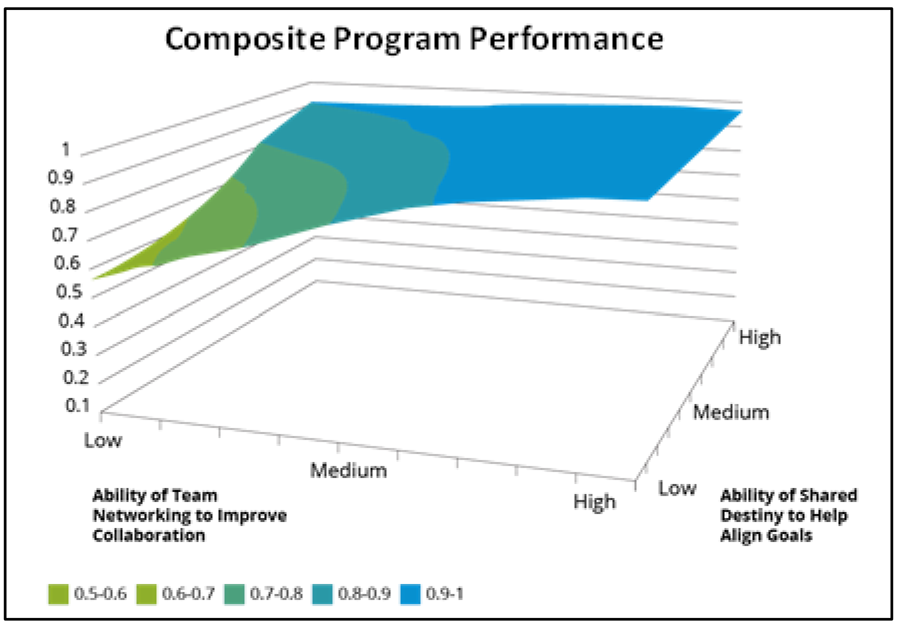


Figure 7. Complementary Effects of Team Networking and Shared Destiny

To some extent, TRIM and Shared Destiny work at different ends of the acquisition problem. TRIM motivates the CP to generate accurate estimates of the schedule, and to keep their development activities focused on meeting that schedule. Shared Destiny motivates meeting the needs of the larger multi-CP acquisition program, providing support for a CP's collaborators by potentially pulling resources away from their own (self-interested) development activities. These incentives can also influence the extent to which a CP accepts integration decisions made by the PMO, especially if those decisions do not favor the CP's own development efforts and local objectives.

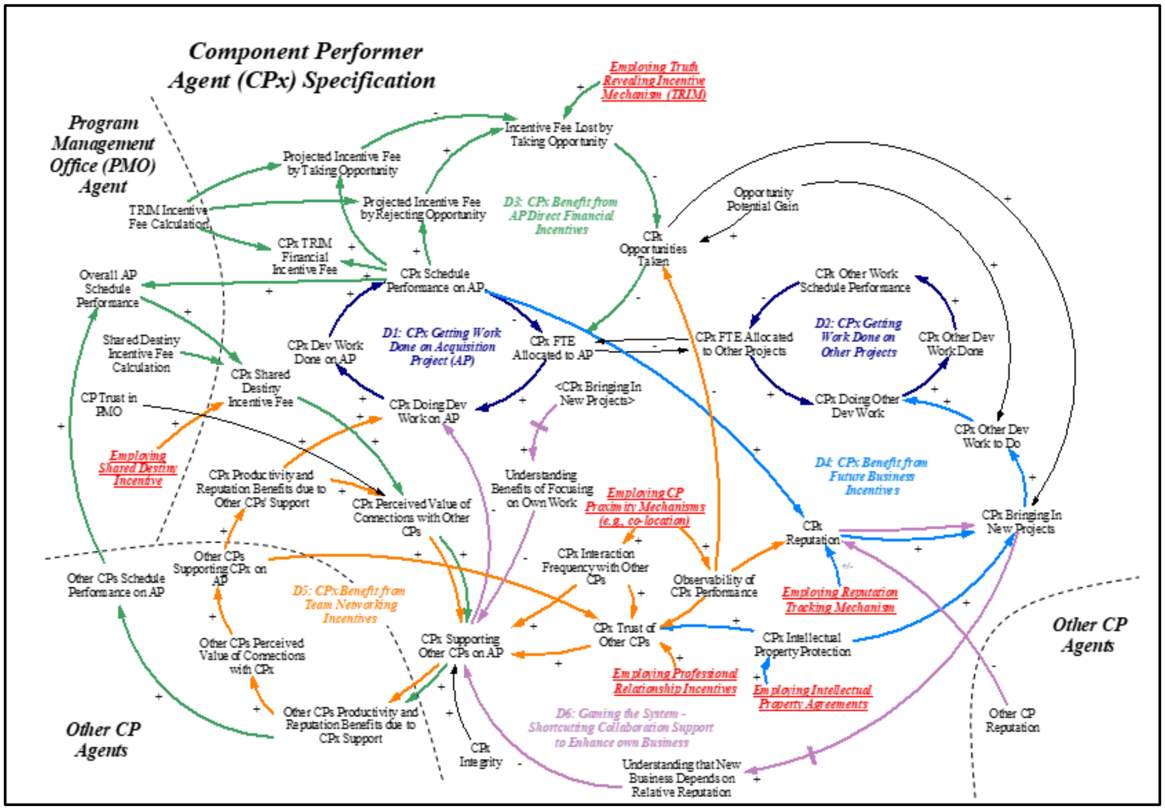


Figure 8. Specification of Component Performer Agent (CPx)

Figure 9 shows Composite Program Performance as the level of collaboration support and the level of acceptance of PMO integration decisions (each varies from low to high). The resulting shape is a gently-sloping ridge with a peak that is not at the endpoint. The bottom line is that improperly balanced TRIM and Shared Destiny incentives can result in too much support being provided to CP collaborators (drawing needed resources away from development) or blind acceptance of PMO integration decisions (when greater CP involvement in integration could produce a better overall system).

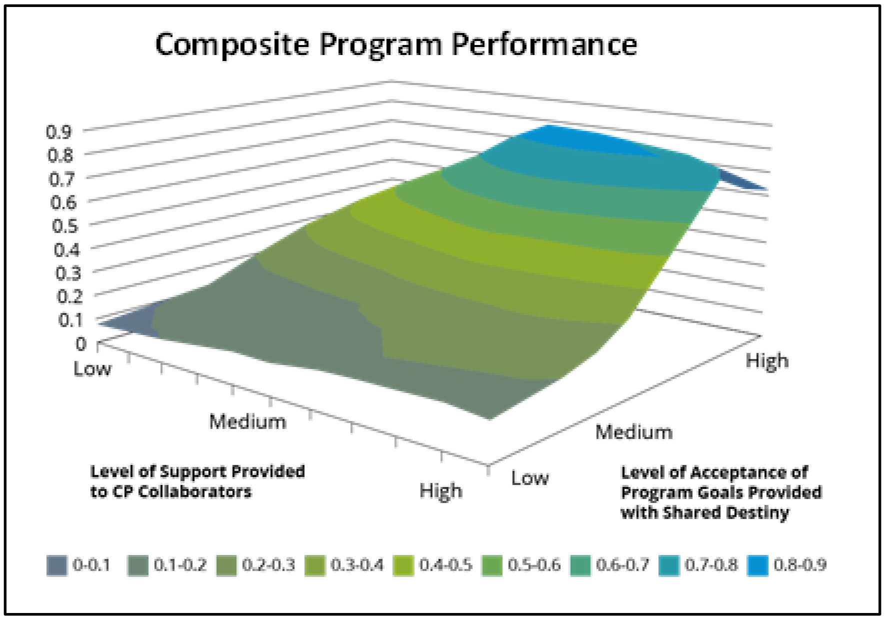


Figure 9. Counterbalancing Effects of TRIM and Shared Destiny

Future Vision of Acquisition Wargaming

An incentive is one weapon in an ongoing war where weapons must evolve. You must plan for the obsolescence of an incentive mechanism as the war escalates. Evolutionary Game Theory defines a framework of strategies and analytics in which Darwinian competition is modeled to allow analysis of the dynamic behavior in a population of acquisition programs in which this escalating conflict occurs among PMOs and their CPs.

It is widely recognized that acquisition programs suffer from recurring cost, schedule, and quality problems, wasting taxpayer dollars, often failing to deliver on-time or with full capability, and delaying or leaving the warfighter without needed capabilities. One key factor behind these issues is that acquisition program managers lack the data, evidence, and tools they need to make complex decisions associated with multi-dimensional acquisition strategy trade-offs, so that they can identify and evaluate acquisition solution options. One area of proposed acquisition research would be to create a virtual acquisition modeling laboratory to improve the quality of acquisition staff decision-making through the ability to analyze the consequences of acquisition decisions, and by designing and testing incentive mechanisms to improve acquisition outcomes over conventional approaches.

Acquisition PMOs historically manage the CP through oversight, gathering costly progress and status information without addressing the fundamental underlying fact that the CP inherently has more complete information about the effort, and as a result may not be acting in the program's interests. This is an instance of a cooperation problem, which exists in many forms within acquisition programs, where decisions can be more focused on maximizing individual returns than on overall outcomes.

Hybrid modeling techniques combining agent-based and system dynamics modeling can model a game theory formalism to characterize the program's behavior. The laboratory could treat acquisition decision-making as an optimization problem searching through the decision space, enabling better decisions with simulated "what if?" questions (e.g., "If A occurs, then in 3 months it will be N% more likely to see outcome X," or "What is the impact of reducing staff by N%?”) related to program decision scenarios, and identifying potential tipping points where key shifts in cost, schedule, and performance occur.

The acquisition program model, after being parameterized to represent a program’s historical performance, would be validated using retrodiction model runs with no incentive mechanism, comparing it to historical performance data, and refining the model to accurately reflect the historical data.

Many incentive mechanisms for mitigating cooperation problems exist in the literature, and could be applied to acquisition contexts. The laboratory would include models of such incentive mechanisms that address the cooperation problem from the relevant acquisition scenario. Explicitly identified assumptions of these incentive mechanisms would be matched with the acquisition context characteristics.

The use of modular models would allow for incremental model development, creating components characterizing many different aspects of modeled acquisition programs that could be plugged into a model architecture to instantiate a model of a single acquisition program with a specific set of characteristics. This would let program-specific acquisition models be built more quickly from existing components that are developed over time, answering PMO questions sooner than older, more monolithic models could.

The objective would be to improve the quality of PMO decision-making by

1. enabling staff to make evidence-based decisions, rather than relying on past solutions and intuition
2. anticipating likely consequences of different decisions, using validated computational analysis
3. mitigating traditionally intractable cooperation problems among program entities/stakeholders that undermine productivity and hinder progress, by using verified mechanisms to align incentives

This technology provides a type of wargaming capability for acquisition programs, allowing them to study “what if” scenarios related to governance and program execution, to anticipate problematic behaviors in program execution for resolution, and generalize and extend acquisition modeling and simulation tools to other program contexts. These tools and methods could provide the DoD with a virtual acquisition laboratory that could be used to simulate the likely response of programs to proposed changes in rules and governance, and thus test the efficacy of proposed policy approaches before their implementation.

This paper has analyzed the interactions of various mechanisms, revealing in the process that the properties of various mechanisms can reveal both compatibilities and counter-intuitive conflicts in their effects when they are combined. We have started to consider how studying the individual properties of mechanisms may be used to predict how well different mechanisms may work together. Believing that this could be done in a rigorous way, we hypothesize that a “mechanism calculus” could be developed as a future research topic, to further assist PMO decision-making in the choice of multiple incentive mechanisms, avoiding counter-productive combinations, and maximizing the intended beneficial effects.

Another area of future work is the characterization of CP preferences for cost, schedule, and quality trade-offs as what is known as a *Pareto surface*. A Pareto surface is a mathematical formalization of the trade-offs agents prefer to make. It represents the result of a mathematical optimization analysis of a problem in which more than one objective must be simultaneously satisfied, and trade-offs must be made among multiple conflicting objectives. CPs know their own Pareto surface (i.e., what trade-offs they would be willing to make, such as money received vs. effort required), but the government PMO does not. Eliciting a CP's Pareto surface reveals ways (e.g., incentives) to influence their behavior.

Summary and Conclusion

The work described in this paper is part of an effort to equip acquisition program managers with more powerful tools to understand and control the behavior of large, complex development efforts. While this work is incomplete, there are some key insights that have been gained from this study of acquisition program analysis through simulation, and the application of a wider range of incentive mechanisms to this domain. Some of the insights from this work can be summarized as follows:

- Incentive mechanisms to promote trust can create a positive self-reinforcing dynamic—but poor relationships can cause this same feedback to be negative, resulting in a downward spiral that reflects the counter-productive dynamics seen in many actual acquisition programs.
- Specific incentive mechanisms can be used to achieve specific objectives (such as helping to keep CPs focused on program work)—but there is no single perfect incentive. They should be used in combination to maximize their positive effect on program performance.
- Different organizational roles are also influenced by different incentive mechanism types (i.e., future business incentives appeal to executives, direct financial incentives appeal to project management, and team networking incentives appeal to engineers and developers). The most effective incentives for a given organization will depend on that organization's values and business priorities.
- Using a combination of different types of incentive mechanisms can achieve greater levels of influence on CP behavior across a range of organizations that employ different business models—especially when information on those preferences is incomplete.
- Using incentives in combination is complex and often non-intuitive. Certain types of incentive mechanisms can undermine the effectiveness of others, such as the way mechanisms promoting cooperation (e.g., Shared Destiny) can undermine those that improve cost and schedule performance (e.g., TRIM).
- Acquisition modeling and simulation could not only help predict the likely results of specific program decisions, but could also analyze the expected responses of acquisition programs to proposed policy and regulation changes, and evaluate their effectiveness.

Perhaps the most significant area of future work would be to provide a virtual acquisition laboratory by constructing a computational model, informed by causal modeling of the relationships, using an extensible, component-based acquisition program model architecture, and validated against historical program behavior data. This model could be used together with appropriate existing incentive mechanisms that could address the cooperation issues involved with the program under study. This capability could be a "wargaming" capability that models acquisition behaviors to help leaders avoid problems through better-informed decision-making. Such a capability could provide acquisition program leadership with a new level of insight that lets them look into the near future of their program's performance to anticipate issues, make *evidence-based* decisions, and thus avoid serious problems.

References

Air Force Federal Acquisition Regulations System (AFFARS). (2013). *Informational guidance: Associate contractor agreements.*

Axelrod, R. (1984). *The evolution of cooperation*. New York, NY: Basic Books.

Axelrod, R. (1997). *The complexity of cooperation: Agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Axelrod, R. (2006). *The evolution of cooperation* (Rev. ed.). New York, NY: Perseus Books Group.

Brandenburger, A. M., & Nalebuff, B. J. (2011). *Co-opetition*. Crown Business.

Buettner, D. (2014). *Acquisition games and software defects: Dynamics of the software developer's dilemma*. Lambert Academic.

Coughlan, P. J., & Gates, W. (2009). *Innovations in defense acquisition: Asymmetric information and incentive contract design* (NPS-CM-09-127). Monterey, CA: Naval Postgraduate School.

Coughlan, P., Gates, W., & Lamping, J. (2008*). Innovations in defense acquisition auctions: Lessons learned and alternative mechanism designs* (NPS-AM-08-013). Monterey, CA: Naval Postgraduate School.

CPARS. (2017). Guid*ance for the Contractor Performance Assessment Reporting System (CPARS).* CPARS Program Office. Retrieved from https://www.cpars.gov/pdfs/CPARS-Guidance.pdf

Dacus, C., & Hagel, S. (2014). A conceptual framework for defense acquisition makers: Giving the schedule its due. *Defense ARJ, 21*(1), 486–504.

Daidj, N. (2017). *Cooperation, coopetition, and innovation*. Hoboken, NJ: John Wiley & Sons.

DePillis, L. (2013, October 24). Government did a poor job on HealthCare.gov. A private firm might've been worse. *Wonkblog*.

Dixit, A. K., & Nalebuff, B. (2008). *The art of strategy: A game theorist's guide to success in business and life.* W.W. Norton & Company.

Gansler, J., Lucyshyn, W., & Spiers, A. (2009). The role of lead system integrator. In the *Proceedings of the Sixth Annual Acquisition Research Symposium* (Vol. II). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

GAO. (2006). *Defense acquisitions: Business case and business arrangements key for future combat system's success* (GAO-06-478T). Washington, DC: Author.

GAO. (2017). *Defense acquisitions: Assessments of selected weapon programs* (GAO-16-329SP). Washington, DC: Author.

Leonard, R. (2010). *Von Neumann, Morgenstern, and the creation of game theory*. New York, NY: Cambridge University Press.

McAllister, D. (1995). Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal, 38*(1), 24–59.

Moore, J. F. (1993). Predators and prey: A new ecology of competition*. Harvard Business Review, 71*(3), 75–86.

Myerson, R. (2007). *Game theory*. Cambridge, MA: Harvard University Press.

Nalebuff, B., & Brandenburger, A. (1997). Co‐opetition: Competitive and cooperative business strategies for the digital economy. *Strategy & Leadership, 25*(6), 28–33.

Padovan, B., Sackmann, S., Eymann, T., & Pippow, I. (2001). A prototype for an agent-based secure electronic marketplace including reputation tracking mechanisms. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences.* IEEE.

Peltoniemi, M., & Vuori, E. (2004). Business ecosystem as the new approach to complex adaptive business environments. In *Proceedings of eBusiness Research Forum* (Vol. 2; pp. 267–281).

Pentland, A. (2012). The new science of building great teams. *Harvard Business Review, 90*(4), 60–69.

Pentland, A. (2015). *Social physics: How social networks can make us smarter*. Penguin.

Rendon, R. G., Huynh, T. V., & Osmundson, J. S. (2012). Contracting processes and structures for systems‐of‐systems acquisition. *Systems Engineering, 15*(4), 471–482.

Senge, P. M. (1990). *The fifth discipline: The art and practice of the learning organization*. New York, NY: Doubleday/Currency.

Smith, J. M. (1982). *Evolution and the theory of games*. Cambridge University Press.

United States DoD Inspector General. (2017). *Summary of audits on assessing contractor performance: Additional guidance and system enhancements needed.*

Ward, M. C., Elm, J. P., & Kushner, S. (2006). *Techniques for developing an acquisition strategy by profiling software risks* (CMU/SEI-2006-TR-002). Pittsburgh, PA: Carnegie Mellon University, Software Engineering Institute.

Young, S. (2010). Lead systems integrator role for government, presentation. *NDIA Systems Engineering Conference.*

Acknowledgments

The authors would like to thank the U.S. Department of Defense for providing funding for this work, and the Software Engineering Institute's Director's Office and Software Solutions Division for their help in encouraging and conducting this research.

Appendix: System Dynamics Modeling Notation

Figure 10 summarizes the notation used in our system dynamics model. The primary elements are variables of interest, stocks (which represent collection points of resources), and flows (which represent the transition of resources between stocks). Signed arrows represent causal relationships, where the sign indicates how the variable at the arrow's source influences the variable at the arrow's target. A positive (+) influence indicates that the values of the variables move in the same direction, whereas a negative (-) influence indicates that they move in opposite directions. A connected group of variables, stocks, and flows can create a path that is referred to as a feedback loop. There are two types of feedback loops: balancing and reinforcing. The type of feedback loop is determined by counting the number of negative influences along the path of the loop. An odd number of negative influences indicates a balancing loop; an even (or zero) number of negative influences indicates a reinforcing loop.

Significant feedback loops identified within the model described here are indicated by a loop symbol and a loop name in italics. Balancing loops—indicated with the label B followed by an identifying number in the loop symbol—describe aspects of the system that oppose change, seeking to drive variables to some equilibrium goal state. Balancing loops often represent actions that an organization takes to manage, or mitigate a problem. Reinforcing loops—indicated with a label R followed by a number in the loop symbol—describe system aspects that tend to drive variable values consistently either upward or downward. Reinforcing loops often represent the escalation of problems, but may include problem mitigation behaviors.

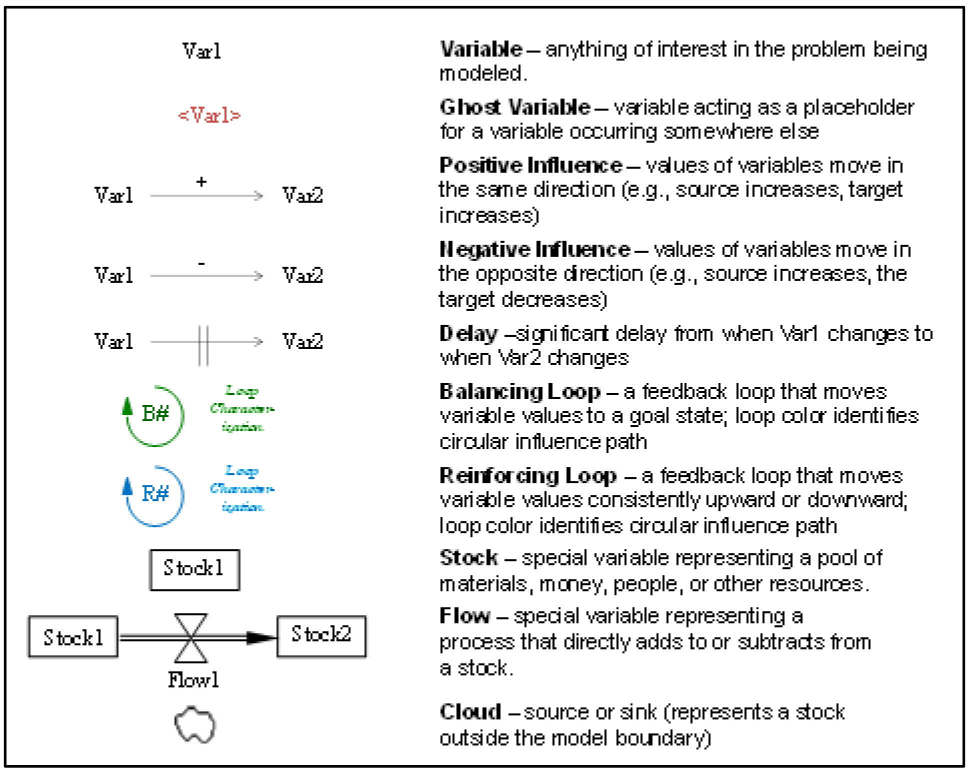


Figure 10. System Dynamics Notation

Panel 6. Leadership Development and Performance Measures for the Acquisition Workforce

Wednesday, May 9, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Rene Thomas-Rizzo,** Director, Human Capital Initiatives, OUSD (A&S) ***The Defense Acquisition Workforce Growth Initiative: Toward an Assessment of Its Impact on Department of the Navy Acquisition Activity*** Ira Lewis, Naval Postgraduate School ***Acquisition Leadership Development and Capabilities for Complexity: Research, Development, Testing, and Evaluation*** Stephen Trainor, Naval Postgraduate School Bruce Rideout, Naval Air Systems Command Paul Shigley, Space and Naval Warfare Systems Command ***Behavior Before Belief: Training for Transformative Change in Defense Acquisition*** Therese C. Bensch, Defense Acquisition University

Rene Thomas-Rizzo—is the Director, Human Capital Initiatives, Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Thomas-Rizzo has also served as the Director of Acquisition Career Management (DACM), the chief of staff for Program Executive Office–Integrated Warfare Systems (PEO IWS), the deputy program manager for Open Architecture and Combat Training Systems at PEO IWS, the Training Systems Development manager/NSST program lead in the PEO IWS 7.0C (Naval Sea Systems Command), and Manpower, Personnel, &Training Branch head (acting) for the Surface Combatants Fleet Support Directorate and LPD 17 Life Cycle Supportability manager in Program Executive Office–Ships organization. Thomas-Rizzo holds a bachelor's degree in business administration and a master's degree in business administration from Mary Washington College and has achieved Level 3 certification in Defense Acquisition Program Management.

Acquisition Leadership Development and Capabilities for Complexity: Research, Development, Testing, and Evaluation

Stephen Trainor—is Visiting Associate Professor of Ethics in the Graduate School of Business and Public Policy at the Naval Postgraduate School in Monterey, CA. [sctraino@nps.edu]

Bruce Rideout—is Director of Continuous Process Improvement (AIR 1.0), NAVAIR, and a doctoral student at Florida Institute of Technology. [bruce.rideout@navy.mil]

Paul Shigley—is Assistant Chief Technology Officer for International Cooperation, SPAWAR, and a doctoral student at the Naval Postgraduate School in Monterey, CA. [paul.shigley@navy.mil]

Abstract

Department of the Navy (DoN) strategies and plans continue to highlight the need for new thinking and innovative approaches to meet the demands of the future force. Navy leadership has called for cultural changes and programmatic improvements to the way the civilian workforce is prepared for leadership roles and responsibilities. While traditional competency models meet some organizational goals and needs, a more responsive approach to leadership development and capabilities may be needed to meet emergent challenges and opportunities. To address the complex challenges facing the DoN workforce, this research project integrates a complexity perspective of leadership development and capabilities with a process model of organizational learning to (1) study how a complexity perspective of leadership development and capabilities contributes to human and social capital strategies of the DoN Research, Development, Testing and Evaluation (RDT&E) workforce, and (2) assess a process model of organizational learning that integrates relevant forecasts of leadership "know-how" needed to meet organizational challenges. To conduct this study, an innovative hybrid-Delphi method of expert forecasting and consensus-building is tested with leadership panels drawn from DoN RDT&E facilities.

Introduction

Department of the Navy (DoN) strategies and plans continue to highlight the need for new thinking and innovative approaches to meet the demands of the future force (DoN, 2016, 2017). For example, Navy leadership has called for cultural changes and programmatic improvements to the way the civilian workforce is prepared for leadership roles and responsibilities (DoN, 2016; DoN Research, Development, Test, & Evaluation [DoN RDT&E], 2017). In fact, the Navy Research and Development Enterprise 30-year plan calls for the creation of a leadership development program focused specifically on the future civilian research and development workforce. However, studies of the federal civilian workforce regularly identify serious challenges related to leadership training and development programs that fail to effectively address organizational needs in a complex and rapidly changing environment (Ingraham & Getha-Taylor, 2004; National Academy of Public Administration, 2017).

Some estimates show that organizations worldwide spend more than $30 billion annually on the selection, training and development of organizational leaders (Hrivnak, Reichard, & Riggio, 2009). This expense reflects the perceived importance of leadership to organizational success. However, the research and assessment behind this strategic investment lags other areas of organizational learning (Boyatzis, 2007). Possible explanations for this lag include the rapid growth of the training industry in response to high organizational demand and a disconnect between the training industry, internal training functions, and the organizational research community (Hrivnak et al., 2009).

This research project focuses on one area of potential disconnect, the work on leadership development and managerial competencies. Many organizations, including the federal government, rely on some form of individual competency model for training programs and role performance (Boyatzis, 2007). The traditional competency model of leader development employs a fixed set of general role competencies that correspond to valued leadership behaviors. Competency models cultivate desired leadership behaviors through standardized training and development programs. However, many general competency models are based upon research and validation efforts conducted 30 to 40 years ago (Boyatzis, 2011). As a result, leader development efforts may not be aligned with the needs of complex and rapidly changing organizational environments. While traditional competency models meet some organizational goals and needs, a more responsive approach to leadership development and capabilities may be needed to meet emergent challenges and opportunities.

Research Problem and Question

The DoN Civilian Workforce Framework highlights the emergence of a “new age of competition” and increasing complexity and pace of change that demands a more effective military and civilian workforce (DoN, 2016). The DoN 30-year Research and Development plan calls for a shift in organization culture that “values learning, collaboration, innovation and the importance of diversity of thought, culture and background in the generation of concepts and proposed solutions” (DoN RDT&E, 2017). Similarly, Navy leadership advocates for the creation of a “learning culture” capable of addressing the organizational and strategic challenges and opportunities facing the Navy (DoN, 2017).

While DoN leadership and other experts seek new leadership strategies, little seems to have changed beyond the use of new tools and technologies that facilitate ease of access and make learning more flexible. However, modern warfare continues to evolve, resulting in demands and impacts on all aspects of the American defense environment, especially the civilian defense acquisition workforce (DAWF; Trainor, 2017). The DAWF is a specialized sub-component of the DoD workforce with key responsibilities to develop, acquire, and deliver warfighting capabilities to the operational forces of the U.S. Armed Forces (Office of the Under Secretary of Defense for Acquisition, Human Capital Initiatives, 2017).

While the challenges of complexity and change impact the whole of the DoD, the DAWF faces its own unique set of problems. The Defense Acquisition Performance Assessment Project (2006) identified some of the factors that contribute uniquely to the work of defense acquisition. Figure 1 depicts a system where values, goals, and functions of defense acquisition often operate in conflict rather than in alignment. These divergent forces combine with the changes in modern warfare to expose important distinctions, or interactions, tensions, and pressures, that influence the thinking about leadership and the development of leadership capabilities in the DAWF. Trainor (2017) suggested that one way to think about these distinctions is to view them as a function of the unique and complex challenges of defense acquisition, the disconnected structure of the acquisition system, and the cultural influences of leadership and learning within the DAWF.

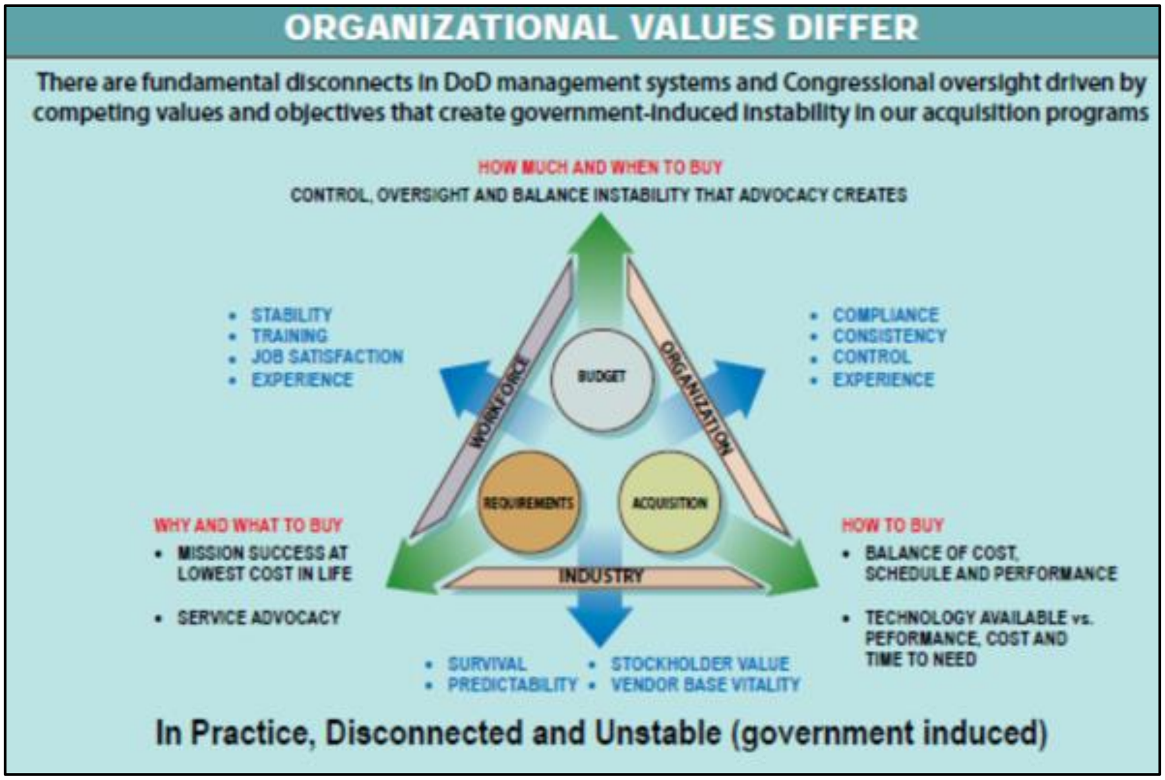


Figure 1. Divergent Forces in Defense Acquisition.
(Defense Acquisition Performance Assessment Project, 2006, p. 4)

To address these unique distinctions and the related demands for different approaches to leadership and learning in the DoN workforce, this research project integrates a complexity perspective of leadership development and capabilities with a process model of organizational learning. The conceptual and methodological design objective of the project is to (1) study how a complexity perspective of leadership development and capabilities contributes to human and social capital strategies of the DoN RDT&E workforce, and (2) assess a process model of organizational learning that integrates relevant forecasts of leadership "know-how" needed to meet organizational challenges.

The research question addressed by the project is, "What organizational capabilities and leader development needs best position the DoN acquisition workforce to meet future challenges and opportunities of a complex environment?" The DoN RDT&E workforce was chosen as the focus of this research because it operates as an integral function within the acquisition system and is impacted by many of the forces and distinctions described above. The output of this project contributes to and extends the emerging field of complexity leadership, adds conceptual rigor to the practice of leadership development and capabilities in the DoN RDT&E civilian workforce, and provides organizational leaders with a practical method for forecasting and prioritizing emergent challenges and needs in a complex environment.

Technical Background and Review of the Literature

The complex global challenges, dynamic nature of the operating environment, and the pace and pervasiveness of technological change demand greater alignment and synergy across the DoN military and civilian workforce (DoN, 2016). The world is operating at machine speed and the workforce must be smarter, more agile, and adaptable to support the mission. The past focus on technological breakthroughs and increased investment in new systems to meet the threat is no longer sufficient to accomplish the mission. An increased emphasis on leadership and innovation is seen as vital to current and future success. The message permeating leadership at all levels of the DoN is the need to cultivate new and different ideas and collaborate and share knowledge that challenges assumptions and provides new perspectives on emerging adaptive problems (DoN, 2017).

In contrast to the call for innovation, rapid learning, and change, the traditional defense civilian workforce model of leadership development emphasizes a stable set of role competencies, broadly applied development opportunities, and specific performance expectations (OUSD[AT&L], 2017). This general role competency model identifies positional competencies and designs training and development necessary for functional success in a role. The role competency and performance model offers predictable and generalizable results and provides scalable education, training, and development programs that benefit a large organization, like the DoN. Despite the clear advantages and efficiencies of the role competency model, leaders continue to call for capabilities of speed, agility, adaptability, and innovation to meet the demands of the operating environment (DoN, 2017; DoN RDT&E, 2017). In response to the call for new ideas on leadership, this research project integrates different perspectives and theories of leadership with a new process of organizational forecasting and consensus-building. The following theoretical and conceptual foundations serve to orient and organize the study.

Complexity leadership theory argues that traditional transactional and hierarchical approaches to organizational leadership are increasingly incapable of delivering new capabilities to solve complex challenges and rapidly shift to capture new opportunities (Uhl-Bien & Arena, 2017). The theory suggests that leadership is more than a role, a style or an approach, but rather an emergent process that occurs as organizations work through the tensions, pressures, and interconnections needed to survive and thrive in a complex environment (Uhl-Bien & Arena, 2017). Complexity leadership theory proposes that organizational effectiveness depends on dynamic, interrelated forms of leadership, which enable creativity and scale innovation into new organizational capabilities (Uhl-Bien & Arena, 2017).

The *dynamic capabilities* literature argues that the current operating environment is marked by globalization, technology, and competition and that organizational success flows from two core capabilities. *Operational capabilities* are the ability to exploit current environments by leveraging existing resources and reinforcing proven operating routines, while *dynamic capabilities* are the simultaneous exploration, creation, and adaptation of organizations to changes in the environment that produces new operational capabilities (Denford, 2013). This research project is focused on *dynamic managerial capabilities*, or the managerial and leadership capacity to search, seize, and transform learning into new operational capabilities (Augier & Teece, 2009).

The *leadership capabilities* literature (Boyatzis, 2007; Boyatzis, 2011) suggests that technical, emotional, and social intelligence competencies are related to the development and performance of effective leaders and managers. The contingency theory of job performance considers aspects of job demands, organizational environment, and the individual in order to determine the intersection, or best fit, for individual performance and organizational success (Boyatzis, 2007).

The literature on *individual and organizational learning* describes the processes by which new knowledge is acquired, transformed, and applied in response to changing circumstances and problems. *Experiential learning theory* describes how individuals engage in a cyclic process of knowledge creation “through the transformation of experience” (Kolb & Kolb, 2009, p. 44). The *organizational knowledge* literature focuses on how leaders “enable, crystalize, and connect” the knowledge created by individuals to organizational knowledge systems (Nonaka et al., 2006, p. 1179).

Research Methodology

This research project uses a mixed mode (qualitative and quantitative) methodology to gather organizational forecasts and prioritize leadership development needs from senior managers and key leaders at two DoN RDT&E facilities. The Delphi method of panel consensus-building is designed to leverage the knowledge of qualified individuals, decision-makers, and stakeholders who are highly trained or experienced in a particular subject area, or who are unique experts in a specified field (Hsu & Sandford, 2007, p. 3).

The Delphi method was developed by the RAND organization for use in the post–World War II field of security studies and the methodology has been used in many settings, including curriculum design and training content in professional development (Linstone & Turoff, 2011). The traditional Delphi technique presents a set of open-ended prompts on a particular topic and gathers individual responses, forecasts, and priorities from participants without any face-to-face interaction. This method offers benefits of simplicity and efficiency while avoiding the influence of group dynamics that tend to negatively impact response quality. However, the individualized design of the Delphi is also a key limitation of the method, because the traditional Delphi fails to capture the shared interaction, experiences, and learning that are key to gaining broader insights and cultivating deeper knowledge about an issue of importance to a group. The lack of specific context and relevance is one reason why the Delphi method has not been used as an organizational or group learning, forecasting, and decision-making process (Landeta, 2006).

To overcome the limitations of the traditional Delphi technique, this research project employs a hybrid-Delphi method consisting of two distinct phases: a facilitated face-to-face Nominal Group Technique (NGT) discovery phase and an online Delphi panel forecasting phase. In the NGT discovery phase, respondents answer and discuss open-ended question prompts. In the online Delphi panel phase, respondents participate in two closed-ended criteria and consensus rating panels based on the aggregated group responses (Landeta et al., 2011). After receiving a report on panel findings, respondents participate in a brief process assessment survey on the effectiveness and impact of the hybrid-Delphi methodology.

The hybrid-Delphi method used in this project is designed to provide robust interactional and individual components, which may be more representative of the types of interactions needed by organizational or group members as they attempt to gather information to solve complex challenges. In particular, the hybrid-Delphi method encourages creativity and openness within the NGT phase, while integrating and focusing diversity of experience and individual perspective in the Delphi phase, which closely resembles the process and thinking often associated with innovation and complex problem-solving (Drucker, 1999; Heifetz, 2006).

In addition to the potential benefits of more relevant, creative, and robust data, the hybrid-Delphi gathers insights and forecasts in a manner less costly and burdensome than individually structured interviews or focus group methods. The use of the hybrid-Delphi methodology also promotes the integration and collaboration of researchers and practitioners, who together add scientific rigor and validity as they seek organizational insights on important human and social capital challenges (Plummer & Armitage, 2007). The specific hybrid-Delphi methodology used in this research project is described in detail below.

Recruitment Phase

The investigators conducted briefings on the research topic with senior organization leaders and gained approval to conduct the research and recruit Research & Development and/or Testing & Evaluation leaders (GS-11 to 15 and SES/SL/ST) within a directorate of the Naval Air Systems Command (NAVAIR) and the Space and Naval Warfare Systems Command (SPAWAR). Senior organization leaders provided a list of potential panel members or allowed investigators access to potential panel members meeting the general criteria. Potential subjects received an email invitation from one of the investigators along with a brief description of the research. Participants were sent a maximum of two follow-up emails/phone calls in the event a potential participant did not respond to the initial recruitment. A target sample size of 10–15 members was sought for each of two panels, totaling 20–30 participants. This sample size satisfies the methodological designs of the Delphi method of forecasting and consensus-building, while accommodating the potential effects of respondent attrition and optimizing the data analysis workload requirements of the research team.

To minimize undue influence during the recruitment process, the research team contacted potential subjects via email or phone. While the invitation mentioned the approval of the organization to conduct the research, the invitation was clear to state that participation was voluntary and in no way was there an expectation to participate by the command. Investigators ensured that during the recruitment and data collection phases, there was no official interaction between senior organizational leaders and individual participants regarding the research project. While it was possible that informal interaction might have occurred between individual subjects who participated in the research, investigators reminded participants that they should respect the privacy and confidentiality of other participants.

Data Collection Phase

The research study subjects participate in three different data collection stages of the project (see the Appendix—Research Protocol and Instrumentation). The first stage of data collection is a facilitated Nominal Group Technique (NGT) discussion with 10–15 leaders from the organization. In this stage, subjects are asked to respond privately and independently to four open-ended questions. The initial ideas are recorded on a whiteboard or flip-chart, and the group conducts a discussion of clarity, relevance, and logic for each of the items. The investigators organized and conducted content analysis on data collected in the NGT discussion and uploaded this data to the Naval Postgraduate School LimeSurvey program for use in the second stage of data collection. The second stage of data collection was the Delphi panel, consisting of two online surveys based on the information gathered in NGT discussion stage. The first survey involves rating items according to defined criteria, and the second survey involves rating the priority, or relative importance, of items from the first survey. Participants received individual email links to the online survey. To protect confidentiality of participant responses, no IP address or personally identifying information (PII) was collected by the survey instrument.

The investigators combined the results the NGT discussion and both surveys in a final report provided for participant review. Following review of the final report, participants received an online survey to assess the effectiveness and impact of the overall process. The NGT was designed to last 90 minutes, and each of the online surveys were designed to take 20–30 minutes to complete. The total estimated time of participation was three hours over the course of one month.

Data Analysis and Next Steps

The process of data collection began at both RDT&E facilities in April 2018 and continued until May 2018. The research team conducted analysis at each stage of data collection using qualitative content analysis following the NGT discussions and standard statistical analysis following the Delphi panel survey stage. Preliminary results will be included in the Acquisition Research Program Symposium presentation of this research project and in the final project technical report.

The analysis and conclusions of this research project are expected to contribute to the public interest by (1) extending the emerging field of complexity leadership, (2) adding conceptual rigor to the practice of leadership development and capabilities in the DoN RDT&E civilian workforce, and (3) providing organizational leaders with a practical method of forecasting and prioritizing emergent challenges and needs in a complex environment.

References

Augier, M. & Teece, D. J. (2009). Dynamic capabilities and the role of managers in business strategy and economic performance. *Organization Science, 20*(2), 410–421.

Baker, A. C., Jensen, P. J., & Kolb, D. A. (2005). Conversation as experiential learning. *Management Learning, 36*(4), 411–427.

Boyatzis, R. E. (2007). Competencies in the 21st century. *Journal of Management Development, 27*(1), 5–12.

Boyatzis, R. E. (2011). Managerial and leadership competencies: A behavioral approach to emotional, social, and cognitive intelligence. *Vision, 15*(2), 91–100.

Boyatzis, R. E., Rochford, K., & Cavanagh, K. V. (2017). Emotional intelligence competencies in engineer's effectiveness and engagement. *Career Development International, 22*(1), 70–86.

Denford, J. S. (2013). Building knowledge: Developing a knowledge-based dynamic capabilities typology. *Journal of Knowledge Management, 17*(2), 175–194.

Defense Acquisition Performance Assessment Project. (2006). *Defense acquisition performance assessment report.* Washington, DC: Deputy Secretary of Defense.

DoD. (2016). *Acquisition workforce strategic plan: FY2016–FY2021*; *Title 10 U.S.C., Sections 115B(D) and 1722B(C).* Retrieved from http://www.hci.mil/docs/Policy/Legal%20Authorities/DOD_Acq_Workforce_Strat_Plan_FY16_FY21.pdf

DoN. (2016). *Navy civilian workforce framework* (Ver. 1.0). Washington, DC: Author.

DoN. (2017). *The future Navy* [White paper]. Washington, DC: Author.

DoN, Research, Development, Test, & Evaluation (DoN RDT&E). (2017). *30-year research and development plan.* Washington, DC: Author.

Dreyfus, C. R. (2008). Identifying competencies that predict effectiveness of R&D managers. *Journal of Management Development, 27*(1), 76–91.

Drucker, P. R. (1999). *Management challenges for the 21st century*. New York, NY: Harper Business.

Heifetz, R. A. (2006). Anchoring leadership in the work of adaptive progress. In F. Hesselbein & M. Goldsmith (Eds.), *The leader of the future 2: Visions, strategies, and practices for the new era*. San Francisco, CA: Jossey-Bass.

Hrivnak, G. A., Reichard, R. J., & Riggio, R. E. (2009). A framework for leadership development. In S. J. Armstrong & C. V. Fukami (Eds.), *The Sage handbook of management learning, education, and development* (pp. 456–475). London, England: Sage.

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment Research & Evaluation, 12*(10). Retrieved from http://pareonline.net/getvn.asp?v=12&n=10

Ingraham, P. W., & Getha-Taylor, H. (2004). Leadership in the public sector: Models and assumptions for leadership development in the federal government. *Review of Public Personnel Administration, 24*(2), 95–112.

Kolb, A. Y., & Kolb, D. A. (2009). Experiential learning theory: A dynamic, holistic approach to management learning, education, and development. In S. J. Armstrong & C. V. Fukami (Eds.), *The Sage handbook of management learning, education, and development* (pp. 42–68). London, England: Sage.

Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting & Social Change, 73,* 467–482.

Landeta, J., Barrutia, J., & Lertxundi, A. (2011). Hybrid Delphi: A methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting & Social Change, 78,* 1629–1641.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting & Social Change, 78*, 1712–1719.

National Academy of Public Administration. (2017). *No time to wait: Building a public service for the 21st century* (Academy Project No. 2214). Washington, DC: Author.

Nonaka, I., von Krogh, G., & Voepel, S. (2006). Organizational knowledge creation theory: Evolutionary paths and future advances. *Organization Studies, 27*(8), 1179–1208.

Office of the Under Secretary of Defense for Acquisition, Human Capital Initiatives. (2017). Mission. Retrieved from http://www.hci.mil/about-us.html#Mission

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2017). *Defense Acquisition Workforce Education, Training, Experience, and Career Development program* (DoD Instruction 5000.66). Retrieved from http:www.esd.whs.mil/dd

Plummer, R., & Armitage, D. R. (2007). Charting the new territory of adaptive comanagement: A Delphi study. *Ecology and Society*, *12*(2), 10. Retrieved from http://www.ecologyandsociety.org/vol12/iss2/art10/

Trainor, S. C. (2017)*. Complex leadership needs in the Defense acquisition workforce* (Acquisition Research Program Technical Report NPS-AN-18-017). Retrieved from http://my.nps.edu/documents/105938399/110483626/NPS-AM-18-017.pdf/c660094b-ecf9-4e15-918f-087ad0f791a9?version=1.0

Uhl-Bien, M., & Arena, M. (2017). Complexity leadership: Enabling people and organizations for adaptability. *Organizational Dynamics, 46,* 9–20.

Yousuf, M. I. (2007). Using experts' opinions through Delphi technique. *Practical Assessment Research & Evaluation, 12*(4). Retrieved from http://pareonline.net/getvn.asp?v=12&n=4

Appendix: Research Protocol and Instrumentation

Research Question

What organizational capabilities and leader development needs best position the DoN acquisition workforce to meet future challenges and opportunities of a complex environment?

Methodology

Nominal Group Technique (NGT) Discovery Stage (Forecasting the Challenges of Complexity Leadership). Respondents are gathered in a room, organized around a U-shaped table with a white board or flip chart for collecting and collating information from the group. Respondents are briefed on the informed consent process and given the opportunity to read and sign the informed form prior to proceeding.

Step 1. Generating Ideas. The facilitator provides a brief overview of the research project and distributes materials to complete the written portion of the NGT. The facilitator presents the four questions to the group in written form and reads the questions aloud to the group. The facilitator invites the group to independently write ideas in brief phrases or statements for each question. Group members independently and privately responds to the following questions, writes the ideas on large post-it notes and then places ideas on flip charts or a white board corresponding to each question prompt.

Question 1 Objective: Identify how leaders experience the tensions, pressures, and interconnections of increasing complexity in their work.

> *"As you work to achieve your organization's current goals and objectives, describe or identify the ways in which you and other leaders experience tensions and interconnections, as well as pressures and demands that are new, different, or unique from those you may have experienced a few years ago."*

Question 2 Objective: Identify how leaders use complexity thinking in their work (e.g., "to catalyze and energize networked interactions that enable emergence and adaptability").

> *"Given the tensions and interconnections, as well as pressures and demands previously identified, envision an organization that fully and effectively achieves its goals and objectives. In a situation such as this, what things do leaders understand, practice, or focus on that are new, different, or unique?"*

Question 3 Objective: Identify how leaders create, facilitate, and manage the adaptive space of complexity (integrate the need to operate, the need to innovate, and the need to adapt in their work).

> *"What things do you and other leaders need to learn, adopt, or change for your organization to fully and effectively achieve its goals and objectives?"*

Question 4 Objective: Identify how leaders gain the understanding and expertise to lead for adaptability.

> *"Envision a leader who has mastered those things needed for your organization to fully and effectively achieve its goals and objectives. In what ways might this leader have acquired the knowledge, skills, and experiences needed for mastery?"*

Step 2. Recording Ideas. The facilitator leads the group members in a round-robin feedback session to concisely capture each idea (without debate). For each question prompt, the facilitator reads aloud ideas on the post-it notes. The facilitator invites group members to offer a different emphasis or variation on ideas, or to clarify meaning if ideas are repeated or unclear. The facilitator invites group members to offer additional ideas that are not included on the list. The facilitator proceeds until all members' ideas have been documented.

Step 3. Discussing Ideas. Each recorded idea is then discussed to determine clarity, relevance, and logic. For each idea, the facilitator asks, "Are there any questions or comments group members would like to make about the item?" The creator of the idea need not feel obliged to clarify or explain the item; any member of the group can play that role. The facilitator then asks, "Are there any organizing themes that appear across the responses for this question?" The process repeats for each question. The session is complete at this point.

Step 4. Content Analysis. The research team conducts a content analysis and categorization of responses to the open-ended questions based upon theory and the NGT session. The research team constructs a list of common themes and individual items from the responses provided in the NGT phase for use in the Delphi panel phase.

Delphi Panel Assessment Stage (Prioritizing Organizational Leadership Development and Capabilities). Panel members receive (via email) a secure link to complete a survey using the NPS LimeSurvey program. Respondents are asked to complete the survey within seven working days of receiving the email. An email reminder is sent to all panel participants after five working days and one working day prior to close of data collection.

Delphi Panel—Round 1 (Criteria Rating—Complexity Leadership Development Needs and Capabilities). Respondents are presented the following forecasting questions in response to the NGT Discovery Phase information.

> *"For each of the following questions, there are corresponding themes and items collected during the group phase. For each question, please review the information and provide your level of agreement for each item associated with a question."*
>
> *(5-point Likert scale, 1=Strongly Disagree, 5=Strongly Agree)*

Question 1 Objective: Respondents recognize and exploit the experience of tensions, pressures, and interconnections into learning about complexity.

> ***Question 1:*** *To what extent are these leadership challenges the result of increasing complexity?*

> 1. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 2. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 3. *Etc.*

Question 2 Objective: Respondents acquire and apply knowledge of complexity in a relevant situation.

> ***Question 2:*** *To what extent are these different or unique capabilities "what it takes" to lead in complexity?*

> 1. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 2. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 3. *Etc.*

Question 3 Objective: Respondents assimilate knowledge about leadership in complexity and transfer it to role-related learning.

> ***Question 3:*** *To what extent will these development activities equip leaders for the challenges of complexity?*

> 1. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 2. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 3. *Etc.*

Question 4 Objective: Respondents internalize and transform development into new forms (identities) of complexity leadership.

> ***Question 4:*** *To what extent will these development opportunities provide the "know-how" to lead effectively in complexity?*

> 1. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 2. *Item (1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)*
> 3. *Etc.*

Delphi Panel—Round 2 (Consensus Rating—Complexity Leadership Priorities). Respondents are presented the following consensus questions in response to criteria means from Delphi Panel Round 1.

> *"For each of the following statements, please review the mean criteria ratings for each item and provide your level of agreement with the priority, or relative importance, of the items in the list. In the space provided, please provide a brief explanation for responses of Strongly Disagree or Disagree."*
>
> *(5-point Likert scale, 1=Strongly Disagree, 5=Strongly Agree)*

Question 1: *The ratings of leadership challenges reflect the priority, or relative importance, of those challenges for this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree) (Open-ended response)

Question 2: *The ratings of different or unique capabilities reflect the priority, or relative importance, of "what it takes" to lead in complexity for this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree) (Open-ended response)

Question 3: *The ratings of development activities reflect the priority, or relative importance, of how to equip leaders for the challenges of complexity for this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree) (Open-ended response)

Question 4: *The ratings of development opportunities reflect the priority, or relative importance, of how to provide the "know-how" to lead effectively in complexity for this organization?*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree) (Open-ended response)

Process Assessment Survey (Methodological Effectiveness and Impact). Panel members receive (via email) a secure link to complete a survey using the NPS LimeSurvey program. Respondents are asked to complete the survey within seven working days of receiving the email. An email reminder is sent to all panel participants after five working days and one working day prior to close of data collection.

Process Effectiveness

"This section asks you to rate the effectiveness of each phase of the process in which you have participated. For each of the activities listed below, please provide your level of agreement. In the space provided, please provide a brief explanation for responses of Strongly Disagree or Disagree."

Question 1. *"The following activity was effective in identifying the leadership development and capabilities needed to best position this organization for future challenges and opportunities of a complex environment."*

(5-point Likert scale, 1=Strongly Disagree, 5=Strongly Agree)

Group Discovery Phase

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Online Complexity Leadership Forecasting Phase

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Online Leadership Development and Capabilities Priorities Phase
(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Process Learning Value

"This section asks you to rate the learning value of the experiences in which you have participated. Please provide your level of agreement with the statements below. In the space provided, please provide a brief explanation for responses of Strongly Disagree or Disagree."

***Question 2**. "The following are valuable learning experiences in this organization."*

Associating with other colleagues outside of our technical roles

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Being able to test your own ideas with other colleagues

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Acquiring new ideas and learning something new

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Contributing your knowledge to solve organizational challenges

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Using a deliberative method to think about organization priorities
(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Developing consensus on important decisions

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Process Impact

"This section asks you to rate the impact of the process on your organization. Please provide your level of agreement with the statements below. In the space provided, please provide a brief explanation for responses of Strongly Disagree or Disagree."

***Question 3**. I am confident that the results of this process reflect group consensus.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

***Question 4**. This process produced timely results for this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

***Question 5**. The process produced actionable priorities for this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

***Question 6**. The process will help this organization transform learning into new operational capabilities.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

***Question 7**. The process will help this organization adapt and connect the knowledge and ideas of individuals to solve complex challenges.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

***Question 8**. I would recommend a process like this to other teams and leaders in this organization.*

(1=Strongly Disagree, 2-Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree)

(Open-ended response)

Final Comments

Question 9. *What comments or suggestions for improvement can you offer about this process?*

(Open-ended response)

Behavior Before Belief: Training for Transformative Change in Defense Acquisition

Therese C. Bensch—is a professor of Engineering for the Defense Acquisition University (DAU). She is a retired U.S. Navy officer with extensive experience in management of high impact, high visibility space and information technology acquisition programs. She has an MS in Meteorology and Physical Oceanography from the Naval Postgraduate School and a PhD in Public Policy and Administration with a specialization in Public Management and Leadership from Walden University. [therese.bensch@dau.mil]

Abstract

For decades, the Department of Defense has been plagued by persistent cost, schedule, and performance problems in defense acquisition programs. Increasing technological complexity, funding instability, and changing requirements are driving the need for transformative change in the acquisition workforce. Although transformational culture change can rarely be made directly, leaders can change behavior that should create positive outcomes, which can then be incorporated into cultural beliefs. The study's theoretical construct was the behavior-before-belief model of organization change. Recent acquisition policy changes were intended to improve efficiency and are demonstrating some improvements, yet little is understood about whether training efforts related to these policies are producing policy-compliant behavior. The purpose was to examine through an ex post facto, cross-sectional study whether there is a significant relationship between learning from Defense Acquisition University (DAU) training in acquisition policy and application of learned policy-compliant behavior, as represented by the variables learning achieved and applied training. DAU data spanned 19 months, included 334,000 training events, were separated into 40 course-type subgroups, and were analyzed through hierarchical regression. The findings confirmed that the independent variable of learning achieved is predictive of policy-compliant behavior change ($p < .001$). Additionally, predictors of learning and application were determined.

Introduction

The Department of Defense (DoD) requires transformative culture change in the acquisition of defense systems to adapt to environmental changes accelerated by globalization, technology, and fiscal instability. Public policy can be an effective and legitimate instrument for implementing needed change in the DoD. Dissemination of public policy that articulates the policymakers' vision and goals can facilitate implementation of organizational change by first creating behavioral changes (Burke, 2011; Schein, 2010; Wedel et al., 2005). Transformative change implementation strategies should focus on creating new behavioral processes that will lead to cultural changes in support of the needed social or organizational change. It is well-documented that culture change in mature organizations like the DoD cannot be successfully implemented directly; however, behavior can be changed by leaders to drive culture change (Burke, 2011; Clawson, 2012; Harris & Ogbonna, 2011; Linn, 2008; Schein, 2010; Weick & Sutcliffe, 2007; Weier, 2008). Little is known about the drivers of behavior change in the defense acquisition workforce. The purpose of this study was to bridge this gap in knowledge by investigating the relationship between mandated training of the defense acquisition workforce and application of policy-compliant behavior.

My study was conducted to address the quantitative research question: To what extent does the Defense Acquisition University (DAU) policy-based training enhance policy-compliant behavior of the DoD acquisition workforce personnel? To find the answer to this question, two additional questions were posed: What are the important predictors of learning new concepts and behaviors in DAU training, and what are the important predictors of application of learned concepts from DAU training? Application of learned concepts from DAU training was the policy-compliant behavior change tested in this study.

I employed a quantitative, ex post facto, longitudinal study design that used multiple regression techniques to analyze 19 months of DAU secondary survey data. The secondary data collected and maintained by the DAU provided the data required for my data analysis effort, which was designed to generate results that are representative of and can be generalized to the defense acquisition workforce population of approximately 150,000 military and civilian personnel (DAU, 2011; GAO, 2012). All acquisition personnel are required to attend DAU career-field specific certification training (Fishpaw, 2010). Eligible study participants were defense acquisition workforce members who responded to DAU online postevent and follow-up surveys following training events during a 19-month period from January 1, 2014, to July 31, 2015. I further divided the DAU sample of more than 334,000 DAU training events into 40 course type subgroups to avoid bias inequality by ensuring internal homogeneity of subgroups.

A probability sampling design allowed me to ensure that all units of the defense acquisition population had an equal probability of being included in the sample. A stratified random sampling technique was used, since subset proportions in the DAU secondary data were known (Field, 2009). I conducted an a priori power analysis to determine appropriate minimum sample sizes of roughly 50 to 790 depending on effect size for a linear multiple regression fixed model with an R-squared deviation from zero (null hypothesis F-test). Actual sample sizes ranged from roughly 180 to 2150. The study found that the important predictors of *applied training* and *learning achieved* have large effect sizes, therefore, all samples were adequately sized for the regression analysis.

Theoretical Foundation

I based the behavior-before-belief model of culture change, used as the framework for this research, on Edgar Schein's three-stage model of learning/change and his theory that behavior changes can lead to changes in culture. The first stage of cultural change is unfreezing the organization by creating the motivation to change. The literature provided that a rapidly changing environment coupled with crises and scandals creates motivation to change, disconfirms dysfunctional assumptions and behaviors, and builds survival anxiety in the defense acquisition workforce (Brown, 2010; Eide & Allen, 2012; Hannay, 2009; Kotzian, 2010; O'Neil, 2011; Weier, 2008). Formal defense acquisition training reduces learning anxiety by creating a psychologically safe environment and an understanding that a new way of doing business is possible, such as transforming competitive relationships into collaboration and teamwork.

The second stage of cultural change is cognitive restructuring through learning new concepts, new meanings for old concepts, and new judgment standards. The DoD has begun the unfreezing process by changing acquisition policies to drive culture change in response to acquisition program crises driven by a rapidly changing external environment (DoD, 2015; GAO, 2017; Under Secretary of Defense, 2015). These changes encourage an internal environment in which cognitive restructuring can come through new learning. Formal training can provide this new learning experience and is required for all acquisition professionals. The third stage of cultural change is refreezing, or internalizing the new concepts, meanings, and standards by incorporating them into the organization's identity and relationships. If the new learned behaviors correct problems and produce better outcomes, the new lessons should stabilize, be internalized as new tacit assumptions, and eventually lead to culture change (Schein, 2010).

For a large, old organization like the DoD, a critical step for managing culture change is missing from Schein's three-stage model. Although evolutionary change in organizational culture happens naturally in response to external environment changes, the literature suggested that rapid changes in the DoD's environment are creating disequilibria that have forced transformational change to occur, which in turn challenges deeper cultural assumptions (Burke, 2011; Eide & Allen, 2012; Hannay, 2009; Kotzian, 2010; O'Neil, 2011; Stevens, Plaut, & Sanchez-Burks, 2010). Schein (2010) argued that existing cultures that have been successful and stable over time cannot be changed directly unless the organization is dismantled, which is not a viable option for the DoD. However, culture change can be launched by behavior change. Changes in behavior that result in better outcomes will encourage personnel to reexamine their beliefs and assumptions and lead them to adopt new beliefs and assumptions.

The behavior-before-belief model of culture change (Table 1) adds a stage between Stages 2 and 3 of the three-stage model of learning/change presented by Schein (2010). The additional stage is *applying new behaviors learned to correct problems and produce better outcomes*. The DAU can teach acquisition policy, but the DAU cannot make acquisition professionals learn new policy-compliant behaviors or apply these learned behaviors on the job. I conducted the research in two parts focusing of Stages 2 and 3 of the expanded model. Part 1 of the study tested student learning of new concepts in DAU policy training courses and determined the predictors of learning. Part 2 of the study examined students' on-the-job application of new behaviors learned following DAU policy training courses and determined the predictors of the students' ability to apply the training.

Table 1. Behavior-Before-Belief Model of Culture Change

(Adapted from Schein, 2010)

	Behavior-Before-Belief Model of Culture Change
1	Unfreezing the organization by creating the motivation to change
2	Cognitive restructuring through learning new concepts, new meanings for old concepts, and new judgment standards
3	Applying new behaviors learned to correct problems and produce better outcomes
4	Refreezing, or internalizing the new concepts, meanings, and standards

Null and Research Hypothesis

Using the behavior-before-belief model of culture change, predictors of State 2, learning new concepts, and Stage 3, applying new behaviors learned, outcomes were tested using statistical analysis of secondary data provided by the DAU. The outcome *learning new concepts* is represented in the data by the variable *learning achieved*. The outcome *applying new behaviors* learned is represented in the data by the variable *applied training*. The 13 hypotheses tested in this study are provided in Table 2.

Table 2. Null and Research Hypothesis

	Null Hypothesis Either that the correlation coefficient is equal to zero or that the slope weight is equal to zero, which means that there is not a correlation, or relationship, between…	**Research Hypothesis** There is a significant positive correlation between…
Hypothesis 1	the predictor, *career benefit*, and the outcome, *learning achieved*	*career benefit* and *learning achieved* and that *learning achieved* can be predicted from *career benefit*
Hypothesis 2	the predictor, *worthwhile investment*, and the outcome, *learning achieved*	*worthwhile investment* and *learning achieved* and that *learning achieved* can be predicted from *worthwhile investment*
Hypothesis 3	the predictor, *exercises value*, and the outcome, *learning achieved*.	*exercises value* and *learning achieved* and that *learning achieved* can be predicted from *exercises value*
Hypothesis 4	the predictor, *examples helped*, and the outcome, *learning achieved*	*examples helped* and *learning achieved* and that *learning achieved* can be predicted from *examples helped*
Hypothesis 5 (Instructor-Led Training [ILT] Only)	the predictor, *instructor enthusiasm*, and the outcome, *learning achieved*	*instructor enthusiasm* and *learning achieved* and that *learning achieved* can be predicted from *instructor enthusiasm*
Hypothesis 6 (ILT Only)	the predictor, *application discussed*, and the outcome, *learning achieved*	*application discussed* and *learning achieved* and that *learning achieved* can be predicted from *application discussed*
Hypothesis 7 (ILT Only)	the predictor, *instructor knowledge*, and the outcome, *learning achieved*	*instructor knowledge* and *learning achieved* and that *learning achieved* can be predicted from *instructor knowledge*
Hypothesis 8 (Self-Paced Web [SPW] Only)	the predictor, *delivery effective*, and the outcome, *learning achieved*	*delivery effective* and *learning achieved* and that *learning achieved* can be predicted from *delivery effective*
Hypothesis 9 (SPW Only)	the predictor, *graphics meaningful*, and the outcome, *learning achieved*	*graphics meaningful* and *learning achieved* and that *learning achieved* can be predicted from *graphics meaningful*
Hypothesis 10	the predictor, *learning achieved*, and the outcome, *applied training*	*learning achieved* and *applied training* and that *applied training* can be predicted from *learning achieved*
Hypothesis 11	the predictor, *task applicability*, and the outcome, *applied training*	*task applicability* and *applied training* and that *applied training* can be predicted from *task applicability*
Hypothesis 12	the predictor, *resources provided*, and the outcome, *applied training*	*resources provided* and *applied training* and that *applied training* can be predicted from *resources provided*
Hypothesis 13	the predictor, *manager involvement*, and the outcome, *applied training*	*manager involvement* and *applied training* and that *applied training* can be predicted from *manager involvement*

Data Collection

The large DAU dataset was divided into 40 subset samples broken out by postevent or follow-up survey type and for the covariates, delivery type and functional topic. The postevent survey data, collected at the end of each course, supported regression analysis of predictors of the *learning achieved* outcome. The follow-up survey data, collected greater than 60 days post course, provided the data needed for regression analysis of predictors of the *applied training* outcome.

The two training delivery type covariates are instructor-led training (ILT) and self-paced web training (SPW). The 10 functional course topic covariates provide required training for the major defense acquisition functional certifications and included acquisition (ACQ); business, cost estimating, and financial management (BCF); contract management (CM); contracting (CON); engineering (ENG); logistics (LOG); program management (PMT); production, quality and manufacturing (PQM); science and technology management (STM); and test and evaluation (TST). All acquisition workforce personnel are required to take online and residency courses for functional certification represented in these samples and are provided the opportunity to respond to postevent and follow-up surveys. Random sampling techniques were used to provide appropriately sized data samples for analysis, as needed. I analyzed multiple samples within the larger data subsets and compared the SPSS outputs to ensure consistent results.

Study Results

I used IBM SPSS Statistics 21 to perform multiple regression analyses on the DAU postevent and follow-up survey data samples to test whether the outcome *learning achieved* and the outcome *applied training* can be predicted by a linear combination of multiple predictor variables. Regression was used to find the best-fitting straight line, or regression line, for the DAU data set. The regression line was then used to predict the outcome value from the value of the predictor variables (Field, 2009).

The regression model must be unbiased for the findings to be generalized to the broader acquisition workforce population, which means that on average the sample and the population models would be the same. To be sure that this is true, necessary underlying assumptions must be met. These assumptions include variable types (independent variables are quantitative or categorical and dependent variables are quantitative, continuous, and unbounded), nonzero variance (independent variables), no perfect multicollinearity, homoscedasticity, independent variables are uncorrelated with external variables, independent errors, normally distributed errors, independence (dependent variable values from separate entity), and linearity (Field, 2009; Green & Salkind, 2011). Each of these assumptions was checked using SPSS validation techniques and these assumptions were met. This means the regression model from the sample is the same, on average, as the regression model from the population (Field, 2009). A comprehensive analysis of the multiple regression results from the samples was performed.

Analysis (Part 1): Predictors of Learning Achieved

Descriptive statistics characterize the 20 samples used for the analysis of Stage 2 of the behavior-before-belief model of culture change to determine important predictors of the *learning achieved* outcome for both the ILT and SPW DAU courses. The descriptive statistics included mean, standard deviation, and sample size. The means of the Likert score (7 = strongly agree and 1 = strongly disagree) responses to the variables indicated how the students in each sample perceived the variable in question. The means of the *learning achieved* outcome for resident ILT courses ranged from a low of 5.76 for ENG to a high of 6.53 for CM. For online SPW courses, the means for *learning achieved* ranged from a low of 5.38 for ENG to a high of 5.88 for CM. These findings indicate resident ILT courses may be more effective in achieving learning than online SPW courses.

For the resident ILT courses, the instructor variables tend to have the highest mean scores even though regression analysis results provided in this paper indicated that the instructor variables are the least important predictors of learning. The variables that measured how worthwhile the training was tended to have the lowest mean scores even though analysis shows them to be the most important predictors of learning. The online courses showed similar results with the most important predictors of learning being scored the lowest on the postevent surveys.

I used the SPSS correlation matrix for each sample as a starting point for exploring the relationships between predictors and the outcome and for an initial check for multicollinearity. The correlation matrix showed the value of Pearson's correlation coefficient between variable pairs. No collinearity was found in the data, because there were no substantial correlations ($r > .9$) between predictors. The findings confirmed that the *career benefit* variable correlates best with the outcome ($p < .001$), so this variable should best predict *learning achieved*. This finding supports the Bontis, Hardy, and Mattox (2011) study which found the strongest driver of learning in DAU courses was whether the student believed that the training was worthwhile. I chose the hierarchical method for variable entry into the model, so summary statistics were repeated for each hierarchy stage.

The SPSS model summaries provided the multiple correlation coefficient (R) between the predictors and the outcome and the value of R-square (data included in Table 3), which measured how much of the outcome variability is accounted for by the predictors (Field, 2009). Model 1 had only the *career benefit* and *worthwhile investment* predictors included and the R-square values for all samples ranged from a high of .695 for LOG SPW to as little as .422 for ACQ ILT. This means that for all samples *career benefit* and *worthwhile investment* accounted for between 42% and 70% of the variation in *learning achieved* depending on functional topic and delivery method. However, when the *exercises value* and *examples helped* predictors are included in model 2, this value increases to as much as .721 or 72% (LOG SPW), and as little as .533 or 53% (ACQ ILT) of the variance in *learning achieved*. When the remaining predictors are added in Model 3, this value increases only slightly to 73% for LOG SPW and 54% for ACQ ILT. These findings indicate that the predictors specific to the training delivery type account for 1% or less of the variability in the outcome, *learning achieved*. The predictors specific to the ILT delivery type are *instructor enthusiasm, application discussed*, and *instructor knowledge*. The predictors specific to the SPW delivery type are *delivery effective* and *graphics meaningful*.

The adjusted R-square was analyzed for all subsets and gives some idea of how well the model can be generalized to the defense acquisition workforce population. For all samples, the adjusted R-square value was the same, or close to, the value of R-square, meaning that testing the population model instead of a sample model would account for the same outcome variance (Green & Salkind, 2011). The change statistics described the difference made when new predictors were added to the model by reporting whether the change in R-square is significant. This was tested using an F-ratio and the change in F was analyzed for all data samples.

The Durbin-Watson statistic was analyzed to determine whether the assumption of independent errors is correct, which means that observation residual terms are uncorrelated. A conservative rule suggests that values less than 1 or greater than 3 could be problematic (Field, 2009). The value should be close to 2. All of the samples met this criterion; therefore, the assumption of independent errors is tenable.

The SPSS ANOVA provided the variance analysis to test whether the regression model was better than using the mean to predict the outcome. For all samples, the three models were highly significant. It is very unlikely for these values to have happened by chance. I found that use of the model provided significant improvement in my ability to predict the outcome variable, *learning achieved*, over using the mean as an estimate of *learning achieved*. These findings mean the null hypothesis that no relationship exists should be rejected (Field, 2009).

For brevity, the following provides analysis examples for specific course types. I provide the results for all course types in the Regression Summary Tables 3 and 4. The SPSS coefficients table (data included in Table 3) shows the model parameters for each step in the hierarchy. The first step in the hierarchy included *career benefit* and *worthwhile investment*. For ACQ SPW, SPSS results provide that B (Y intercept constant) is 1.678 and this can be interpreted as meaning that when no benefit to career or employer occurs (when X = 0), the model predicts very low *learning achieved* scores will result. The B values of .449 for *career benefit* and .271 for *worthwhile investment* represent the outcome change associated with a unit change in the predictor. If the predictor variable is increased by one on the Likert scale for *career benefit*, then the model predicts that *learning achieved* increases by 0.449 on the Likert scale following acquisition web-based training of acquisition professionals.

These results indicate that the regression model is useful, because it significantly improves the ability to predict learning from defense acquisition policy training. To make predictions for ACQ SPW, I would define the model as follows:

learning achieved = 0.945 + (0.308career) + (0.187worthwhile) + (0.079exercises) + (0.107examples) + (0.111delivery)

For comparison, the model for ACQ ILT would be defined as

learning achieved = -0.027 + (0.253career) + (0.044worthwhile) + (0.374exercises) + (0.109examples) + (0.100enthusiasm) + (0.047application)

This allows a prediction about *learning achieved* for online SPW and resident ILT acquisition courses to be made by replacing the predictors with values of interest.

For the ACQ ILT model, the *career benefit* (t(1818) = 10.692, p < .001), *worthwhile investment* (t(1818) = 2.212, p < .05), *exercises value* (t(1818) = 10.640, p < .001), *examples helped* (t(1818) = 2.831, p < .01), *instructor enthusiasm* (t(1818) = 2.380, p < .05), *application discussed* (t(1818) = 2.376, p < .05), and *instructor knowledge* (t(1818) = 1.253, not sig.) are all significant predictors of *learning achieved*, except for *instructor knowledge*. The magnitude of the t-statistics indicates that the *career benefit* and *exercises value* predictors had the greatest impact and that *instructor knowledge* had no significant impact on the *learning achieved* outcome. Although all course topic and delivery combination results provided that *career benefit* was the most important predictor, the other predictors varied greatly in their importance in predicting *learning achieved* in DAU classes across delivery types and functional topics.

For ACQ ILT, the standardized beta values for *career benefit* (Beta = .336) and *exercises value* (Beta = .293) are more than three times that of any other predictor and are, therefore, of much greater importance than any of the other variables in the model. Most of the ACQ ILT model predictors have relatively tight confidence intervals that do not cross zero; however, the *instructor knowledge* predictor confidence interval does cross zero, which supported the finding that this variable is not a significant predictor of *learning achieved* for the ACQ ILT model.

The coefficients tables for the samples showed no collinearity in the data. The VIF values for all samples were well less than 10 indicating no cause for concern. The average VIF values were not substantially greater than 1, so the regression is assumed to be unbiased. No tolerance values fell below 0.2. Based on these results, I concluded that there is not a collinearity problem within the data.

The data samples were also examined for extreme cases that have a standardized residual less than -2 or greater than 2 using the summary table of the residual statistics. When analyzing the 20 samples, I expected 95% of the cases to have standard deviation residuals within about + or – 2. The cases that had standardized residuals greater than 3 were large enough to warrant further investigation. SPSS residuals statistics and case summaries provide that none of the cases had a Cook’s distance greater than 1 (the worst case was .097); therefore, none of the cases had an undue influence on the model. The Mahalanobis distance values of greater than 25 also supported the conclusion that these cases may be problematic and further investigation was warranted.

For each of the outlier cases, I analyzed the survey scores for the outcome and significant predictors (*learning achieved, career benefit, worthwhile investment, exercises value,* and *instructor knowledge*) and the response to the variables “what percent of your total work time requires the knowledge or skills presented in this training?” and “the participant materials (manual, presentation handouts) will be useful on the job.” The additional variables associated with some of these cases indicated that the students’ work required 0% of the training provided and they strongly disagreed that the material was useful on the job. For the other cases, the additional variables indicated that the student’s work required only 10% of the training provided and the student strongly disagreed that the material was useful on the job. It is likely that learning did not occur because the training was not useful in the student’s current job, which aligns relatively well with the regression model that has training value as a primary predictor of learning. The cases examined are likely a problem with “having the wrong butts in seats,” or students for whom the defense acquisition policies taught do not apply in their workplace. The model appears to be reliable without undue influence by outlier cases.

Histograms, standardized residuals (*ZRESID) against standardized predicted values (*ZPRED) plots, and normal probability plots of the residuals were also analyzed to check that all assumptions have been met. All sample scatterplots showed a relatively even dispersion with no funneling or curvature, so the assumptions of linearity and homoscedasticity were likely met. The sample histograms showed relatively normal distributions or bell curves and deviations from normality were not seen in the normal probability plot for any of the samples, which indicated that the normality of residuals assumption has likely been met. Partial plots were analyzed to confirm homoscedasticity and linear relationships. For all samples, there are few obvious outliers on the plots and the dots appear to be relatively evenly spaced around a gradient line, which is an indicator of homoscedasticity.

I provide the key results from the regression analysis of the predictors of *learning achieved* in the Regression Summary Table 3. The findings from my analysis of the data indicated that the model appears to be accurate for the samples tested and generalizable to the defense acquisition workforce.

Table 3. Regression Summary—Predictors of Learning Achieved

Model		ACQ						BCF					
		ILT (*N*=1826)			SPW (*N*=1532)			ILT (*N*=1474)			SPW (*N*=1366)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
1	(Constant)	3.394	.083		1.678	.089		2.526	.093		1.456	.106	
	I will benefit from what I learned in the course for my career/professional development.	.377	.024	**.502***	.449	.027	**.477***	.327	.025	**.392***	.516	.033	**.531***
	This training was a worthwhile investment for my employer.	.115	.022	**.169***	.271	.024	**.324***	.284	.023	**.371***	.212	.031	**.233***
		***R*-square = .422**			***R*-square = .588**			***R*-square = .522**			***R*-square = .545**		
2	(Constant)	.812	.150		1.044	.104		.916	.136		.842	.125	
	I will benefit from what I learned in the course for my career/professional development.	.269	.022	**.358***	.318	.029	**.338***	.241	.023	**.289***	.417	.034	**.428***
	This training was a worthwhile investment for my employer.	.055	.020	**.082****	.218	.024	**.260***	.184	.022	**.241***	.152	.031	**.168***
	The exercises added value to my learning.	.380	.035	**.298***	.128	.034	**.125***	.318	.034	**.275***	.200	.040	**.179***

Model		B	Std. Error	Beta	B	Std. Error	Beta	B	Std. Error	Beta	B	Std. Error	Beta
	The examples presented helped me understand the content.	.166	.037	.120*	.158	.035	.146*	.105	.035	.085**	Not Sig.		
		***R*-square change = .111**			***R*-square change = .030**			***R*-square change = .072**			***R*-square change = .025**		
3	(Constant)	-.027	.249		.945	.105		.263	.269		.727	.129	
	I will benefit from what I learned in the course for my career/professional development.	.253	.024	.336*	.308	.029	.328*	.231	.026	.276*	.411	.034	.423*
	This training was a worthwhile investment for my employer.	.044	.020	.065***	.187	.024	.224*	.176	.022	.230*	.139	.031	.153*
	The exercises added value to my learning.	.374	.035	.293*	.079	.035	.078***	.312	.034	.269*	.161	.042	.144*
	The examples presented helped me understand the content.	.109	.039	.079**	.107	.036	.099**	Not Sig.			Not Sig.		
	The instructor's energy and enthusiasm kept the participants actively engaged.	.100	.042	.057***				.089	.033	.058**			
	On-the-job application of each class objective was discussed during the course.	.047	.020	.055***				Not Sig.					
	The instructor was knowledgeable about the subject.	Not Sig.						Not Sig.					
	This delivery method was an effective way for me to learn the material.				.111	.023	.123*				Not Sig.		
	The graphics and illustrations used were meaningful and within context.				Not Sig.						.106	.039	.092**
		***R*-square change = .007**			***R*-square change = .010**			***R*-square change = .004**			***R*-square change = .004**		
		*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)		

Note. Dependent Variable: I learned new knowledge and skills.
DAU Postevent Surveys

Model		CM						CON					
		ILT (*N*=1668)			SPW (*N*=1462)			ILT (*N*=2000)			SPW (*N*=1588)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
1	(Constant)	2.059	.107		.734	.130		2.040	.081		.971	.101	
	I will benefit from what I learned in the course for my career/professional development.	.405	.027	.423*	.517	.034	.459*	.412	.021	.460*	.545	.026	.536*
	This training was a worthwhile investment for my employer.	.284	.024	.338*	.320	.031	.304*	.286	.018	.362*	.264	.024	.275*
		***R*-square = .520**			***R*-square = .527**			***R*-square = .609***			***R*-square = .593**		
2	(Constant)	1.465	.114		.474	.162		1.455	.096		.660	.109	
	I will benefit from what I learned in the course for my career/professional development.	.307	.027	.320*	.477	.035	.424*	.320	.021	.357*	.436	.030	.429*
	This training was a worthwhile investment for my employer.	.210	.023	.249*	.296	.032	.281*	.232	.018	.294*	.230	.024	.240*
	The exercises added value to my learning.	.084	.017	.116*	.145	.044	.114**	.180	.023	.193*	.158	.039	.149*

	The examples presented helped me understand the content.	.181	.024	**.186***		**Not Sig.**		.055	.026	**.052*****		**Not Sig.**	
		***R*-square change = .046**			***R*-square change = .005**			***R*-square change = .030**			***R*-square change = .014**		
	(Constant)	1.107	.159		.382	.168		.804	.151		.542	.110	
	I will benefit from what I learned in the course for my career/professional development.	.305	.030	**.318***	.470	.035	**.417***	.249	.023	**.279***	.414	.029	**.407***
	This training was a worthwhile investment for my employer.	.204	.023	**.243***	.288	.033	**.274***	.222	.018	**.281***	.200	.025	**.209***
	The exercises added value to my learning.	.083	.017	**.114***	.139	.044	**.110****	.171	.023	**.183***	.123	.039	**.117****
3	The examples presented helped me understand the content.	.164	.026	**.168***		**Not Sig.**		-.003	.026	**-.003**		**Not Sig.**	
	The instructor's energy and enthusiasm kept the participants actively engaged.		**Not Sig.**					.072	.021	**.064****			
	On-the-job application of each class objective was discussed during the course.		**Not Sig**					.112	.019	**.121***			
	The instructor was knowledgeable about the subject.	.084	.035	**.061*****				.059	.029	**.036*****			
	This delivery method was an effective way for me to learn the material.					**Not Sig.**					.128	.025	**.129***
	The graphics and illustrations used were meaningful and within context.					**Not Sig.**						**Not Sig.**	
		***R*-square change = .003** *($p < .001$) **($p < .01$) ***($p < .05$)			***R*-square change = .001** *($p < .001$) **($p < .01$) ***($p < .05$)			***R*-square change = .013** *($p < .001$) **($p < .01$) ***($p < .05$)			***R*-square change = .009** *($p < .001$) **($p < .01$) ***($p < .05$)		

Note. Dependent Variable: I learned new knowledge and skills.
DAU Postevent Surveys

Model		ENG						LOG					
		ILT (*N*=1484)			SPW (*N*=1417)			ILT (*N*=1489)			SPW (*N*=1558)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
	(Constant)	2.052	.085		1.079	.097		2.788	.087		1.162	.074	
1	I will benefit from what I learned in the course for my career/professional development.	.462	.027	**.527***	.392	.029	**.389***	.391	.024	**.499***	.529	.027	**.563***
	This training was a worthwhile investment for my employer.	.207	.024	**.266***	.391	.027	**.427***	.186	.022	**.255***	.254	.025	**.296***
		***R*-square = .581**			***R*-square = .607**			***R*-square = .520**			***R*-square = .695**		
	(Constant)	.779	.106		.728	.114		.600	.137		.624	.091	
	I will benefit from what I learned in the course for my career/professional development.	.316	.025	**.361***	.322	.031	**.320***	.256	.023	**.327***	.402	.028	**.429***
2	This training was a worthwhile investment for my employer.	.130	.022	**.168***	.353	.027	**.386***	.138	.020	**.189***	.198	.024	**.231***
	The exercises added value to my learning.	.253	.028	**.252***	.102	.038	**.095****	.228	.034	**.194***	.256	.035	**.243***

	The examples presented helped me understand the content.	.163	.031	**.146***	.063	.038	**.055**	.279	.036	**.215***		**Not Sig.**	
		***R*-square change = .081**			***R*-square change = .010**			***R*-square change = .099**			***R*-square change = .027**		
	(Constant)	.075	.187		.636	.117		.107	.262		.443	.093	
	I will benefit from what I learned in the course for my career/professional development.	.279	.026	**.318***	.321	.031	**.318***	.220	.025	**.280***	.387	.027	**.413***
	This training was a worthwhile investment for my employer.	.117	.022	**.151***	.334	.027	**.365***	.130	.020	**.179***	.177	.024	**.205***
	The exercises added value to my learning.	.236	.028	**.235***		**Not Sig.**		.206	.034	**.176***	.180	.036	**.171***
3	The examples presented helped me understand the content.	.121	.031	**.108***		**Not Sig.**		.263	.037	**.203***		**Not Sig.**	
	The instructor's energy and enthusiasm kept the participants actively engaged.		**Not Sig.**						**Not Sig.**				
	On-the-job application of each class objective was discussed during the course.	.097	.023	**.093***				.081	.021	**.092***			
	The instructor was knowledgeable about the subject.		**Not Sig.**					.152	.061	**.064*****			
	This delivery method was an effective way for me to learn the material.				.124	.027	**.123***				.148	.024	**.143***
	The graphics and illustrations used were meaningful and within context.					**Not Sig.**						**Not Sig.**	
		R-square change = .008			**R-square change = .007**			**R-square change = .006**			**R-square change = .010**		
		*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)			*(*p* < .001) **(*p* < .01) ***(*p* < .05)		

Note. Dependent Variable: I learned new knowledge and skills.
DAU Postevent Surveys

Model		PMT						PQM					
		ILT (*N*=1847)			SPW (*N*=1377)			ILT (*N*=1832)			SPW (*N*=1401)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
	(Constant)	2.058	.098		1.127	.112		2.769	.092		1.148	.109	
1	I will benefit from what I learned in the course for my career/professional development.	.456	.024	**.489***	.440	.033	**.429***	.379	.024	**.451***	.536	.031	**.526***
	This training was a worthwhile investment for my employer.	.226	.022	**.273***	.316	.031	**.335***	.201	.022	**.259***	.229	.028	**.247***
		***R*-square = .523**			***R*-square = .532**			***R*-square = .453**			***R*-square = .549**		
	(Constant)	1.177	.118		.838	.129		.267	.164		.582	.132	
	I will benefit from what I learned in the course for my career/professional development.	.365	.024	**.391***	.390	.034	**.381***	.253	.023	**.301***	.439	.033	**.431***
2	This training was a worthwhile investment for my employer.	.150	.021	**.180***	.263	.032	**.278***	.155	.020	**.200***	.168	.029	**.182***
	The exercises added value to my learning.	.186	.027	**.183***	.158	.037	**.153***	.213	.033	**.158***	.188	.051	**.161***

	The examples presented helped me understand the content.	.116	.030	**.103***		**Not Sig.**		.324	.035	**.217***		**Not Sig.**	
		***R*-square change = .042**			***R*-square change = .012**			***R*-square change = .081**			***R*-square change = .019**		
	(Constant)	.509	.163		.907	.133		-.660	.238		.483	.136	
	I will benefit from what I learned in the course for my career/professional development.	.325	.026	**.348***	.393	.034	**.384***	.222	.025	**.264***	.428	.033	**.420***
	This training was a worthwhile investment for my employer.	.138	.021	**.167***	.238	.032	**.252***	.150	.020	**.193***	.158	.029	**.171***
	The exercises added value to my learning.	.173	.026	**.170***	.158	.038	**.153***	.182	.033	**.135***	.154	.052	**.132****
3	The examples presented helped me understand the content.		**Not Sig.**			**Not Sig.**		.256	.037	**.171***		**Not Sig.**	
	The instructor's energy and enthusiasm kept the participants actively engaged.		**Not Sig.**						**Not Sig.**				
	On-the-job application of each class objective was discussed during the course.	.063	.023	**.066****				.057	.022	**.063*****			
	The instructor was knowledgeable about the subject.	.108	.035	**.071****				.129	.054	**.061*****			
	This delivery method was an effective way for me to learn the material.				.141	.027	**.149***				.084	.033	**.076*****
	The graphics and illustrations used were meaningful and within context.				-.118	.037	**-.104****					**Not Sig.**	
		***R*-square change = .010**			***R*-square change = .009**			***R*-square change = .009**			***R*-square change = .003**		
		*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)		

Note. Dependent Variable: I learned new knowledge and skills.
DAU Postevent Surveys

Model		STM						TST					
		ILT (*N*=878)			SPW (*N*=0)			ILT (*N*=1213)			SPW (*N*=1371)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
	(Constant)	1.595	.125					1.701	.103		1.062	.116	
1	I will benefit from what I learned in the course for my career/professional development.	.486	.034	**.505***				.539	.027	**.576***	.445	.032	**.418***
	This training was a worthwhile investment for my employer.	.247	.030	**.300***				.188	.023	**.240***	.349	.029	**.366***
		***R*-square = .585**			***R*-square =**			***R*-square = .608**			***R*-square = .553**		
	(Constant)	.781	.173					.465	.145		.529	.136	
	I will benefit from what I learned in the course for my career/professional development.	.398	.036	**.413***				.425	.028	**.455***	.364	.033	**.342***
2	This training was a worthwhile investment for my employer.	.209	.029	**.254***				.125	.022	**.160***	.290	.029	**.304***
	The exercises added value to my learning.	.171	.038	**.153***				.245	.034	**.210***	.139	.031	**.132***

	The examples presented helped me understand the content.		**Not Sig.**					.111	.037	**.087****	.088	.036	**.072*****
		***R*-square change = .024**			***R*-square change =**			***R*-square change = .044**			***R*-square change = .020**		
3	(Constant)	.117	.263					.141	.251		.462	.138	
	I will benefit from what I learned in the course for my career/professional development.	.375	.038	**.389***				.392	.030	**.420***	.356	.033	**.335***
	This training was a worthwhile investment for my employer.	.198	.029	**.240***				.119	.022	**.152***	.245	.030	**.257***
	The exercises added value to my learning.	.157	.039	**.141***				.232	.034	**.200***	.115	.031	**.109***
	The examples presented helped me understand the content.		**Not Sig.**					.087	.038	**.068*****		**Not Sig.**	
	The instructor's energy and enthusiasm kept the participants actively engaged.		**Not Sig.**						**Not Sig.**				
	On-the-job application of each class objective was discussed during the course.		**Not Sig.**					.073	.028	**.068****			
	The instructor was knowledgeable about the subject.	.122	.054	**.067*****					**Not Sig.**				
	This delivery method was an effective way for me to learn the material.										.124	.028	**.126***
	The graphics and illustrations used were meaningful and within context.											**Not Sig.**	
		***R*-square change = .006**			***R*-square change =**			***R*-square change = .003**			***R*-square change = .008**		
		*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)			*($p < .001$) **($p < .01$) ***($p < .05$)		

Note. Dependent Variable: I learned new knowledge and skills.
DAU Postevent Surveys

Analysis (Part 2)—Predictors of Applied Training

I conducted the analysis for Part 2 of the study in the same manner; however, the data used was from the follow-up surveys provided to students greater than 60 days after training. All regression assumptions were met. Descriptive statistics characterize the 20 samples used for the analysis of Stage 3 of the four-stage culture change model to determine important predictors of the *applied training* outcome for both the ILT and SPW DAU courses. The descriptive statistics include mean, standard deviation, and sample size. The means of the Likert score (7 = strongly agree and 1 = strongly disagree) and percentage score responses to the variables indicate how the students in each sample perceive the variable in question. The *applied training* outcome means for resident ILT courses range from a low of 5.23 for ENG to a high of 6.11 for CM. For online SPW courses, the means for *learning achieved* range from a low of 4.98 for LOG to a high of 5.81 for CM. A review of the means shows that resident ILT courses appear to be more effective in driving workplace application of behavior learned from training compared to online SPW courses.

The means of *learning achieved* from the follow-up survey responses align relatively well with the means of *learning achieved* from the postevent surveys. The follow-up survey means for resident ILT courses ranged from a low of 5.74 for ENG and STM to a high of

6.34 for PMT. For online SPW courses, the means for *learning achieved* from follow-up surveys ranged from a low of 5.53 for LOG to a high of 6.00 for CM. The results showed higher *learning achieved* scores for resident ILT courses than for the online SPW courses for all functional areas.

For the resident ILT courses, the *learning achieved* variable has the highest mean scores, and regression analysis results indicate that the *learning achieved* variable is the most important predictor of application of behavior learned in training. The *manager involvement* variable has the lowest mean scores, but *manager involvement* is the least important of the tested predictors of *applied training*. The online SPW courses showed similar results with the most important predictor of *applied training* having received the highest Likert scores.

The findings confirmed that out of the four predictors across all data subsets, the *learning achieved* variable correlates best with the outcome ($p < .001$), so this variable should best predict *applied training*. Learning accounts for 76% for ACQ ILT, so t*ask applicability, resources provided*, and *manager involvement* account for 7% of outcome variation. For PMT ILT, learning accounts for 29%, so *task applicability, resources provided,* and *manager involvement* account for 20% of the variation in *applied training*. PMT ILT is unique in providing 400 level courses, however, even with those courses removed, the results are nearly the same.

I summarized the findings of this analysis in the Regression Summary Table 4, which indicated that the model appears to be accurate for the samples and generalizable to the defense acquisition workforce. For all of the samples, *learning achieved* is the most important predictor of *applied training*; however, *task applicability* is also important in predicting the acquisition professional's ability to apply what was learned in acquisition policy training courses on the job. Functional topic and delivery method must be factored in when determining the importance of *resources provided* and *manager involvement* as additional predictors of *applied training*. The multiple regression assumptions appear to have been met, so this model should generalize to the acquisition workforce.

Table 4. Regression Summary—Predictors of Applied Training

Model		ACQ						BCF					
		ILT (*N*=1317)			SPW (*N*=1783)			ILT (*N*=646)			SPW (*N*=919)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
1	(Constant)	-.076	.093		.021	.116		.088	.279		.281	.163	
	I learned new knowledge and skills from this training.	.962	.015	**.871***	.898	.020	**.735***	.872	.045	**.606***	.861	.028	**.716***
		***R*-squared = .759**			***R*-squared = .541**			***R*-squared = .367**			***R*-squared = .513**		
2	(Constant)	-.385	.083		-.022	.104		-.247	.240		.099	.153	
	I learned new knowledge and skills from this training.	.801	.015	**.725***	.657	.020	**.538***	.580	.041	**.403***	.640	.028	**.532***
	What percent of your total work time have you spent on tasks that require the knowledge/skills presented in the training?	.007	.001	**.146***	.014	.001	**.242***	.016	.002	**.280***	.015	.001	**.247***

	I was provided adequate resources (time, money, equipment) to successfully apply this training on my job.	.142	.012	**.167***	.051	.016	**.056****	.207	.031	**.220***	.063	.024	**.064****
	After training, my manager and I discussed how I will use the learning on my job.	.043	.008	**.071***	.155	.013	**.201***	.087	.024	**.119***	.138	.019	**.183***
		***R*-squared Change = .067** *(*p* < .001)			***R*-squared Change = .122** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .202** *(*p* < .001)			***R*-squared Change = .121** *(*p* < .001) **(*p* < .01)		

Note. Dependent Variable: I have been able to successfully apply the knowledge/skills learned in this class to my job.
DAU Follow-Up Surveys

Model		CM						CON					
		ILT (*N*=416)			SPW (*N*=297)			ILT (*N*=1624)			SPW (*N*=1894)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
1	(Constant)	.514	.253		.934	.265		.350	.166		.471	.100	
	I learned new knowledge and skills from this training.	.884	.040	**.740***	.813	.043	**.737***	.871	.026	**.638***	.871	.017	**.770***
		***R*-squared = .547**			***R*-squared = .543**			***R*-squared = .638**			***R*-squared = .593**		
2	(Constant)	.162	.231		.840	.253		.225	.139		.319	.093	
	I learned new knowledge and skills from this training.	.672	.040	**.562***	.622	.054	**.564***	.569	.024	**.417***	.678	.018	**.600***
	What percent of your total work time have you spent on tasks that require the knowledge/skills presented in the training?	.009	.001	**.222***	.009	.002	**.208***	.013	.001	**.284***	.010	.001	**.205***
	I was provided adequate resources (time, money, equipment) to successfully apply this training on my job.	.130	.035	**.149***			**Not Sig.**	.207	.031	**.213***	.019	.249	**.077***
	After training, my manager and I discussed how I will use the learning on my job.	.069	.025	**.105****	.139	.029	**.203***	.045	.013	**.067****	.089	.011	**.130***
		***R*-squared Change = .103** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .090** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .198** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .078** *(*p* < .001) **(*p* < .01)		

Note. Dependent Variable: I have been able to successfully apply the knowledge/skills learned in this class to my job.
DAU Follow-Up Surveys

Model		ENG						LOG					
		ILT (*N*=726)			SPW (*N*=2148)			ILT (*N*=1196)			SPW (*N*=2033)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
1	(Constant)	.289	.178		.316	.091		-.087	.179		-.127	.097	

	I learned new knowledge and skills from this training.	.861	.030	**.727***	.864	.016	**.764***	.909	.029	**.674***	.923	.017	**.768***
		***R*-squared = .528**			***R*-squared = .584**			***R*-squared = .455**			***R*-squared = .590**		
	(Constant)	.126	.169		.207	.086		-.122	.148		-.114	.088	
	I learned new knowledge and skills from this training.	.634	.030	**.535***	.631	.017	**.558***	.552	.027	**.409***	.655	.018	**.545***
	What percent of your total work time have you spent on tasks that require the knowledge/skills presented in the training?	.015	.001	**.278***	.013	.001	**.223***	.019	.001	**.327***	.015	.001	**.228***
2	I was provided adequate resources (time, money, equipment) to successfully apply this training on my job.	.104	.025	**.107***	.071	.014	**.073***	.185	.022	**.190***	.030	.015	**.030***
	After training, my manager and I discussed how I will use the learning on my job.	.077	.018	**.110***	.135	.011	**.181***	.095	.016	**.126***	.189	.013	**.228***
		***R*-squared Change = .120** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .101** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .197** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .119** *(*p* < .001) **(*p* < .01)		

Note. Dependent Variable: I have been able to successfully apply the knowledge/skills learned in this class to my job.
DAU Follow-Up Surveys

Model		PMT						PQM					
		ILT (*N*=338)			SPW (*N*=548)			ILT (*N*=476)			SPW (*N*=746)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
	(Constant)	1.528	.374		.450	.166		.240	.270		.245	.172	
1	I learned new knowledge and skills from this training.	.690	.058	**.541***	.875	.028	**.799***	.884	.043	**.683***	.877	.029	**.744***
		***R*-squared = .293**			***R*-squared = .638**			***R*-squared = .467**			***R*-squared = .554**		
	(Constant)	.681	.341		.102	.163		-.206	.222		.075	.164	
	I learned new knowledge and skills from this training.	.496	.055	**.388***	.714	.029	**.652***	.605	.038	**.468***	.675	.030	**.573***
2	What percent of your total work time have you spent on tasks that require the knowledge/skills presented in the training?	.014	.002	**.338***	.010	.001	**.204***	.013	.001	**.269***	.011	.001	**.201***

	I was provided adequate resources (time, money, equipment) to successfully apply this training on my job.	.180	.046	**.182***	.075	.025	**.080****	.244	.030	**.285***	.065	.026	**.066*****
	After training, my manager and I discussed how I will use the learning on my job.	.053	.025	**.092*****	.088	.018	**.136***		**NOT SIG.**		.126	.019	**.177***
		***R*-squared Change = .199** *(*p* < .001) ***(*p* < .05)			***R*-squared Change = .077** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .210** *(*p* < .001)			***R*-squared Change = .099** *(*p* < .001) ***(*p* < .05)		

Note. Dependent Variable: I have been able to successfully apply the knowledge/skills learned in this class to my job.
DAU Follow-Up Surveys

Model		STM						TST					
		ILT (*N*=182)			SPW (*N*=0)			ILT (*N*=212)			SPW (*N*=257)		
		B	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta	*B*	Std. Error	Beta
	(Constant)	.680	.347					.247	.413		.734	.324	
1	I learned new knowledge and skills from this training.	.811	.059	**.714***				.873	.068	**.661***	.788	.055	**.670***
		***R*-squared = .510**			**No SPW Classes**			***R*-squared = .437**			***R*-squared = .449**		
	(Constant)	.304	.326					.325	.368		.255	.312	
	I learned new knowledge and skills from this training.	.608	.059	**.536***				.530	.070	**.401***	.604	.055	**.514***
	What percent of your total work time have you spent on tasks that require the knowledge/skills presented in the training?	.011	.003	**.216***				.014	.002	**.281***	.015	.002	**.288***
2	I was provided adequate resources (time, money, equipment) to successfully apply this training on my job.	.166	.047	**.192****				.148	.055	**.152****		**Not Sig.**	
	After training, my manager and I discussed how I will use the learning on my job.		**Not Sig.**					.129	.037	**.187****	.116	.033	**.163****
		***R*-squared Change = .129** *(*p* < .001) **(*p* < .01)						***R*-squared Change = .166** *(*p* < .001) **(*p* < .01)			***R*-squared Change = .136** *(*p* < .001) **(*p* < .01)		

Note. Dependent Variable: I have been able to successfully apply the knowledge/skills learned in this class to my job.
DAU Follow-Up Surveys

Interpretation of Findings

This study found DAU training to be a key contributor to implementing defense acquisition change by driving policy-compliant behavior change in the defense acquisition workforce. I interpreted the findings from the two-part study in the context of the behavior-before-belief model of culture change (Table 1). The first stage of the behavior-before-belief model is unfreezing the organization by creating the motivation to change. The literature strongly supports that defense acquisition problems, fiscal crises, and complex, rapid environmental changes are driving the need for culture change in the defense acquisition workforce (Edison & Murphy, 2012; Eide & Allen, 2012; Gates, 2010; Kratz & Buckingham, 2010; Lee, 2013; Masciulli, 2011). For DoD leadership and personnel, my review of the literature has shown that the motivation to change exists, which should unfreeze the status quo and prepare the organization to start the process of developing new cultural assumptions (Schein, 2010). Strategic management efforts used by DoD and other organizations to drive change through policy change planning and implementation were also well supported in the literature (Boyne & Walker, 2010; Bryson, 2011; Burke, 2011; Linn, 2008; Poister, 2010; Wedel et al., 2005).

The second stage in the behavior-before-belief model is cognitive restructuring through learning new concepts, new meanings for old concepts, and new judgment standards. The organization and culture change literature supports the use of training to facilitate behavioral change (Bontis et al., 2011; Burke, 2011; Bryson, 2011; Eide & Allen, 2012; Knowles, 1980; Kotzian, 2010; Ng'ang'a & Otii, 2013; Nissen, 2012). The literature also strongly suggested that transformational, collaborative, active-learning strategies enhance learning and the likelihood of change success (Bass & Riggio, 2010; Beattie, Thornton, & Laden, 2013; Boyne & Walker, 2010; Burns, 2010; Eide & Allen, 2012; Kotzian, 2010; O'Neil, 2011; Rendon, Apte, & Apte, 2012; Stevens et al., 2010). All DAU training courses teach complex defense acquisition policies and best practices tailored to the functional topic. Part 1 of the study tested whether Stage 2, cognitive restructuring, occurred by students learning new concepts in DAU policy training courses and determined the predictors of learning. Part 2 of the study tested whether Stage 3 of the change model occurred by examining students' on-the-job application of new behaviors learned following DAU policy training courses. Part 2 of the study also tested for the predictors of the students' ability to apply new concepts learned in training after the students had returned to the workplace.

Summary of Key Findings

This study found that students learned new concepts in all DAU policy training courses and that the most important predictor of *learning achieved* is *career benefit*, meaning that how beneficial the training is to the acquisition professional's career drives learning of new concepts in all DAU course types. Whether the training was a worthwhile investment for the employer was also a significant predictor of learning. These findings support the Bontis et al. (2011) study that found the worthwhile investment construct, which combined benefit to the student's career and employer, to be the most significant predictor of individual learning for DAU courses. This means that important factors in students' learning the defense acquisition policy and best practices taught in DAU courses are how worthwhile the training is to their career and employer.

The study also found that for resident courses, the *exercises value* variable was a highly significant predictor of *learning achieved*. The *exercises value* variable is a measure of the learning value of collaborative, scenario-based, team exercises that provide students with hands-on experience in applying acquisition policy to real-world problems. This predictor was less important for online SPW courses, likely due to the absence of collaborative teaming experience in addressing scenario-based problems presented in training. The study found that conditional relationships exist between the predictor variables *examples helped*, *instructor enthusiasm*, *application discussed*, *instructor knowledge*, *delivery effective*, and *graphics meaningful* and the outcome, *learning achieved*, dependent on course type.

The study also confirmed that application of concepts learned from DAU training occurs in the defense acquisition workplace and that the most important predictor of this application of learning was the *learning achieved* variable, which measured whether the student learned new knowledge and skills from the DAU training. This variable was found to be a highly significant predictor and the most important predictor of the *applied training* outcome for all DAU courses, accounting for greater than 50% of the variability in the *applied training* outcome for most courses. Increasing learning achieved in DAU training increases application of the policy-compliant behavior learned in the defense acquisition workplace. These findings support acceptance of the research hypothesis that there is a highly significant positive correlation between *learning achieved* and *applied training* and that *applied training* can be predicted from *learning achieved* for all DAU training courses. Application of learned concepts from DAU policy training was the policy-compliant behavior change tested; therefore, this study found that the DAU training does enhance policy-compliant behavior of the DoD acquisition workforce personnel.

Another highly significant predictor of *applied training* for all DAU courses was the *task applicability* variable, which measured the percentage of total work time spent on tasks that required the knowledge/skills presented in the training. This finding indicates that to increase application of training on the job, the DoD needs to ensure that the personnel who can use the training on the job are the personnel who are given the training. This variable also supports the worthwhile construct and adds further support to the importance of “having the right butts in seats” in DAU courses to increase policy-compliant behavior in the defense acquisition workplace. Conditional relationships exist between *resources provided* and *manager involvement* and the outcome, *applied training*, dependent on the type of course.

Conclusion

In the DoD, transformative change is implemented across the acquisition workforce in part by DAU training to enhance understanding of acquisition policy and best practices and to facilitate policy-compliant behavior in the defense acquisition military and civilian workforce. The findings from Part 1 of the study (Table 3) confirmed that the second stage in the behavior-before-belief model for culture change (Table 1) took place in DAU training. These findings showed that cognitive restructuring through learning new concepts, new meanings for old concepts, and new judgment standards occurred during DAU scenario-based training of cross-functional teams. These findings further confirm the knowledge found in the literature that suggests that transformational, collaborative, active-learning strategies enhance learning.

The findings from Part 2 (Table 4) of the study confirmed that the third stage in the behavior-before-belief model for culture change (Table 1) took place following DAU training. These findings showed that students applied the new behaviors learned following DAU training courses and determined important predictors of the students’ ability to apply these new concepts after the students had returned to the workplace. The findings from this study confirm that use of training facilitates behavioral change and that transformational, collaborative strategies enhance the likelihood of change success. Learning achieved in policy courses predicted application on-the-job of behaviors learned. If the new behaviors correct problems and produce better outcomes, then culture change as described in Stage 4 of the behavior-before-belief model (Table 1) should occur (Schein, 2010).

The study results extend knowledge by providing a better understanding of policy change implementation in the DoD, using DAU training to facilitate policy-compliant behavior change that should lead to needed culture change. For each DAU course type, the findings provide key drivers of learning and behavior change following DAU courses. Further confirming knowledge found in the literature, the results indicate that once the value to the student and employer is established, the greatest learning and behavior change occurs following resident courses that provide collaborative teaming experiences not found in online courses. These findings confirm that transformative, collaborative training techniques provided in a psychologically safe training environment facilitate behavioral change required to enhance the likelihood of successful implementation of complex policy changes, as suggested by the literature (Bass & Riggio, 2010; Boyne & Walker, 2010; Hackman, 2010; Kotzian, 2010; Masciulli, 2011; Schein, 2010; van Eeden, Cilliers, & van Deventer, 2008).

The literature provides that environmental change has been accelerated by globalization and technology, requiring transformative culture change to adapt. Changes in culture, or tacit assumptions, of mature organizations like the DoD cannot, in all likelihood, be successfully implemented and institutionalized directly; however, behavior can be changed by leaders to drive culture change (Burke, 2011; Schein, 2010). DAU training is required for all defense acquisition workforce personnel, so behavior change across the workforce should facilitate Stage 4 of the behavior-before-belief model, which is refreezing, or internalizing the new concepts, meanings, and standards in the defense acquisition workforce. This means that the DoD's efforts to implement complex defense acquisition policy changes should be successful using DAU training to address the complexity of the acquisition processes involved, the hyper-turbulent environment, and the change-resistant culture of the DoD acquisition workforce. This study established that a positive relationship exists between training and policy-compliant behavior; therefore, training is likely an effective contributor to policy change implementation in the DoD's defense acquisition workforce.

References

Bass, B., & Riggio, R. (2010). Leadership. In G. Hickman (Ed.), Leading organizations: *Perspectives for a new era* (2nd ed.; pp. 76–86). Thousand Oaks, CA: SAGE.

Beattie, J., Thornton, B., Laden, R., & Brackett, D. (2013). 21st century challenges in higher education: Strategic changes and unintended consequences. NCPEA International *Journal of Educational Leadership Preparation, 8*(1), 62–71.

Bollen, K. (1989). Structural equations with latent variables. New York, NY: Wiley.

Bontis, N., Hardy, C., & Mattox, J. (2011). Diagnosing key drivers of job impact and business results attributable to training at the Defense Acquisition University. *Defense Acquisition Research Journal, 18*(4), 348–367.

Boyne, G. A., & Walker, R. M. (2010). Strategic management and public service performance: The way ahead. *Public Administration Review, 70*, 185–192. doi:10.1111/j.1540-6210.2010.02271.x

Brown, T. (2010). The evolution of public sector strategy. *Public Administration Review, 70,* s212–s214. doi:10.1111/j.1540-6210.2010.02275.x

Bryson, J. M. (2011). *Strategic planning for public and nonprofit organizations: A guide to strengthening and sustaining organizational achievement* (4th ed.). San Francisco, CA: Jossey-Bass.

Burke, W. (2011). *Organization change: Theory and practice* (3rd ed.). Thousand Oaks, CA: SAGE.

Burns, J. (2010). Leadership. In G. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (2nd ed.; pp. 66–75). Thousand Oaks, CA: SAGE.

Clawson, J. (2012). *Level three leadership: Getting below the surface* (5th ed.). Upper Saddle River, NJ: Pearson/Prentice Hall.

Defense Acquisition University. (2011). *Delivering best value: 2011 annual report*. Retrieved from http://www.dau.mil

DoD. (2015). *Operation of the defense acquisition system* (DoDI 5000.02). Retrieved from http://www.acq.osd.mil

Edison, T., & Murphy, A. (2012). A new look at enablers and barriers to performance based life cycle product support (PBL) implementation. *Defense Acquisition Research Journal, 19*(4), 376–393.

Eide, P., & Allen, C. (2012). The more things change, acquisition reform remains the same. *Acquisition Research Journal, 19*(1), 99–120. doi:10.1145/1961189.1961194

Field, A. (2009). Discovering statistics using SPSS (3rd ed.). London, UK: SAGE.

Fishpaw, E. (2010). *Defense Acquisition Workforce Improvement Act (DAWIA) annotated—Current law* (Vers. A-3.2). Retrieved from http://www.library.dau.mil

Frankfort-Nachmias, C., & Nachmias, D. (2008). *Research methods in the social sciences* (7th ed.). New York, NY: Worth.

GAO. (2012). *Defense acquisitions: Assessments of selected weapon programs* (GAO-12-400SP). Retrieved from http://www.gao.gov

GAO. (2017). *Defense acquisitions: Assessments of selected weapon programs* (GAO-17-333SP). Retrieved from http://www.gao.gov

Gates, R. M. (2010). *Defense spending* [Speech]. Retrieved from http://www.defense.gov

Green, S., & Salkind, N. (2011). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hackman, J. (2010). Leading teams. In G. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (2nd ed.; pp. 209–238). Thousand Oaks, CA: SAGE.

Hannay, M. (2009). The cross-cultural leader: The application of servant leadership theory in the international context. *Journal of International Business & Cultural Studies, 1*, 1–12.

Harris, L., & Ogbonna, E. (2011). Antecedents and consequences of management-espoused organizational cultural control. *Journal of Business Research, 64*, 437–445. doi:10.1016/j.jbusres.2010.03.002

Knowles, M. (1980). *The modern practice of adult education: From pedagogy to andragogy.* New York, NY: Cambridge.

Kotzian, M. (2010, April). Acquisition leadership: An opportunity lost for acquisition excellence? *Defense Acquisition Review Journal*, 158–181.

Kratz, L., & Buckingham, B. (2010). Achieving outcomes-based life cycle management. *Defense Acquisition Review Journal, 53*, 45–66.

Lee, J. (2013). Environmental legislative standstill and bureaucratic politics in the USA. *Policy Studies*. doi:10.1080/01442872.2013.804176

Linn, M. (2008). Organizational culture: An important factor to consider. *The Bottom Line, 21*(3), 88–93. doi:10.1108/08880450810912844

Masciulli, J. (2011). Global public leadership in a technological era. *Bulletin of Science, Technology & Society, 31*(2), 71–80. doi:10.1177/0270467611402812

Ng'ang'a, S. I., & Otii, L. O. (2013). Constructivism and the Likert scale on the perception of teaching/learning creativity at the university level. *Journal of Sociological Research, 4*(1), 19–48. doi:10.5296/jsr.v4i1.3159

Nissen, M. (2012). *Toward an operational proxy for acquisition workforce quality: Measuring dynamic knowledge and performance at the tactical edges of organizations*. Retrieved from http://calhoun.nps.edu

O'Neil, W. (2011). Cost growth in major defense acquisition: Is there a problem? Is there a solution? *Defense Acquisition Review Journal, 59,* 277–294.

Poister, T. H. (2010). The future of strategic planning in the public sector: Linking strategic management and performance. *Public Administration Review*, 246–254. doi:10.1111/j.1540-6210.2010.02284.x

Rendon, R., Apte, U., & Apte, A. (2012). Services acquisition in the DoD: A comparison of management practices in the Army, Navy, and Air Force. *Acquisition Research Journal, 19*(1), 3–32.

Schein, E. H. (2010). *Organizational culture and leadership* (4th ed.). San Francisco, CA: Jossey-Bass.

Stevens, F., Plaut, V., & Sanchez-Burks, J. (2010). Unlocking the benefits of diversity: All-inclusive multiculturalism and positive organizational change. In G. Hickman (Ed.), *Leading organizations: Perspectives for a new era* (2nd ed.; pp. 101–121). Thousand Oaks, CA: SAGE.

Under Secretary of Defense (AT&L). (2015). *Implementation directive for Better Buying Power 3.0—Achieving dominant capabilities through technical excellence and innovation*. Retrieved from http://www.acq.osd.mil

van Eeden, R., Cilliers, F., & van Deventer, V. (2008). Leadership styles and associated personality traits: Support for the conceptualization of transactional and transformational leadership. *South African Journal of Psychology, 38*(2), 253–267. doi:10.1177/008124630803800201

Wedel, J., Shore, C., Feldman, G., & & Lathrop, S. (2005). Toward an anthropology of public policy. *The ANNALS of the American Academy of Political and Social Science, 600*(1), 30–51. doi:lO.1177/0002716205276734

Weick, K., & Sutcliffe, K. (2007). *Managing the unexpected: Resilient performance in an age of uncertainty* (2nd ed.). San Francisco, CA: Jossey-Bass.

Weier, M. H. (2008). Collaboration and the new product imperative. *InformationWeek, 1195*, 26–32.

Panel 7. Containing Cost in Major Defense Acquisition Programs

Wednesday, May 9, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Wendy Kunc,** Director, Naval Center for Cost Analysis, OASN (FM&C) ***Quantifying Annual Affordability Risk of Major Defense Programs*** David M. Tate, Institute for Defense Analyses Thomas J. Coonce, Institute for Defense Analyses Michael R. Guggisberg, Institute for Defense Analyses ***Have Changes in Acquisition Policy Influenced Cost Growth of Major Defense Acquisition Programs?*** David McNicol, Institute for Defense Analyses ***GAO Observations on Defense Acquisition Performance*** Christopher Durbin, U.S. Government Accountability Office J. Andrew Walker, U.S. Government Accountability Office

Wendy Kunc—is the Director of the Naval Center for Cost Analysis (NCCA). In this capacity, she advises DoN leadership on cost issues, develops defensible independent cost estimates and assessments for major acquisition programs, provides cost analysis tools, and performs special studies. Kunc chairs the DoN Cost Review Board and Cost Estimating Stakeholders Group. Previously, Kunc led NCCA's Cost Analysis Tools Division, managing the Naval Visibility and Management of Operating and Support Costs (VAMOSC) management information system and the Operating and Support Cost Analysis Model suite. Kunc also spent 15 years with the Department of the Air Force. In 2000, she led the Cost Factors Branch within the Air Force Cost Analysis Agency's Forces Analysis Division, developing the multibillion-dollar Cost per Flying Hour program. In 1993, Kunc managed Air Force VAMOSC and led the expansion into the more comprehensive Air Force Total Ownership Cost (AFTOC) management information system.

Kunc served as an operations research analyst at what is now the Air Force ISR Agency in San Antonio. Various positions included Chief of Software Support, where she developed Air Force Cryptologic Support Center software applications.

Kunc began her government career as a cartographer with the Defense Mapping Agency. Kunc holds a bachelor's degree in mathematics from the University of Missouri and a Master of Science in Computer Information Systems from St. Mary's University, San Antonio, TX. She received a Master of Science degree in National Resource Strategy from the National Defense University and completed the Industrial College of the Armed Forces Senior Acquisition Course. She completed the National Defense University's CIO certification program in 2005.

Kunc received the Department of the Navy Superior Civilian Service award in 2009 and the Air Force Headquarters civilian award for Outstanding Contribution to Financial Management and Comptroller in 1998. She received the OSD Comptroller team award for Innovative use of Technology in Financial Management in 1999 and 2002. She is a Certified Defense Financial Manager, is Level III certified in the Defense Acquisition Workforce, and is a member of the Acquisition Corps.

Quantifying Annual Affordability Risk of Major Defense Programs

David M. Tate—joined the research staff of the Institute for Defense Analyses' (IDA) Cost Analysis and Research Division in 2000. Prior to that, he was an Assistant Professor of Industrial Engineering at the University of Pittsburgh and the Senior Operations Research Analyst (Telecom) for Decision-Science Applications, Inc. At IDA, he has worked on a wide variety of resource analysis and quantitative modeling projects related to national security. These include an independent cost estimate of Future Combat Systems development costs, investigation of apparent inequities in Veterans' Disability Benefit adjudications, and modeling and optimization of resource-constrained acquisition portfolios. Tate holds bachelor's degrees in philosophy and mathematical sciences from the Johns Hopkins University, and MS and PhD degrees in Operations Research from Cornell University. [dtate@ida.org]

Thomas J. Coonce—is an Adjunct Research Staff Member with the Institute for Defense Analyses (IDA). He performs cost and schedule research for various clients, including the Office of the Secretary of Defense (OSD), the Department of Homeland Security (DHS), and the Congressional Section 809 Panel. Prior to joining IDA, Coonce led NASA's cost analysis activities and served as a cost/schedule analyst for the OSD's CAIG, The MITRE Corporation, and the Government Accountability Office (GAO). [tcoonce@ida.org]

Michael R. Guggisberg—is a Research Staff Member in the Strategy, Forces, and Resources Division at the Institute for Defense Analyses. He earned his PhD in Economics at the University of California, Irvine, in 2017. Prior to that, he was a squad leader in the Marine Corps Reserve, with tours in Iraq and Afghanistan. [mguggisb@ida.org]

Abstract

To a first approximation, acquisition programs never spend what they originally said they would spend when they began. In fact, the error bars around an initial cost estimate are much larger than is generally understood, once program cancellations, restructurings, truncations, and block upgrades have been accounted for. Worse yet, all of this uncertainty arises in a context where programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it *when* you said you would, or problems ensue.

We have developed a methodology that uses historical program outcomes to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs and procurement costs and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows Resource Managers to estimate annual budget risk levels, required contingency amounts to achieve a target probability of staying within a given budget, and many other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Planning Is About Not Being Surprised

A Hypothetical New Program

Suppose you are the Resource Manager for a portfolio of acquisition programs. A new helicopter program in your portfolio—call it "H-99"—has just received milestone approval. The H-99 Acquisition Program Baseline (APB) for development looks like Figure 1.

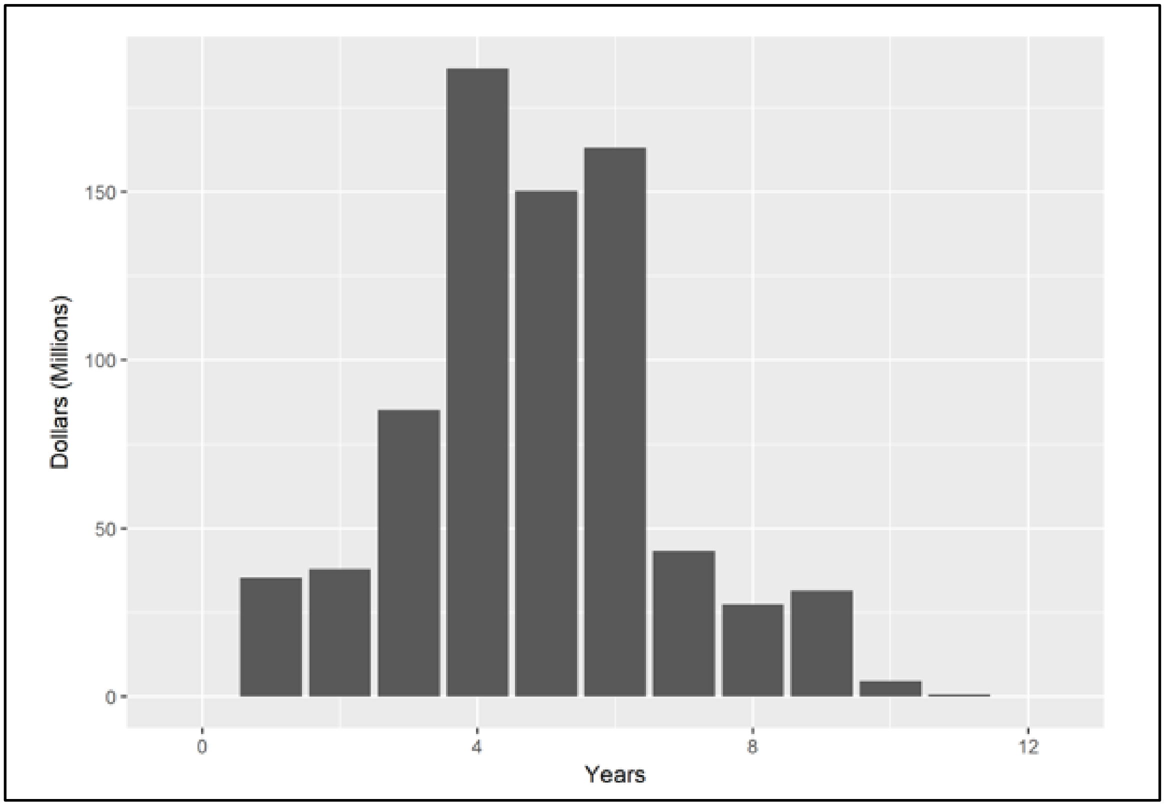


Figure 1. Proposed H-99 Development Costs

This is the official Service position about how many research, development, test, and evaluation (RDT&E) dollars this program will receive in each of the next dozen years—but it is not what will actually happen. What should you expect? Or, more precisely, what range of outcomes should you be prepared for, and how likely should you think those outcomes are? You don't want to be surprised.

Some Past Examples

Figure 2 shows the planned and actual development costs of the MH-60S Seahawk (a.k.a. "Knighthawk") helicopter program, as reported in the program's first Selected Acquisition Report (SAR) in 1998 and its final SAR in 2015. The original plan was to spend $71 million (constant 1998 dollars) over four years of RDT&E. The program actually spent more than $670 million over 19 years. The original requirement was for a ship-based cargo helicopter that could also support search-and-rescue and torpedo recovery. After the first 50 helicopters were built, the program evolved through several "block upgrades" to add airborne mine detection and countermeasures and shifted from an unarmed platform to an armed combat search and rescue (CSAR) and maritime interdiction platform—all within the same program of record.

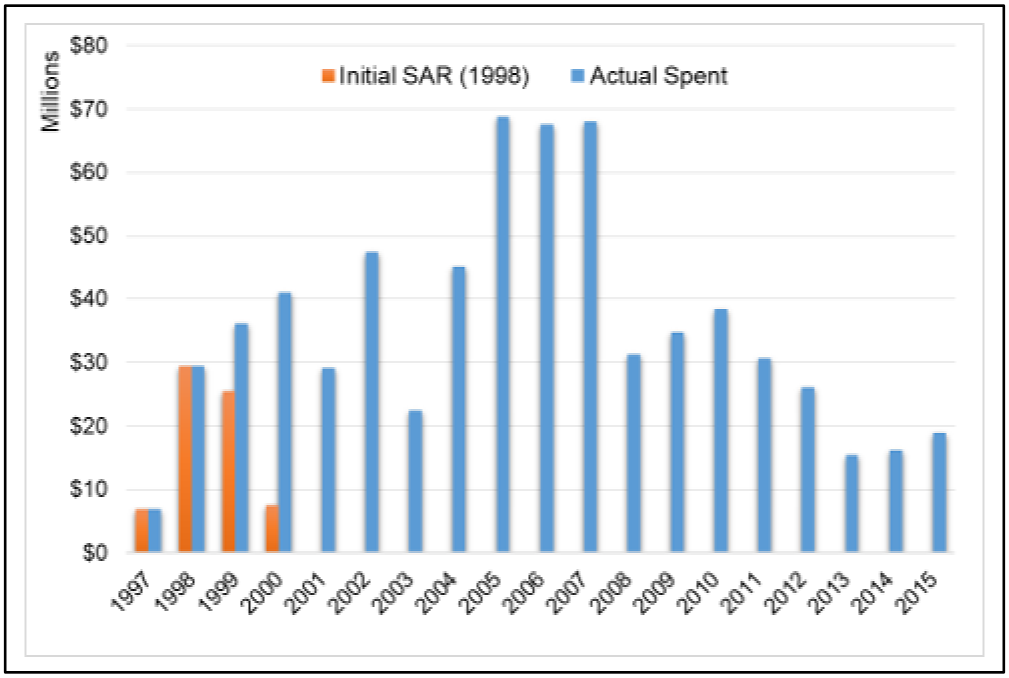


Figure 2. MH-60S Development Costs

Conversely, Figure 3 shows the planned versus actual procurement costs for the Armed Reconnaissance Helicopter (ARH) program. Here, the error is in the opposite direction; because the program was cancelled before procurement spending even began, none of the original estimated $3 billion (constant 2005 dollars) was ever spent.

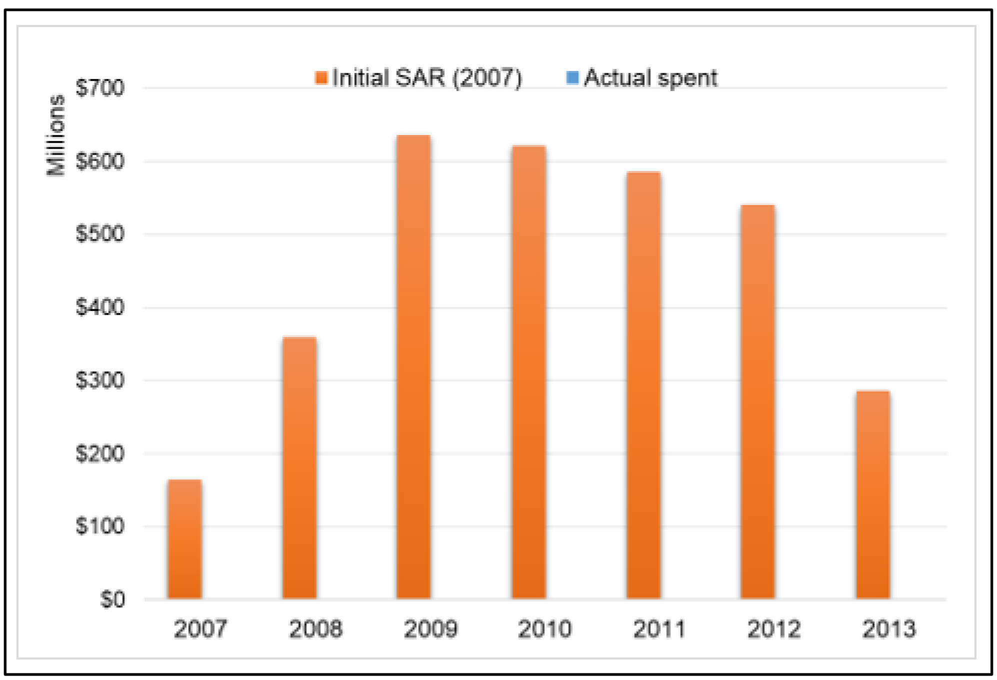


Figure 3. ARH Planned Procurement Costs (Zero Actually Spent)

Clearly, the dollars actually spent on development or procurement of a given program can vary wildly—either up or down—from what was originally planned. In order to manage acquisition portfolios sensibly, Resource Managers need to have some idea of just how much actual future resource usage might differ from its original estimate. We compiled original plans and actual outcomes from 115 historical major defense acquisition programs (MDAPs) in order to attempt to answer this question.

The Resource Manager's Challenge

We saw in the previous examples that what actually happens can be vastly different from what was predicted. Given the range of possible program outcomes, what should Resource Managers be prepared for? Are there tools that could help Resource Managers quantify the year-by-year affordability risk they face, based on actual historical outcomes and observable program characteristics, so that they can allocate contingency funds wisely?

Yes, there are tools—or at least there could be. That is the purpose of this research.

Affordability means having a high probability of being able to buy the thing you want with the funds you have designated for that purpose. We often think of this as a question of total cost or unit cost, but in practice it is a year-by-year question of whether there is enough money in this year's budget to cover this year's costs. The Resource Manager's goal is to balance the need for every program to have enough funds to execute efficiently this year against the demands of other programs and the overall budget. Planning sensibly for future years' demands is a major part of this. Programs that get stretched for lack of immediate funding in the budget year not only take longer to finish, but they also grow in total cost (due to inefficient use of labor and a higher proportion of fixed costs). On the other hand, providing enough contingency funding to make every program highly likely to stay within budget every year would be tremendously wasteful—especially since overfunded programs will tend to find something to spend that money on, rather than giving it back. Major differences between planned and actual demands for funding lead to budget instability, inefficient acquisition, and cancelled programs.

Existing Literature on Cost Growth Doesn't Help

What Question Are We Asking?

Most of the literature on cost growth is focused on trying to identify *causes* of cost growth in order to avoid them. That is a noble goal, but as a Resource Manager, you are probably aware that people have been trying to do that for decades, with limited success.

Similarly, most of the literature on cost growth looks at *unit* cost growth—either actual unit cost growth or unit cost growth adjusted to control for changes in the quantity procured (Asher & Maggelet, 1984; McNicol, 2017; McNicol et al., 2013). That makes sense if you're trying to understand the dynamics of why things cost more than expected. But as a Resource Manager, you don't care nearly as much about unit costs as you do about how many actual dollars will be needed.

In addition, nearly all cost growth studies describe cost growth (be it unit cost or total cost) in terms of a single program cost growth factor—the ratio of *final* total (or unit) cost to the originally estimated total (or unit) cost (Arena et al., 2006). To a Resource Manager, eventual total cost is not nearly as important as <u>annual</u> cost—that is, how many dollars will the program consume *this year*, and next year, and for each year in the Future Years Defense Plan (FYDP)?

Finally, nearly all past efforts are aimed at *predicting* costs, with those predictions expressed as an expected value or "50%" estimate (Anderson & Cherwonik, 1997; Bitten &

Hunt, 2017). This is in part because the literature has been driven by the needs of cost estimators, who are expected to produce point estimates that can be used for budgeting purposes. For cost estimators, the quality of the prediction is given by its *accuracy* (how much bias) and its *precision* (in terms of narrow error bars). Resource Managers prefer accurate (unbiased) forecasts, too, but do not really care about the mean value *per se.* Instead, they care greatly about being able to assess whether a program is likely to exceed yearly budgets and by how much, so that adjustments can be made before the program is forced to shed capabilities or be stretched out in order to "fit" within the budget profile. This is not a question of expected total cost or expected cost growth—it is a question of how the annual needs can vary and how to plan for those possibilities.

The authors are not aware of any past work that has addressed this need for annual resource requirement risk analyses. The Congressional Budget Office (CBO) uses commodity-specific cost growth factors when predicting future budget implications of current and predicted programs, but it does this in terms of average total program cost growth factors applied uniformly across all years of the planned program (CBO, 2017). As we shall see later, this is not an accurate estimate of *when* the cost growth in a given program would manifest itself. It is also just an average—it gives no information on the range of possible outcomes, and how relatively likely they are. This is the gap that our research aims to fill.

Desirable Outputs of a Model

Given a planned program (or set of programs—we'll get to that later) and a budget, Resource Managers would very much like to answer questions such as

- What is the distribution of funding the program will receive in year N = 1, 2, …?
- What is the probability that the program will receive more funding in year N than is currently budgeted, for N = 1, 2, …?
- How many total contingency dollars would be enough to achieve a given percent certainty that the current budget plus the contingency is enough to fund the program over the FYDP?
- What is the probability that the program will use at least $X less than planned over the FYDP, for various values of X?
- …etc.

The goal of our research is to develop empirical models based on historical program attributes, environments, and outcomes that will allow us to answer questions like these.

What Tools Do We Need?

In this section, we describe the tools that are needed to attempt to answer the kinds of questions that Resource Managers care about. These tools are

1. a way to describe funding profiles mathematically;
2. a list of program attributes and environmental factors that help predict program outcomes;
3. a statistical model to estimate the probability distribution of final funding profile shapes given the initial funding profile, environmental factors, and other program attributes;
4. a mathematical characterization of how well the shape tends to fit actual data; and
5. historical data on program initial plans and final outcomes.

A detailed discussion of each of these tools follows.

Step 1: Modeling Funding Profiles

It is an essential (and bothersome) fact that the various years of an acquisition program are not independent. If the actual RDT&E obligation authority in year 5 of development turns out to be higher than originally planned, it is very likely that year 6 and year 7 (etc.) will also be higher than originally planned. In fact, we are not interested only in the distribution of outcomes for year N; we are interested in the *joint* distribution of outcomes in all years—including years that were not part of the original baseline plan at all. That means we can't just build separate risk models for year 1, year 2, and so forth, and then use that collection of models to understand the risk over the entire planning horizon. We have to account for the ways cost *profiles* change over time.

Instead of treating the year-by-year outcomes as having some complicated joint distribution, we will instead use the techniques of *Functional Regression* to treat the individual year-by-year outcomes as having been generated by some (noisy) underlying functional form and then think about probability distributions over the parameters of those generating functions.

Development Costs

For the development portion of program funding, there is already a history of fitting functional models to the yearly funding requirements. In particular, Weibull distributions (or Rayleigh distributions, which are a special case of the Weibull) have often been used to describe both the shape a development program ought to have and the observed actual funding profiles of historical programs. Brown, White, Ritschel, and Seibel (2015) provide a good summary of past approaches.

Weibull distributions can be parametrized in several ways. We base our model of development cost profiles on this version, which has two parameters, λ and α:

$$W(t|\alpha,\lambda) = \frac{\alpha}{\lambda}\left(\frac{t}{\lambda}\right)^{\alpha-1} \exp\left(-\left(\frac{t}{\lambda}\right)^{\alpha}\right) 1(t \geq 0) \quad (1)$$

In this parametrization, λ is a time-scaling parameter that determines how much the profile changes from year to year, while α is a shape parameter. The term $1(t \geq 0)$ is an indicator function that equals 1 if t is greater than or equal to 0 and 0 else. Figure 4 shows the flexibility of the Weibull for various (λ, α) pairs. The black $(\lambda = 1, \alpha = 2)$ and red $(\lambda = 2, \alpha = 2)$ profiles show how the scaling parameter λ affects how quickly the spending profile changes. The red profile evolves more slowly than the black one because it has a larger λ value. The black $(\lambda = 1, \alpha = 2)$, blue $(\lambda = 1, \alpha = 1.5)$, and teal $(\lambda = 1, \alpha = 4)$ profiles show how the shape parameter α affects the peakedness. The teal profile is the most peaked, because it has the largest α value. Values of α less than 1 (e.g., the green curve) are not appropriate for modeling spending profiles, since $\alpha < 1$ implies that peak spending occurs in year 0.

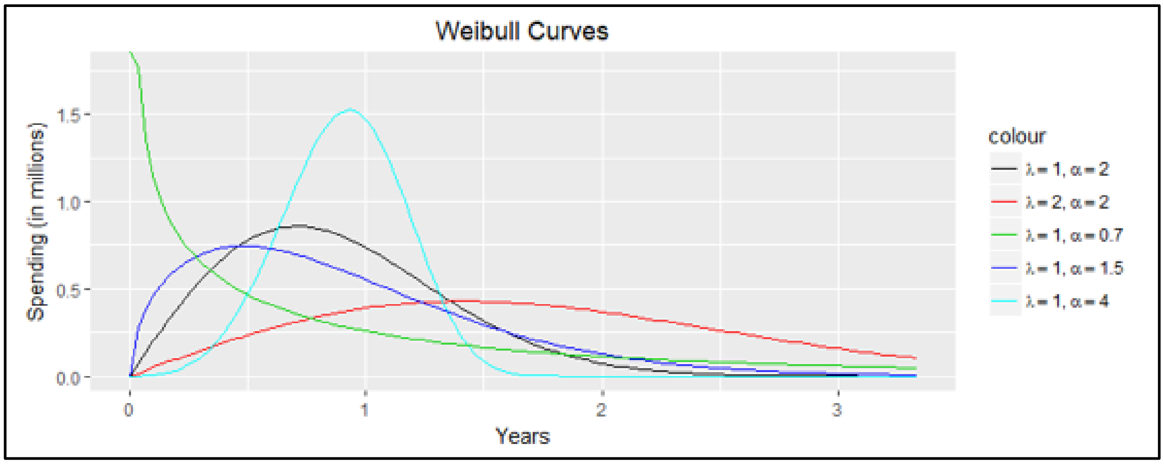


Figure 4. Weibull Distribution Shapes for Various Parameter Values

There are two major discrepancies between the Weibull distribution and the data we observe. First, the data we observe is annual, and therefore discrete, while the Weibull distribution is continuous. Second, some programs are cancelled during development, while the Weibull distribution assumes they are completed. To address these issues, we discretize and truncate the Weibull distribution using two additional parameters:

- the total development cost, denoted *C*
- the number of years of nonzero spending, denoted *T*

Let *C(t)* denote the costs in year t. The resulting functional form fit to the incurred costs in year t is

$$C(t) = K \cdot W(t|\alpha, \lambda) + \epsilon(t), t = 1, \ldots, T \quad (2)$$

where $\epsilon(t)$ is the independent random error in year *t* and the constant *K* is chosen such that $\sum_{t=1}^{T} C(t) = C$. This scaled, discretized truncation of the Weibull profile lets us accurately fit the final outcomes of programs cancelled in mid-development, as well as typical development profiles that taper off more slowly.

Procurement Costs

For procurement, there is no similarly traditional functional form to use as the basis of a functional regression. One possibility is to use a simple two-segment piecewise linear shape, consisting of an initial ramp-up, followed by a linear trend line. This distribution could be parametrized in various ways using four parameters. The choice of functional form here is purely descriptive—unlike the theory behind the application of Weibull distributions to development profiles, there is no underlying reason why procurement profiles should be roughly linear following their initial low-rate initial production (LRIP) ramp-up. For purposes of risk characterization, it is enough that historical programs do (approximately) follow this pattern.

For the remainder of this paper, we will ignore procurement costs and focus on modeling development cost profiles for simplicity of exposition.

Step 2: Identifying Potential Predictor Variables

Given choices for functional forms, the next challenge is to somehow characterize how the distribution of possible actual outcome profiles could be derived for a given initial plan. It seems obvious that different programs involve different levels of cost risk. There is a long literature attempting to identify specific factors that are correlated with program cost and schedule growth. Some factors that have been found by past researchers to be correlated with (unit) cost growth and/or total program cost growth risk include

- Commodity type (e.g., helicopter, satellite, MAIS, missile, or submarine) (Arena et al., 2006; Drezner et al., 1993; Tyson, Harmon, & Utech, 1994)
- Acquiring Service (Army, Navy, Air Force, Joint, Department of Energy [DoE]) (Drezner & Smith, 1990; Jessup & Williams, 2015; Light et al., 2017; McNicol, 2004)
- New design vs. modification of existing design (Arena et al., 2006; Coonce et al., 2010; Drezner et al., 1993; Jimenez et al., 2016; Marshall & Meckling, 1959)
- New build vs. remanufacture of existing units (Tyson et al., 1989)
- Budget climate at Milestone B (Asher & Maggelet, 1984; McNicol, 2017)
- Number of years of spending prior to Milestone B (Jimenez et al., 2016; Light et al., 2017)
- Schedule optimism (Arena et al., 2006; Asher & Maggelet, 1984; Glennan et al., 1993; Tate, 2016)
- Technology maturity of the program (Adoko, Mazzuchi, & Sarkani, 2015; GAO, 2006)
- Investment size (Bliss, 1991; Creedy, Skitmore, & Wong, 2010)

Because we are only trying to understand and characterize risk, on the assumption that the past is a reasonable guide to the future, we do not distinguish here between risks arising from discretionary choices, environmental factors, or intrinsic program features.

Step 3: Describing Changes in Cost and Schedule as Changes in Profile Functions

We saw previously that we can model development costs or production costs as being generated by an underlying functional form, with variation about that smooth curve treated as independent random noise. Mathematically,

$$C(t) = f(t|\theta) + \varepsilon(t) \qquad (3)$$

where *C(t)* is the cost in year *t, θ* is the vector of fitted parameter values for the family of curves being used, $f(\cdot)$ is the functional form we are assuming for the generating function, and $\varepsilon(t)$ is a random error whose distribution may depend on the year t. Let θ_0 denote the parameters that best fit the program's original profile and θ_1 denote the parameters that best fit the program's final profile. What we need to estimate is the conditional (joint) distribution of θ_1 given the appropriate program and environmental attributes and the fact that the program's original estimate was best fit by the curve $f(t|\theta_0)$.

There are several possible approaches to this and many choices of how to parametrize the family of curves being fit, but the general method will be the same in all cases. We estimate the distribution of θ_1 as a function of the best fit parameters θ_0 and the historical program characteristics *X*:

$$\log(\theta_1) = (X, \log(\theta_0))\beta + \eta, \qquad (4)$$

where X includes factors such as initial estimated cost, Service, budget climate, and so forth. The matrix element X_{jk} gives the value of predictor k for historical program j. The matrix θ_0 has one column for each parameter and one row for each historical program. $(X, \log(\theta_0))$ denotes the block matrix obtained by appending the componentwise natural logarithm of θ_0 as additional columns of X, one column per parameter.

This linear regression model implies a functional fit and distribution over the annual cost profile function $C(t)$. Rather than attempting to predict eventual actual cost as a function of initial estimated cost and other predictors, we instead attempt to predict the distribution of the parameters of a function that *generates* eventual cost, given program-specific attributes and the parameters that generate the initial estimate. Note that this is a multiple output regression—we are simultaneously estimating all of the best-fit parameters θ_1 and the covariance matrix that describes how those parameters are correlated.

We use a Bayesian estimation framework, starting with a weakly informative prior distribution $F_{prior}(\theta_1)$ and using Markov Chain Monte Carlo estimation to derive a posterior distribution $F_{posterior}(\theta_1)$, including the covariance matrix (Chib & Greenberg, 1995). We do this separately for development costs and procurement costs, using different families of profile-generating functions and treating their changes in shape and size as independent. Treating development and procurement jointly is a potential area for future research.

Step 4: Accounting for Noisy Curve Fits

The posterior distribution on θ_1 accounts for the uncertainty in the generating function parameters for the eventual profile of the program, but it does not capture the variability corresponding to the original error term $\varepsilon(t)$ when we fit truncated Weibull distributions to profiles. In order to capture all of the uncertainty in actual yearly costs, we need to also add in yearly random error terms $E(t)$. We derive the distribution of $E(t)$ from the observed $\varepsilon(t)$ in the best fits for actual costs of historical programs. In other words, we look at how much our curve fits to actual outcome profiles that tended to be off in each year, and we add corresponding random yearly error terms to any final profile generated from $F_{posterior}(\theta_1)$ in order to capture that additional source of uncertainty.

Step 5: Regression Data

The data for the regression are the initial estimate and final actual cost profiles for 155 completed historical MDAPs. The earliest program in the data set passed Milestone B in 1982. The data are taken from Selected Acquisition Reports (SARs), together with compiled attributes and environmental factors (as described previously in Step 2) for each program. In this paper, we will focus solely on characterization of development (RDT&E) cost risk. The methodology for procurement cost risk is similar, differing only in which predictor values are used and in the functional form of the generating function.

The specific predictor variables used in this paper are

- $\log(\alpha_0)$—natural logarithm of the shape parameter of the original estimate Weibull fit
- $\log(\lambda_0)$—natural log of the scale parameter of the original estimate Weibull fit
- $\log(C_0)$—natural log of the original total planned spending
- $\log(T_0)$—natural log of the original planned number non-zero spending years
- The Service overseeing the program (Navy, DoD, Air Force, Army, DoE)

- A commodity type (Air; Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR); Ground; Ordnance; Sea; Space; other)[1]
- A measure of relative Service budget tightness compared to two years ago
- A measure of relative Service budget tightness over the last 10 years
- A measure of budget optimism—planned spending divided by the mean historical actual spending for this commodity type
- A measure of schedule optimism—planned duration divided by the mean historical actual duration for this commodity type
- Whether the program is based on a modification of a preexisting design (binary)

The measures of relative budget tightness were based on the year the program passed Milestone II/B.

Regression Methodology

Let $i = 1,2, \ldots, I$ index over the historical programs in our data set; let the subscript $l = 0$ denote an original profile estimate and $l = 1$ denote an actual realized profile. We compiled original estimates and actual outcomes for $I = 115$ historical programs. For each historical program i, we fit scaled, discretized Weibull distributions[2] to the original estimated and actual development cost profiles $C_{i0}(t)$ and $C_{i1}(t)$:

$$C_{il}(t) = K_{il}\, W(t|\alpha_{il}, \lambda_{il}) + \epsilon_{il}(t), t = 1, \ldots, T \quad (5)$$

As before, the constants K_{il} are chosen so that $\sum_{t=1}^{T} C_{il}(t) = C_{il}$, the total cost of the original/final profile for program i.

[1] More precise commodity categories (e.g., distinguishing helicopters from fixed-wing aircraft) might be useful, given enough data. We found that increasing the sample size in each category led to better results than increasing the precision of the categories.

[2] For historical profiles, we set T = estimated years to reach Milestone C and C = total development cost through year T. We added fictitious years of zero spending in year 0 and year T + 1, then fit a Weibull distribution to the scaled yearly costs. For actual profiles, we set C = actual total cost and T = actual years of spending. We then fit a Weibull to the scaled annual costs, holding the sum of costs through year T equal to C. We used the Levenberg-Marquardt algorithm to perform the nonlinear least-squares optimizations to find the best fits.

Let $\theta_{il} = (C_{il}, T_{il}, \alpha_{il}, \lambda_{il})$ be the parameters of those best-fit curves. Then θ_{i0} are the best fit parameters to the initial profiles and θ_{i1} are the best fit parameters to the actual outcomes. We model the distribution of θ_{i1} as a function of θ_{i0} and a set of predictor variables X_i simultaneously over all programs, where X includes the program-specific and environmental factors previously listed. Parametric linear models are simultaneously fit to obtain a predictive model for the final profile parameters θ_1. The parameters that uniquely identify a profile are $\theta_{i1} = (C_{i1}, T_{i1}, \alpha_{i1}, \lambda_{i1})$ where $C_{il} = \sum_{t=1}^{T_{it}} C_{il}(t)$ and the other parameters are as previously defined. The models are

$$\log(C_{il}) = (X; \log(\theta_0))\beta_C + \eta_C, \tag{6}$$

$$\log(T_{il}) = (X; \log(\theta_0))\beta_T + \eta_T, \tag{7}$$

$$\log(\alpha_{il}) = (X; \log(\theta_0))\beta_\alpha + \eta_\alpha, \tag{8}$$

$$\log(\lambda_{il}) = (X; \log(\theta_0))\beta_\lambda + \eta_\lambda\,, \tag{9}$$

where the error terms are assumed to be jointly normally distributed. The covariates X include information about previously finished programs that had initial planned spending profiles and actual final profiles. Using these historical data, the model is fit to predict final actual profiles using only information available from a program's Milestone B date. The parameters $\beta = (\beta_C, \beta_T, \beta_\alpha, \beta_\lambda)$ are jointly estimated using a Bayesian Seemingly Unrelated Regressions model with prior distributions on the parameters β and $Var[\log(\theta_{i1})\,|X] \equiv \Sigma$.

The prior for β has a multivariate normal distribution, calibrated such that prior belief is that there is no change in the profile from initial estimate to final actual profile and no other traits of the initial profile are predictive of the final actual profile. This prior belief is fairly strong in order to induce regularization. This prior choice balances the bias-versus-variance tradeoff to produce better predictions.

The prior for Σ has an inverse Wishart distribution, chosen to account for the errors in the initial profile fits to the Weibull curve. In addition, it includes central limit theorem estimates of the final length and size of actual programs. It also includes correlation between the final length, scale, and size parameters.

The joint posterior distribution of β and Σ incorporates the prior beliefs and the historical data to arrive at an updated posterior belief. The Bayesian machinery is especially useful for our purposes because it allows us to obtain random draws from the posterior distribution of β and Σ, which in turn allows us to generate random draws of a final profile distribution $\hat{\theta}_1$ for any program with known initial profile characterized by θ_0 and covariates $X = x$. This lets us estimate the complete (posterior) distribution of final profiles, rather than just a point estimate and variance measure.

Applying the Model to a Specific Program

Experimentation with the predictors listed previously and the 115 programs in our data set indicated that only the coefficients for $\log(C_{i0}), \log(T_{i0}), \log(\alpha_{i0}), \log(\lambda_{i0})$ and Service = Army were significant predictors of final RDT&E development profile shape and size.[3] However, all predictors were retained in the model and strongly regularized to improve predictive performance.

Returning to the hypothetical H-99 program we introduced early on, assume that the program passed Milestone B in 2014. The resulting *X* covariate values for estimating the distribution of eventual actual development expenditure profiles are

- Commodity = Aircraft
- $\alpha_0 = 3.3$
- $\lambda_0 = 5.3$
- $C_0 =$ = $766.2 Million
- T_0 = 12
- Two-year tightness = -0.073
- 10-year tightness = 1.0
- Service = Army
- Commodity Size Optimism = 0.18
- Commodity Length Optimism = 1.11

Using these inputs, we compute the posterior distribution $F_{posterior}(\theta_1)$ of the parameter vector θ_1 that describes the best Weibull fit to the eventual actual development cost profile. In the Bayesian paradigm, this is not a point estimate for a best-fit curve, but instead an updated joint probability distribution that summarizes both our new beliefs about what those parameters are likely to be and our uncertainty about them. We can think of this posterior distribution as a probability distribution over funding profiles.

We also need to account for the fact that the best-fit curve isn't a perfect fit. We also estimate the year-by-year error distributions $E(t)$ using our historical data on how well Weibull distributions fit actual development profiles.

We now have all of the machinery we need to characterize year-by-year cost risk. We use Monte Carlo simulation to sample repeatedly from the $F_{posterior}(\theta_1)$ distribution, generating a large number of possible funding profiles. For each profile generated, we then perturb the annual values using offsets generated from the $E(t)$ distributions. We repeat this process tens of thousands of times, collecting year-by-year statistics on how frequently the required dollars in that year exceeded any given threshold. The result is a set of annual "dollars required" distributions that can be compared against both the original estimate and a hypothetical budget.

[3] Using type 2 sum of squared errors.

Monte Carlo Risk Analysis

Approach

Suppose that we have budgeted a program at some level, possibly different from its predicted cost profile. Let $B(t)$ be the budgeted funds in year t, and let $C_0(t)$ be the predicted cost that will be incurred in year *t.* There are many questions we might wish to ask about the program's affordability risk:

- In how many years will the program exceed the planned budget?
- How many total dollars over budget will the program spend?
- What is the probability of exceeding the budget at least once over the FYDP?
- How much contingency funding would be needed to achieve 90% confidence of staying within budget, assuming unspent contingency carries over to the next year?

These are all questions of potential interest to both Program Managers and Resource Managers. Using the posterior final profile distribution derived from the original profile C_0, we can perform many counterfactual Monte Carlo analyses to answer these kinds of questions. The general pattern for these analyses is as follows:

1. Given the initial development estimate for a program…
2. Define a yearly budget level $B(t)$, and a contingency fund size (if any).
3. Use the regression described previously to determine the posterior distribution on the parameters of the best fit to the final actual development profile for the program.
4. Define outcomes or events of interest.
5. For s = 1,…,S (indexing over iterations of the Monte Carlo algorithm):
    a. "Draw" random parameter vector $\theta_1^{(s)}$ from the posterior distribution.
    b. Compute the corresponding yearly values by evaluating the best fit curve at t = 1, …, $T_1^{(s)}$ and computing $K^{(s)}\,W\left(t|\theta_1^{(s)}\right)$..
    c. Add random noise $E(t)$ drawn from the estimated distribution of the historical errors in fitting curves to final development cost profiles to get a final spending profile $C_1^{(s)}(t), t = 1, \ldots, T_1^{(s)}$.
    d. Evaluate and store any events or outcomes of interest.

Note that the value of $T_1^{(s)}$ used in step 5b is determined as part of $\theta^{(s)}$ in step 5a.

After S iterations, calculate the statistics of interest over the stored events or outcomes. For example, count the number of times N that $C_1^{(s)}(t) < B(t)$ for t = 1…5 and compute $N/_S$. This is the estimated probability of staying within budget for the first five years. The Monte Carlo framework can also allow comparison of different management policies. For example, one could compare the effect of pre-allocating contingency to specific program years versus maintaining a contingency fund to be spent down over time as needed.

In general, we would do this not only for development profiles, but also for procurement spending. In that case, policy makers might be interested in how much difference it would make to be able to manage both RDT&E and Procurement using a single combined budget and/or a single program contingency fund, rather than having to manage separate budgets and contingency amounts due to "color of money" prescriptions.

Example

Consider the H-99 program. Figure 5 shows the mean of the final estimated profile distribution, shown with the original (blue) estimated profile for comparison. Note that not only is the average predicted outcome longer and more expensive in total, but it also ramps up more slowly than the estimate, so that the year-by-year errors change sign over time.

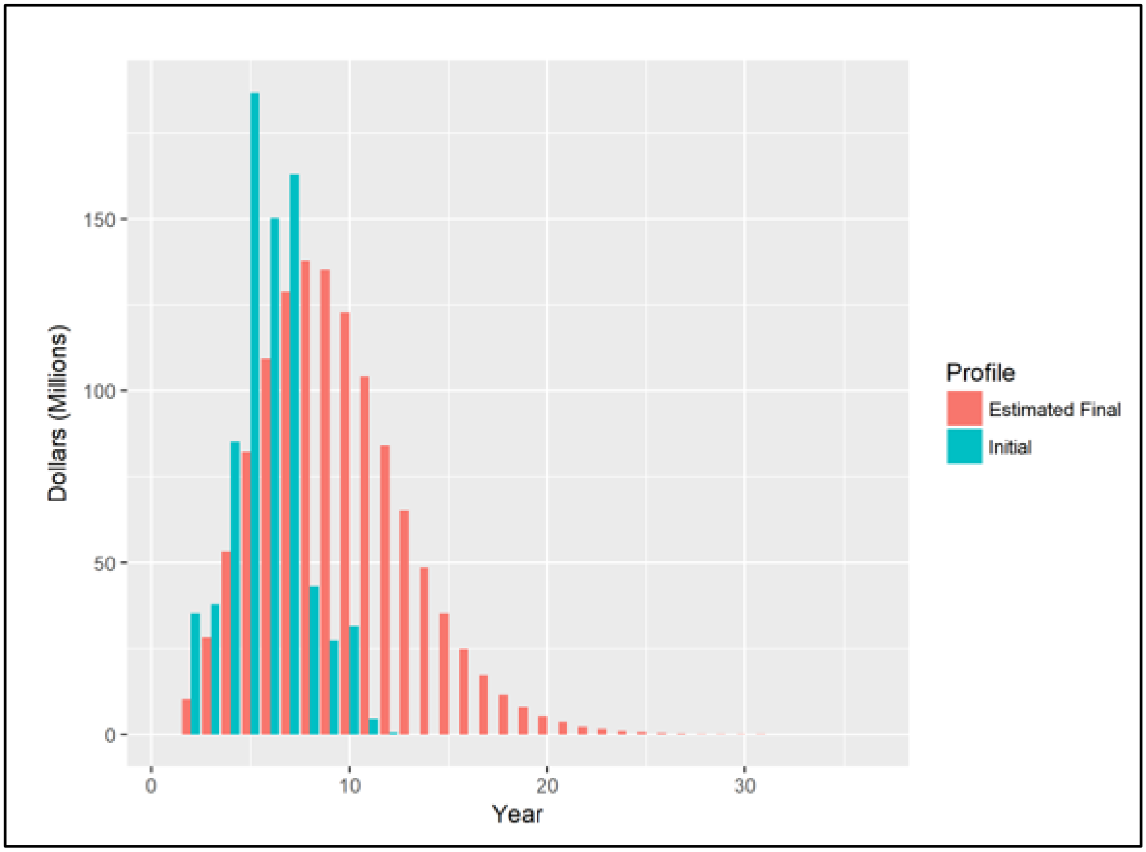


Figure 5. H-99 Original Estimate and Mean Estimated Final Profile

Figure 6 plots the initial planned profile (solid blue) against three different yearly quantile estimates. The dashed gray line shows the median predicted actual expense in each future year estimated from the Monte Carlo simulation. By "median profile," we mean the year-by-year median spending levels $\underset{\theta_1}{\text{median}}(C_1(t))$ over the distribution of all profiles; the shape of the dashed gray line does not correspond to any one set of θ_1 values. The solid black lines in Figure 6 show the original cost estimate profile against upper and lower 10% probability bands. The interpretation of these bands is that, in a given year, there is a 10% chance that the H-99 program will receive less funding than the lower band, and a 10% chance that the H-99 program will receive more funding than the upper band. Note that these probabilities are conditional on looking forward at Milestone II/B—if you already know that a program fell below the 10% band in year 4, this plot provides no information about what to expect in years 5 and beyond.

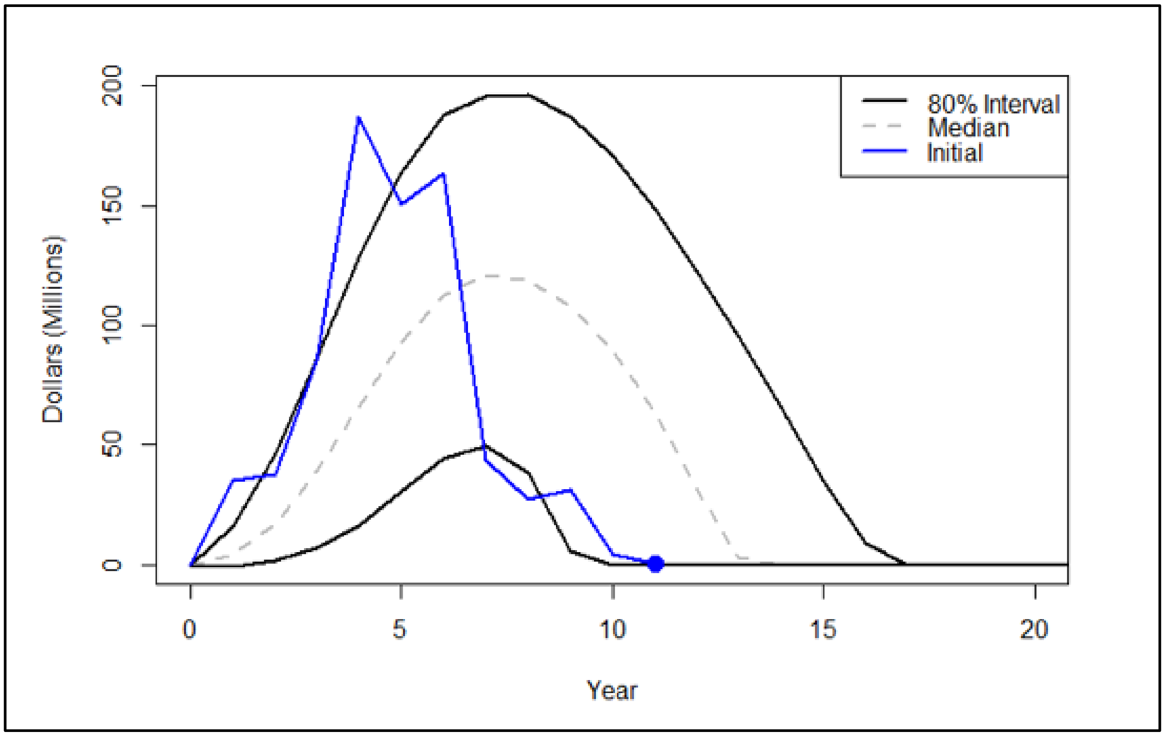


Figure 6. Quantiles of Predicted Annual H-99 Costs

Suppose you, as the Resource Manager for H-99, are interested in how the program will perform in its first FYDP, that is, the first five years of its profile. You wish to know how much additional contingency funding you should expect to require. Table 1 gives the expected overages for consecutive five-year periods.

Table 1. Expected Budget Overages in Five-Year Bins

Overage (Millions)	2.6	336.6	333.4	67.0	9.2	1.4
Years	1–5	6–10	11–15	16–20	21–25	26–30

The table shows that, on average, you would need $2.6 million more than was originally budgeted in years 1 through 5. In contrast, in years 6 through 10, you would need $336.6 million more than planned, on average. Note that even though this table goes out to 30 years, it is unlikely that spending would continue for 30 years. These averages include spending of $0 for outcomes where the program ends prior to that time frame.

It is important to remember that these single estimates of overage are summaries of distributions. Figure 7 shows a box-and-whisker plot that helps to visualize the distribution of the FYDP overages. In each five-year bin, there is a large point mass at zero overage and a highly-skewed distribution of nonzero overages. In all but the second and third bins, the median overage is zero.

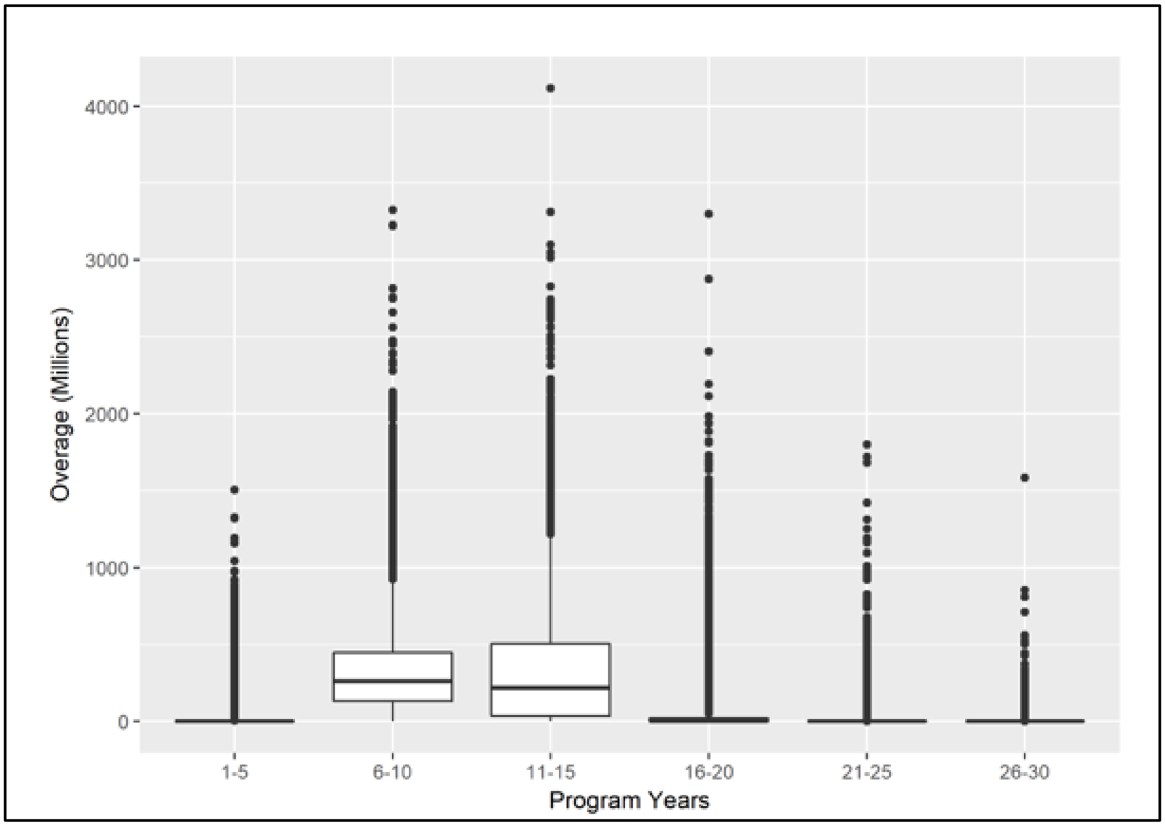


Figure 7. Distribution of Total Cost Overages in Five-Year Bins

Characterizing Mid-Life Programs

We have shown how our model can characterize the affordability risk of a new program's development budget. However, programs are only new once. For a program in the middle of its development, we would like to be able to take advantage of the program's history to date to make an updated, more precise characterization of the remaining cost risk.

A straightforward approach to this might be to add new predictive variables to the functional regression model, reflecting factors such as the age of the development program in years, the relative cost and schedule growth to date (compared to the original estimate), and the relative growth in the program's estimated cost at completion. These could be combined with a revised best-fit functional curve reflecting the program's actual history. In this approach, the same regression model would be used for all points in the development life cycle, with the original estimated profile being a special case with program age = 0 and cost growth factor = 1.

A second possible approach would be to use different regression models for programs at different points in their life cycle. This approach would have the advantage that different underlying functional forms could be used for profiles (or for remaining profiles), depending on the actual outcomes thus far. However, each prediction would then be based on a smaller historical data set, with corresponding loss of statistical power.

Using either of these approaches and the Monte Carlo framework, it would be possible to characterize the future year-by-year development cost risk of every program. Extending the technique to procurement cost risk is also (again) straightforward—the relevant predictive factors might be how many years the program has been in production, how actual unit costs compare to predicted unit costs, any change in planned quantity, etc. As we discuss below, this would allow defense resource analysts to completely characterize the collective behavior of the entire acquisition portfolio, or any subset of it.

Portfolios: More Than One Program at a Time

We have shown how our model can characterize the affordability risk of a single program's development budget. In practice, it would be even more useful to be able to characterize the affordability risk of a group of projects or programs being managed with a common contingency pool. If the outcomes of these programs are approximately independent, this is not much more complicated than the single-program case.

If we have estimated the $F_{posterior}(\theta_1)$ and $E(t)$ distributions for each of a set of programs, we can apply the same kind of Monte Carlo analysis to the sum of their annual costs, compared against a collective portfolio budget and contingency fund. This could be done separately for RDT&E and Procurement, each with its own budget, or it could be done using a combined investment budget. This would enable true affordability analysis of portfolios as envisioned by the Better Buying Power initiatives,[4] but with considerably more realism than current affordability analyses that are based on point-estimate cost profiles assuming fixed program content and quantities.

One potential use of such a model would be to quantify the benefits of portfolio-level contingency funding versus program-level contingency funding. It is well known in the project management world that allocating reserve funds to specific cost areas before you actually know where the cost growth is going to occur leads to less efficient use of those reserve funds. However, it is often politically impossible to protect funds that are not part of the base budget for some cost element. In the DoD, apart from a highly limited ability to reprogram funds from one program element or line item to another, there is currently no ability to reserve funds for contingency use outside of a specific program's budget.

Potential Criticisms of the Method

At this point, it would not be unreasonable to object that this modeling approach assumes that the DoD is incapable of learning to estimate costs more accurately or to contain cost growth more effectively. Worse yet, if Resource Managers were to actually use tools like these to manage portfolios of programs more efficiently, the resulting changes in program outcomes might invalidate the models.

There are perhaps two points of optimism:

- If use of these tools improves the efficiency of contingency funding of portfolios, and therefore reduces cost and schedule growth due to program stretches and funding instability, that will tend to cause users of the model to overestimate the required contingency. That isn't a bad thing; the opportunity cost of allocating a little too much contingency is far less than the marginal cost of allocating too little contingency. As noted previously, even small funding shortfalls lead to additional cost growth due to schedule stretches, increased fixed costs and overhead, and inefficient use of resources.

[4] DoD, *Better Buying Power*, http://bbp.dau.mil/index.html.

- Over time, as the outcomes of recent programs are fed back into the regression models, the posterior distributions produced by the model will correctly adjust to the new reality. If the changes are substantial, one could include “year of program start” as a predictor variable to capture the trend of improvement.

In the meantime, if the drivers of program outcomes are systemic—built into the incentive structure of defense acquisition—the models will continue to capture the likely results of those incentives.

A different potential criticism of this work is that it offers no insights into why costs and schedules deviate from their original estimates or how this could be “fixed.” That is correct—we take cost and schedule changes as given. We do not distinguish between “good” and “bad” causes. If a program is cancelled after seven years, that is just another source of negative cost and schedule growth. If a system turns out to be so useful that the original buy is tripled, or successive block upgrades continue for 30 years, that is just another source of positive cost and schedule growth. The difference between those is very important to warfighters, but it is not relevant to the question of how many dollars we can expect to spend within these programs over the next N years.

Finally, we note that this method explicitly models how much funding a program *will receive* in a given year—not how much it needs, or ought to receive, or would receive if there was more money to go around. As such, the model data incorporate the history of negotiations between the Services, the Office of the Secretary of Defense, and the Congress regarding how much to fund programs year by year, and when to cancel them. If there were to be a fundamental change in the dynamic of how those decisions are made, then that, too, might invalidate the link between historical outcomes and future program outcomes, at least until enough new data could be collected.

Conclusion

Quantifying Annual Resource Risks for a Program or Portfolio

To a first approximation, acquisition programs never spend what they said they would when they began. In fact, the error bars around an initial cost estimate are much larger than is generally understood once program cancellations, restructurings, truncations, and block upgrades have been accounted for. Worse yet, all of this uncertainty arises in a context where programs must fit within annual budgets—it is not enough to only spend as much as you said you would; you must also spend it <u>when</u> you said you would, or problems ensue.

We have developed a methodology to characterize the year-by-year budget risk associated with a major acquisition program. This methodology can be applied to both development costs and procurement costs and can be extended to understand the aggregate affordability risk of portfolios of programs. The method allows Resource Managers to estimate annual budget risk levels, required contingency amounts to achieve a specified probability of staying within a given budget, and a host of other relevant risk metrics for programs. It also allows policy makers to predict the impact on program affordability of proposed changes in how contingency funds are managed.

Future Research

This technique is currently in the prototype stage and is based on a relatively sparse set of historical program outcome data. There is still much work to be done on improved statistical techniques for the functional regressions, modeling of procurement profile risk, conditional modeling of procurement given development outcomes, and characterization of the distribution of residuals around the best-fit functional curve. There is also a great deal to be learned about how managers could best use the information provided by this method to manage actual programs and portfolios, and what the implications might be for recommending changes to acquisition law and regulations.

References

Adoko, M. T., Mazzuchi, T. A., & Sarkani, S. (2015). Developing a cost overrun predictive model for complex systems development projects. *Project Management Journal, 46*(6), 111–125. doi:10.1002/pmj21545

Anderson, T. P., & Cherwonik, J. (1997, Summer). Cost estimating risk and cost estimating uncertainty guidelines. *Acquisition Review Quarterly*, 339–348. Retrieved from www.dtic.mil/get-tr-doc/pdf?AD=ADA487532

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (TR-343-AF). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/technical_reports/2006/RAND_TR343.pdf

Asher, N. J., & Maggelet, T. F. (1984). *On estimating the cost growth of weapon systems* (IDA Paper P-1494). Alexandria, VA: Institute for Defense Analyses.

Bitten, R. E., & Hunt, C. (2017). *Assessing the impact of confidence levels in funding and budgeting NASA science missions.* Presentation at the ICEAA Professional Development & Training Workshop, Portland, OR.

Bliss, G. (1991, June). *The accuracy of weapon systems cost estimates*. Presentation at the 59th Military Operations Research Symposium, U.S. Military Academy, West Point, NY.

Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time phasing aircraft R&D using the Weibull and Beta distributions. *Journal of Cost Analysis and Parametrics, 8*(3), 150–164. doi: 10.1080/1941658X.2015.1096219

Chib, S., & Greenberg, E. (1995). Hierarchical analysis of SUR models with extensions to correlated serial errors time-varying parameter models. Journal of Econometrics, 68, 339–360. doi:10.1016/0304-4076(94)01653-H

Congressional Budget Office (CBO). (2017). *An analysis of the Obama administration's final future defense program* (CBO Report No. 52450). Retrieved from https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/52450-fydp.pdf

Coonce, T., Bitten, R., Hamaker, J., & Hertzfeld, H. (2010). NASA productivity. *Journal of Cost Analysis and Parametrics, 3*(1), 59–78. doi:10.1080/1941658X.2010.10462228

Creedy, G. D., Skitmore, M., & Wong, J. K. W. (2010). Evaluation of risk factors leading to cost overrun in delivery of highway construction projects. *Journal of Construction Engineering and Management, 136*(5), 528–537. Retrieved from https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29CO.1943-7862.0000160

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Hough, P. G., & Norton, D. (1993*). An analysis of weapon system cost growth.* Santa Monica, CA: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/monograph_reports/2006/MR291.pdf

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (R-3937-ACQ). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/reports/R3937.html

GAO. (2006). *Space acquisitions: DOD needs to take action to address unrealistic initial cost estimates of space systems* (GAO-07-96). Washington, DC: Author. Retrieved from https://www.gao.gov/products/GAO-07-96

Glennan, T. K., Bodilly, S. J., Camm, F., Mayer, K. R., & Webb, T. (1993). *Barriers to managing risk in large scale weapons system development programs* (MR-248-AF). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/monograph_reports/MR248.html

Jessup, R., & Williams, J. (2015). *The cost of commonality: Assessing value in joint programs* (MBA professional report). Monterey, CA: Naval Postgraduate School.

Jimenez, C., White, E. D., Brown, G. E., Ritschel, J. D., Lucas, B. M., & Seibel, M. J. (2016). Using pre-Milestone B data to predict schedule duration for defense acquisition programs. *Journal of Cost Analysis and Parametrics, 9*(2), 112–126. doi:10.1080/1941658X.2016.1201024

Light, T., Leonard, R. S., Pollak, J., Smith, M., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for major defense acquisition programs (MDAPs).* Santa Monica, CA: RAND. doi:10.7249/RR1723

Marshall, A. W., & Meckling, W. H. (1959). *Predictability of the costs, time, and success of development* (P-1821). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/papers/P1821.html

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (2nd ed.). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017). *Influences on the timing and frequency of cancellations and truncations of major defense acquisition programs* (IDA Paper P-8280). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., Gould, B. R., Wu, L., Fuchs, E., & Bronson, P. (2013). *Evidence on the effectiveness of the Defense Acquisition Board process* (IDA Paper P-5011). Alexandria, VA: Institute for Defense Analyses.

Tate, D. M. (2016). *Acquisition cycle time: Defining the problem* (IDA Document NS D-5762). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Harmon, B. R., & Utech, D. M. (1994). *Understanding cost and schedule growth in acquisition programs* (IDA Paper P-2967). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., Om, N. I., & Palmer, P. R. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectivenes*s (IDA Paper P-2201). Alexandria, VA: Institute for Defense Analyses.

Tyson, K. W., Nelson, J. R., & Utech, D. M. (1992). *A perspective on acquisition of NASA space systems* (IDA Document D-1224). Alexandria, VA: Institute for Defense Analyses.

Acknowledgments

This work was originally conceived and funded by the congressionally established Section 809 Panel (https://section809panel.org/), and in particular by Commissioners Dr. Allan Burman and David Ahern. Darren Harvey of the panel technical staff provided valuable guidance and feedback. Dan Cuda and James Bishop of the IDA were technical reviewers, and their observations and suggestions improved the paper significantly.

Have Changes in Acquisition Policy Influenced Cost Growth of Major Defense Acquisition Programs?

David L. McNicol—joined the Department of Defense (DoD) in 1982, where, from 1988 until 2002, he was a Deputy Director of Program Analysis and Evaluation (PA&E). Earlier, Dr. McNicol taught at the University of Pennsylvania and the California Institute of Technology. He holds a BA in Economics from Harvard and an MS (Management) and PhD (Economics and Finance) from MIT. Employed at the Institute for Defense Analyses (IDA) since his retirement from the DoD, he became director of the Cost Analysis and Research Division in 2006. Still at IDA, Dr. McNicol stepped down in 2012 to return to his previous role as a Research Staff Member. [dmcnicol@ida.org]

Abstract

This paper is part of a series, several previous papers of which explored whether unit cost growth for major defense acquisition programs (MDAPs) is statistically associated with changes in acquisition policy over the period FY 1965–FY 2009. The project is now substantially completed and is being assembled into a final report. This paper presents the project's three main conclusions on the effects of changes in acquisition policy on MDAP cost growth. First, changes to the acquisition process implemented in 1969 reduced average growth in unit cost. These changes remained in place through the end of the period considered in this study (FY 2009) and, taking changes in funding climate into account, average unit cost growth remained at about the lower level stemming from the 1969 reforms. Second, the OSD-level oversight process has not been fully successful in responding to the increased pressures on cost growth during bust—as opposed to boom—funding climates. Third, again taking account of bust and boom funding climates, the experiments undertaken post-1969 on different contract types and relaxation of acquisition regulations seem not to have reduced either the cost of systems or growth in unit costs.

Introduction

McNicol, Tate, Burns, and Wu (2016) and McNicol (2017a) explored whether unit cost growth for major defense acquisition programs (MDAPs) is statistically associated with changes in acquisition policy over the period Fiscal Year (FY) 1965–FY 2009. Parts of this work were presented to the NPS Acquisition Research Symposia in 2016 and 2017 (McNicol, 2017b; McNicol & Tate, 2016). The project is now substantially completed and is being assembled into a final report. This paper presents the project's main conclusions on the effects of changes in acquisition policy on MDAP cost growth.

Framework

An early discovery of the project was that average unit cost growth of programs that pass Milestone (MS) B during a bust period is significantly higher than that of programs that passed MS B during a boom climate (McNicol & Wu, 2014).[1] For that reason, the analysis distinguishes between bust and boom funding climates. During the 45 years covered by this study (FY 1965–FY 2009), there were two complete bust-boom cycles in Department of Defense (DoD) procurement funding: (1) The bust climate for modernization of weapon

[1] The most developed explanations of funding climate and acquisition policy configuration are provided in McNicol (forthcoming, IDA, Chapter 1).

systems that began in the mid-1960s and lasted until the Carter-Reagan buildup of the early to mid-1980s, and (2) the long post–Cold War bust climate followed by the post-9/11 boom.

Where a bust funding climate may provide an upward pull on cost growth, the acquisition oversight process should provide a restraining push. Accordingly, it is necessary also to recognize changes over time in acquisition policy and process configurations. Five policy and process configurations are distinguished:

- McNamara-Clifford (FY 1964–FY 1969);
- Defense Systems Acquisition Review Council (DSARC; FY 1970–FY 1982);
- Post-Carlucci DSARC (P-C DSARC; FY 1983–FY 1989);
- Defense Acquisition Board (DAB; FY 1990–FY 1993 and FY 2001–FY 2009); and
- Acquisition Reform (AR; FY 1994–FY 2000).

"Policy" and "process" tend to be intertwined; process typically is required to implement policy, and the most successful and durable policies tend to be embedded in process. For this reason, and to avoid constant repetition of "process and policy," the term *acquisition policy* is used here in a broad sense to encompass both policy on particular topics (for example, contract types) and the Office of the Secretary of Defense (OSD)–level oversight process (for example, definition of the milestones).

Finally, a measure of cost growth is required. The measure used is based on Program Acquisition Unit Cost (PAUC). PAUC is the sum of Research, Development, Test, and Evaluation (RDT&E) cost and procurement cost, divided by the number of units acquired. For this paper, PAUC growth is computed by comparing the MS B baseline value of PAUC in program base-year dollars—which can be thought of as a goal or a prediction—to the actual PAUC reported in the program's last Selected Acquisition Report (SAR) in program base-year dollars and adjusted to the MS B baseline quantity. Appendix B of McNicol (2017a) describes the conventions used in assembling the database, the sources of the data used, and the quantity adjustment computations. The unit cost growth estimates were updated to the December 2015 SARs. Only completed programs (defined as programs with an end date of FY 2016 or earlier) are used in this analysis because some costs associated with a program may not be fully reflected in its SAR until the program is completed. To be clear, in what follows, the term *PAUC growth* means PAUC growth as defined above, that is, growth from MS B through the end of procurement, adjusted to the MS B quantity.

Success of the Milestone Review Process

Studies done in the past 20 years found no evidence that changes in DoD acquisition policy made after 1970 reduced cost growth on major systems acquisitions[2] (Christensen, Searle, & Vickery, 1999; McNicol & Wu, 2014; O'Neil, 2011). The conventional wisdom

[2] Under Secretary of Defense (Acquisition, Technology, and Logistics; USD[AT&L]; 2016) reports evidence that growth in the RDT&E portion of MDAP costs was lower than would be expected for a bust period (see, in particular, p. 13). This report and the earlier reports in the series (USD[AT&L], 2013, 2014, 2015) provide a comprehensive review of MDAP outcomes and changes in acquisition policy over roughly FY 2010 through FY 2016.

seems to have transformed this finding into a much stronger assertion—that OSD-level oversight of MDAPs has had no effect on their outcomes, or at least on cost growth.

The studies cited do not in fact reach this conclusion and could not, if only because they do not consider data prior to FY 1970.[3] In July 1969, David Packard, then Deputy Secretary of Defense, introduced a set of reforms. These reforms may well have reduced PAUC growth and they did remain in place over time. At the cartoon level, the 1969 Packard reforms may then have been analogous to a light switch. Of course, light comes on only if the bulb is good and circuit hot—that is, that decision makers to a reasonable extent embrace the relevant acquisition policies.

The light switch metaphor of the 1969 Packard reforms is examined in this section, first statistically and then historically.

Statistical Analysis

The first of the questions posed by the light switch metaphor is whether average PAUC growth after the introduction of the Packard reforms in July 1969 (the start of FY 1970) was significantly lower than that of the preceding McNamara-Clifford period. During the 10 years following the Packard reforms (FY 1970–FY 1980), average PAUC growth was 37%; average PAUC growth for McNamara-Clifford was twice that, 74%. There is more to the issue than just this comparison, however. In addition to the 1969 Packard reforms, three other factors may have had significant effects on the difference in average PAUC growth between the two periods:

- Program duration,
- Funding climate at MS B, and
- Proportion of programs that passed MS B in the period that entered a boom funding climate post MS B.[4]

In fact, these factors do not explain the higher PAUC growth of the McNamara-Clifford period:

- MDAPs in the database for the bust portion of the DSARC period had a longer average duration (15.1 years) than did those of the McNamara-Clifford period (13.1 years).
- The comparison is between the bust phase of DSARC (FY 1970–FY 1980) and McNamara-Clifford (FY 1965–FY 1969), which also was a bust climate.

---

[3] The key point is made most clearly in McNicol and Wu (2014): “We have no fully comparable [Program Acquisition Unit Cost] PAUC growth data for the periods before the DSARC was established. Consequently, the statistical analysis leaves open the possibility that the DSARC and its successors provided a useful discipline on acquisition programs” (p. 7). Dews et al. (1979, Chapter IV) found that the 1969 Packard reforms led to lower PAUC growth. Drezner et al. (1993, pp. 28–30) found that this conclusion did not hold when account was taken of program duration. Using a model that accounts for funding climate and time spent in boom and bust periods (and therefore program duration), McNicol (forthcoming, Acquisition Research Symposium) finds that PAUC growth during the decade following the 1969 Packard reforms was significantly lower than that during the McNamara-Clifford period.

[4] On this, see McNicol (2017a).

- A higher proportion of programs that passed MS B in the bust portion of the DSARC period later entered a boom period (42 of 49); in comparison, only four of 16 McNamara-Clifford programs went on to enter a boom period.

In addition, a model that includes acquisition policy, funding climate, duration, and boom effects finds that average PAUC growth in the DSARC period was significantly less than it was in the McNamara-Clifford period (McNicol, forthcoming, IDA).

The second question posed by the light switch metaphor is whether the effects of the 1969 Packard reforms persisted. The statistical results indicate that they did. In particular, the model mentioned in the preceding paragraph finds that the average PAUC growth in the bust funding climates of each of the acquisition policy periods after the DSARC (P-C DSARC, DAB, and AR) was significantly lower than that of McNamara-Clifford. Apparently, the light switch remained in the "on" position.

Finally, are there statistically significant differences in cost growth between DSARC, P-C DSARC, DAB, and AR? As is discussed below, some initiatives on contract policy and regulation did affect PAUC growth. These effects do not come through in the averages. The averages, however, do not present an entirely consistent picture. On the one hand, the model cited above does not show any statistically significant differences in average PAUC growth among the four post McNamara-Clifford policy periods. One the other hand, when the acquisition policy bins (DSARC, P-C DSARC, DAB, and AR) are dropped from the analysis, we find a statistically significant decreasing trend (of about four-tenths of a percentage point annually) in PAUC growth.[5] Taking all of the evidence together, the safe conclusion is that given funding climate, PAUC growth did not increase over the period FY 1970–FY 2009 and may have shown a modestly decreasing trend.

Historical Evidence

The statistical results present straightforward historical questions: (1) Did the 1969 Packard reforms differ substantially from what came before? (2) Did they persist through the end of the period considered in this study (FY 2009)? These questions require that we look at what came before Packard—what his reforms reformed, the substance of the Packard reforms themselves, and changes in the Packard reforms over the four decades that followed.

Although the fact seems to have been dropped from the historical memory of the DoD acquisition community, the 1969 Packard reforms were reforms to a process established in February 1964 by then Secretary of Defense Robert McNamara. The initial version of this process, set out in Department of Defense Directive (DoDD) 3200.9, *Project Definition Phase*,[6] specified only a single milestone. A revision of this directive issued in July 1965 established a second milestone.

The DoDD 3200.9 process was built around Total Package Procurement (TPP), the use of which the directive required whenever it was feasible. In instances in which TPP was judged to be infeasible, use of a Fixed Price (FP) development contract was strongly

---

[5] See McNicol (forthcoming, IDA, Chapter 3).
[6] For a discussion of DoDD 3200.9, see Glennan (1965, p. 12).

encouraged.[7] A TPP contract covered Engineering and Manufacturing Development (EMD), procurement, and usually some aspects of Operations and Maintenance (O&M), each on a fixed price basis. These contracts were awarded after a competition. Approval at the first of the revised DoDD 3200.9's milestones authorized the Component to fund engineering development work sufficient to define the project at a level of detail that permitted the contractors (usually at least two) to write TPP contracts for EMD, procurement, and—often—aspects of O&M as well. This limited engineering development phase was to last at most six months.[8] With the proposals then in hand, the Service would (at the second milestone) seek authority to select one of the competing contractors and award a TPP contract.

The 1969 Packard reforms retained the basic architecture of the DoDD 3200.9 process but made major changes in three aspects of acquisition policy that directly influence PAUC growth:

- Policy on contract types,
- Definitions of the milestones, and
- The OSD-level milestone review process.

In addition, Packard stated more clearly policies on realistic costing and full funding, and changed the OSD process for monitoring cost growth during program execution.

Packard ruled out use of TPP: "[New complex defense] systems will not be procured using the total package procurement concept or production options that are contractually priced in the development contract." Packard also discouraged the use of FP development contracts: "Cost type prime and subcontracts are preferred where substantial development effort is involved." As a general matter, Packard's policy was that "contract type shall be consistent with all program characteristics including risk."[9]

Absent the insistence on the use of TPP, DoDD 3200.9's two milestones no longer made sense. New milestone definitions were then required. Packard's reforms defined three milestones:

- MS I—authorization to begin technology development,
- MS II—authorization to enter EMD, and
- MS III—authorization to begin Full Rate Production (FRP).[10]

---

[7] Fox (2011) reports that McNamara "abandoned the TPP concept in 1966" (p. 38). This may be the case, but the source Fox cites is for the facts stated earlier in the paragraph. There is some evidence that TPP continued to be used through the end of the McNamara-Clifford period (see Poole, 2005, p. 83).
[8] DoDD 3200.9 (July 1, 1965, p. 9, para. VI.F.7).
[9] DoDD 5000.1 (July 13, 1971, p. 5, para. C.7).
[10] DoDD 5000.1 (January 18, 1977) was the first to give the milestones numbers. DoDI 5000.2, issued October 23, 2000, formally established MSs A, B, and C (in place of MSs I, II, and III) as the main decision points for an MDAP. The definitions are such that MS B is placed several months earlier in the process than MS II. At different times, MS C has been defined as the start of Low Rate Initial Production (LRIP) (MS IIIa until 2000) or FRP (MS III.)

Technology validation (or technology development) was not a new activity. The change made by Packard was that entry into the Validation Phase (i.e., technology development) required Milestone Decision Authority (MDA) approval (what was then called MS I—MS A post-2000). The important point, however, was not the requirement for MDA approval as such, but that the purpose of the Validation Phase was to ensure the technologies that a system would use were sufficiently mature to proceed into EMD. One of Packard's signature policies was "fly-before-you-buy." He encouraged building and testing a prototype during the Validation Phase but did not require it. Nevertheless, it is a reasonable guess that on average, MDAPs devoted more time and funding to technology development than was the case before the Packard reforms. Introduction of the Validation Phase probably did then reduce the risk of programs that came forward for MS II. Moreover, under the Packard reforms, EMD became a contractually distinct phase that the firm(s) were required to complete before they could gain authority (at MS III) to enter FRP. This, again, had the effect of embedding an aspect of "fly-before-you-buy" into the acquisition process.

In addition to the basic policy changes on contracting and the milestone definitions, Packard sought to better codify and regularize the OSD-level acquisition process. An important part of this was his establishment of the DSARC. The DSARC replaced the more ad hoc coordination process of DoDD 3200.9. DoDD 3200.9 itself was replaced with DoDD 5000.1 (July 13, 1971), issued after Packard had left the DoD. In 1975, the first version of DoDI 5000.2, *The Decision Coordinating Paper (DCP) and the Defense Systems Acquisition Review Council (DSARC),* was released. This instruction served mainly to define the process in more detail.

The question is whether it is plausible to attribute the lower average PAUC growth during the 1970s to the Packard reforms. The answer to this question offered here is "Yes." The main factors were Packard's change in contract policy and his introduction of a more extensive technology development and risk reduction phase before MS II. This phase embedded in the milestone review process Packard's policy of "fly-before-you-buy," thereby presumably on average reducing the risks remaining in MDAPs that sought MS II authority.

The second historical question is whether the Packard reforms persisted. For the period covered by this study, none of the Packard reforms was reversed or reduced to a dead letter or overtaken by other changes. For example, like the original DoDI 5000.2, the version in force in FY 2009 (1) required a robust Technology Development phase, (2) required realistic costing of the program proposed at MS B and provided for an independent cost estimate by what is now CAPE-CA, and (3) required full funding at MS B of the cost estimate adopted by the MDA. Other examples could be provided. On a historical basis, then, it is not at all farfetched to conclude that the effects of the Packard reforms persisted because the reforms themselves continued in force.

This is a remarkable conclusion. There is a Darwinian "survival of the fittest" aspect to changes in OSD processes. Many changes do not survive the administration that introduces them. Those that do generally are abraded until they fit well with the other OSD processes. The DSARC/DAB process lost none of its parts, over four decades, and in fact was strengthened. The historical evidence is, then, consistent with the statistical finding that average PAUC growth (within a funding climate) has remained below its level in the McNamara-Clifford period.

Evidence of a Limitation of the Milestone Review Process

Consideration of PAUC growth in boom and bust funding climates points to a limitation of the OSD-level MDAP oversight process. Table 1 provides a summary of average PAUC growth of MDAPs that passed MS B in the bust and boom phases of each of the two bust-boom cycles in the database. Average PAUC growth of MDAPs that passed MS B during the bust phase of the first cycle was about twice that of MDAPs that passed during the boom phase; the difference was nearly a factor of 10 for the second cycle. More intense competition for funding in bust climates is a major part of the explanation for these facts, as it would provide the Services with a stronger incentive to propose programs with relatively greater risk in their MS B baselines. It is not a sufficient explanation, however. DoDD 5000.1 and DoDI 5000.2 do not permit MDAPs that passed MS B in bust periods to be riskier, and therefore have higher PAUC growth on average than those that passed in boom periods. Accordingly, it is necessary to ask why the DSARC/DAB process did not prevent the higher average PAUC growth in bust periods.

Table 1. Average PAUC Growth in Boom and Bust Phases for Completed Programs

Cycle	Period (Fiscal Years)	Bust Climates	Period (Fiscal Years)	Boom Climates
First Bust-Boom Cycle‡	1970–1980	37% (49)	1981–1986	18% (35)
Second Bust-Boom Cycle	1987–2002	37% (45)	2003–2009	2% (11)

‡ Excluding McNamara-Clifford.

One possible explanation is that in bust periods, the greater frequency and severity of problems with programs that came to an MS B review pushed the OSD-level oversight process to a capacity constraint. For example, if the workload involved in milestone reviews increases significantly in bust periods, the staff could be stretched to the point that it fails to identify to the MDA significant problems in the proposed baseline.[11] A possibly more important constraint is the greater intensity of Service opposition to any changes in proposed programs that would delay programs or add to funding requirements.

Another possibility challenges the premise that the DSARC/DAB process failed to check the PAUC growth of MDAPs that passed MS B in bust climates. This challenge is prompted by the statistical finding that MDAPs that passed MS B in bust periods and later went into boom periods had significantly higher PAUC growth than those completed in a bust period. Stripped of all qualifications, the challenge is this: In bust periods, program ambitions are scaled back so as to be consistent with the tighter funding constraint and their PAUC growth is attributable to the costs of program changes—that is, enhancements—adopted in a later boom period. In this case, the DSARC/DAB process would be judged to be a success in that programs that passed MS B in bust climates had relatively modest ambitions and were structured as evolutionary acquisitions. In short, given the way the SARs and some statutes are structured, it is possible to have significant PAUC growth without failures in the acquisition process. This possibility is only a partial explanation, however, since less than one-third of PAUC growth of MDAPs that passed MS B in bust periods was due to program changes (McNicol, forthcoming, IDA).

[11] Fewer MDAPs tend to pass MS B annually in bust years, but they might each have a larger number of problems with their baseline.

Finally, there is a more subtle challenge to the premise that the DSARC/DAB process failed to check the PAUC growth of MDAPs that passed MS B in bust climates—that the MDA deliberately, with adequate information and at least tacit support from the Secretary of Defense, decided to accept greater risks in MDAPs that came to MS B reviews in bust climates. The underlying point here is that bust climates presented senior officials in the OSD and those in the Services with the same menu of unappealing choices. Case by case and overall, there was no option that did not have serious undesirable consequences.

Each of the Services has a portfolio of programs across mission areas and commodity types, extending from efforts in the technology base through programs nearing the end of production. When a program is completed, it opens a resource “hole” that programs emerging from EMD can occupy. In turn, programs earlier in the acquisition cycle can move forward. When funding for acquisition turns down, these holes get smaller, or close entirely, or require cuts in funding for ongoing programs. The alternatives available in this circumstance are cancellations of programs, delays in new starts, programs that are more austere than is cost-effective on a long-term view, stretches, and unrealistic baselines—in particular, unrealistic cost and schedule estimates. Taking DoDD 5000.1 and DoDI 5000.2 at face value, one role assigned to the DAB is that of precluding one class of options—unrealistic baselines. Doing so would not address the underlying problem, which is an inconsistency between force structure, the capabilities that the Department was expected to provide, and funding. These factors almost certainly were inconsistent during the 1970s and for more than a decade after the end of the Cold War. That inconsistency is the context in which high average PAUC growth and most cancellations arise, and presumably is a major factor to be considered in designing proposals for improved outcomes.

The three explanations offered here are not mutually exclusive. It seems likely that each is accurate in some cases but that none is clearly satisfactory as an overall explanation of why the OSD-level oversight process was not fully successful in limiting cost growth of MDAPs that passed MS B in bust climates.

Policy Initiatives on Contract Types and Regulations

Starting with the Reagan Administration in 1981, Secretaries of Defense early in their tenures typically announced changes in acquisition policy. Most of these were directed at objectives connected to PAUC growth only very loosely if at all. The main exceptions to these statements are changes in policy on contract types and experiments with relaxation of DoD acquisition regulations and, in some instances, both DoD acquisition regulations and some statutory provisions on acquisition of major weapon systems.

Initiatives on Contract Types

As was noted previously, McNamara required the use of TPP when it was feasible to do so and encouraged the use of an FP development contract when it was not. Packard reversed both of these policies, but both TPP and FP development contracts were again tried in the 1980s. During the AR period (FY 1994–FY 2000), several MDAPs were acquired using an approach called Total System Performance Responsibility (TSPR).

Table 2 lists the MS B year and the PAUC growth for completed MDAPs in the database acquired with TPP. Four McNamara-Clifford programs were acquired using TPP. Probably because they were grandfathered, three programs in our sample for the early 1970s used a TPP contract. The prohibition on TPP did not appear in the next update of DoDD 5000.1, dated January 18, 1977, and three additional programs in our sample that passed MS B in the Reagan boom years also used a TPP contract. Only one of these 10 MDAPs (AGM-65A Maverick [TV]) had a quantity-adjusted PAUC growth of less than 50%. The average PAUC growth of the 10 programs is 86.2%, and median PAUC growth is 68%.

This is among the clearest and strongest results to come out of the literature on cost growth of MDAPs and one for which the underlying causes are reasonably well understood (McNicol, Tate, Burns, & Wu, 2016, p. 7).[12]

Table 2. MS B and PAUC Growth for 10 MDAPs Procured With TPP Contracts

MDAP	MS B (FY)	PAUC Growth
C-5A Galaxy	1966	77%
AGM-65A Maverick (TV)	1968	1%
Landing Ship Assault (LHA) Tarawa class	1969	57%
SRAM	1967	263%
FIM-92 Stinger Missile	1973	110%
AGM-84A Harpoon	1973	56%
SURTASS/T-AGOS	1975	68%
T-45 Goshawk	1984	70%
JSTARS USAF	1985	123%
C-17A Globemaster	1985	57%

Note: The identifications are based on Tyson et al. (1992, Chapter X and Appendix A, Table A-10) and McNicol (2004, pp. 53, 57–59). Tyson et al. includes the Spruance Class destroyer among the TPP programs. The lead ship of the class may have been acquired on a TPP contract but the class as a whole seems not to have been.

The FP development contract was used in the 1960s and again in the 1980s. The conventional wisdom associates it with high PAUC growth, but this opinion is not supported by the data in Table 3. Note, however, that five of the six MDAPs identified as using an FP development contract passed MS B during a boom climate, which may account for their low PAUC growth. It is also relevant that the RDT&E portion of most MDAPs acquired with TPP contracts were fixed price, and growth in their RDT&E cost certainly contributed to their high PAUC growth.

[12] For further discussion of TPP and FP development contracts, see Tyson et al. (1992, Chapter X); McNicol (2004, pp. 53, 57–59); and O’Neil and Porter (2011, pp. 9–31).

Table 3. MS B and PAUC Growth for 10 MDAPs Procured With FP Development Contracts

MDAP	MS B (FY)	PAUC Growth	EMD Growth
F-14A	1969	29%	45%
E-6A	1983	0%	9%
JTIDS	1982	not available	not available
Stinger RMP	1983	not available	not available
T-AO 187	1984	-3%	24%
F-14D	1986	-6%	-2%

Note: The identifications are based on Tyson et al. (1992, Ch. X and Appendix A, Table A-10).

The third problematic contracting approach is TSPR, which was used primarily during the AR period and is one of the signature experiments of that period.[13] TSPR was a clause included in contracts; it was a way of structuring contracts, not a type of contract, and could be used with different contractual forms. The term *performance* in TSPR was understood in a specialized way. It referred to metrics that characterized the ability of the system to accomplish certain missions. For example, one aspect of performance of a cargo aircraft might be the tons of cargo of a specified type that a given number of the aircraft could deliver in 24 hours under specified conditions. The idea was to cast contracts in terms of such performance metrics, rather than the usual statements of work and technical specifications. The contractor would be responsible for delivering a system that met the performance specifications, while the government would do only a limited number of “inherently governmental” functions (primarily contract management, specification of the performance metrics, budgeting and financial management, and acceptance testing).

Table 4 provides a list of TSPR MDAPs. The list may not be complete—it can be hard to tell whether TSPR was used to acquire any particular system. For example, the Advanced Extremely High Frequency (AEHF) satellite is sometimes discussed with TSPR programs. Note that all but one of the MDAPs in Table 4 (AGM-158 Joint Air-to-Surface Standoff Missile) is a satellite system.

[13] AR also encouraged the use of three other contracting initiatives: Alpha contracting, Price-Based Acquisition (PBA), and Best Value contracting. Hanks et al. (2005) provides a useful listing of acquisition reform initiatives between 1991 and 2001 at least nominally accepted by the DoD. Contrary to what might be inferred from some descriptions of PBA, none of these was problematical insofar as PAUC growth is concerned. See Quander and Woppert (2010); Hawkins and Cuskey (2011, pp. 240–274); and Rapka et al. (2006, pp. 34–37). On PBA, see Lorell, Graser, and Cook (2005), especially Chapter 2.

Table 4. MDAPS Acquired Using a TSPR Strategy

MDAP	MS B	PAUC Growth
Global Positioning System IIF (GPS-IIF)[a]	1996	high[b]
Space Based Infrared Sensor-High (SBIRS-H (baseline)[a]	1997	299%
AGM-158 Joint Air-to-Surface Standoff Missile (JASSM)[c]	1998	73%
Evolved Expendable Launch Vehicle (EELV)[a]	1998	251%
National Polar-orbiting Operational Environmental Satellite System (NPOEES)[a]	2002	Cancelled

[a] GAO (2006, p. 8).

[b] GPS-IIF was not an MDAP; it was part of the NAVSTAR GPS program. GAO reported a program office estimate—apparently from about 2009—that implied a cost growth from the MS B baseline of 119%. See GAO (2009, p. 13). It is not clear that the program office estimate cited by GAO was adjusted for quantity changes.

[c] GAO (2010, p. 4).

It was anticipated that TSPR would reduce the number of people employed in government program offices,[14] but the main source of cost reductions was expected to stem from the freedom TSPR gave contractors to make trades that reduced cost while maintaining performance. Those expected cost reductions were built into the MS B baselines.[15] The savings failed to materialize and the result was high PAUC growth in five of the six cases (the sixth was cancelled). In 2002, the USD(AT&L) Aldridge stated that TSPR would no longer be used (Hanks et al., 2005, pp. 19–20). More generally, during 2001–2009, there were no further major experiments with different contracting approaches.

A TSPR arrangement provides the contractor with the authority to make trades that reduce cost while maintaining performance. Whether the contractor is incentivized to make these trades depends on the contract type. An FP contract does, among other things, provide such an incentive, but there is no reason to think that an FP TSPR contract for a major EMD effort would not have the same flaws as an ordinary FP development contract.[16] In contrast, a cost-type contract tends to incentivize a capability-cost spiral and for that

[14] For critiques of TSPR, see Defense Science Board (2003, pp. 3, 10); Lorell, Leonard, and Doll (2015, p. 31); Kim et al. (2015, pp. 33–34); GAO (2006, p. 10); and Temple (2013, pp. 269–271). In some cases, the government did not require the provision of the data needed to understand the state of a program. Moreover, government staff, particularly systems engineering staff, was reduced to a point that compromised their ability to establish baseline requirements and monitor the programs' progress. One major reason for failure of TSPR programs apparently was the lack of sufficiently strong engineering expertise in both government and industry.

[15] Apparently, in at least some cases, this was done over the objections of the independent cost analysts in the OSD and the Air Force. See, in particular, GAO (2006) and Defense Science Board (2003).

[16] On the limitations of FP development contracts in space programs, see Arnold et al. (2013). Lorell et al. (2015, p. 7) seems to equate TSPR with TPP. TPP does not imply a "hands off" stance, but the government probably did generally place total system responsibility on the contractor. TSPR, however, amounts to TPP only if it uses an FP contract that extends beyond EMD to production.

reason probably requires the government to exercise a degree of oversight that obviates the advantages sought by a TSPR arrangement.

To put these results in context, we have a PAUC growth estimate for 110 MDAPs that passed MS B in a bust climate and were completed by the end of FY 2016. Forty of these had a PAUC growth of at least 50%; one of these was a Defense Acquisition Pilot Program (DAPP) program and nine were acquired using a TPP or TSPR contract. Of the 46 MDAPs in the database that passed MS B during a boom climate, only the three early 1980s TPP programs and Titan IV had a PAUC growth of at least 50%. It is reasonable to conclude that high cost growth is more common for these TPP and TSPR acquisitions.

Relaxation of Regulations and Statutes

The Congress explored the consequences of relaxing acquisition regulations and statutes through the Defense Enterprise Program (DEP) and subsequently, the DAPP. Although Other Transaction Authority (OTA) was enacted for somewhat different reasons, its use of in the development phase of an MDAP acquisition also permitted relaxation of most acquisition regulations and statutes.

Table 5 shows for each DEP, DAPP, or OTA MDAP the fiscal year in which the program passed MS B and its PAUC growth. These will be discussed in the order stated.

Table 5. Fiscal Year in Which the Program Passed MS B and Quantity Adjusted PAUC Growth for DEP Programs, DAPP Programs, and Programs Acquired With OTA

MDAPs by Category	MS B (FY)	% PAUC Growth
DEP Programs[a]		
TOW II	1984	13%
Trident D-5 Missile	1984	15%
SSN-21†	1985	8%
Mobile Subscriber Equipment†	1986	1%
Army Tactical Missile System†	1986	13%
Medium Launch Vehicle	1990	n/a
SRAM II	1987	cancelled
T45-TS† ‡	1984	70%
C-17‡	1985	57%
Titan IV	1985	212%
DAPP Programs[b]		
JDAM	1995	12%
JPATS	1995	42%
C-130J§	1996	83%
OTA Programs		
UCAV	*	n/a
RQ-3 DarkStar	*	n/a
Arsenal Ship	*	n/a
Evolved Expendable Launch Vehicle (EELV)#	1998	251%
Global Hawk	2001	n/a^
Future Combat Systems (FCS)	2003	cancelled
DDG-1000	2006	Truncated

[a] See footnote 19 for references that identify the DEP MDAPs.

[b] Federal Acquisition Streamlining Act of 1994, Pub. L. No. 103-355, Section 5064.

† Milestone funding authorized. National Defense Authorization Act (NDAA) for Fiscal Years 1988 and 1989, Pub. L. No. 100-180, Section 106, 101 Stat. 1019 (1987).

‡ TPP program.

§ Only DoD regulations waived.

* Did not pass MS B.

\# TSPR program.

^ The database does not include a cost growth estimate for Global Hawk that is quantity-adjusted and in common base-year dollars. It is clear from the SARs, however, that cost growth for Global Hawk was high.

DEP was established by the FY 1987 National Defense Authorization Act (NDAA).[17] DEP programs were exempt from DoD regulations other than those specified by the Service Acquisition Executives (SAEs). We do not know what DoD regulations the SAEs elected to retain. In addition, DEP programs could be granted Milestone Authorization, that is, for authorization of funding through the end of their then current acquisition phase.[18] Ten MDAPs were nominated by DoD as DEP programs; the Congress accepted all of these and granted Milestone Authorization to four of them.[19] No other programs were added to the DEP after this initial group.

The MS B baselines for DEP were established before the initiatives were built into the programs. Consequently, PAUC growth (adjusted for quantity change) is equal to the growth in the acquisition cost of the programs (in program base year dollars). Those who believe that acquisition regulations are a major contributor to both high weapon system cost and PAUC growth would expect the DEP programs should have below average PAUC growth. A skeptic who believes that the regulations waived served a good purpose would expect above average PAUC growth.[20]

Each of the DEP programs for which we have a PAUC growth estimate passed MS B during the Reagan boom climate. The average PAUC growth of the DEP programs was 48.6%; the average for all of the programs in the database that passed during the Reagan boom was less than half that, 20%. If the programs acquired with TPP are dropped, the average PAUC growth for the DEP programs was 42% and that for the Reagan boom climate programs, 12%. These data do not make a case for DEP. DoD found that the DEP programs "were more trouble than they were worth … and … allowed it [DEP] to lapse by 1990" (Fox, 2011, p. 159).

DAPP was established in the NDAA for FY 1991.[21] From the DoD's perspective, the key difference between the DEP and the DAPP probably was that the latter permitted the Secretary of Defense to waive not only DoD regulations but also acquisition statutes and regulations. The Federal Acquisition Streamlining Act of 1994 authorized five programs to participate in the DAPP.[22] Of these, four were MDAPs, but two of these did not continue as

---

[17] NDAA for FY 1987, Pub. L. No. 99-661, § 905 (1986).
[18] Some accounts of the DEP state that its establishment was a recommendation of the Packard Commission. This is not accurate, in that the Packard Commission reports did not specifically include such a recommendation. The Packard Commission, however, did recommend the use of Milestone Authorizations. See President's Blue Ribbon Commission (1986, pp. xxiv–xxvii, xix).
[19] The requests were made in a letter from Deputy Secretary of Defense William H. Taft, IV to the Honorable Les Aspin, Chairman of the Committee on Armed Services of the House of Representatives, March 30, 1987 (following Radice, 1992). The Army Mobile Subscriber Equipment, Army Tactical Missile System, Navy's Trident II Missile, and the Navy's T-45 TS were granted Milestone Authorization. See the NDAA for FY 1988 and 1989, Pub. L. No. 100-180, § 106, 101 Stat. 1019 (1987).
[20] The one exception to this statement is Medium Launch Vehicle, for which we do not have a PAUC growth estimate.
[21] NDAA for FY 1991, Pub. L. No. 101-510, § 809, 104 Stat. 1594 (1990).
[22] Federal Acquisition Streamlining Act of 1994, Pub. L. No. 103-355, § 5064.

MDAPs after 1994.[23] Another MDAP was included in the DAPP in 1995 (Reig, 2000, Appendix A, p. 43).[24] Just when the DAPP ended is not clear, but no indication was found that any additional programs were added after 1995.

There does not seem to be anything to be made of the data for three DAPP MDAPs. PAUC growth for JDAM is notably low for a program that passed MS B during a bust climate, but PAUC growth figures for the JPATS and C-130J programs are somewhat high even for programs that passed during a bust climate. The average PAUC growth for the three DAPP programs was 45.7%. The average for completed programs from the AR period is 31%.

An Other Transaction (OT) is

> a special vehicle used by federal agencies for obtaining or advancing research and development (R&D) or prototypes. An OT is not a contract, grant, or cooperative agreement. … Only those agencies that have been provided OT authority may engage in other transactions. (Halchin, 2011, Summary)

MDAPs whose development was funded under OTA are included in this subsection because some procurement statutes do not apply to such arrangements and OTAs typically are not required to comply with DoD procurement regulations. The Defense Advanced Research Projects Agency (DARPA) was granted OTA in 1989.[25] The DoD as a whole received OTA in 1994.[26]

According to a RAND study, the DoD entered into 72 OTs during 1994–1998. Nearly 60% of these OTs had total funding of less than $10 million, and only seven had funding greater than $100 million. The study entailed a detailed assessment of 21 of the 72 OTs. Based on this assessment, Smith, Drezner, and Lachow (2002, pp. iii, 7, 31) offered a favorable assessment of OTAs; they were found to have limited risks and to provide broad benefits.

Table 5 includes only the seven OTAs with funding greater than $100 million; these programs were MDAPs or, perhaps with one exception, intended to become MDAPs. In contrast to the OTs that Smith et al. (2002) judged to work well, these seven projects had little or no commercial potential and to a substantial extent used technology developed by the companies involved under previous DoD contracts. They do not make a good surface case for OTAs for projects with those characteristics—two high cost growth programs, one cancellation, one truncation, and three programs that never went to MS B.

Concluding Comment

There is no difficulty in placing the TPP and FP development contracts within the context of the DSARC process. Packard reversed McNamara on the use of TPP and FP

[23] The Fire Support Combined Arms Tactical Trainer seems to have been an ACAT II or ACAT III program. The Commercial Derivative Engine and the Commercial Derivative Aircraft appear to have been part of the 1994 competition of the C-17 and commercial derivative aircraft and probably did not continue after that competition was concluded.
[24] Hanks et al. (2005, p. 25, note 41) indicates that only regulations were waived for the C 130J.
[25] NDAA for FY 1990 and 1991, Pub. L. No. 101-189, § 251, 103 Stat. 1352 (1989).
[26] Federal Acquisition Streamlining Act of 1994, Pub. L. No. 103-355, 108 Stat. 3243.

development contracts, and that accounted for a substantial portion of the lower average PAUC growth of the 1970s. Packard's policy on TPP and FP development contracts was in turn reversed in the early 1980s, the Reagan boom climate. As a result, average PAUC growth during this period was substantially higher than it was for the subsequent post-9/11 boom. TSPR is more complicated. The high PAUC growth associated with the TSPR contracts seems to reflect some combination of flawed implementation and inherent flaws in the TSPR concept.

At the broad brush level, there also is an obvious connection between the comparative success the DSARC/DAB process had in maintaining PAUC growth below its level in the McNamara-Clifford period and the lack of success of the DEP and DAPP experiments and the use of OTA on large programs. From the start, the DSARC process was the actualization of regulations embodied in DoDD 5000.1, DoDI 5000.2, and subsidiary regulations. These became more extensive over time, and some were required or augmented by statutes. DEP, DAPP, and OTA all relaxed some regulations and, in the case of DAPP and OTA, some statutory restrictions. This did not result in lower PAUC growth, which seems to indicate that the regulations and statutes relaxed play a useful role in this respect. Although perhaps accurate, this argument is facile. To be convincing, it would be necessary to go much further than this paper has into just which regulations and statutes were relaxed and how those relaxations were connected to cost growth.

References

Arnold, S. A., Horowitz, S. A., Patel, P. R., & Davis, G. A. (2013). *Lessons from literature on the effects of contract type on satellite acquisition* (IDA Document NS D-4859). Alexandria, VA: Institute for Defense Analyses.

Christensen, D. S., Searle, D. A., & Vickery, C. (1999, Summer). The impact of the Packard Commission's recommendations on reducing cost overruns on defense acquisition contracts. *Acquisition Review Quarterly*, 251–262.

Defense Science Board and the Air Force Scientific Advisory Board Joint Task Force. (2003). *Acquisition of national security space programs*. Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics.

Dews, E., Smith, G. K., Barbour, A., Harris, E., & Hesse, M. (1979). *Acquisition policy effectiveness: Department of Defense experience in the 1970s* (R-2516-DR&E). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/reports/R2516.html

Drezner, J. A., Jarvaise, J. M., Hess, R. W., Norton, D. M., & Hough, P. G. (1993). *An analysis of weapon system cost growth* (MR-291-AF). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/content/dam/rand/pubs/monograph_reports/2006/MR291.pdf

Fox, J. R. (2011). *Defense acquisition reform, 1969 to 2009: An elusive goal*. Washington, DC: U.S. Army Center of Military History.

Glennan, T. K., Jr. (1965). *Policies for military research and development* (RAND Paper P-3253). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/papers/P3253.html

GAO. (2006). *Space acquisitions: DOD needs to take more action to address unrealistic initial cost estimates of space systems* (GAO-07-96). Washington, DC: Author. Retrieved from https://www.gao.gov/new.items/d0796.pdf

GAO. (2009). *Global positioning system: Significant challenges in sustaining and upgrading widely used capabilities* (GAO-09-325). Washington, DC: Author. Retrieved from https://www.gao.gov/new.items/d09325.pdf

GAO. (2010). *Defense acquisitions: DOD needs to reassess Joint Cruise Missile costs before starting new production phase* (GAO-11-112). Washington, DC: Author. Retrieved from https://www.gao.gov/assets/320/311277.pdf

Halchin, L. E. (2011). *Other Transaction (OT) Authority* (CRS Report for Congress RL34760). Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/misc/RL34760.pdf

Hanks, C. H., Axelband, E. I., Lindsay, S., Malik, M. R., & Steele, B. D. (2005). *Reexamining military acquisition reform: Are we there yet?* (MG-291). Santa Monica CA: RAND. Retrieved from https://www.rand.org/pubs/monographs/MG291.html

Hawkins, T. G., & Cuskey, J. R. (2011). Do the baby and the bathwater deserve the same fate? An exploratory study of collaborative pricing in the U.S. Department of Defense. *Journal of Public Procurement, 11*(2), 240–274. doi:10.1108/JOPP-11-02-2011-B004

Kim, Y., Axelband, E., Doll, A., Eisman, M., Hura, M., Keating, E. G., … Shelton, W. (2015). *Acquisition of space systems, volume 7: Past problems and future challenges* (MG-1171/7-OSD). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/monographs/MG1171z7.html

Lorell, M. A., Graser, J. C., & Cook, C. R. (2005). *Price-based acquisition: Issues and challenges for Defense Department procurement of weapon sy*stems (MG-337-AF). Santa Monica CA: RAND. Retrieved from https://www.rand.org/pubs/monographs/MG337.html

Lorell, M. A., Leonard, R. S., & Doll, A. (2015). *Extreme cost growth: Themes from six U.S. Air Force major defense acquisition programs* (RR-630-AF). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/research_reports/RR630.html

McNicol, D. L. (2004). *Cost growth in major weapon procurement programs* (2nd ed.). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017a). *Post-Milestone B funding climate and cost growth in major defense acquisition programs* (IDA Paper P-8091). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (2017b). Post-Milestone B funding climate and cost growth in major defense acquisition programs. *Proceedings of the 14th Annual Acquisition Research Symposium* (SYM-AM-17-34; pp. 86–97). Retrieved from http://acqnotes.com/wp-content/uploads/2017/08/SYM-AM-17-034_Wednesday-Vol-1_5-1-2017.pdf

McNicol, D. L. (forthcoming). *Acquisition policy, funding climate, and cost growth of major defense acquisition programs* (IDA Report R-8396). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L. (forthcoming). Further evidence on program duration and unit cost growth. *Proceedings of the 15th Annual Acquisition Research Symposium*, Naval Postgraduate School, Monterey, CA.

McNicol, D. L., & Tate, D. M. (2016). Further evidence on the effect of acquisition process and policy on cost growth. *Proceedings of the 13th Annual Acquisition Research Symposium* (SYM-AM-16-019; pp. 135–146). Retrieved from http://www.acquisitionresearch.org/files/FY2016/SYM-AM-16-019.pdf

McNicol, D. L., Tate, D. M., Burns, S. K., & Wu, L. (2016). *Further evidence on the effect of acquisition policy and process on cost growth of major defense acquisition programs (revised)* (IDA Paper P-5330 (R)). Alexandria, VA: Institute for Defense Analyses.

McNicol, D. L., & Wu, L. (2014). *Evidence on the effect of DoD acquisition policy and process on cost growth of major defense acquisition programs* (IDA Paper P-5126). Alexandria, VA: Institute for Defense Analyses.

O'Neil, W. D. (2011, July). Cost growth in major defense acquisition: Is there a problem? Is there a solution? *Acquisition Research Journal,* 278–293. Retrieved from http://dau.dodlive.mil/files/2014/11/Oneil_ARJ59.pdf

O'Neil, W. D., & Porter, G. H. (2011). *What to Buy? The Role of Director of Defense Research and Engineering (DDR&E): Lessons from the 1970s* (IDA Paper P-4675). Alexandria, VA: Institute for Defense Analyses.

Poole, W. S. (2005). Acquisition in the Department of Defense, 1959–1968: The McNamara legacy. In S. Brown (Ed.), *Providing the means of war: Historical perspectives on defense acquisition, 1945–2000* (pp. 79–96). Washington, DC: United States Army Center of Military History and Industrial College of the Armed Forces. Retrieved from https://history.army.mil/html/books/070/70-87-1/CMH_Pub_70-87-1.pdf

President's Blue Ribbon Commission on Defense Management. (1986). *A quest for excellence: Final report to the president.* Washington, DC: Author.

Quander, A. Y., & Woppert, J. H. (2010). *Analysis of Alpha contracting from three perspectives: Government contracting, the government program office, and industry* (Master's thesis, Naval Postgraduate School). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a531488.pdf

Radice, M. R. (1992). *The Defense Enterprise Program: A managerial assessment* (Master's thesis, Naval Postgraduate School). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a255174.pdf

Rapka, R. J., Heyward, E., Boyle, B., & Godin, S. (2006, January–March). A successful alpha contracting experience. *Army AL&*T, 34–37.

Reig, R. W. (2000, Winter). Baselining acquisition reform. *Acquisition Review Quarterly.* Retrieved from https://www.dau.mil/library/arj/ARJ/arq2000/reig.pdf

Smith, G., Drezner, J., & Lachow, I. (2002). *Assessing the use of "Other Transactions" authority for prototype projects* (DB-375-OSD). Santa Monica, CA: RAND. Retrieved from https://www.rand.org/pubs/documented_briefings/DB375.html

Temple, L. P., III. (2013). Quality on the horizon. *In Implosion: Lessons from national security, high reliability spacecraft, electronics, and the forces which changed them.* Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118487105.ch11

Tyson, K. W., Om, N. I., Gogerty, D. C., & Nelson, J. R. (1992). *The effects of management initiatives on the costs and schedules of defense acquisition programs, Vol. I: main report* (IDA Paper P-2722). Alexandria, VA: Institute for Defense Analyses.

Under Secretary of Defense (Acquisition, Technology, and Logistics). (2016). *Performance of the Defense Acquisition System: 2016 annual report.* Washington, DC:Author.

Acknowledgments

David M. Tate (the dean of the reviewers and co-author of one of the papers in the series), David A. Sparrow, Daniel Cuda, Prashant Patel, and Philip Lurie, all of the Institute for Defense Analyses (IDA), provided insightful comments on successive drafts. I am similarly grateful to Gregory Davis for his support as critic and task leader for the past two years. Brian Gladstone and Sarah Burns (co-author of one of the papers) of IDA, and D. Mark Husband of the Defense Acquisition University provided valuable comments on the first three papers of the series. Linda Wu, formerly of IDA, managed data acquisition and the database. More recently, J. M. Breuer provided helpful assistance on the statistical work. The research presented in this paper was sponsored by the Director, Office of Performance Assessments and Root Cause Analyses.

Panel 8. Innovations in Services Contracting

Wednesday, May 9, 2018	
11:15 a.m. – 12:45 p.m.	**Chair: Brigadier General Cameron Holt, USAF,** Commander, Air Force Installation Contracting Agency ***Category Management of Services: A Methodology for Strategically Clustering DoD Installations*** Lt Col Karen A. F. Landale, USAF, Air Force Installation Contracting Agency Aruna Apte, Naval Postgraduate School Capt Austin Clark, USAF, Naval Postgraduate School Capt Corey Arruda, USAF, Naval Postgraduate School ***A Review of Alternative Methods to Inventory Contracted Services in the Department of Defense*** Nancy Y. Moore, RAND Corporation Molly Dunigan, RAND Corporation Frank Camm, RAND Corporation Samantha Cherney, RAND Corporation Clifford A. Grammich, RAND Corporation Judith D. Mele, RAND Corporation Evan D. Peet, RAND Corporation Anita Szafran, RAND Corporation ***MARAD's Maritime Security Program: Exemplary Innovation in Acquisition Policy?*** Marvin Phaup, George Washington University

Brigadier General Cameron G. Holt, USAF—is the Commander, Air Force Installation Contracting Agency, Air Force Installation and Mission Support Center, Air Force Materiel Command, Wright-Patterson Air Force Base, OH. He leads an over 700-personnel agency with a total contract portfolio of $33 billion. In this capacity, he directs enterprise-wide installation strategic sourcing efforts for the Air Force and oversees $3.9 billion in annual obligations in mission and installation requirements. His contracting authority extends worldwide across AFICA in support of eight major commands and their 77 units. Additionally, he is designated as the Commander of a Joint Theater Support Contracting Command upon activation. General Holt also directs the contract execution in support of the Defense Technical Information Center, Air Force Medical Support Agency, Air Force Medical Operations Agency, and Air Force Civil Engineer Center.

Prior to this assignment, he served as the Director of Staff, Air Force Installation Contracting Agency, Air Force Installation and Mission Support Center, Air Force Materiel Command, Wright-Patterson AFB, OH. General Holt oversaw a team of 87 officers, enlisted and civilian personnel who coordinated the agency workflow, policies, procedures, resources, and directives for 77 units worldwide with over 800 personnel. Additionally, he led the team responsible for finding strategic sourcing opportunities within the Air Force.

General Holt received his commission through the Reserve Officer Training Corps at the University of Georgia in 1990. He has experience in the full spectrum of acquisition and contract management across four major commands, Headquarters U.S. Air Force, U.S. Air Forces Central Command, and the Joint Staff. General Holt is a joint qualified officer with multiple deployments in support of Operation Enduring Freedom.

Category Management of Services: A Methodology for Strategically Clustering DoD Installations

Lt Col Karen A. F. Landale, USAF—is the Commander of the 773d Enterprise Sourcing Squadron, Joint Base San Antonio–Lackland, San Antonio, TX. At the time of this research, she was an Assistant Professor at the Naval Postgraduate School. Landale has a BS in International Business from the University of Tampa, an MBA in Strategic Purchasing from the Naval Postgraduate School, and a PhD in Marketing from the Kenan-Flagler Business School at the University of North Carolina-Chapel Hill.

Aruna Apte—is a tenured Associate Professor in the Graduate School of Business and Public Policy (GSBPP), Naval Postgraduate School (NPS). Dr. Apte received her PhD in Operations Research from Southern Methodist University. Her accomplishments include over 50 research articles, including 26 peer-reviewed journal articles (12 out of these are in HADR), and one patent; the Excellence in Research award from the GSBPP, NPS, for her defense-focused research; selection to present her research on NPS day; briefing several admirals, diplomats, and senior government officials in HADR; the Outstanding Teaching award at the Cox School of Business, SMU; the Liskin award for outstanding teaching at GSBPP; and the Hamming award for excellence in teaching at NPS.

Capt Austin Clark, USAF—is an Air Force Contracting Officer. He received a Bachelor of Science degree in Construction Management and a commission through the Reserve Officer Training Corps, East Carolina University, NC. Prior to NPS, Capt Clark served three years in operational contracting at Joint Base Pearl Harbor-Hickam, HI. He is currently assigned to the National Reconnaissance Office, Chantilly, VA.

Capt Corey Arruda, USAF—is an Air Force Contracting Officer. He studied management and information systems at Park University, where he received a Bachelor of Science degree. He was commissioned through the Officer Training School at Maxwell Air Force Base, AL. Prior to NPS, Capt Arruda served three years in operational contracting at Joint Base Charleston, SC. He is currently assigned to the Military Satellite Communications Directorate at the Space Missile and Systems Center, Los Angeles Air Force Base.

Abstract

In an increasingly budget-constrained environment, the Department of Defense (DoD) must maximize the value of fiscal resources obligated to service contracts. According to the Government Accountability Office (GAO) report *Strategic Sourcing*, published in 2013, over half of procurement spending between 2008 and 2013 was obligated to service contracts. Therefore, this research focused on identifying rate, process, and demand savings for common, recurring DoD service requirements. We developed a methodology to standardize analyses of service requirements to identify relevant cost drivers. Furthermore, a clustering continuum was created to organize services based on proximity between the customer and the supplier base. Utilizing commercial business mapping software, we analyzed the cost driver data, produced visualizations, and illustrated strategic opportunities for category management initiatives. Requirements for Integrated Solid Waste Management (ISWM) within the southern California area were evaluated using the software and methodology to demonstrate practical application.

Introduction

Background

U.S. economic spending has dramatically evolved over recent decades as the country has moved from a goods-consuming society to a service-consuming society. This cultural movement has led to a substantial increase in the demand for services over tangible goods. In 1968, economist Victor R. Fuchs published findings that more than half of the employed population in the United States was working in the services sector and thus was "not involved in the production of food, clothing, houses, automobiles, or other tangible goods" (Church, 2014). The U.S. economy, he argued, had become a "service economy" (Church, 2014).

Fifty years later, Fuchs' analysis stands the test of time, as services continue to comprise a significant portion of consumer spending. In early fiscal year (FY) 2017, U.S. citizens consumed nearly $9 trillion in services, up nearly $2.5 trillion from FY 2007. In comparison, spending on goods increased approximately $1 trillion, to a total of $4.2 trillion for FY 2017. This recent data suggests that the trend toward spending on services is expected to remain the same or, more than likely, increase in the foreseeable future.

Procurement in the Department of Defense (DoD) has mirrored consumer spending behavior—agencies have reported a trend toward service-related requirements. In February 2017, the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]) reported to congressional committees that the DoD had obligated over $149 billion to service-related defense contracts in FY 2016 (see Figure 1). Accounting for over half of defense spending, service-related contracts deliver an exhaustive list of critical defense-sustaining capabilities, such as maintaining installations, information technology (IT) security services, and medical services. The DoD has consistently spent more than three times the fiscal resources on services than on supplies and equipment (S&E), such as investments in aircraft, ships, submarines, and land vehicles (OUSD[AT&L], 2017).

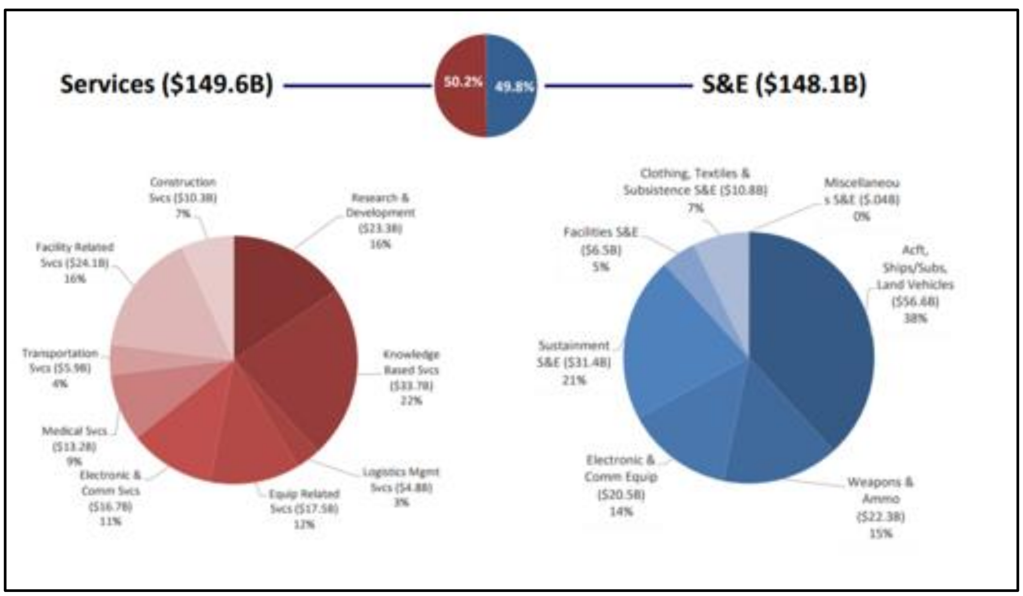


Figure 1. FY 2016 Spend—DoD as Contracting Department
(OUSD[AT&L], 2017)

Historically, the DoD has struggled with acquiring services due to the inherently complex nature of services, relative to the seemingly straightforward procurement of commodities. This complexity, paired with the DoD's growing portfolio of services, has gained the attention of multiple government watchdog agencies, including the Government Accountability Office (GAO). In 2017, the GAO released its biannual high-risk report to Congress, which "identifies government operations with greater vulnerabilities to fraud, waste, abuse, and mismanagement or the need for transformation to address economy, efficiency, or effectiveness challenges" (p. 2). In the report, the GAO noted that "Improving DOD's Acquisition of Services" is a recurring high-risk category that should be addressed immediately by DoD officials.

The DoD continues to take action to improve how it manages services acquisitions, with demonstrable progress. In January 2016, the DoD issued a new instruction for service acquisitions that provides a management structure for acquiring services and identifies the roles and responsibilities of key leadership positions; however, the DoD still lacks an action plan that will enable it to assess progress toward achieving its goals, and efforts to identify goals and associated metrics are still in the early stages of development (GAO, 2017, p. 491).

Numerous initiatives, such as Better Buying Power, have emerged to educate DoD stakeholders on best practices to improve tradecraft in services. These initiatives have inspired grassroots efforts that have led to a few attempts at enterprise-sourced, cost-saving solutions. For example, the Building Maintenance & Operations (BMO) service contracts of the General Service Administration (GSA) are an attempt at a regional-based, enterprise-sourced contract solution. However, those familiar with category management principles would argue that "big contracts," while offering process-related savings, may not be the optimal, or comprehensive, solutions for enterprise-wide services.

To promote strategic cost saving initiatives in the acquisition of services, the Air Force Installation Contracting Agency (AFICA) partnered with the Graduate School of Business and Public Policy (GSPBB) at the Naval Postgraduate School (NPS) to identify a methodology that optimally groups DoD installations for enterprise-sourced solutions.

Research Objective

This research develops a methodology that optimally clusters DoD installations based on known cost drivers of common, enterprise-wide installation services. A service contract is defined in the Federal Acquisition Regulation (FAR) as "a contract that directly engages the time and effort of a contractor whose primary purpose is to perform an identifiable task rather than to furnish an end item of supply" (FAR 37.101). Our method targets the Army, Navy, Air Force, and Marine Corps requiring activities within the continental United States (CONUS) engaged in contracting for common, recurring, installation-level service requirements. For this research, we use Integrated Solid Waste Management (ISWM), which is essentially garbage collection, as our example recurring service. ISWM services are acquired by most CONUS DoD installations and consist of identifiable tasks that are similar in nature.

While we specifically focus on ISWM services, our method is versatile and can be adapted to many other service requirements. Strategically clustering DoD installations that acquire like services allows the DoD to manage its portfolio in a way that yields the greatest rate, process, and demand savings achievable. As such, we aim to answer the following research question: Are there potential cost savings (rate, process, demand) through strategically clustering common DoD service contracts?

Literature Review

Category Management—Private Sector

Category management is the latest evolution of the private sector's attempts to control costs in order to achieve competitive advantages. Its original form was strategic purchasing, which achieved rate and process savings by aggregating purchases (leveraging spend) of similar requirements. Following strategic purchasing, the next cost control method was strategic sourcing, which achieved rate and process reductions using market-focused techniques like partnering with suppliers in the research and development stage of new product development, monitoring and measuring supplier performance, implementing supplier relationship management techniques, etc. For all its advancement, strategic sourcing remained focused on acquisition-related solutions. Category management added an additional layer of analysis to the concepts included in strategic sourcing by

> incorporate[ing] many familiar aspects of business improvement processes and change management … it is not an approach that is confined to purchasing but typically requires the active participation of and engagement with stakeholders, functions and individuals across the business to make it successful. (O'Brien, 2015, p. 5)

Category management is a functionally-led (i.e., end-user led) process, whereas strategic purchasing and strategic sourcing tend to be acquisition- or purchasing-led processes.

Historically, many organizations viewed their purchasing function as an operational entity responsible solely for handling routine transactions. Peter Kraljic (1983) asserted that organizations' top management must change this viewpoint and recognize the strategic value of its purchasing function. His philosophy was based on the practice of strategic purchasing. Kraljic (1983) asserted,

> A company's need for a supply strategy depends on two factors: (1) the strategic importance of purchasing in terms of the value added by product line, the percentage of raw materials in total costs and their impact on profitability, and so on; and (2) the complexity of the supply market gauged by supply scarcity, pace of technology and/or materials substitution, entry barriers, logistics cost or complexity, and monopoly or oligopoly conditions. (p. 110)

He used these two factors (importance or impact on the Y-axis, and complexity of market or supply risk on the X-axis) to create a matrix to categorize an organization's purchases (see Figure 2).

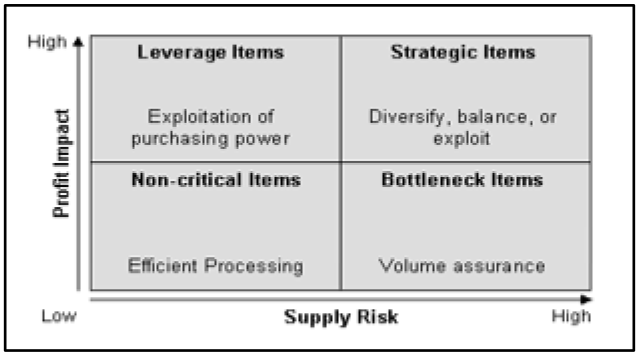


Figure 2. Kraljic Portfolio Matrix (KPM)
(Holt, 2017)

The matrix categorized supply items into Non-Critical, Bottleneck, Leverage, and Strategic quadrants. The Leverage quadrant is the most relevant to this study because it is composed of items that have a high importance to the organization and minimal supply risk. High importance to the organization and minimal supply risk are characteristics that describe most of the common, recurring DoD service requirements. Kraljic (1983) claimed,

> On items where the company plays a dominant market role and suppliers' strength is rated medium or low, a reasonably aggressive strategy ("exploit") is indicated. Because the supply risk is slight, the company has a better chance of achieving a positive profit contribution through favorable pricing and contract agreements. (pp. 113–114)

This statement suggests that supply items in the Leverage quadrant of the Kraljic Matrix provide the best opportunity to aggregate purchases by clustering common, recurring service requirements to "exploit" the enormous purchasing power of the DoD over its many, less powerful suppliers in order to achieve savings. Note that while the terminology is aggressive, the DoD is not in the business of throwing its weight around in order to put companies asunder. However, budget constraints demand the DoD sharpen its pencil when it comes to leveraging its strengths—one of which is its large, enduring buying power—to achieve reasonable savings.

To develop and implement supply management strategies as large as category management, an organization must have the proper governance structure in place, starting at the top (i.e., the strategic level). Category management decisions and policy-making need to be centralized in order to leverage the organization's buying power. While there is a need for centralized control, the goal is to simplify decentralized execution for lower level units (i.e., those closest to the requirement owner). In an article written in 2005, David L. Reese and Douglas W. Pohlman stated that

> today's commercial procurement community is leaning heavily toward the organizational concept of centralized procurement. Although the large and medium corporations around the globe that are centralizing their purchasing efforts use several different organizational constructs, the overarching objective is typically the same. To the maximum extent possible, the entire organization should be corporately leveraging its purchasing volume and customer and supplier relationships through strategic planning and execution.

> Indeed, companies that are striving to ensure supply of critical goods and services are finding a decentralized strategy that promotes fragmented processes is fundamentally detrimental to their goal. (p. 6)

In short, category strategy and planning must be done centrally, while execution remains tactical and decentralized.

Category Management—Public Sector

In 2005, the Office of Management and Budget (OMB), via memorandum, officially charged all federal government agencies to begin implementing strategic sourcing (Johnson, 2005). The OMB issued additional guidance in 2012, establishing the Strategic Sourcing Leadership Council (SSLC) and placing additional responsibilities upon the GSA for helping to implement federal-wide strategic sourcing (Zients, 2012). While these efforts led to several small strategic purchasing wins, the federal government, and the DoD specifically, still lacked a comprehensive, coordinated approach to managing its spend.

The most recent guidance, issued by the OMB in 2014, declared that category management was common in industry practices and would be the future approach to the federal government's acquisitions of goods and services (OMB, 2015). Using category management concepts and processes, spend associated with commonly purchased goods and services is owned and managed by assigned category managers. Category managers are charged with managing enterprise-level spend in a way that aligns to the way industry produces and delivers the goods and services that fall into their category in order to achieve rate, process, and demand savings. Category managers are also responsible for organizing multi-functional teams to research and understand how the DoD acquires and uses the goods and services that fall into their category, how industry creates and delivers those goods and services, and any best practices implemented by near-peer organizations. The 2014 memo appointed the GSA as the lead organization for implementing government-wide category management (Rung, 2014).

Finally, in 2015, the OMB published the *Government-Wide Category Management Guidance Document*, which provided agencies with direction for successful implementation of category management processes and established procedures for federal-wide category management operations. Importantly, the OMB established a logical grouping of goods and services purchased throughout the federal government. This logical grouping, known as the OMB taxonomy, aligns product service codes (PSCs) into Level II categories, which roll up into 10 federal-wide Level I categories. For example, ISWM is a service that falls under Facility Related Services (4.4), which falls under Facilities and Construction (4.0) in the OMB taxonomy. See Figure 3 for a more detailed view of the Level I and Level II categories created by the OMB.

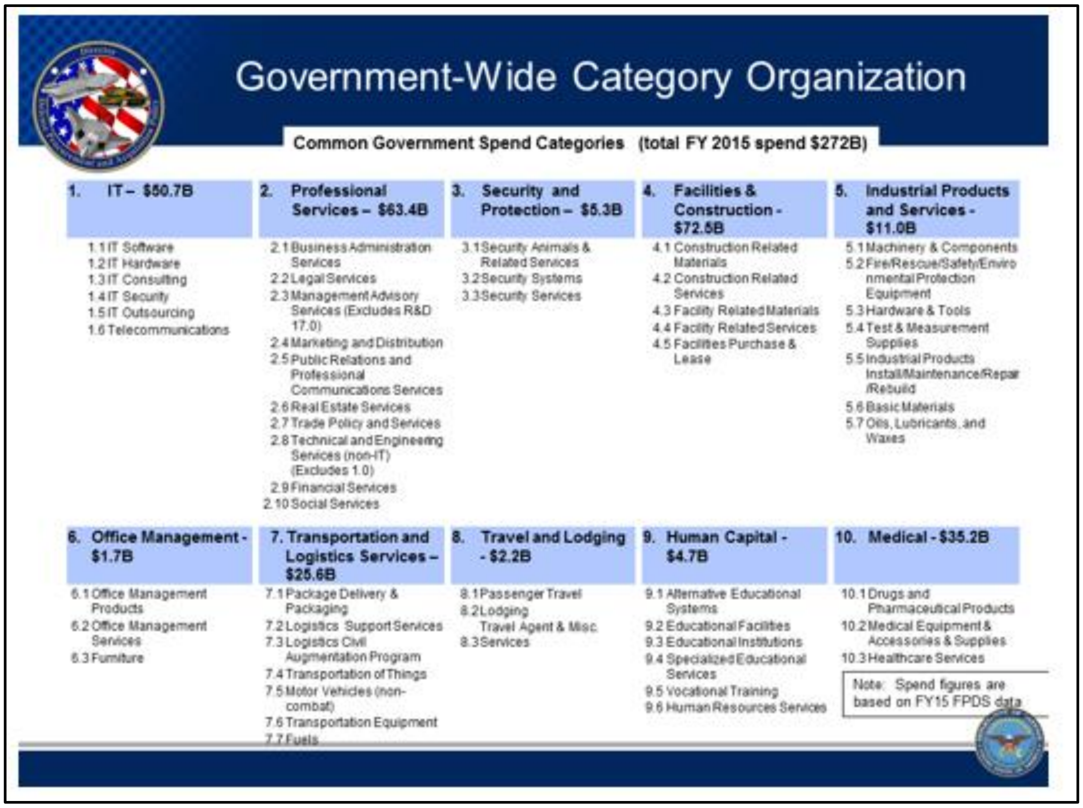


Figure 3. Government-Wide Category Organization.
(Defense Procurement and Acquisition Policy [DPAP], n.d.)

Critique of the DoD's Management of Services

In 2013, the GAO issued a report that critiqued many strategic sourcing initiatives and provided examples of commercial best practices in acquisition. For example, many companies conduct spend analyses to understand their supply chain portfolios. A spend analysis involves identifying the number of suppliers, the number of contracts, prices paid, etc. to target inefficiencies, such as paying different rates for similar services and suppliers, or not consolidating purchases across the company to achieve lower prices. This knowledge allows companies to leverage their buying power, reduce costs, and better manage their suppliers. Following a spend analysis, many companies make structural changes (with top leadership support), to establish commodity managers who are responsible for purchasing services within a category, thus leveraging their buying power to achieve substantial savings (GAO, 2013). By 2013, however, the DoD had still not performed a comprehensive spend analysis in order to highlight inefficiencies and target where to commit their limited strategic sourcing resources.

The GAO (2013) made the following observations and recommendations to overcome key challenges and improve (at the time) strategic sourcing efforts:

1. Agency officials noted that they have been reluctant to strategically source services (as opposed to goods) for a variety of reasons, such as difficulty in standardizing requirements or a decision to focus on less complex commodities that can demonstrate success.
2. For less complex services, such as housekeeping and telecommunications, agencies could consolidate purchases to leverage

buying power. Standardizing requirements could also help drive down costs.

3. For complex services, such as professional services, … agencies could apply company tactics to understand cost drivers and prequalify suppliers. (GAO, 2013, pp. 19–20, 25)

In this research, we focus on the observations and recommendations made by the GAO in 2013, specifically that agencies are reluctant to strategically source services and that opportunities to achieve savings in service requirements exist, even when the complexity of the service varies. To tackle this problem, we first identified all DoD installations that are engaged in contracting for common, recurring service requirements. Second, we analyzed existing market intelligence on several common DoD service requirements to identify an optimal service to demonstrate the potential opportunities associated with clustering DoD installations. Finally, we created a model using ISWM as an example service to demonstrate the concept of clustering to achieve category management goals.

Methodology

The Clustering Continuum

We sought to develop an elementary framework for how PSCs (referred to as services in this report) could be classified and organized to align with the category management framework. From our research, we began to understand that proximity of suppliers to their service location is often a limiting factor in developing clusters. For example, it is reasonable to assume that for a service like ISWM, suppliers would be opposed to taking on long-haul regional or interstate ventures because of the high costs of fuel, dumping fees, and maintenance on their truck fleets. Suppliers seeking ISWM contracts typically favor a short-range business model; that is, the service has proximity dependence between its supplier base and the location at which the service is performed. Granted, there are large businesses capable of covering large regions; however, even those businesses need a local office and local employees to deliver the service.

Conversely, we believe some services can be classified as exhibiting characteristics of proximity independence, a polar opposite to proximity-dependent services. Proximity-independent services are those groups of services that have limited correlation between supplier location and place of performance. For example, information technology (IT) services encompass a wide range of services such as day-to-day protection of base network security, network troubleshooting, and over-the-air software updates; many of these services are conducted in remote, centralized locations throughout the United States, or even worldwide. Figure 4 identifies a few services that we have organized on the clustering continuum. The continuum allows us to logically organize services to help determine appropriate cluster size, which is discussed later in the chapter.

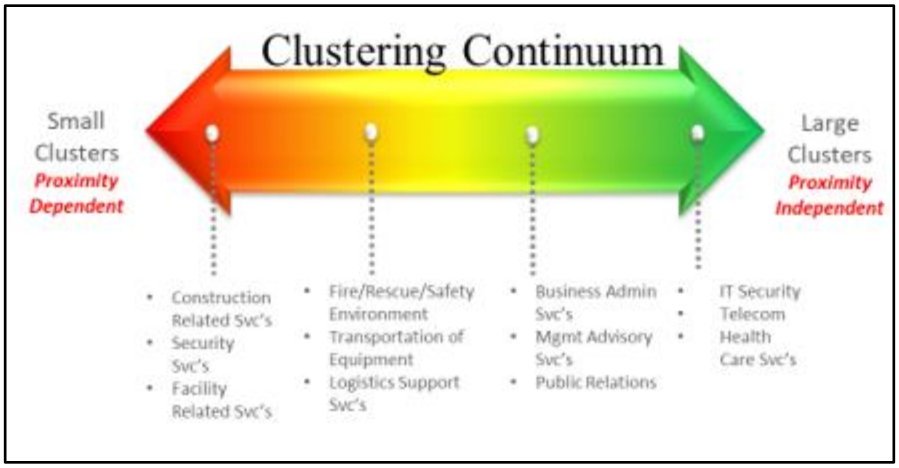


Figure 4. Clustering Continuum

Commercial Business Mapping Software

To facilitate the visual representation of clustering DoD installations and to employ the Category Management goal of mirroring commercial-sector best practices (or, as the case may be, developing new public-sector best-in-class practices), we explored numerous commercially available software options used for business analysis functions. One of the best contenders was the Maptitude Geographic Information Software (GIS). Maptitude GIS is robust, easy-to-use, professional business mapping software that businesses use for in-depth geographic analysis to make data-driven decisions (Maptitude, n.d.). Maptitude provides an array of functions, such as data-integrated heat mapping, drive-time rings, geographic census data analysis, and territory creation; and it contains expandable functions to include other third-party software. We believe Maptitude is a promising suite of capabilities that would likely yield the greatest opportunities for scalable clustering analysis. We believe the capabilities of Maptitude are promising when compared to Microsoft Excel–based mathematical clustering because Maptitude provides greater information integration, including the ability to layer information (see Figure 5, which illustrates the robust functions of Maptitude).

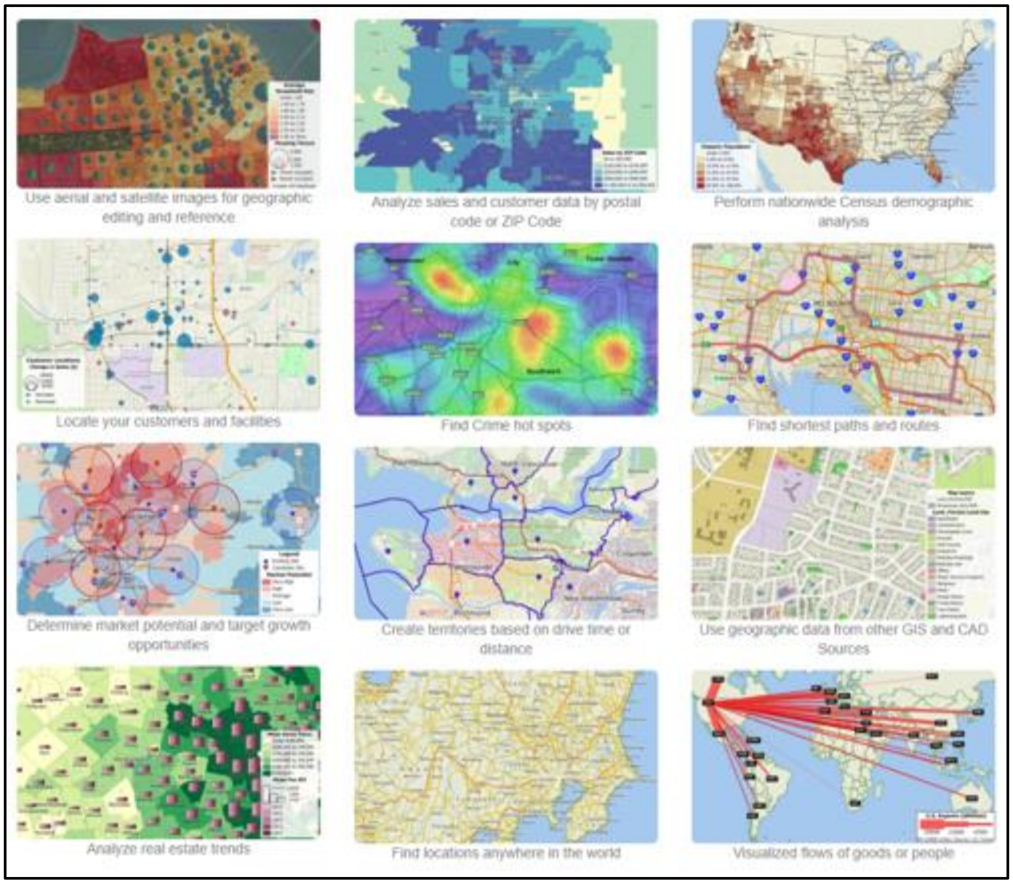


Figure 5. Maptitude Capabilities
(MPCluster, n.d.)

Integrating Installation Data

Prior to embarking on clustering analysis, we determined which requiring agencies would be involved in the procurement of services. Numerous DoD reorganizations over the past few decades, including multiple Base Realignments and Closures (BRAC), consolidated commands and assets, to include procurement units. We needed some way to decipher which areas within the United States are relevant to our analysis.

We were provided an opportunity to integrate data from the DoD Base Structure Report for FY 2015 (DoD, 2015). This valuable report provides a snapshot of real property within the DoD, including an exhaustive list of installation specifics, such as building square footage, owned acreage, and personnel assigned. We saw opportunities for comprehensive clustering analysis. To integrate the data, we converted the report into a readable database with Maptitude and incorporated each reported installation by pinning the location on a map. We also attached personnel assigned to each pinned location, which has benefits for the ISWM analysis (we discuss these benefits later).

MPCluster

MPCluster is a commercially available third-party, add-in software application that provides cluster analysis capabilities within the Maptitude GIS suite. MPCluster identifies groups or clusters in the Maptitude data and then creates new layers, drawing the clusters as boundary shapes and centroid points. MPCluster makes it possible to find clusters (natural groupings) in your Maptitude data. Although it is typically used with point layers, it can also be used with area layers (e.g., shaded area territories). Applications include market research, the determination of supply territories, and finding potential sales territories (MPCluster, n.d.).

Incorporating MPCluster into our analysis was a simple solution to add depth to our analysis by considering potential parameters, such as minimum and maximum number of installations within a cluster, distance from a DoD installation within a cluster, and centroid weighting based on a certain factor, such as distance from a refuse collection station for ISWM requirements.

Employing the concepts of the clustering continuum within the Maptitude GIS suite with MPCluster software gives us the necessary toolset to develop a scalable methodology for clustering analysis for any service-related category management efforts. This advanced toolset offers an innovative approach that provides category management teams a way to make informed, data-driven decisions. Data-driven decisions derived from clustering not only align with category management strategies, but also correspond with the contracting officer's duty to "take the lead in encouraging business process innovations and ensuring that business decisions are sound" (FAR 1.102-4).

Model Application

The challenges we encountered with our initial approach revealed that clustering based on distance alone does not account for all of the complexities of a given service requirement. Therefore, we developed a model that determines which installations should be clustered based on the market intelligence collected for the given service requirement. The methodology has four main steps:

1. Identify DoD requiring activities for a given service.
2. Identify cost-driver market intelligence relevant to developing clusters.
3. Integrate cost-driver market intelligence into commercial mapping software.
4. Use cost-driver market intelligence to determine optimal cluster size.

To develop and demonstrate this methodology, we selected ISWM as a common, recurring DoD service requirement. ISWM provides a viable service for analysis because it is a service requirement that is common across all three service components of the DoD, and significant cost-driver data and market intelligence is available in the ISWM Category Intelligence Report (CIR; Brady et al., 2016). We narrowed the scope of our ISWM test case to the southern California region for feasibility. We lacked the raw data required to import all landfill and transfer stations into Maptitude, but future research should find and include these data points in order to produce more robust clusters.

In Step 1, we identified DoD requiring activities in southern California that purchased ISWM services in FY 2012–FY 2016. We included the entire service contract spend data on the most recent five-year span to ensure that we captured all contract awards during that time period. We filtered that spend data down to contracts awarded under PSC S205 for "Trash/Garbage Collection Services" within the state of California to give us an accurate picture of which requiring activities had a valid ISWM requirement. After we identified all DoD requiring activities, we integrated them into the Maptitude software by geocoding all DoD installations contained in the 2015 Base Structure Report (DoD, 2015).

For Step 2, we identified cost-driver market intelligence relevant to developing clusters intended to target rate, process, and demand savings. AF CIRs are composed from extensive market research and provide significant insight into common, recurring DoD service requirements. The ISWM CIR highlighted the industry cost structure, cost drivers, and other factors that were of interest for clustering decisions. Figure 6 shows the cost structure of the ISWM industry on the right, as compared to the overall service industry on the left.

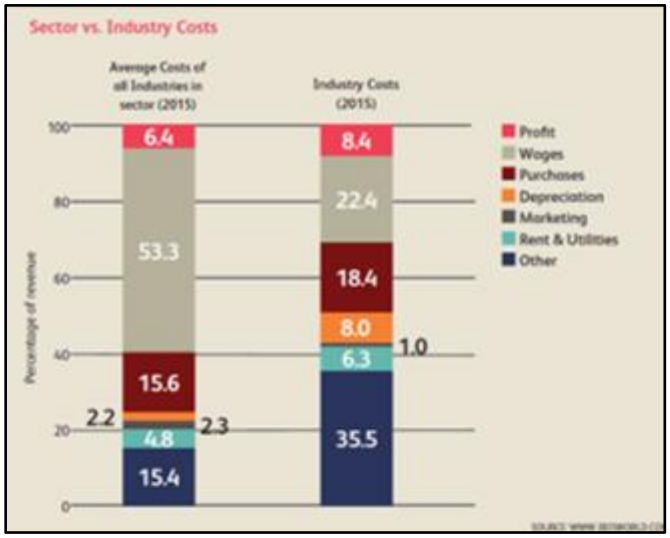


Figure 6. ISWM Cost Structure
(Brady et al., 2016, p. 32)

ISWM is less wage-driven, compared to the overall service sector, because it is a capital-intensive service requirement, requiring significant investment in fixed assets, such as trucks, equipment, and dumpsters. The cost structure shows that 18.4% of the industry costs are attributed to purchases and an additional 8% to depreciation, as compared to 15.6% and 2.3% respectively for the overall service sector (Brady et al., 2016, p. 32). Additionally, the garbage trucks employed during ISWM performance are not fuel-efficient. Fuel costs represented a significant amount of variable costs in the industry. Last, “other fees/expenses” are high in the industry because of the use of landfills and transfer stations, which charge fees based on usage. Fuel costs and “other fees/expenses” were referred to as “Other” in Figure 6 and represented 35.5% of the ISWM cost structure. “Other” costs were substantially higher for ISWM, whereas “Other” costs accounted for only 15.4% of costs in the overall service sector (Brady et al., 2016, p. 32).

Clustering DoD requiring activities based on cost drivers potentially drives efficiency in the utilization of fixed assets and generates savings related to fuel costs and “other fees/expenses.” Small, dense clusters would offer opportunities to utilize excess capacity of fixed assets for a proximity-dependent service like ISWM. Additionally, these small cluster sizes would allow contractors to design optimal routes that minimize fuel expenses, as well as labor costs of employees sitting in traffic or taking unnecessarily long routes. Finally, a cluster of bases could allow contractors to negotiate more favorable rates or fees for dumping waste at one landfill or transfer station that is centrally located among requiring activities, which could also subsequently minimize fuel expenses.

Step 3 in the methodology is integration of cost drivers and market intelligence into the Maptitude software. For ISWM, this step entailed mapping all of the landfill and transfer stations in close proximity to the DoD installations. After we mapped those locations, we utilized a feature in Maptitude referred to as “drive time rings.” We dropped pins on the map around each grouping of installations and then applied the “drive time rings” feature. This feature produced three 25-minute drive time rings emanating from each pin, as shown in

Figure 7. This allowed us to see how many DoD installations, landfills, and transfer stations are within a 75-minute radius of each pin.

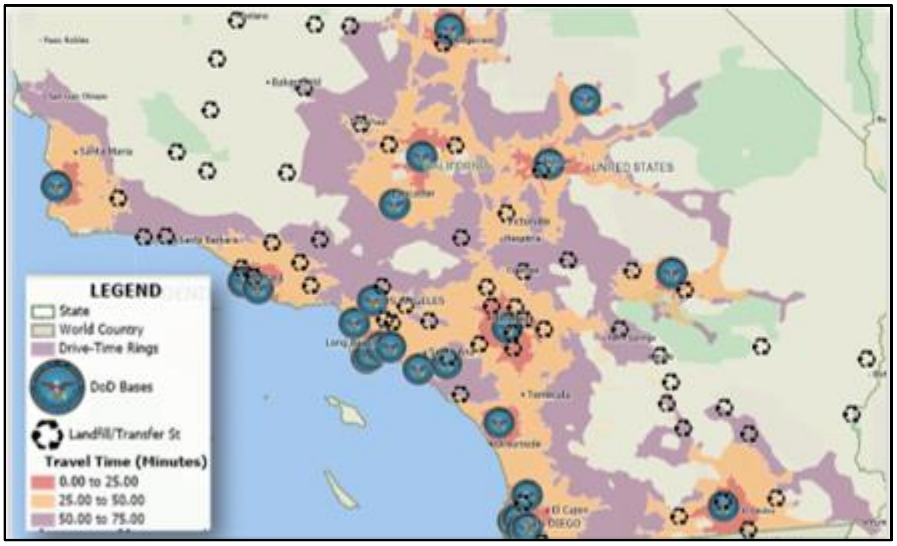


Figure 7. ISWM Drive Time Rings

Step 4 is to use cost-driver market intelligence to determine the optimal cluster size for the given service requirement. The inherently complex nature of contracting for services makes determination of the optimal cluster size challenging. This step in the methodology allows the flexibility that is required to develop an acquisition strategy that makes the most sense for the given service requirement. The category management team should use outputs from Maptitude and MPCluster to make data-driven decisions for the optimal cluster size. These decisions may require them to adjust Maptitude and MPCluster outputs to account for factors like small business participation, competition, and other public policy goals. Distance, travel time, or geographic features are examples of factors that could affect optimal cluster sizes. Distance and travel time were relevant for ISWM due to the proximity-dependent nature of the service, but also because of the specific cost drivers associated with ISWM. Category managers could achieve some of the savings discussed in Step 3 by consolidating requirements.

However, consolidating requirements is not the only way to lower costs for ISWM. Clustering could also have other applications, like demand management. For example, ISWM best practices could be applied through clustering, such as the use of weight sensors to trigger dumpster service, higher capacity dumpsters, or the use of kitchen waste dehydrators.

With the information outlined within the Category Intelligence Report for ISWM service, we derived two key cost drivers that could be used for our model: range constraints and wage constraints (determined by the Department of Labor for all federal contracts).

To better understand the key cost driver of ISWM range constraints, we needed to understand general distance capabilities of suppliers in terms of fuel consumption constraints. As a benchmark, we estimated that the farthest distance a refuse truck could service was approximately 200 miles roundtrip. This estimate was derived from research led by Dr. Sandhu at North Carolina State University, which found that traditional refuse trucks have a "typical fuel economy of 2 to 3 mpg of diesel" (Sandhu et al., 2014). Because our model was applied to the southern California area, we determined that it would be reasonable to err on a conservative estimate of 2 mpg, considering the congested traffic environment. Peterbilt, an industry manufacturer of refuse trucks, advertises fuel tank capacities ranging from 50 gallons to 150 gallons (Peterbilt, n.d.). On average, we estimated 100 gallons available for use on a typical refuse truck. Therefore, our calculated range of a refuse truck was 200 miles roundtrip, which means that all servicing locations must be within 100 miles of a central location.

Results

Using our map of DoD installations and landfill and transfer stations, we overlaid service range constraints and wage constraints to create clusters. Figure 8 shows a consolidated geographic output visualizing driving distance rings in 25-mile increments from El Segundo, CA. We chose El Segundo as a point of reference because of its relatively central location in the southern California area. Furthermore, El Segundo is also home to Los Angeles Air Force Base (AFB), which provides an opportunity to showcase the number of procuring agencies and places of performance contracting for ISWM. Examining FPDS-NG system data from FY 2012 to FY 2016, we were able to identify and validate 15 DoD installations procuring refuse collection services under the PSC S205 “Trash/Garbage Collection Services” within a 100-mile radius of El Segundo’s Los Angeles AFB (DPAP, n.d.). Moreover, the data shows an additional 151 DoD locations listed as procuring “waste collection services” working in commercial office space (DPAP, n.d.). We assume these other locations may be various program offices that consist of DoD employees. In addition, 166 adjacent locations have procured the same service over a five-year period (DPAP, n.d.). Finally, within a 100-mile driving range, 27 landfill or transfer active stations are available for use. This provides great opportunity for competition or negotiated rate savings, which are discussed later in this chapter.

Figure 8. PSC S205 Data Shown in 25-Mile Distance Intervals From El Segundo

To add depth to our research, we decided to test our model with a rough round-table estimate that assumes a refuse truck would remain within a two-hour roundtrip drive time from its base of operations. This estimate is used strictly to test our model’s clustering in terms of labor or hourly wage constraints. For example, suppose industry practice suggests the most efficient routes for refuse collection has a truck remain in a centralized location versus servicing areas spanning large geographic distances. Viewing the data in terms of drive time provides an analysis of locations available to be serviced within a relatively congested location like southern California, as compared to rural areas.

Figure 9 shows drive time in minutes from El Segundo, CA. The rings are divided into 20-minute increments, up to 60 minutes from El Segundo. From this output, we were able to derive a total of 38 locations procuring ISWM services within a 60-minute driving time. Additionally, nine landfill or transfer stations are available for use for ISWM services within the same area.

Figure 9. PSC S205 Data Shown in 20-Minute Drive Time Intervals From El Segundo

For comparative analysis, Figure 10 and Figure 11 show differences in driving distance analysis and driving time analysis. Determining which output to utilize depends on the type of cost savings being targeted. For example, if market intelligence leans toward the assumption that fuel costs are a significant factor to overall ISWM costs, then the driving distance clusters may provide a better solution. If labor costs are a more heavily weighted cost driver, the driving time output may provide a better solution.

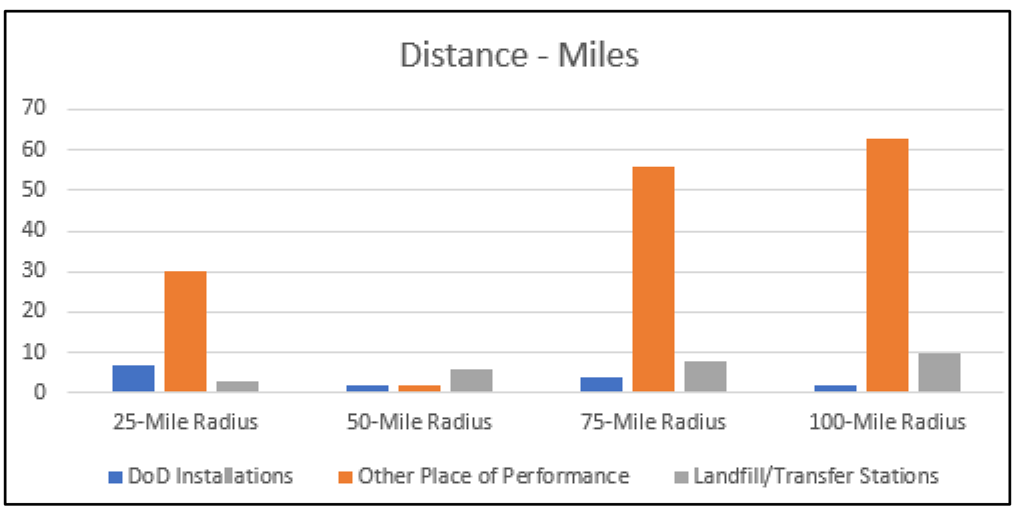


Figure 10. PSC S205 Cost Driver: Driving Distance (Miles) From El Segundo

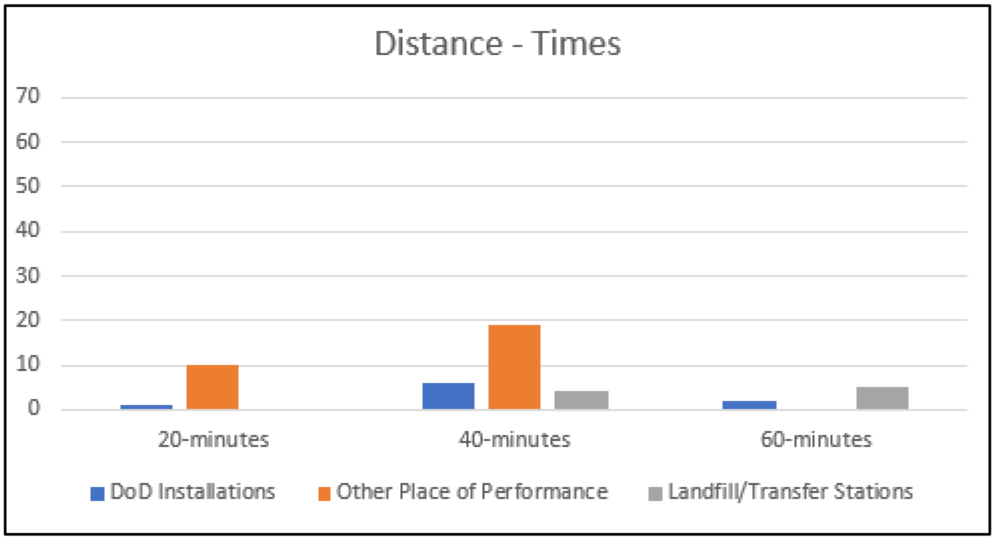


Figure 11. PSC S205 Cost Driver: Driving Time (Minutes) From El Segundo

Discussion and Conclusion

Service contract requirements need to garner additional focus in future DoD category management efforts. Service-related contracts deliver critical defense-sustaining capabilities and account for over half of defense spending. Historically, the DoD has struggled with the acquisition of services due to the inherently complex nature of services, compared to the seemingly straightforward procurement of commodities. This reality makes improvements in category management of service contracts vital to future mission success of the DoD.

We developed a methodology that clusters installations strategically, based on relevant cost drivers of a specific service. We recognize that recurring, common DoD service-related requirements yield the greatest opportunity for implementing strategic initiatives to achieve rate, process, and demand savings. A one-size-fits-all mathematical model will not generate optimal clusters for every service acquisition scenario. Rather, a flexible solution, like our model, allows category management teams to uphold their charge to innovate and enact best-in-class solutions for their category. Our solution is a versatile, commercial off-the-shelf software solution that provides the capability to map DoD requiring activities and cluster them based on virtually any type of data inputs.

As a reminder, our research question was: Are there potential cost savings (rate, process, demand) through strategically clustering common DoD service contracts?

Rate Savings

AFICA published the most recent version of the *Cost Savings Tracker Guidebook* in February 2017, which outlines how organizations should verify rate, process, and demand savings achieved through category management initiatives (AFICA, 2017). Clustering DoD-requiring activities based on cost drivers specific to the service requirement may lead to rate savings. We used ISWM to demonstrate the potential for rate savings by promoting efficient utilization of fixed assets, saving fuel costs, and reducing labor costs and “other fees/expenses” associated with landfills and transfer stations.

We are unable to state the achievement of rate savings with certainty in this study because we do not have the proper data to make a quantifiable claim. Due to the varying levels of service quality and the scope of work performed at various requiring activities across the DoD, we are unable to quantify levels of service quality or scope of work performed by looking at the collected spend data, which only provides a total contract price. We are unable to discern the number of containers serviced on base, the volume of waste produced, or other factors related to cost (e.g., hazardous waste disposal). Without a higher level of data granularity, we are unable to make an "apples-to-apples" comparison, which prevents us from stating the rate savings that could be achieved with certainty. The DoD could implement a pilot test at a few locations to estimate potential savings before undertaking an enterprise-wide approach.

Process Savings

Clustering DoD installations to develop large acquisition solutions like indefinite quantity/indefinite delivery (ID/IQ) contracts would create significant process savings in contract formation and administration. The use of IDIQs, where practical, would decrease the number of contract awards and subsequent administrative actions required to provide common services to DoD installations.

One of the requiring activities in close proximity to Los Angeles AFB is Edwards AFB, which has fulfilled its ISWM requirement using an ID/IQ contract since 2009. The remaining requiring activities awarded their ISWM requirement under individual definite delivery/definite quantity contracts with one base year and four option years. The metrics from the AFICA Cost Savings Tracker prove that substantial process savings are possible. The Cost Savings Tracker uses a 2014 Operational Contracting Air Force Manpower Standard developed by the Fifth Manpower Requirements Squadron (5MRS) to measure process savings by establishing standard process times for the award and administration of various contract types (AFICA, 2017). This manpower standard requires 615.08 hours to award a "definite" service contract and 219.66 hours to award a service task order off an "indefinite" contract vehicle (AFICA, 2017). This suggests that the DoD could potentially realize 5,535.88 hours of process savings—395.42 hours per contract over a five-year period—should the 14 other DoD installations in the southern California area fulfill their ISWM requirements using the ID/IQ awarded at Edwards AFB. These savings are even more substantial when extrapolated to include clusters encompassing all CONUS DoD installations.

Demand Savings

Lastly, clustering common, recurring DoD service requirements would result in standardized levels of service at all installations. The demand savings from clustering would promote the implementation of best practices for that service requirement across the DoD, which would eliminate non-value-added activities currently performed at some installations. The Air Force Civil Engineer Center (AFCEC) gathers ISWM sub-Activity Management Plan (sub-AMP) data for Air Force ISWM requirements—data points on the number of containers at each base, tons of waste generated, and cost per ton to remove the waste (CIR; Brady et al., 2016). However, ISWM data for the other DoD installations in the southern California area was not available. Therefore, based on the lack of data availability/granularity, we are unable to validate any demand savings for ISWM services.

Our findings suggest that there are substantial opportunities to achieve process savings through strategic management of common, recurring DoD service requirements. Additional research and application are needed to prove rate and demand savings. We narrowed the scope of this research to ISWM to provide depth of analysis and to demonstrate a methodology for a common, recurring DoD service. It was not feasible to discuss all common, recurring DOoD service requirements in this research. However, ISWM spend during FY 2016 was less than 1% of the $149.6 billion spent on all DoD service contracts. This suggests that our research barely scratched the surface of total spend on DoD service contracts.

Our research revealed a significant number of complexities associated with Category Management of service contracts that prevented us from recommending a "one-size-fits-all" model. We recommend additional data be gathered on service requirements procured within the DoD for future research related to category management of services.

Additionally, future research should focus on services that fall on the proximity-independent end of the continuum. We suspect there are several proximity-independent services in the IT category. Our model allows future researchers to collect data and develop visualizations that inform category management decisions for proximity-dependent and proximity-independent services.

References

Air Force Installation Contracting Agency (AFICA). (2017). *Cost savings tracker guidebook*. Wright-Patterson Air Force Base, OH: Author.

Brady, S., Briden, D., Carper, N., Dunham, S., Herrmann, P., Lovejoy, A., & Turnipseed, C. (2016). *Category intelligence report—Appendix A: Integrated Solid Waste Management*. Air Force Installation Contracting Agency (AFICA).

Church, J. D. (2014, April). Explaining the 30-year shift in consumer expenditures from commodities to services, 1982–2012. *Monthly Labor Review.* Retrieved from https://www.bls.gov/opub/mlr/2014/article/explaining-the-shift-in-consumer-expenditures.htm

Defense Procurement and Acquisition Policy (DPAP). (n.d.). Government-wide category management organization [Figure]. Retrieved September 28, 2017, from http://www.acq.osd.mil/dpap/ss/images/Government-wide_Category_Structure1.png

DoD. (2015). *Base structure report—Fiscal year 2015 baseline*. Retrieved from http://www.acq.osd.mil/eie/Downloads/BSI/Base%20Structure%20Report%20FY15.pdf

Defense Procurement and Acquisition Policy (DPAP). (n.d.). Inventory of services contracts. Retrieved October 10, 2017, from http://www.acq.osd.mil/dpap/cpic/cp/inventory_of_services_contracts.html

Federal Acquisition Regulation (FAR), 48 C.F.R. 1.102-4 (2017).

Federal Acquisition Regulation (FAR), 48 C.F.R. 37.101 (2017).

GAO. (2013). *Leading commercial practices can help federal agencies increase savings when acquiring services*. Retrieved from https://www.gao.gov/assets/660/653770.pdf

GAO. (2017). *High-risk series: Progress on many high-risk areas, while substantial efforts needed on others*. Retrieved from http://www.gao.gov/assets/690/682765.pdf

Holt, C. (2017, January). *AFICA business analytics*. Presented during conference at Wright-Patterson Air Force Base, OH.

Johnson, C. (2005). *Implementing strategic sourcing*. Washington, DC: Office of Management and Budget.

Kraljic, P. (1983, September). Purchasing must become supply management. *Harvard Business Review,* 109–117.

Maptitude. (n.d.). Maptitude business mapping software. Retrieved October 10, 2017, from http://www.caliper.com/Maptitude/BusinessMap/default.htm

MPCluster. (n.d.). MPCluster for Maptitude: Features. Retrieved from http://www.mpcluster.com/features.php

O'Brien, J. (2015). *Category management in purchasing*. London, UK: Kogan Page Limited.

Office of Management and Budget (OMB). (2015, April). *Government-wide category management guidance document.* Washington, DC: Author.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2017). *FY16 spend—DoD as contracting dept*. Retrieved from http://www.acq.osd.mil/dpap/sa/Learn-More/images/ServicesSpendingFY16.pdf

Peterbilt. (n.d.). Peterbilt vocational model 320 [Specification sheet]. Retrieved August 27, 2017, from http://www.peterbilt.com/products/vocational/320/#specifications

Reese, D., & Pohlman, D. (2005, Spring). Centralized purchasing power: Why Air Force leadership should care. *Air Force Journal of Logistics*, 2–12.

Rung, A. (2014). *Transforming the marketplace: Simplifying federal procurement to improve performance, drive innovation, and increase savings*. Washington, DC: Office of Management and Budget.

Sandhu, G., Frey, H. C., Bartelt-Hunt, S., & Jones, E. (2014, April). *Real-world activity and fuel use of diesel and CNG refuse trucks* [PowerPoint slides]. Retrieved from http://www.cert.ucr.edu/events/pems2014/liveagenda/25sandhu.pdf

Zients, J. (2012). *Improving acquisition through strategic sourcing.* Washington, DC: Office of Management and Budget.

A Review of Alternative Methods to Inventory Contracted Services in the Department of Defense[1]

Nancy Y. Moore—RAND Corporation

Molly Dunigan—RAND Corporation

Frank Camm—RAND Corporation

Samantha Cherney—RAND Corporation

Clifford A. Grammich—RAND Corporation

Judith D. Mele—RAND Corporation

Evan D. Peet—RAND Corporation

Anita Szafran—RAND Corporation

Preface

Title 10, Section 2330a, of the U.S. Code requires the Secretary of Defense to "submit to Congress an annual inventory of the activities performed during the preceding fiscal year pursuant to contracts for services." Persistent concerns regarding both the methods for collecting these data in the Inventory of Contracted Services (ICS) and the utility of the data led the conferees for the National Defense Authorization Act for Fiscal Year 2016 to direct the Secretary of Defense to examine the approach that the U.S. Department of Defense (DoD) is taking to comply with this statutory requirement. Congress directed the Secretary of Defense, as part of this examination, to determine whether the ICS produced by the DoD enhances oversight of contracting activities and to submit a report to the congressional defense committees explaining the results of that examination, outlining efforts to better manage contractor and civilian personnel costs within the DoD, and outlining potential alternative methods of meeting ICS requirements.

To assist the Secretary of Defense in making this determination, the Principal Deputy Assistant Secretary of Defense for Manpower and Reserve Affairs asked the RAND Corporation to conduct the mandated research. This final report builds on an interim report delivered in advance of the March 1, 2016, deadline for reporting to Congress. It should be of interest to policymakers concerned with DoD purchases of services as well as to DoD officials charged with ensuring better oversight of purchased services.

This research was sponsored by the Principal Deputy Assistant Secretary of Defense for Manpower and Reserve Affairs and conducted within the Forces and Resources Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community. For more information on the RAND Forces and Resources

[1] Library of Congress Cataloging-in-Publication Data ISBN: 978-0-8330-9672-2. Published by the RAND Corporation, Santa Monica, CA.

Policy Center, see www.rand.org/nsrd/ndri/centers/frp.html or contact the director (contact information is provided on the webpage).

Summary

Since the late 1940s, U.S. Department of Defense (DoD) purchases of services have increased consistently, from less than 30% to more than 60% of the department's overall budget. This increase reflects both the growth of services in the overall economy and the initiatives of political administrations over time to procure services from the private sector on behalf of the DoD to the greatest extent possible. Nevertheless, such growth has led to concerns regarding contracting of inherently governmental functions, contract oversight, contractor accountability, and contract waste, fraud, and abuse.

Concerns about the growth in the DoD's purchases of services have led Congress to institute several policies aimed at strengthening oversight of such purchases. These policies have included 2001 legislation requiring the DoD to collect and track data on the procurement of services, 2002 and 2008 congressional language expressing an interest in spend analyses that might be used to increase buying leverage and improve contractor performance, and a 2008 requirement in Title 10, Section 2330a, of the U.S. Code establishing the DoD Inventory of Contracted Services (ICS) to collect information on activities performed under DoD service contracts.

Concern regarding both the methods for collecting data in the ICS and the utility of these data led Congress to request that the Secretary of Defense review the methods used to create the ICS, as well as the products resulting from these efforts. Congress specifically requested that the Secretary of Defense examine the extent to which the ICS provides data on service contracts that are useful to the DoD and congressional stakeholders, the extent of gaps between ICS data and data that the DoD and Congress would find most useful, whether existing databases or other information technology systems could provide a timely solution and data that are relevant to workforce planning, and the strengths and weaknesses of different methods for reporting on the DoD's use of contractor personnel. The DoD asked RAND to assist the Secretary of Defense in fulfilling this congressional mandate.

This report documents the final results of that research. It explores the congressional intent underlying the ICS requirement, gaps between the ICS data and data most useful to the DoD and congressional stakeholders, insights on the issues that Congress seeks to address through the ICS requirement that can be derived from analyses of non-ICS data found in alternative databases, and the strengths and weaknesses of different methods for estimating and reporting contractor personnel use.

Research Methods

This study employed multiple research methods and was conducted in a compressed time frame. The bulk of the data collection and analysis was completed between mid-December 2015 and mid-February 2016 to produce an interim report in advance of the Secretary of Defense's March 1, 2016, deadline for reporting to Congress. During that time, we reviewed relevant legislation and literature; analyzed relevant data from the ICS, the Federal Procurement Data System—Next Generation (FPDS-NG), and the System for Award Management (SAM); and interviewed key stakeholders in Congress, the DoD, non-DoD federal agencies, and the offices of relevant service contractors. Over the course of the project, we interviewed 83 individuals and reviewed more than 80 documents, focusing on the legislative and historical context underlying the ICS, as well as insights from the economics literature. We also analyzed ICS and FPDS-NG data to develop distribution and trend data on spending, contracts, business size, and type of service, as well as to identify contractors to interview. Finally, we devised and tested several alternative metrics for calculating contractor full-time equivalents (FTEs) using existing non-ICS data sources.

What Does the Current ICS Look Like?

The current ICS is produced approximately one year after the end of the fiscal year (FY) for which data are reported and is captured in two publicly available formats: a report to Congress and 37 different defense-component spreadsheets on the Defense Procurement Acquisition Policy (DPAP) website. The ICS is produced using the Contractor Manpower Reporting Application (CMRA) system. The Army first developed the CMRA system, but now there are four separate "instances," or versions, of the system—one each for the Army, Air Force, and Navy, and a combined one for the other defense agencies. As currently planned, the different instances of CMRA will be combined into one "enterprise-wide" system (eCMRA) in the next several years, and all instances are now being moved under Defense Manpower Data Center stewardship.

We were unable to gain access to the raw CMRA data for this study, as access is limited in an attempt to protect contractors' proprietary data from competitors. However, it is critical to note that even without access to restricted CMRA data, we were able to link ICS-reported direct labor hours to particular service contractors using contract number information publicly available on the FPDS-NG website and the publicly available ICS data published on the DPAP website (which reports contract number as well as direct labor hours information). When we analyzed the ICS data and compared them to FPDS-NG data, we also found shortfalls in completeness and quality. These analyses reinforce some of what we heard in our interviews with various stakeholders and subject-matter experts.

How Well Does the ICS Meet Congressional Objectives and DoD Needs?

In our interviews with congressional staff and DoD stakeholders, we found that the current ICS falls short of meeting the needs of Congress and the DoD. Many congressional staff suggest that the format in which ICS data are reported to Congress is not useful and hinders assessment of the data. Several commented that the data, as reported, are too detailed and would be more useful if they were synthesized before reporting. Ultimately, it appears that Congress seeks analysis—not raw data—from the DoD, but this is not well specified in the statute.

The views of DoD stakeholders, meanwhile, vary based on the interests of their functional communities. Manpower and personnel, budgeting, and acquisition officials require different information to do their jobs most effectively. This, in turn, shapes their views of the utility of the ICS. Stakeholders who focus on manpower and personnel planning, for example, seek data on contractor FTEs and level of effort needed to enable strategic workforce planning and insourcing decisions. Those in the budgeting community seek data on total costs and data that integrate well into budget considerations, allowing them to budget more effectively. Meanwhile, those in the acquisition community seek data on level of performance and total costs to enable smart acquisition decision-making. Such variation in the preferred types of data on service contracts makes it difficult to determine what data need to be collected and why. Understanding the goals of collection is critical in making this determination.

The characteristics and types of data that appear to be most relevant to congressional and DoD stakeholders are (1) processed, analyzed data; (2) forward-looking data that can be integrated into budget processes; (3) data on contractor FTEs to compare with civilian FTEs in making sourcing decisions; (4) auditable and verifiable data; and (5) data distinguishing types of contracts by total costs, contractor FTEs, and other values of interest. By contrast, the ICS includes data that are unprocessed, retrospective, and can largely be found elsewhere, with the exception of contractor direct labor hours. Moreover, the direct labor hours data included in the ICS were, at the time this research was conducted, largely estimated rather than contractor-reported, making them difficult to verify or even distinguish among contracts.

Meanwhile, our interviews with service contractors indicated that CMRA reporting can be burdensome for the contractor and that contractors are subject to a multiplicity of reporting requirements, some mandating that they enter overlapping data points into CMRA and other systems, such as SAM. Moreover, contractors questioned the utility of collecting direct labor hours data and were concerned about the exposure of their proprietary data and how that may affect their success in competing for future contracts.

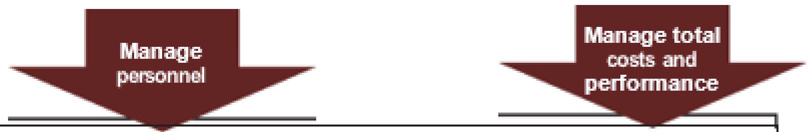
Why Are There Gaps Between the Current ICS and What Congress and the DoD Envisioned?

To understand the shortcomings of the ICS and the challenges in meeting congressional intent related to the ICS requirement, it is critical to note that service contractors' production functions vary, so comparing metrics across these firms can be misleading. Yet the ICS is structured to measure contractors using equivalent inputs, as though they all pro- duce equivalent services. This has the potential to distort results, as there is extensive variability between service contractors in the *types* of services they provide and, particularly, the degree to which the services they provide replace or simply *augment* governmental functions. Furthermore, service contractors demonstrate great variability in how they produce outputs, specifically in terms of the degree to which they substitute capital for labor and their various types of labor input. Indeed, collected labor input data show that although direct labor accounts for about half of total contract costs, the direct labor fraction varies greatly by type of service, from about one-fourth to three-fourths of total costs.

Table 1 illustrates the spectrum of contracting activities in which the DoD may engage, ranging from staff augmentation contracting (also known as "labor contracting") to complete contracting, with mixed contracting lying between the two extremes. In instances of staff augmentation contracting, the DoD provides the facilities, materials, equipment, technologies, and other inputs to production. Meanwhile, in complete contracting, the DoD provides only contractor management. Because of the distinction in how these levels of contracting are managed, collecting direct labor hours for all DoD service contracts without distinction in terms of the types of services provided is problematic. Even assuming that data on direct labor hours are valid and precise, collecting them for complete contracting is inappropriate because each contractor engaged in complete contracting makes distinct decisions regarding the inputs, processes, and practices used to provide the service. Because direct labor hours do not account for distinctions between the various types of contracting activities, they are insufficient to inform strategic workforce planning or DoD budget decision-making and acquisition planning.

Exacerbating the insufficiency of direct labor hours for informing strategic workforce planning is the fact that substitutions between different components of the total force—military, civilian, and contractor—cannot always be exchanged one-for-one within and across sectors because of individual-, organization-, and sector-level variations and gaps in productivity. For instance, different organizations tend to hire workers from different backgrounds, motivate them in different ways, and train them to have different skill sets using distinct methods. Maximizing labor productivity would clearly be ideal. However, without precise measures of productivity, and with legal constraints on sourcing decisions and governmental influence in contractor labor decisions—such as a moratorium on outsourcing competitions and constraints on military and civilian personnel hiring—the ability to use proxy measures of productivity correctly and appropriately is key to informing strategic workforce management. The collection of direct labor hours in the ICS is not an appropriate proxy measure of productivity, especially when these data make no distinction between the various types of contracting activities being performed.

Table 1. Distinct Contracting Activities Require Different Management
(Allen & Chandrashekar, 2000)



Provider	Staff Augmentation Contracting	Mixed Contracting	Complete Contracting
Host firm/ buyer	Some employees Materials Processes and systems Technology and equipment Facilities Management/ supervision	Some or all of the following: • Employees • Materials • Processes and systems • Technology and equipment • Facilities • Management/ supervision	Program management
Contractor	Some employees	Some or all of the following: • Employees • Materials • Processes and systems • Technology and equipment • Facilities • Management/ supervision	Employees Materials Processes and systems Technology and equipment Facilities Supervision

Insights on DoD Service Contracting Provided by Data Systems Other Than the ICS

Our work exploring the potential to meet congressional intent for the ICS with the use of other data systems focused primarily on data from the FPDS-NG (and, to a lesser extent, on budget data). While FPDS-NG data may contain some errors in data submission, it is the authoritative system for federal contract reporting, and the quality of its data has improved over time. The FPDS-NG provides, for contractions of at least $3,000, information on the amount of the contract action, identification codes indicating whether the firm providing the service is a small business, the North American Industry Classification System (NAICS) code for the firm, the Treasury Account Symbol for the transaction funding (which can be linked to budget categories), and the Product or Service Code (PSC), a more finely grained indicator than the NAICS code regarding the exact nature of goods and services purchased. Though subject to some delay in publication due to security measures and verification, these data can provide numerous insights on the services the DoD has recently purchased and, in doing so, can assist in addressing the various congressional concerns underlying the ICS requirement—namely, enabling the production of spend analyses, trend analyses, and forecasting to inform budgeting and acquisition decisions. As we discuss in greater detail in the next section, FPDS-NG data can also be used to produce alternative metrics for calculating contractor FTEs in an effort to inform strategic workforce planning.

In terms of their contribution to spend and trend analyses, FPDS-NG data indicate that half of DoD service spending falls under three PSC categories: Support (Professional/Administrative/Management), Research and Development, and Maintenance, Repair, and Rebuilding of Equipment. Further probing of FPDS-NG data shows that four specific types of services—including engineering and technical services and general health care services—were significant drivers of increases in DoD support service spending. FPDS-NG data indicate some opportunities to leverage purchases (that is, to consolidate contracts or purchases across offices so as to increase buying power), but they also point to possible difficulties in doing so. These potential challenges include the large proportions of small businesses and the wide array of industries (denoted by NAICS codes) providing these services, each of which is likely to vary along a number of dimensions. Finally, FPDS-NG data also help to illustrate the extent to which current service purchases are open to competition, as well as the contract types used to purchase services.

Coupling FPDS-NG data with budget-category projections can yield insights regarding likely future trends in overall spending for services. Most spending (59%) for services is related to operations and maintenance (O&M), one of the categories Congress uses for budgeting. Current budget projections indicate that O&M spending will continue to decrease, meaning spending on contracted services is likely to decrease as well. Congress stated that it wanted the DoD to achieve a reduction in service spending of $4.1 billion by FY 2017, relative what it was spending in FY 2012 ($186 billion). This amount of reduction in services spending—$4.1 billion—is equivalent to a parallel reduction in military basic pay resulting from reductions in military end strength in the same period. Calculating actual spending reductions using FPDS-NG data indicated that the DoD had already more than met this goal in FY 2015, reducing service spending by $38 billion. Using the President's budget projections, and assuming that the DoD out-year spending matches these budget projections and a constant percentage use of service spending occurs in each budget category over time, we estimated that the reduction in service spending will continue along the same trend, decreasing by $60 billion between FY 2012 (when total service spending was $186 billion) and FY 2021 (when we project total service spending to be $126 billion).

Risks and Benefits of Different Methods for Estimating and Reporting Contractor Personnel Use

In our interviews, we found that one of the key motives underlying the collection of data on direct labor hours associated with a contract is to use this information to assess the scale of the contracted services relative to the size of comparable DoD in-house activities. However, due to the shortcomings of relying on direct labor hours data for strategic workforce planning and insourcing decisions, as discussed earlier, the DoD might consider alternative measures that do not require collecting, validating, auditing, and protecting proprietary data reported by contractors.

We identified three alternative metrics to estimate contractor manpower numbers, in addition to the current ICS metrics (both actual contractor-reported direct labor hours and direct labor hours calculated using Army algorithms that are based on previously reported data on firms providing similar services). These are as follows:

1. the number of civilian FTEs that could be hired with the contract dollars ("civilian labor FTE per contract")
2. the number of industry or location-average employees per contract dollars ("contractor labor FTE per contract")

3. contract employees as a proportion of overall contractor revenue

These metrics may be calculated from data available through the FPDS-NG, the Bureau of Labor Statistics, and the U.S. General Services Administration–owned SAM, which consolidates the Catalog of Federal Domestic Assistance and various federal procurement systems.

Because these alternatives draw on available, in-house federal data or publicly available data, they do not require the DoD to collect, validate, audit, and protect proprietary data from contractors as the current metrics do. This, in turn, would likely generate cost savings, as the expenses incurred by contractors to collect and report direct labor hours on a given contract are included in the overall price of that contract. The use of these alternative metrics in lieu of contractor-reported or estimated direct labor hours could also assist the DoD in producing an ICS in a more timely manner, as they might not be as time-consuming to generate. The common disadvantage of these three alternative metrics is that they assume equal productivity across employees, industries, and sectors. Nevertheless, our comparative analyses of the results of the current ICS-derived metrics and these alternative metrics for determining the relative importance of contracted versus noncontracted labor across functions—based on calculations performed using each respective metric on “case studies” of particular PSCs—indicate that these alternatives are close proxies for the ICS metrics.

Conclusions

Our findings suggest that the ICS products, and the processes used to create them, are not meeting either congressional or DoD stakeholder needs. Several factors led us to this conclusion. First, the congressional intent underlying the ICS requirement is multifaceted and not always clearly specified in statute. Second, different ICS stakeholders are based in distinct functional communities, each of which has its own interests and needs driving its purpose for utilizing ICS data—and these needs and purposes do not always align across these divergent communities. Third, opinions differ both inside and outside the DoD on the utility and quality of the current ICS data, with some stakeholders finding the data more valuable and some finding them less valuable. Fourth, because the majority of ICS data through FY 2014 (the most recent year for which ICS data were available during the period of research) are derived using algorithms developed by the Army that are based on unverified contractor-reported data, their validity is questionable from the outset—particularly for contracts held by military services and defense components other than the Army. Moreover, the ICS data do not currently support spend analyses, trend analyses, forecasting, or strategic sourcing, and more information would be needed to conduct effective labor comparisons to inform insourcing decisions. Finally, much of the information Congress seeks to allow oversight of service contracts is available in other systems.

These findings led to several recommendations. First, policymakers should institutionalize the development and reporting of DoD-wide spend analyses of services, including analyses of trends, forecasts, and FTEs. This would entail issuing a detailed requirement for an institutionalized capability to analyze data on DoD service contracts and providing the necessary funding for its development. The DoD would also likely need to employ dedicated research programmers or statistical analysts in long-term positions to produce ICS-related analyses.

Second, ICS-related statutory requirements could be refined to better distinguish between different types of contracting and, accordingly, to require the collection of different data elements for each. Our research found that DoD contracting practices vary with both the types of services purchased and the level of oversight the DoD expects over such purchases. ICS requirements could be revised to identify and distinguish among staff augmentation, mixed contracting, and complete contracting arrangements. For staff augmentation contracts, ICS requirements could be revised to specify the use of multiple alternative metrics relying on existing data sources, such as the FPDS-NG, to estimate a likely range of contractor FTEs. For mixed and complete contracting, the ICS requirement could be rewritten to focus on measuring total cost and performance rather than direct labor hours. Finally, for operational support contracts—for which Congress wants increased oversight of the number of deployed contractors on the ground—reporting requirements should focus on the number of actual deployed contractors, not FTEs.

Third, the DoD should periodically perform sourcing analyses of selected commercial services to determine whether civilians or contractors deliver the required level of performance at the lowest total costs. Doing so will ensure continuous adjustment of task assignments across the total force, where necessary, to maintain the lowest cost and most effective staffing solutions for a diverse set of defense functions.

Reference

Allen, S., & Chandrashekar, A. (2000, March). Outsourcing services: The contract is just the beginning. *Business Horizons, 43*(2).

Distribution Statement

MARAD's Maritime Security Program: Exemplary Innovation in Acquisition Policy?

Marvin Phaup—is Professorial Lecturer and Research Scholar at the Trachtenberg School, The George Washington University. His research focuses on budget concepts and budget process reform for sovereign governments. He received a PhD in economics from the University of Virginia and a BA from Roanoke College. [mphaup@gwu.edu]

Abstract

The potential of real call options to reduce the cost of meeting unpredictable variations in demand for support assets and services by government motivates this examination of an apparently successful instance by the U.S. Maritime Administration's Maritime Security Program (MSP). This case study, however, fails to find corroborating evidence of efficiency gains. The MSP's financing and structure obscures and understates the total cost of the acquired service and likely fails to minimize costs. Identified program modifications could increase transparency and strengthen program management.

Introduction

The principal defense-related function of the U. S. Department of Transportation's Maritime Administration (MARAD) is to ensure an adequate supply of U.S.-flag vessels to meet the DoD's sealift requirements in unpredictable defense and national emergencies. It does so through (1) the Maritime Security Program (MSP) and (2) maintenance of a Ready Reserve fleet (RRF). The MSP consists of MARAD operating agreements with owners of 60 privately-owned, commercially-active vessels in international commerce that are U.S.-flagged and operated by U.S. crews. Most of the MSP ships are owned by "documentation citizens," companies chartered and headquartered in the United States, but wholly-owned subsidiaries of foreign shipping companies[1] (Frittelli, 2015). The Ready Reserve Fleet (RRF) is made up of 46 government-owned, older vessels, partially crewed, loaded with support cargo, and berthed in accessible harbors (e.g., Baltimore, on stand-by call from the DoD).

Maritime Security Program

Under the MSP, MARAD pays an annual fee (currently about $5 million per ship) to owners of 60 militarily-useful vessels. In exchange, owners commit to make participating ships, crews, and inter-modal transport networks available to the DoD at pre- established ("preference cargo") shipping rates on the call of the Secretary of Defense. MARAD also provides no-premium, war-risk insurance against losses of ships, crews, and cargos for merchant ships operating in war zones in support of DoD missions. The DoD reimburses MARAD for insured losses after the loss event.

When I submitted a proposal to the NPS organizers of this symposium, I assumed that the MSP retainers could be described accurately as the price/cost of a specified call

[1] Participation of large integrated international companies, via documentation citizenship, is one means government gains access to established inter-modal transportation networks. Small, U.S.-owned and -flagged shippers are unlikely to carry sufficient commercial cargo to warrant the development of inter-modal systems.

option on ship's services, that is, that the $5 million compensated ship owners for the increase in cost from operating subject to the constraint of fulfilling the commitment to divert shipping capacity to the DoD on short notice. I also assumed that, absent this payment, the DoD would not have reliable access to the services of those ships when needed.

My interest and proposal were motivated by possible cost savings from the use of call options compared with the alternative of stockpiling and holding inactive assets to meet unpredictable and uncertain demand. Indeed, MARAD reports that the costs of maintaining a ship in the RRF is more than $8 million per year, exclusive of capital costs. Compared with the $5 million per year cost of an MSP vessel and the avoided up-front capital cost—in excess of $200 million per ship—the MSP looks to be a cost-effective choice. Similarly, proposals to replace the aging RRF with new ships are non-starters given the reported lower cost of an MSP alternative.[2]

However, as closer examination has shown, framing the MSP as a purchase of a real call option on commercially viable international shipping services is a fundamental mischaracterization of the underlying reality.

First, U.S.-flag merchant ships, with the associated requirement for U.S. crews, cannot compete in international shipping markets without substantial government subsidies; their operating costs are about three times those of foreign-flag vessels (MARAD, 2011). U.S. flag ships can operate in international shipping only because the federal government has implicitly committed to absorb the losses of all 81 U.S. flags[3] currently operating in international trade, even if they are not enrolled in the MSP. It is also a logical stretch to regard U.S.-flag ships as privately-owned, if government equity must be available to absorb operating losses. The principal risk facing U.S.-flag private equity appears to be political: that government might withdraw its implied commitment without notice.

Second, the annual MSP retainer is only a part of the government subsidy to U.S.-flags. Most of the balance is provided in premium rates paid to U.S. flags transporting U.S. government (preference) cargo.[4] Preference cargos, including those shipped by the DoD during emergencies, are actively sought by all U.S. flags, including the 21 not enrolled currently in the MSP program, because it is their primary means of covering operating losses. MARAD has little need to purchase a call on asset services that are readily available to the government when needed.

Thus, I conclude that the MSP is not an exemplary use of a real call option to meet unpredictable demands for durable goods and services because the MSP is a form of

---

[2] The case for expecting the MSP's use of commercial vessels for military sealift to be less costly than a government-owned fleet—notably the revenues that may be earned from commercial shipping when not required by the DoD—is developed in Herberger et al. (2015). A second related source of expected lower cost is the advantage of private managers in attracting and servicing commercial trade. One difficulty with this claim is the absence of public information showing that commercial shipping rates are sufficient to yield marginal commercial revenues in excess of the associated marginal cost for U.S.-flag ships. Further, the recognition of capital costs can be deferred with MSP annual payments, but not avoided

[3] U.S. fleet numbers are as of March 1, 2018, but are subject to month-to-month variation.

[4] Federal income tax expenditures are also provided to U.S.-flags. For details, see Joint Committee on Taxation, 2017.

government ownership that, among other important features, avoids recognizing the up-front capital cost and debt financing of an asset purchase. It has more in common with a government-sponsored enterprise or a capital lease-purchase than a call option.[5]

If so, what are we to make of the MSP from an acquisition or budgetary perspective? Is it a cost-effective innovation to be considered for wider use? The remainder of this paper argues to the contrary: The financing and budgetary treatment of the MSP is inconsistent with good acquisition and budget policy because it fails to provide policymakers, administrators, and the public with a transparent and timely measure of the cost of this activity. Ironically, the MSP could be the most cost-effective of available alternatives, but under current practice, it is not possible to know if that is the case. The MSP is more clearly an example of the perils of acquisition without known prices.

The Operating Cost Differential of U.S. Flag Ships and Its Financing

MARAD (2011) in its last published estimate (for 2010) of the average operating cost of a U.S.-flag ship compared with the cost of a foreign-flag ship, found that U.S. flags had operating costs 2.7 times higher than a foreign-flag. Operating costs include labor, supplies, maintenance & repair (M&R), insurance, and overhead. This measure omits voyage cost, including fuel and port fees, and capital costs, meaning depreciation and financing costs, on grounds that those would be equal for all flags. (U.S.-flag ships do not have to be built in U.S. shipyards to be eligible for MSP or preference cargos. Jones Act ships that operate in U.S. domestic trade, without competition from foreign-flag ships, must be built in the United States.) The big drivers of the operating cost difference are labor cost and, a distant second, M&R. Repairs of U.S.-flag ships performed in foreign, rather than U.S. shipyards, are subject to a 50% ad valorem U.S. tax.

In dollar terms, the average U.S.-flag ship annual operating cost gap was $4.6 million greater than the operating cost of a foreign-flag ship. In 2010, however, the MSR annual payment was only $2.9 million, or 63% of the total operating cost difference, leaving an average cost gap of $1.7 million for the 60 ships enrolled in the MSP. For non-MSP-U.S. flags, the full $4.6 million had to be financed by other means.[6] For both MSP and non-MSP vessels, that financing gap is covered mostly by a second subsidy, premium shipping rates for government preference cargo. Preference cargo rates must be “fair and reasonable,” but that standard may be defined in terms of the operating costs of U.S.-flag ships rather than international rates (for details, see Frittelli, 2015). Under current law and policy, most water-borne “government-impelled cargos” (100% of military and Ex-Im Bank-financed shipments and 50% of humanitarian food aid, i.e., USAID and USDA) must be transported in U.S.-flag ships.

[5] Ship owners use the MSP agreements as collateral to obtain financing for ship construction or purchase (Econometrica 2009, p. 32), that is, the agreements are construed as an implied federal guarantee of the debt of MSP participants. An argument can be made that these agreements should be disclosed, if not recognized, as a debt of the U.S. government.
[6] The 21 non-MSP ships are said to be niche carriers: for example, dry bulk (grain) carriers, which have no military use or, alternatively, operate under contract to the DoD.

What Is the Size of the Preference Cargo Premium?

Although MARAD documents acknowledge that preference cargo rates are "significantly higher than commercial rates" (MARAD, 2011), MARAD officials disavow knowledge of its magnitude. That missing information is problematic because the preference cargo premium and cargo volume are required to estimate the annual MSP retainer.

Absent a MARAD estimate, it seems reasonable to assume that, as a lower bound, the 81 U.S.-flag ships receive preference cargo premium rates sufficient to cover their operating cost differential, net of the MSP payment (Transportation Research Board, 2016).

In 2010, the MSP payment covered 63% of the operating cost differential. If we assume that the payment of $5 million covers the same share of the cost gap, then the total current gap would be $8 million per ship per year for 81 ships, or $648 million. Payments to owners/operators of MSP ships were (60 x $5 million) or $300 million, which leaves $348 million to be financed by the preference cargo premiums.

In Fiscal Year (FY) 2012, the last year for which data are available (MARAD, 2013, Appendix 3), carriage of U.S. government cargos produced total revenues of $3.718 billion for U.S.-flag ships. Assuming no change in the volume of government cargo, a preference rate premium of about 10% (348/3718-348) would have been sufficient to cover the operating cost differential of U.S.-flag ships, if the allocation of preference cargo were aligned to match the unfinanced cost differential for each owner/operator. In fact, MARAD offers a type of brokerage service for government agencies to assist in identifying commercial carriers likely able to provide shipping capacity as needed. In any case, the allocation of preference cargo cannot be assumed to match exactly the operating cost differential for each owner/operator. Thus, some excess of aggregate preference cargo subsidy is likely, even at the lower bound.

An upper bound for preference cargo rates is far more difficult to identify because of the presence of factors that can be expected to hold actual preference rates close to the minimum as well as others that suggest preference shipping rates might be well above the minimum.

First, there are many potential entrants into the U.S.-flag market, even though "flagging in" requires hardware modifications; a dry-dock, Coast Guard hull inspection; and fire and safety drills that may entail costs in the range of $500,000–$1 million (Transportation Research Board, 2016). Entry into the market is facilitated by the five documentation citizen MSP providers who have both U.S.- and foreign-flag ships in their fleets. For those companies, the ongoing relative costs and potential gains of operating in the U.S.-flag market are observable and easily exploited if available.[7] Second, over 500 U.S.-owned vessels are currently engaged in international service, but foreign-flagged, mostly in the Marshall Islands, Singapore, and Liberia. Those operators could be expected to reflag in, if it were sufficiently profitable to do so. Finally, the DoD accounts for about 90% of all preference cargo shipments. That market dominance appears to offer sufficient negotiating leverage to minimize the premium in preference rates.

---

[7] Flagging in/flagging out of U.S. registry is frequently observed. In the period of January 1, 2016–September 17, 2017, 20 ships were flagged in and nine ships flagged out (MARAD, 2017).

However, there are also market and financial indicators consistent with a substantial gap between the lower bound and actual rates. For one, ownership of U.S.-flag vessels is concentrated: Four participating documentation citizen companies own more than 45 of the 60 ships in the MSP. MARAD provides frequent opportunities for consultation and collaboration among MSP stakeholders, including carriers, labor unions, and the DoD. Legislation provides some exemptions from U.S. anti-trust laws for the industry. Further, estimated average charter rates reportedly paid in 2008 by the DoD (Econometrica, 2009, pp. 29–30) were more than 160% of 2010 daily operating costs. Even a generous allowance for voyage and capital costs suggests that the preference cargo premium could be several times the 10% lower bound. Further, in 2017, following years of decline in the U.S.-flag fleet, the number of vessels increased by nine, from 72 to 81, which is consistent with a preference premium sufficient in expectation to compensate operators for the costs of flagging in/flagging out and the operating cost differential.

Another factor that may tend to boost the preference cargo premium is the uncertainty of U.S. subsidies. MSP cash payments to participants are appropriated annually and subject to the uncertainties of congressional action, including dollar amounts. Indeed, the Trump administration has proposed to reduce the MSP retainer to $3.6 million per ship in FY 2019, from its current level of $5 million (MARAD, 2018). In addition, the amount of preference cargo revenue that a U.S. flag receives each year depends on both the volume of cargo it can secure and the size of the premium. For non-MSP vessels, revenue uncertainty may be especially high because preference cargo is its only means of covering operating losses.

Finally, under long-standing policy—usually referred to as “commercial first”—the DoD has used U.S.-flag commercial shipping despite its higher cost, to the extent feasible from a military perspective, rather than U.S. government vessels for the transport of military cargo (Frittelli, 2015; Herberger et al., 2015). This policy limits the ability of the DoD to use its monopsony market power to restrain increases in the preference cargo premium.

Those competing, potentially offsetting drivers of the preference cargo rate premium and the paucity of data that would permit assessment of their net effect obscures the true cost of the MSP program. Budgeting and acquisition decisions must be made without salient, full-cost prices, which is equivalent to the phenomenon of market failure in the face of external costs.

Getting Better Cost Information and (Possibly) Increasing Value for Money

MARAD could significantly improve the measurement of the full cost of current policy—using proprietary information it receives from U.S.-flags as a condition of MSP participation and eligibility for preference cargo, and industry data—by

- updating its 2011 estimate of the daily average operating cost for U.S.- and foreign-flag ships;[8]
- preparing a comparable estimate of the daily average operating revenue for U.S.-flag and foreign-flag ships by source, i.e., commercial and government cargo, distinguishing MSP and non-MSP U.S.-flag vessels; and
- reporting daily average net income for U.S.-flag ships.

The objectives of acquisition and budget policy, however, go beyond simply providing a defensible estimate of the cost of a current service to offering some assurance to policymakers and the public that government is getting the best value for the money. At present, the use of cargo preference to deliver a subsidy muddies budget transparency, is costly to deliver to intended recipients, and hides part of the cost of a DOT program in the budgets of other agencies. It also adds a costly and unnecessary element of uncertainty to the expected revenues of U.S.-flag vessels.

To achieve the more important objectives of acquisition policy, the financing of the MSP needs to be modified to reduce the uncertainty of the MSP subsidy and to increase the incentives of service providers to operate efficiently and minimize costs.

One means of advancing those objectives would be to replace the current dual subsidy system of annual cash payments and preference cargo premia with annual cash payments over the 10-year life of the operating agreement, with up-front, full-budget authority scored at contract agreement. In effect, the cargo preference component of the subsidy to U.S.-flags would be cashed-out through more reliable and predictable annual payments, as originally proposed by the Eisenhower administration (Frittelli, 2015).[9]

It is also undesirable to place the onus for determining a fair and reasonable price of a purchased service on the purchaser, who can never know as much about the minimum costs of delivery as the provider. Accordingly, MARAD should solicit bids from all U.S. international shipping companies for the limited number of MSP slots. Bidders would be required to demonstrate the capability to provide a U.S.-flag vessel of the desired condition and type (container, roll-on/roll off, tanker) with U.S. crews, as well as the inter-modal network services normally provided by commercially-active international carriers.[10]

Use of real call options to meet unpredictable demands for support assets and services is a promising alternative to government stockpiling, but the MSP, in its current

---

[8] They could also disclose the estimated average voyage and capital costs which are excluded from the reported operating cost measure.

[9] The drawdown of U.S. forces in Afghanistan and Iraq and the shift in humanitarian aid from commodities to cash have reduced the volume of preference cargo in recent years. This development may provide a convenient opportunity to terminate the practice completely.

[10] Exceptions to the inter-model network requirement might be provided for carriers with few ships active in international trade. Equivalent services could be purchased by the DoD through fee-for-service agreements.

form, does not provide an informative trial. Indeed, it is a poster child for the antithesis of good budgeting/good acquisition policy: acquiring goods and services without knowing the price of the chosen alternative.

If the U.S. government increases the number of U.S.-flag and U.S.-crewed ships beyond those the market will support, at best the United States will have to give up the value lost by diverting U.S. labor and capital from their higher-valued uses. A worse case is that costs will be significantly higher than the minimum. A loss cannot be avoided by obscuring its cost, but making cost transparent can enable it to be managed more effectively.

References

Econometrica Inc. (2009). *Maritime Security Program impact evaluation*. Bethesda, MD. Retrieved from https://www.marad.dot.gov/wp-content/uploads/pdf/MSP_Revised_Final_Report_Transmitted_07-24-09.pdf

Frittelli, J. (2015, October). *Cargo preferences for U.S.-flag shipping* (CRS Report R 44254). Washington, DC: Library of Congress. Retrieved from https://fas.org/sgp/crs/misc/R44254.pdf

Herberger, A. J., Gaulden, K. C., & Marshall, R. (2015). *Global reach: Revolutionizing the use of commercial vessels and intermodal systems for military sealift, 1990–2012.* Annapolis, MD: Naval Institute Press.

Joint Committee on Taxation. (2017). Estimates of federal tax expenditures. Retrieved from https://www.jct.gov/publications.html?func=startdown&id=4971

Maritime Administration, U.S. Department of Transportation (MARAD). (2011, September). *Comparison of U.S. and foreign-flag operating costs*. Retrieved from https://www.marad.dot.gov/wp-content/uploads/pdf/Comparison_of_US_and_Foreign_Flag_Operating_Costs.pdf

Maritime Administration, U.S. Department of Transportation (MARAD). (2014). *Annual report, FY 2013*. Retrieved from https://www.marad.dot.gov/wp-content/uploads/pdf/2013-ANNUAL-REPORT-Final.pdf

Maritime Administration, U.S. Department of Transportation (MARAD). (2017). US flag, privately-owned merchant fleet: Summary of changes, 2016 onward. Retrieved from https://www.marad.dot.gov/resources/data-statistics/

Maritime Administration, U.S. Department of Transportation (MARAD). (2018). *Budget estimates, FY 2019*. Submitted for use by the Committees on Appropriations, Washington, DC. Retrieved from https://www.transportation.gov/sites/dot.gov/files/docs/mission/budget/304521/marad-fy-2019-cj-final-508-compliant.pdf

Transportation Research Board of the National Academies of Science, Engineering and Medicine. (2016, February 25). *Impact of US Coast Guard regulations on US flag registry* [Letter report]. Retrieved from http://www.trb.org/Main/Blurbs/173981.aspx

Acknowledgment

I gratefully acknowledge helpful comments and suggestions received from the Trachtenberg Faculty Seminar, especially Leah Brooks and Kathryn Newcomer. However, the views expressed are the author's and should not be attributed to anyone whose work and comments have contributed to the development of this paper.

Disclaimer

Preliminary Draft: Please do not quote without author's permission.

Panel 9. The Role of Innovation in Improving Defense Acquisition Outcomes

Wednesday, May 9, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Colonel Eric Ropella, USMC,** Program Manager, Presidential Helicopters Program ***Meeting Warfighter Needs Through Innovation*** Cheryl Andrew, Government Accountability Office ***Pushing the Acquisition Innovation Envelope at the Office of Naval Research*** Anthony C. Santago, II, The MITRE Corporation Mike J. Arendt, The MITRE Corporation Jeffrey Colombe, The MITRE Corporation Timothy B. Bentley, Office of Naval Research Lisa L. Lalis, The MITRE Corporation ***Bridging the Gap: Improving DoD-Backed Innovation Programs to Enhance the Adoption of Innovative Technology Throughout the Armed Services*** Amanda Bresler, PW Communications

Colonel Eric Ropella, USMC—Hailing from Green Bay, Wisconsin, Col. Eric Ropella, graduated from the U.S. Naval Academy in 1993, commissioning as a second lieutenant. Ropella earned a Master of Business Administration in Systems Acquisition and Program Management from the Naval Postgraduate School in 2004 where he was awarded the Assistant Secretary of the Navy Research, Development and Acquisition Excellence Award.

After earning his wings as a Naval Aviator in 1996 he flew the CH-46E Sea Knight, serving with HMM-163 "Evil Eyes" from 1997-2002. He completed two WestPac deployments with the 15th MEU during Operation Enduring Freedom as a weapons and tactics instructor, planning and leading contingency and combat operations into Pakistan and Afghanistan. As Air Mission Commander, Ropella served as a liaison officer between Task Force-58 and Special Operations Forces during the seizure of "Camp Rhino" Forward Operating Base.

From 2004 to 2007, he served as the Project Manager for the Performance Based Logistics Implementation and the Low Rate Initial Production/Sustainment Integrated Product Teams (IPT) at the Joint Strike Fighter (JSF) Program.

In 2007, Ropella transitioned to the Acquisition Management Professional Community. In August 2007, he assumed the role of Survivability IPT lead for the V-22 Program Office (PMA-275). During his assignment he received the Commandant's Acquisition Excellence Award. In October 2009, Ropella became the Level I MV-22 IPT lead; responsible for USMC V-22 programmatic issues and execution of the USMC V-22 acquisition budget.

He attended the International College of the Armed Forces at the National Defense University from August 2011 to June 2012, graduating with a Master of Science in National Resource Strategy.

In June 2013, Ropella re-joined the V-22 Program Office managing negotiation of the program's second multi-year procurement production contract award, worth over $5.6B, resulting in his team's selection for the FY13 Dr. Somoroff Department of the Navy Acquisition Excellence Award. In September 2013, Colonel Ropella assumed the role of Program Manager for International Programs, building the Foreign Military Sales organization from the ground up and completing the first international sale of the V-22 in June 2015.

From June 2016 through July 2017, Ropella served as the Military Assistant to the Assistant Secretary of the Navy (Research, Development, and Acquisition).

Reporting to Presidential Helicopters Program (PMA-274) as the lead for the VH-92A program in August 2017; Ropella assumed the position of Program Manager in March 2018.

Ropella awards and decorations include the; Legion of Merit, Defense Meritorious Service Medal with oak leaf cluster, Meritorious Service Medal, Air Medal (First and Second Strike/Flight Awards), Joint Service Commendation Medal, Navy Commendation Medal, Navy Achievement Medal and numerous campaign and unit awards.

Pushing the Acquisition Innovation Envelope at the Office of Naval Research[1]

Anthony C. Santago—received his BS in Mechanical Engineering from NC State University and his MS and PhD in Biomedical Engineering from the Virginia Tech–Wake Forest School of Biomedical Engineering and Science. His research spanned injury and musculoskeletal biomechanics, which included both experimental and computational modeling techniques. Since joining The MITRE Corporation's Emerging Technologies Department in June 2015, he has been supporting Department of Defense programs that are researching, developing, and setting future direction for human body computational models. [asantago@mitre.org]

Mike J. Arendt—is a subject matter expert in innovative acquisition and contracting strategies. Over the past 12 years, he has supported numerous military and civilian agencies in their efforts to implement forward-leaning business processes. He has authored and co-authored many cutting-edge studies and reports including the MITRE Innovative Contracting Implementation Framework and the MITRE Challenged-Based Acquisition Handbook. Dr. Arendt was a consultant with IBM and a member of the research faculty at the University of Maryland's Center for Public Policy and Private Enterprise. He holds a PhD in Policy Studies from the University of Maryland, an MS from Missouri State University in Defense & Strategic Studies, and two BA degrees from The Ohio State University with majors in Economics, Political Science, and Sociology. [marendt@mitre.org]

Jeffrey Colombe—has a BS in Biomedical Engineering from the University of Illinois at Chicago and a PhD in Neurobiology from the University of Chicago. He leads the Neurotechnology and Biomechanics research group in The MITRE Corporation's Emerging Technologies Department, working to advance biomedicine, human performance enhancement, and biologically inspired approaches to artificial intelligence and high-performance computing. [jcolombe@mitre.org]

Timothy B. Bentley—has a BS in Aquatic Sciences from Cornell University and an MS and PhD in Biological Oceanography from the Rosenstiel School at the University of Miami. He is the Program Officer leading Force Health Protection research in the Warfighter Performance Department at the Office of Naval Research (ONR). He funds and guides research developing cutting-edge technologies for the Fleet and other military services to improve warfighter health, performance, injury mitigation, and enhanced casualty care capabilities. [timothy.b.bentley@navy.mil]

Lisa L. Lalis—is a MITRE Project Leader for the Army Medical Department, and Office of Naval Research Code 34 Warfighter Performance. Lisa has supported numerous military and civilian agencies in software centric acquisitions and implementations. Lisa has assisted financial, intelligence, and healthcare private and public sectors in vision, strategy, and execution of programs with varying degrees of complexity. Lisa has a BS in Computer Science with a minor in Physics from Drexel University. [llalis@mitre.org]

Abstract

Developing prototypes may require performers, all with different areas of expertise, working together to address the complexity required for a successful development effort. Current Federal Acquisition Regulation (FAR) policy makes it difficult for these collaborations to assemble efficiently. Complex research projects, such as the Office of Naval Research's Incapacitation Prediction in Expeditionary Domains: An Integrated

[1] This technical data deliverable was developed using contract funds under Basic Contract No. W15P7T-13-C-A802; Approved for Public Release; Distribution Unlimited. 18-0917 © The MITRE Corporation.

Software Tool (I-PREDICT) project, which seeks to develop a computational model to predict human injury and functional incapacitation as a result of military hazards, often face difficulty when attempting to transition across the “valley of death” from development to adoption. A decision framework was developed and implemented for I-PREDICT to select the appropriate acquisition strategy aligned with the technical needs of the program. A three-phase implementation strategy was also designed, which included the use of an Other Transaction Authority (OTA) and the use of a Technical Committee to promote communication between performers. The resulting decision framework and implementation strategy may be used Navy-wide or across other military Services for R&D programs requiring acquisition flexibility coupled with collaborative technology development. Additionally, the research produced a customizable method for leveraging OTAs as a mechanism for development of complex prototypes depending on disparate kinds and sources of expertise.

Introduction

Background

Developing prototypes in many research & development (R&D) fields may be adequately addressed by one or merely a few performers from industry or academia with few dependencies among them, while other fields require a more widely distributed and collaborative approach. In some cases, several performers with different areas of expertise must work together to address the development of a complex prototype under the guidance of the funding agency. Certain aspects of the Federal Acquisition Regulation (FAR) that are motivated by fair competition requirements may extend time for contract awards, implement inflexible vendor payment processes, and impose a lack of coordination across contracting vehicles between vendors, among other limitations. In addition, when developing an innovative solution, highly complex research projects often face significant difficulties when attempting to transition across the so-called “research valley of death” from development to fielding. The R&D aim of leveraging computational models to predict and prevent battlefield injuries for the warfighter is one of these highly complex research fields that can yield enormous benefits if the contracted performers can collaborate with each other and the funding agency after a solicitation but prior to awards, and if the gap between development and fielding can be bridged.

Injury and incapacitation estimates for combat scenarios are currently educated guesses at best. Estimates may be based on simplified injury risk thresholds on hazard parameters like pressure, stress, strain, or force applied to an organ or tissue. Increasingly, such knowledge is incorporated into scientific simulations that can be run many times over to explore variations in hazards and to assign statistical confidence to predictions of injury risk. Current modeling and simulation methods for predicting injury can be inaccurate, regional rather than whole-body, not validated appropriately, and may not be based upon physiologically or operationally relevant data. Injury prevention standards are needed to protect warfighters from injuries based on a scientific understanding of hazardous conditions typical of military service, and of the vulnerability of tissues, organs, and bodily functions to those hazards. Such standards will inform the development of personal protective equipment (PPE), safe vehicles, and safe weapons systems, as well as tactics, techniques, and procedures (TTPs) to protect against injury. The development of a high-quality, whole human body computational model of injury is needed to inform such standards and to act as a pivotal part of operational mission planning and risk assessment.

Addressing the Problem

The Future Naval Capabilities (FNC) program was initiated in 2002 by the Department of the Navy to develop a prototype and to transition cutting-edge technologies, at a technology readiness level (TRL) 6, to acquisition program managers within a five-year time frame. Recent changes to the FNC program have placed an increased emphasis on accelerating the transition of Office of Naval Research (ONR) developed solutions to the fleet/force by requiring up-front financial contributions from stakeholders to cover transition costs. Stakeholders commit, via a Technology Transition Agreement (TTA), to develop, transition, and deploy a product delivered by a specific FNC project to the fleet/force. For FNC products that involve a high degree of technical complexity, the use of FAR-based acquisition tools may limit the likelihood of successful product development and/or transition, thus promoting the need to explore non-traditional acquisition methods.

ONR's Code 34 Force Health Protection initiated the Incapacitation Prediction in Expeditionary Domains: An Integrated Computational Tool (I-PREDICT) pre-FNC project to provide an in silico integrated computational model of the warfighter's body to use for injury prevention and treatment, medical response planning, and equipment design including tradeoff analysis, validation, and testing. Warfighter injury in combat and training has high financial and personal costs, and interferes with the ability to complete mission objectives. Accurate prediction of injuries and resulting functional incapacitation under varying hazard conditions would provide the ability to design safe equipment and behavioral practices, and to allow commanders to weigh operational risks during the planning and execution of missions and to allocate resources appropriate to those risks. Faster transition to the field would result in more timely realization of benefit to the warfighter.

Way Ahead

To overcome both the collaboration and transition barriers, R&D programs such as I-PREDICT may leverage Other Transaction Authority (OTA) contractual vehicles to support development of prototype technologies. OTAs are not subject to the FAR, permit the use of commercial-like, negotiated agreements that can be awarded in as little as 90 days, allow highly flexible use of intellectual property, and promote unique public/private partnerships to achieve program objectives. Moreover, upon completion of prototype development, solutions may be transitioned from the OTA vehicle to a sole-source FAR-based procurement production contract which is permissible under 10 U.S.C. 2371b and accelerates the timeline from development to fielding (U.S.C. Code § 2371b).

This paper provides a **description of a decision framework** that was developed to allow full evaluation of technical and acquisition options to meet project needs, building and evaluating potential program strategies, and developing a process for execution of the selected strategy that included leveraging OTA and the Medical Technology Enterprise Consortium (MTEC). The project has executed this decision framework, which is outlined below in later sections of this paper. Additionally, a **three-phase implementation strategy was developed for the execution of the selected project strategy**. The three phases are outlined in the section titled Implementation of the Single Model–Multiple Performer Strategy and have not yet been implemented by the project. The decision framework and three-phase implementation strategy are outlined in Figure 1.

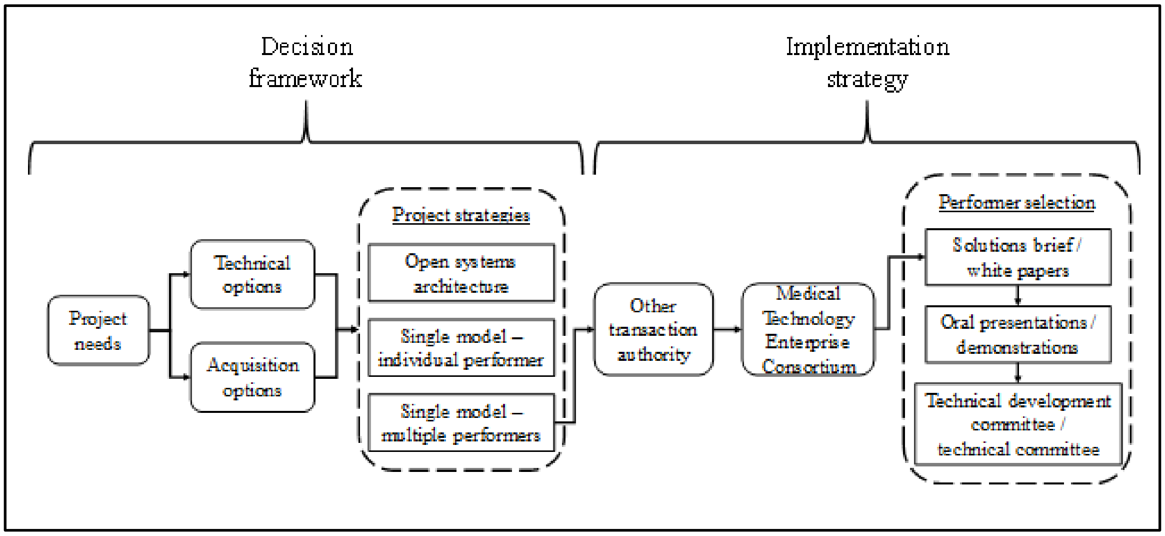


Figure 1. Decision Framework and Implementation Strategy Developed for the I-PREDICT Project

Beyond the decision framework and three-phase selection process, this paper also provides a **tailorable method for leveraging OTAs** as a mechanism for development of prototypes that require many disparate kinds and sources of expertise. This method may be used Navy-wide or across other military Services for any R&D program that requires acquisition flexibility coupled with highly collaborative complex technology development. The implementation strategy grants programs the ability to leverage the innovative Technical Committee (TC) construct, outlined in a later section (Phase III Technical Development Team/Technical Committee Selection), permitting an increased level of collaboration and communication between performers than is typically accessible under the FAR. When the decision framework and the supporting three-phased OTA method are combined, programs can benefit from a unique partnership with performers from industry and academia while streamlining deployment and fielding. As a result, the warfighter may more quickly gain access to effective technological solutions to enhance operations and safety.

Technical Options for I-PREDICT

Considerations regarding the technical options outlined below were essential to ensuring that the I-PREDICT computational model could be successfully developed, and because the technical decisions provide the foundations upon which the modeling capability will be designed, constructed, and employed by the end users. However, decisions regarding each technical option were fraught with complexity because they significantly altered both programmatic scope and the skill sets required to achieve the project goals. The following discussion maps out application and technical needs of the program to specifically illustrate the challenges.

The I-PREDICT program's technical goal is to develop a deformable finite element model (FEM) with detailed human anatomy and accurate human body responses to military hazards (e.g., blunt force impacts and blast shockwave pressure effects). To construct and use a whole-human deformable FEM, several highly interdependent data products are required: **experimentation** is used to gather biomechanical responses of human tissue on scales ranging from small volumes of tissues to organs and large body regions**, digital anatomy** is needed to computationally represent the human body as a group of computer aided design (CAD) components that are converted to finite element mesh components,

interfaces between the components must be defined to mathematically represent how the response of each structure is dependent upon the surrounding structures, management of **anthropometric variability** (different body shapes and sizes) will allow exploration of vulnerability risk dependent on body parameters, creation of additional lower **fidelity** components will result in reduced computational runtimes, and the FEM needs to be **validated** against experimentation at larger scales of organs or large body regions.

A challenge facing the I-PREDICT program is the choice of an optimal set of sources and/or performers for the array of needed data products. For example, if the project decided to pursue multiple CAD anthropometries versus a model that is morphable to multiple anatomical variations, the project would be asking for expertise in CAD development from biomedical imaging data instead of expertise in the development and implementation of morphing technologies. The major topical requirement categories identified for this project where technical options exist are (1) biomechanics **experimentation** in support of model development, (2) software used to simulate the physics (commercial equation **solver**), (3) generation of **digital anatomy** including CAD and subsequent finite element meshing, (4) management of **anthropomorphic variability**, (5) mathematical **interfaces** between component body structures, (6) deliberate variations in component **fidelity**, (7) verification, **validation**, and accreditation (VV&A) of models based on experimental biomechanics, and (8) **pre- and post-processing tools**. The options are described in greater detail below.

Experimentation

Computational models of human injury require experimental validation datasets at a succession of anatomical scales to calibrate and validate biomechanical response properties of the model. These involve small-volume tests of homogeneous tissue types (e.g., liver, muscle, cortical bone), isolated anatomical structure tests (e.g., segments of tendon or ligament, humerus, clavicle), or large-scale tests such as cadaver crash tests using crash sleds, impact pendulums, and/or blast tubes. Biomechanical responses may include measurement of physics parameters such as stress, strain, and force relevant to tissue injury, collected using precision material testing systems. Analysis of movement corridors for whole-body responses to stimuli may calibrate such global parameters as kinematics, and are typically measured with precision high-speed video recording of landmarks and load cells. Ideally, support for experimental decisions should be motivated by knowledge of the military hazard environment, with specific references to experimental data from hazard environments. Biomechanical experiments performed under the I-PREDICT project should also primarily be in support of anatomical components that are most frequently injured during the hazard conditions prioritized by the project, and use cases outlined by I-PREDICT stakeholders and end-users. Determining the appropriate types and quantities of experimental test is necessary to the successful parameterization and validation of the I-PREDICT model. Three technical options were identified: (a) Government provided methodology in which Government would dictate the experiments, (b) performer developed methodology where the performers indicated the experiments they wanted to perform, and (c) a combination approach where both the Government and performers were involved in collaborative decision-making.

Solver

There currently exist several software systems (equation solvers) that are used to mathematically calculate the response of the human body to dynamic hazards. Examples of some of the most prominent solvers include LS-DYNA, Abaqus Explicit, Velodyne, CTH Sandia Shock Wave Physics, and CoBi. These solvers use numerical techniques to calculate a variety of physical variables (e.g., stress, strain, strain rate, and flow rates) within the human body at discrete time points following the onset of the hazard. The finite element analysis method, the most common method used to study human injuries from blunt impact hazards, represents small physical volumes of material, each referred to as a finite element, with a simple equation. A finite element solver then creates a system of these equations for an entire physical structure of coupled finite elements that are all solved simultaneously over discrete time steps. Selecting the appropriate solver was necessary to ensuring compatibility between the I-PREDICT model sub-components and between the I-PREDICT model and other computational models such as vehicles. Three technical options for selecting a solver were identified: (a) solver independence where multiple solvers would be able to be used simultaneously, (b) a Government-selected solver, and (c) a performer-selected solver.

Anatomy

CAD anatomy is required to accurately model the response of the entire human body to a military hazard. The CAD anatomy is essential because it provides the bounding box for modelers to create not only individual anatomical component models (e.g., liver, spleen, ribs) but also to model the interactions between anatomical components. Although multiple CAD anatomies exist that could be purchased by the project, typically, licenses restrict the distribution of any models developed from them. Therefore, there were two options the project could pursue for obtaining CAD anatomy: (a) a Government-provided CAD and (b) a performer-provided CAD.

Anthropomorphic Variability

It is well known that variations in anthropometry and posture can influence the risk of sustaining injuries. Accurately representing these variations is paramount to understanding how injury risk across the entire representative warfighter population ought to influence design decisions or mission planning. Therefore, the I-PREDICT FNC must be able to represent warfighters of differing anthropometries. As was outlined above in the introduction to the technical options, there were two technical options the project can use to represent multiple anthropometries and postures: (a) development of multiple CAD anatomies that represent multiple body shapes, sizes, and genders in multiple postures; and (b) morphing a single model to multiple anthropometries and/or postures.

Interfaces Between Component Pieces

The whole human body model is constructed of multiple component level models (e.g., heart, lungs, vessels, rib bones), requiring that significant consideration be given when designing the interfaces between the component-level models to avoid excess computational expense, while ensuring that the model accurately represents the response of the human body to the hazard. These interfaces represent the most computationally expensive portion of the simulation. However, models can be constructed to minimize these types of interfaces. Two technical options were identified to address model component interfaces: (a) the development of interface standards that explicitly define the interfaces between the anatomical component pieces and (b) allow the performers to define the interfaces.

Variation in Component Fidelity

Simulations of the human response to dynamic hazards are computationally expensive, with typical full body simulations taking between 12-48 hours using high performance computing (HPC) resources. To achieve model outputs in a more timely manner, recent work has focused on reductions in fidelity of models, or of selected model components. Allowing for the judicious reduction in fidelity of the I-PREDICT FNC in areas of the body that are of little interest to specific hazard scenarios, or are not typically injured as part of the hazard scenario, may result in improved run-time with minimal effect on the accuracy of the results. The project identified two technical options to address deliberately varied fidelity of model components: (a) the development of fidelity standards that explicitly define discrete fidelity levels of the anatomical component pieces, including regional components (e.g., thorax, abdomen) and more detailed components (e.g., blood vessels, bones, nerves); or (b) performer-defined fidelity levels.

Verification, Validation, and Accreditation

According to DoD Instruction 5000.61, *DoD Modeling and Simulation (M&S) Verification, Validation, and Accreditation (VV&A)*, it is DoD policy that (1) models, simulations, and associated data used to support DoD processes, products, and decisions shall undergo verification and validation (V&V) throughout their life cycles; (2) models, simulations, and associated data used to support DoD processes, products, and decisions shall be accredited for an intended use; and (3) VV&A results shall be documented and made accessible to the DoD Components, other Government agencies, and non-Governmental activities, as applicable and in accordance with DoD Directive 8320.02, *Data Sharing in a Net-Centric Department of Defense.*

Initiation of the VV&A process early in the project will help to ensure model accuracy and thoroughness, and assist in rapid fleet integration, as much of the necessary work will already be underway. As part of the plan, V&V should be performed throughout the period of performance so that required knowledge gaps can be filled, thus minimizing additional labor needed for FNC deployment to the fleet. The project identified four technical options for V&V: (a) Government develops the V&V plan and executes all of the V&V; (b) developer-initiated V&V in which the model developer(s) would be responsible for creating their own V&V plan and executing the V&V on the model(s) they are developing; (c) alternate developer V&V where V&V of the model components and whole-body model are executed by performers who did not develop the models being tested; and (d) combination V&V where the project would pursue a mixture of Government V&V, developer V&V, and alternate developer V&V.

Pre- and Post-Processing Tools

Pre- and post-processing tools are used to prepare a human body model for specific simulations and to gather outputs following the simulations. Pre-processing tools may include selection of model components, integration of model components via interfaces into a whole human body model, morphing the model to desired anthropometric parameters, altering the posture of the model, and deliberately varying the fidelity of certain model components. Post-processing tools should include the ability to extract injury and incapacitation risk from standard physical parameters such as stress, strain, velocity, and strain energy. The project identified two technical options for the development of pre- and post-processing tools: (a) Government-provided tools where the Government would develop the pre- and post-processing tools and (b) performer-developed tools.

Acquisition Options for I-PREDICT

Considerations regarding the acquisition options for I-PREDICT were critical to ensure that the appropriate technical requirements could be achieved. The technical options, described previously, have several inherent impacts on the acquisition options that may be selected. For example, if an existing contract vehicle such as a GSA schedule were to be chosen, subject matter expertise would be limited to those on that particular contract vehicle who may not possess the depth and breadth of skills required. Likewise, if the most flexible intellectual property approach isn't open and competitive, it would hamper the ability for the model to have free communications between the relative component pieces. In support of these types of concerns, the consideration of acquisition options and their potential impact on available technical options was paramount. As a result, the project examined several acquisition options that could be leveraged to help build the overall project strategy for I-PREDICT. The major topical categories identified for this project where acquisition options exist are (1) leadership structure, (2) contracting approach, (3) source-selection/evaluation approach, (4) incentive approach, and (5) intellectual property approach. Each of these options is defined below and is later incorporated in the Project Strategies section.

Leadership Structure

Quality project leadership is imperative to delivering a technically sound solution such as the I-PREDICT FNC. There are several leadership structures that have been used to create whole human body models, many of which focus upon the need for collaboration. The Total Human Model for Safety (THUMS) was created independently by Toyota Central R&D Labs. Other whole human body models, including the Human Model for Safety (HUMOS) and the Global Human Body Models Consortium (GHBMC), have used a consortium of model developers to create their whole human body models. Within the GHBMC program, technical leads were assigned to each body region with an overall technical lead responsible for the whole program. For I-PREDICT, there were four leadership structure options that were considered:

- *Government integration with multiple contracts*: In this approach, the Government will be responsible for integrating model components created by the performers under contract into one cohesive model. This provides additional assurances that the I-PREDICT FNC meets the needs of the Government stakeholder-defined use cases.
- *Industry/academia integration with multiple contracts*: In this approach, a designated performer will be responsible for integrating model components created by the other performers under contract into one cohesive model. This allows the project to leverage existing subject matter expertise and removes Government burden.
- *Industry/academia integration and development*: In this approach, a single performer will be responsible for creating the model components and integrating the model components into one cohesive model, potentially subcontracting and supervising components of the modeling. This allows performer flexibility to alter model construction during the period of performance.
- *Technical committee (TC):* In this approach, the Government will assume the management and administration of a TC, including standing up the committee and ensuring that the committee meets project goals. Technical directors will be assigned for each body region who are responsible for the experimental work and model component creation within that body region. The integrator will be a separate performer and part of the TC. In addition, the TC is structured to allow input from consultation with Government advisors and SMEs. This approach allows the project to leverage expertise across industry and academia while promoting communication among performers, and has been used successfully to create the GHBMC model.

Contracting Strategy

The contracting approach provides the rationale for the desired contract vehicle type chosen to acquire integration services, model components, CAD anatomy, and V&V for the I-PREDICT project. Contracting strategies to be considered may include those which are FAR-based (Federal Acquisition Regulation) and non-FAR-based (such as Other Transaction Authority or OTA). There were five contracting approaches evaluated for use for I-PREDICT:

- *Indefinite delivery/indefinite quantity (ID/IQ) contract:* Indefinite delivery, indefinite quantity contracts provide for an indefinite quantity of services for a fixed time. Awards are usually for base years and option years. The Government places delivery orders (for supplies) or task orders (for services) against a basic contract for individual requirements. Minimum and maximum quantity limits are specified in the basic contract as either number of units (for supplies) or as dollar values (for services). Each time a requirement under the scope is identified, individual delivery orders or task orders require a separate contracting action beyond the initial base contract award (GSA, 2017b).
- *C-Contract:* General term for contracts of all types except basic purchasing agreements, basic ordering agreements, indefinite delivery contracts, facilities contracts, sales contracts, and contracts placed with or through other Government departments or agencies or against contracts placed by such departments or agencies outside the DoD (Acquisition Guides, n.d.).
- *Other Transaction Authority (OTA)/Other Transactions (OTs):* OTs are legal binding agreements between the U.S. Government and industry, including traditional and non-traditional Government contractors, small businesses, and academia. Because they are not subject to the FAR, OTs are, by design, more flexible and responsive to atypical Government procurement requirements. Indeed, Congress provides the authority in recognition that, from time to time, boilerplate procurement methods are at odds with the Government's need to innovate. Consequently, OTs are primarily associated with some form of research, development, test, and evaluation (RDT&E; Arendt et al., 2018). The common theme of OT use is the primary goal is to reduce barriers to participation by firms not typically willing to subject themselves to the typical Government acquisition bureaucracy. In particular, the Competition in Contracting Act, Bayh-Dole & Rights in Technical Data, Truth in Negotiations Act, Contract Disputes Act, Procurement Protest System, and the Procurement Integrity Act (OUSD[AT&L], 2002) do not apply. Consequently, agencies can streamline competition and cost accounting, and agree to forgo intellectual property considerations. OTs require some level of cost sharing between Government and industry, or some other "in-kind" consideration in lieu of cost share. OTs are used much less frequently, and are much less constrained, than the FAR. For these reasons, anecdotally, procurement via OTA is typically considered "riskier" than procurement under the FAR. It is, therefore, not surprising that procurement professionals who are familiar with contracting under the FAR benefit from additional training regarding why and how OTs may be applied (Arendt et al., 2018).
- *Broad Agency Announcement*: The Broad Agency Announcement (BAA) is a competitive solicitation procedure used to obtain proposals for basic and

applied research and that part of development not related to the development of a specific system or hardware procurement. The BAA is described in FAR 6.102, Use of Competitive Procedures, and FAR 35.016, Broad Agency Announcements. The type of research solicited under a BAA attempts to increase knowledge in science and/or to advance the state of the art as compared to practical application of knowledge ("Broad Agency Announcements," 2017).

- *Existing contractual vehicle (GSA Schedule/GWAC):* GSA Schedules are fast, easy, and effective contracting vehicles for both customers and vendors. For GSA Schedules, GSA establishes long-term Government-wide contracts with commercial companies to provide access to millions of commercial products and services at volume discount pricing (GSA, 2018). The Government can also buy cost-effective, innovative solutions for information technology (IT) requirements through Government Wide Acquisition Contracts (GWACs). GWACs provide access to IT solutions such as systems design, software engineering, information assurance, and enterprise architecture solutions (GSA, 2017a).

Source Selection/Evaluation Approach

The evaluation strategy consists of the rationale used to evaluate the performance of an I-PREDICT Offeror who is proposing to work on the project. The evaluation strategy is used to ultimately make a source-selection decision and award the offeror a contract or agreement to perform. For this project, we examined four options:

- *White paper/paper proposal*: A white paper or paper proposal is a written persuasive argument that is used to respond to a Government solicitation. White papers are defined as shorter, more tailored written responses to a Government solicitation than a traditional full paper proposal which may be anywhere from dozens of pages to hundreds of pages in length. White papers/paper proposals may be written in response to a Request for Proposal (RFP), Statement of Work (SOW), Statement of Objectives (SOO), BAA, or Request for Project Proposal (RPP).
- *Oral proposal/demonstration*: "Oral presentations (or demonstrations) by offerors as requested by the Government may substitute for, or augment, written information. Use of oral presentations as a substitute for portions of a written proposal can be effective in streamlining the source-selection process. Oral presentations may occur at any time in the acquisition process, and are subject to the same restrictions as written information, regarding timing (see FAR 15.208) and content (see FAR 15.306). Oral presentations provide an opportunity for dialogue among the parties" (FAR 15.208, 2005; FAR 15.306, 2005).
- *Challenge event:* Challenges are related to demonstrations but are issued in terms of operational needs. Challenges are accompanied by mechanisms for evaluating proposed solutions and contractual terms for provider participation. Any challenge should be transparent and understandable. It should let challengers prove that their solution is the capability sought by the Government. This forces the Government to design a challenge that, if met, proves that the offered solution provides the needed capability. Typically, solutions take the form of simplified implementations, and evaluations assess how well a solution satisfies the need in a real-world operational environment. A well-crafted challenge, accompanied by clear and effective assessment

methodologies and appropriate contracting vehicles, leads to sound and effective acquisitions (Arendt et al., 2018).

- *Combination/hybrid*: A combination or hybrid approach may be any grouping of white paper, proposal, oral proposal, demonstrations, and/or challenge event used to make I-PREDICT award decisions to vendors for integration, component models, and/or biomechanical experiments.

Incentive Approach

The incentive approach is the rationale used to motivate a potential I-PREDICT integrator, model component providers, and biomechanical experimentalists to achieve cost, schedule, and performance requirements. Incentives may be monetary or non-monetary in nature. There were six incentive options that were considered for this project:

- *Cost-plus-fixed-fee contract:* "A cost-plus-fixed-fee contract is a cost-reimbursement contract that provides for payment to the contractor of a negotiated fee that is fixed at the inception of the contract" (FAR 16.3).
- *Cost-plus-incentive-fee contract:* "The cost-plus-incentive-fee contract is a cost-reimbursement contract that provides for the initially negotiated fee to be adjusted later by a formula based on the relationship of total allowable costs to total target costs. This contract type specifies a target cost, a target fee, minimum and maximum fees, and a fee adjustment formula. After contract performance, the fee payable to the contractor is determined in accordance with the formula" (FAR 16.4).
- *Time and materials contract:* "A time-and-materials contract provides for acquiring supplies or services on the basis of: (1) Direct labor hours at specified fixed hourly rates that include wages, overhead, general and administrative expenses, and profit; and (2) Actual cost for materials (with exceptions)" (FAR, 16.6).
- *Firm Fixed Price:* "A Firm-Fixed-Price (FFP) contract provides for a price that is not subject to any adjustment on the basis of the contractor's cost experience in performing the contract. This contract type places upon the contractor maximum risk and full responsibility for all costs and resulting profit or loss. It provides maximum incentive for the contractor to control costs and perform effectively and imposes a minimum administrative burden upon the contracting parties" (FAR 16.2).
- *Data rights*: "Within Government, the concern for intellectual property (IP) is primarily focused on the issue of 'data rights.' The term 'data rights' is a short-hand way to refer to the license rights that the Government acquires in two types of deliverables: technical data and computer software" (DoD OSA—Data Rights Team, 2014). For the I-PREDICT project, IP rights could be used as incentive for participants to deliver a successful model on time and budget. For example, the project could allow these participants to continuing using the I-PREDICT model even after the project was over for their own internal purposes. Such an arrangement would be of mutual benefit to the participants and the Government.
- *Combination/hybrid*: A combination/hybrid incentive strategy includes any grouping of cost plus fixed fee, cost plus incentive fee, time and materials, firm fixed price, and intellectual property as a part of an overall incentive package for a given contract or agreement. A combination strategy allows for the use of multiple approaches for varying tasks throughout the performance

period depending upon the performer and scope of the work being performed. The combination strategy allows the Government to take advantage of the benefits of multiple incentive approaches while mitigating their independent risks.

Intellectual Property Approach

"IP broadly refers to intangible 'creations of the mind'—inventions, literary and artistic works, unique business names and symbols, and so forth. Owners are granted certain exclusive rights to control the use and dissemination of their intellectual properties. ("Intellectual Property," 2017).

The IP strategy for a project is used to identify and develop a plan managing IP and related issues from the inception of the project throughout the life cycle. The key question that must be answered when developing an IP strategy is the following: What IP does the project need to maximize opportunities for competition and acquisition flexibility throughout the life cycle?

When the IP such as technical data or computer software are not available for the Government to distribute to a third party throughout the life cycle, it creates vendor lock. Vendor lock is where the Government finds itself inexorably tied to a vendor for key aspects of a project, thus giving the vendor a "monopoly" over the Government following contract award. As a result, the IP strategy must be identified and negotiated prior to contract award and evaluated during source selection. This is also a key factor when consideration is made for use of IP as part of an incentive package as described in the previous sub-section (DoD OSA—Data Rights Team, 2014). There were two IP options, the restricted/proprietary model and the open/competitive model, that were considered for the I-PREDICT project.

- *Restricted/proprietary model (DoD OSA—Data Rights Team, 2014):* When EITHER the data rights, OR the data deliverables do not allow the data to be used or released for competitive development or sustainment activities.
    - **Data Rights**: Standard License rights for technology developed 100% private expense: Limited Rights (LR), Restricted Rights (RR), or customary commercial license (CCL) for commercial computer software (CCS).
    - **Data Deliverables**: No contract requirements for delivery of necessary data or delivered data lacks technical information needed for development/sustainment or delivered with restriction.
- *Open/competitive model (DoD OSA—Data Rights Team, 2014):* When BOTH the data rights AND the data deliverables allow the data to be used or released for competitive development or sustainment activities.
    - **Data Rights**: Standard License rights for technology developed 100% Government funds or mixed funding: Unlimited Rights (UR), or Government Purpose Rights (GPR), respectively. Form, Fit, and Function (FFF) and Operation, Maintenance, Installation, and Training (OMIT) data qualify for UR regardless of funding.
    - **Data Deliverables**: Must have both a contract requirement to deliver the data, and deliverable data with the level of technical detail necessary for the desired development/sustainment activity.

Project Strategies

Upon identification and definition of the respective technical and acquisition options for the I-PREDICT project, they were combined to develop a set of project strategies for the project to consider before moving ahead. A total of three project strategies were developed to address the needs and complexity of the I-PREDICT project: Open Systems Architecture (OSA) strategy, Single model–individual performer strategy, and a single model–multiple performers strategy. Each strategy was defined and then each technical and acquisition option was assessed for its usability within that particular strategy. Strengths and weakness of the strategies were then outlined based on the usability of the technical and acquisition options and a final strategy selected. We describe each strategy below, providing strengths and weakness of each and providing justification for the chosen strategy.

Open Systems Architecture (OSA) Strategy

Definition and Overview

An Open Systems Architecture (OSA) is a technical architecture that adopts open standards supporting a modular, loosely coupled, and highly cohesive system structure. An OSA ensures that key interfaces within the system and relevant design disclosure are openly published and available for all. The key enabler for open architecture is the adoption of an open business model (OBM) that permits the collaborative innovation of numerous participants across the enterprise. The OBM permits shared risk, maximizes reuse of assets, and reduces total ownership costs. The combination of open architecture and an OBM permits the acquisition of an OSA that promise to yield modular, interoperable systems. OSA systems, by definition, allow components to be added, modified, replaced, removed, and/or supported by different vendors throughout the life cycle to afford opportunities for enhanced competition, innovation and maximize opportunities for acquisition flexibility (DoD OSA—Data Rights Team, 2013). Strengths and weaknesses of the strategy are outlined in Table 1. If this project strategy is selected, to ensure this flexibility, the project will use a solver-independent language, an open CAD anatomy to be used by all performers, and documented open standards for component interface requirements and variable component fidelity.

Table 1. Strengths and Limitations of the OSA Project Strategy

Strengths of the OSA Strategy	Limitations of the OSA Strategy
Future development of human body model components can be successfully and easily integrated into a full human body model.	This strategy risks delivering a modeling framework while under-delivering on an actual model due to focus on modeling framework.
Development of open standards for human body modeling may promote competition among model developers and drive future model development.	Development of the framework will add substantial complexity to the pre- and post-processing tools.
Solver independence will promote additional flexibility for the end users by allowing the users to leverage strengths of each solver.	A standardized language for human body modeling that results in identical simulation results across multiple solvers will require buy-in from the solver developers, which may require changes to the structure of their software.
	Limited coordination and communication between model developers and experimentalists may limit the required cooperation between these two roles.
	Potential for a lack of coordination and communication among project performers, which may impact the creation of OSA standards and may result in discrepancies in the capabilities and accuracy of model components.
	OSA strategy may prove to be difficult to execute with respect to overall contract management, as well as the associated incentive structure for performers due to the sheer number of variables the OSA strategy needs to consider.

Single Model–Individual Performer Strategy

Definitions and Overview

The single model–individual performer strategy was defined as a single performer executing or sub-contracting all the tasking related to the development of the I-PREDICT FNC. This strategy was designed to ensure the delivery of a functioning model that meets a set of pre-defined, Government-supplied requirements outlined in a statement of objectives. Strengths and weaknesses of the strategy are outlined in Table 2. If this strategy is implemented, solver selection will be made *a priori* to avoid the eventual performer delivering an I-PREDICT model that is incapable of integrating with existing DoD models or hazards, PPE, and vehicles. Freedom will be granted to the performer to use or acquire component-level models that they think are best suited for the full body model and to implement interfaces and fidelity levels they believe are most appropriate to accomplish the requirements. Anatomy can either be given to the performer by the Government or the performer would create their own anatomy. Experimental data gathered throughout the project will assist in informing these decisions. The delivery at the end of the period of performance will be a turn-key model that will be able to selectively alter fidelity, morph anatomy, and change posture as needed to accurately quantify human responses to military hazards.

Table 2. Strengths and Limitations of the Single Model–Individual Performer Project Strategy

Strengths of the single model-individual performer strategy	Limitations of the single model–individual performer strategy
Reduces Government project management burden as the Government will be only interacting with a single performer.	Selection of single performer may prioritize one aspect of the project over the other, with the effect of under-delivering on the needs of the Government stakeholder defined use cases.
Grants flexibility during the period of performance to rapidly alter fidelity levels, interfaces, pre- and post-processing tools, or other technical products.	A single performer is unlikely to be the premiere subject matter expert in development of each model component piece.
One performer streamlines deployment to the fleet.	Development of requirements to vet performers early in the project may limit performer flexibility later in the period of performance.

Single Model–Multiple Performers Strategy

Definitions and Overview

The single model–multiple performer strategy was defined as a group of performers executing explicitly defined tasking to deliver the I-PREDICT FNC. The strategy was designed to ensure some future flexibility while safeguarding against under-delivery. Strengths and weaknesses of the strategy are outlined in Table 3. If this project strategy is selected, the integrator role will be responsible for the delivery of the final model, they will be beholden to additional performers that will be delivering anatomy (a single representative human from a single performer), component level models based on project standard anatomy, and experimental results on the biomechanical response to inform the development of these models. Multiple performers will allow for the use of technical leads for different body regions that will be responsible for oversight over the model development and experimentation within that region, helping to ensure that the model is delivered with the state-of-the art technology. Technical leads will also ensure appropriate integration of experimental data gathered throughout the project into the component models and model validation. Development of interface and fidelity definitions via consultation between model component developers and the whole-body integrator will allow for model complexity where it is needed but simplicity where it is not, decreasing unnecessary computational expense.

Table 3. Strengths and Limitations of the Single Model–Multiple Performers Project Strategy

Strengths of the single model–multiple performers strategy	Limitations of the single model–individual performer strategy
Provides flexibility for rapid model updates as challenges arise while also ensuring that model component development is handled by subject matter experts.	Possibility of indecision if the performers disagree and no consensus can be reached.
Potential performers are already familiar with this leadership structure and the outlined technical options because of their exposure during the development of GHBMC human body model.	Managerial role by Government adds burden and shifts responsibility for under-delivery away from project performers and onto the Government.
Technical leads for body regions will help to ensure that the experimental data is being used to parameterize and validate a robust and accurate model.	Multiple model component providers open the possibility for component level models with varying degrees of accuracy.
The strategy allows for the development of fidelity and interface standards via the appropriate subject matter experts, granting flexibility to model developers and validation by the TC.	Government learning curve to stand up and manage a TC using this strategy.
Establishment of the TC provides the Government with an organizational structure to go back to if/when the model requires updates or maintenance	
Grants flexibility to leverage innovation from a wide range of partners from industry and academia while residing under a structure to enable efficient Government communication and collaboration with performers.	

Analysis of Alternatives

Figure 2 provides the technical option usability summary, and Figure 3 provides the acquisition options usability summary. Green highlighting indicates that an option can be used within the strategy with minimal limitations, gold highlighting indicates that an option is usable but has limitations that are considerable, and orange highlighting indicates that the limitations of the option supersede the strengths or that the option is not feasible for the strategy.

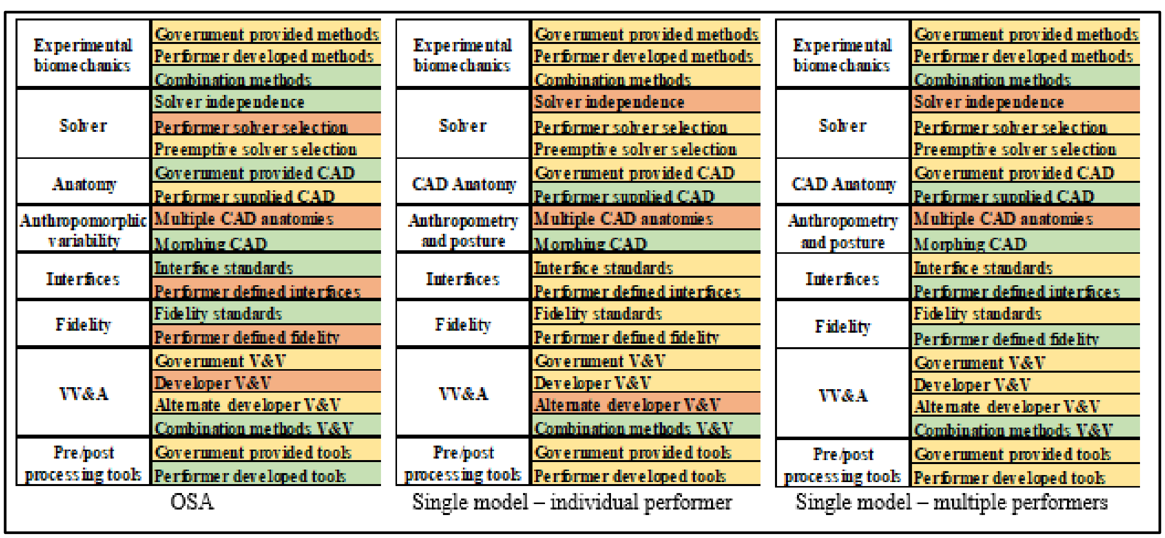


Figure 2. Assessment of Technical Options for I-PREDICT

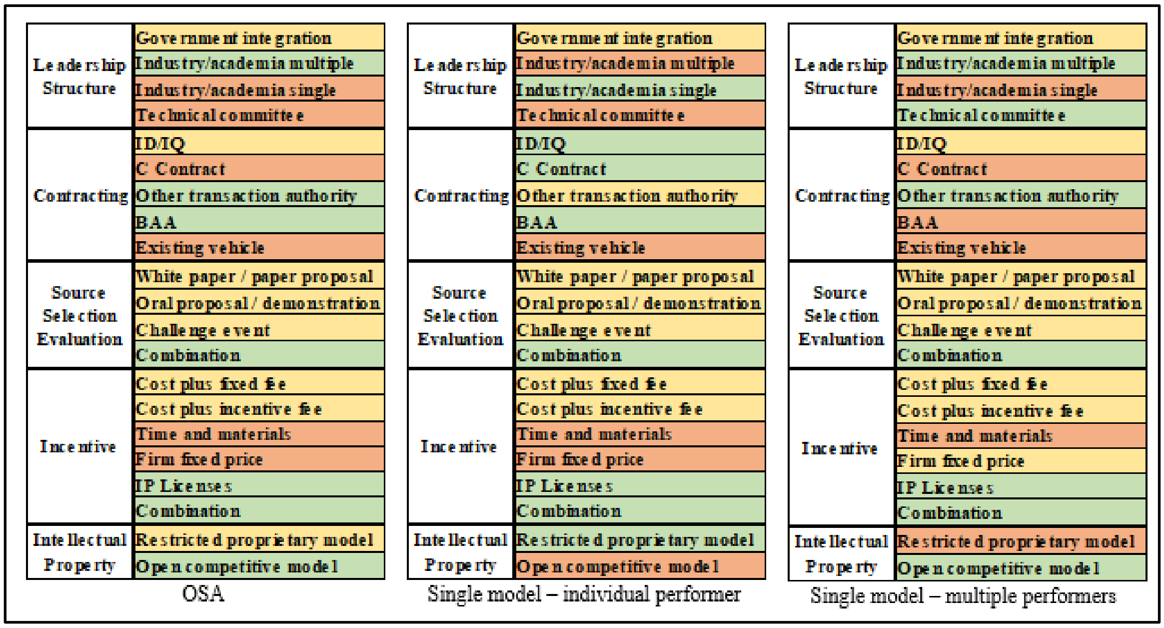


Figure 3. Assessment of Acquisition Options for I-PREDICT

Preferred Project Strategy

Based on the evaluation of the three strategies, I-PREDICT chose the single model–multiple performer strategy. Select technical and acquisition options within the strategy are indicated in Table 4. The single model–multiple performer project strategy increases communication and collaboration among potential performers which is expected to result in a higher quality, more robust FNC. The strategy accomplishes this in two ways. The first is by leveraging the MTEC OTA acquisition vehicle, which allows for the Government and performers to interact and collaborate more frequently and freely than traditional FAR-based acquisition vehicles. The second is by establishing a TC, which the other two approaches cannot use. In this capacity, the TC meets regularly to discuss project progress and oversees the model development and experimental work for each body region. The single model–multiple performer acquisition strategy will allow the project to be agile and adaptable as the requirements are updated throughout the period of performance. Increased communication resulting from the use of both the MTEC OTA and TC will allow rapid changes to the modeling and experimental work. The OTA vehicle allows additional RPPs to be posted and awarded on reduced time scales that traditional FAR-based acquisition approaches simply cannot achieve. Use of the TC and the OTA acquisition vehicle within this strategy also provides benefits beyond the other two approaches for future updates to the model throughout the life cycle as warfighter needs and potential use cases evolve. Having the TC in place with the OTA allows for the I-PREDICT model to live on in perpetuity, granting the Navy or any future Government user the ability to quickly release RPPs under the OTA and award performers for model updates as needed.

Implementation of the Single Model–Multiple Performer Strategy

To implement the single model–multiple performer strategy, the project seeks to leverage the Medical Technology Enterprise Consortium using OTA and a three-phased execution strategy. The steps below describe how ONR created a business relationship with MTEC, how the consortium model can be used when working under an OTA, and the three-phased strategy for bringing performers on contract to stand up the project.

Table 4. Selected options for single model–multiple performer strategy

Technical options	Experimental biomechanics	Combination methods
	Solver	Preemptive solver selection
	CAD Anatomy	Performer supplied CAD
	Anthropomorphic variability	Morphing CAD
	Interfaces	Performer defined interfaces
	Fidelity	Performer defined fidelity
	VV&A	Combination methods
	Pre/post processing tools	Performer developed tools
Acquisition options	Leadership structure	Technical committee
	Contracting	Other transaction authority
	Source Selection Evaluation	Combination/hybrid
	Incentive	Combination/hybrid with IP incentives
	Intellecutal Property	Open competitive model

Medical Technology Enterprise Consortium (MTEC)

MTEC is a collaboration between industry and academia to enable R&D, in cooperation with the U.S. Army Medical Research and Materiel Command (USAMRMC) and other Government agencies in the biomedical sciences. The purpose of MTEC is to protect, treat, and optimize the health and performance of U.S. military personnel. MTEC is a nonprofit corporation with the following principal objectives: (1) biomedical research and prototyping, (2) exploration of private sector technology opportunities, (3) technology transfer, and (4) deployment of intellectual property (IP) and follow-on production (Medical Technology Enterprise Consortium, 2018).

The scope of MTECs R&D falls into six primary scoping categories that fall within the scope of the OTs they execute. These categories include (1) Prevention, Diagnosis, and Treatment of Infectious Diseases; (2) Care of Combat Casualties; (3) Clinical and Rehabilitative Medicine; (4) Military Operational Medicine; (5) Medical Simulation and Information Sciences; and (6) Advanced Medical Technologies (Medical Technology Enterprise Consortium, 2018). In order to access the OTA and ultimately seek to leverage MTEC, ONR began a formal relationship with USAMRMC, for which ONR was required to do the following: (1) Completion of a Department of the Treasury Interagency Agreement (2700a Instructions and 2700b form); (2) Completion of an ONR Inter-Service Support Agreement (DD1144); (3) Acceptance of Department of the Navy General Terms & Conditions (GT&C); (4) Completion of an Annual Contracting/Assistance Agreement Workload Estimate for MTEC; (5) Submission of a pOTA–Project description overview for approval and acceptance by MTEC which included the following information about the project: (a) definition of the prototype to be developed and collaboration plans; (b) detailed requirements for the MTEC solicitation; (c) funding plan and any specific cost-share or private funding requested; (d) evaluation plan; criteria and plan for whitepaper/proposal evaluation; (e) project management plan with a Sponsor Office Technical Representative

(SOTR); (f) description of the end goal with the requirement with MTEC and any anticipated follow-on actions (Medical Technology Enterprise Consortium, 2018).

Consortium Model Using OTA

By gaining access to MTEC, ONR can leverage OTA in furtherance of the I-PREDICT single model–multiple performers project strategy granting access to numerous members of industry and academia to perform R&D in a highly streamlined manner relative to traditional FAR-based contracts. A consortium is defined as "an association of two or more individuals, companies, organizations, or Governments (or any combination of these entities) with the objective of participating in a common activity or pooling their resources for achieving a common goal" (Eilenberger, 2016). Consortia are open to all entities and entrance and participation is based on an entity's approval of an application, payment of a small annual fee, and the execution of a Consortium Member Agreement. This agreement provides rules and operating procedures that govern activity within the consortium to include procedures for handling intellectual property and data rights (Eilenberger, 2016). Consortia are often established for conducting shared research and development on technologies for the consortium's member companies, and in this case, also for the Government (Arendt et al., 2018).

The consortium model gets its statutory authority from the National Cooperative Research and Production Act (NCRPA) of 1993 (15 U.S.C. § 4301-06), which encourages innovation and collaboration between industry, academia, and the Government. The act also facilitates trade and helps to promote competition within the marketplace and is "aimed at reducing Governmental obstacles to the commercialization of new technology" (Bianco, n.d.).

The consortium model helps participants (e.g., Government, industry, and academia) to avoid duplication of effort and to be more efficient by sharing resources, information, resources, talent, and expertise. Furthermore, results of the research within the consortium are typically shared, making all members more competitive within the marketplace. It can be said that industry starts consortia for the same reasons that the Government does. John M. Eilenberger, Jr., Chief of the Contracting Office at the Army Contracting Command–New Jersey, noted some additional benefits of this consortium approach. These include that it creates relationships where they may not have otherwise occurred, allows for ease of communication, leverages capabilities, provides for clearer communication of needs and priorities, and can more easily obligate funds (Eilenberger, 2016). The Government establishes consortia for performing work within a given area of interest, technology profile, or capability gap. The Government's relationship with a consortium is typically solidified through a business agreement using OTA with a single point of contact: the Consortium Agent, a non-profit business entity. The Consortium Agent, or prime contractor, has a direct relationship with consortium members (industry, academia, small businesses, and non-traditional suppliers), or sub-contractors, typically through a Consortium Member Agreement and makes payment to these entities through a commercial or technology initiative agreement. The Consortium Member Agreement is referenced within the OTA, but it is not part of it. The Consortium Agent works directly with the consortium members, as shown in Figure 4. Once a consortium model using OTA is established, the Government can start work. The Consortium Agent earns a small administrative fee and is paid for the work accomplished by its members. The Consortium Agent then passes the remaining funds on to the consortium entity that "wins" the work through a commercial or technology initiative agreement. It is important to note that the Government can utilize both RDT&E and Operations and Maintenance (O&M) funding, which offers flexibility in choosing the work and initiatives to be accomplished, executed, and, ultimately, funded. Using this model, the

Government can purchase prototypes, conduct intensive R&D, and even execute a sole-source follow-on procurement for additional products (Arendt et al., 2018).

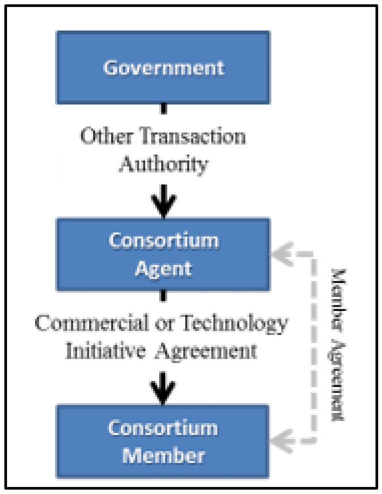


Figure 4. Government -Consortium Relationships

The method enabled by the consortium model lowers the barriers to entry for industry, non-traditional suppliers, small businesses, and academia that tend to be very innovative but may shy away from the bureaucracy of Government acquisition. This model allows the Government to tap into colleges and universities, laboratories, and small innovative companies, experts, and teams without the typical barriers put forth by federal regulations and policies that do not apply when using OTA. Furthermore, this model incentivizes innovation, collaboration, and communication, and has proven to be a win-win for both the Government and member entities of the consortium. Using this model, the Government can purchase prototypes, conduct intensive R&D, and even execute a sole-source follow-on procurement for additional product (Arendt et al., 2018).

I-PREDICT's Three-Phase Implementation Strategy

To maximize use of the OTA-Consortium model afforded to ONR via MTEC, a three-phased implementation of the single model–multiple performers implementation strategy was developed. This strategy can easily be tailored and applied for others where OTAs are being used to perform R&D and develop prototypes with highly complex technical requirements. The approach presented below allows for maximum collaboration not only between the Government and performers, but also amongst the performers so that they may work together in a highly flexible environment to deliver cutting-edge solutions to the sponsor. The three phases for this acquisition strategy begin with Phase 1, consisting of a simple white paper selection. Those offerors who receive favorable evaluations in Phase 1 are down-selected for participation in Phase 2. Phase 2 is an oral proposal/demonstration. Those proposers who receive favorable evaluations in Phase 2 are down-selected for Phase 3. Phase 3 is the use of a Technical Development Team (TDT) to collaboratively develop requirements for a statement of work. All participants then become part of the TC and thereby are eligible to submit a full proposal in response to the statement of work. Those members who are not selected as performers following full proposal evaluations would remain as participants on the TC to serve in an advisory capacity to the Government, with opportunities to bid against future work on I-PREDICT as opportunities arise. Each of these three phases of the acquisition strategy are addressed in more detail below.

Phase I Solution Brief/White Paper

Offerors for the I-PREDICT project will be required to submit a Solution Brief, which describes the overall technical concept and approach along with the viability toward achieving stated outcomes of the I-PREDICT project. The value of using the solution brief under the OTA versus a traditional paper-based proposal is the streamlined format (limited to only 10 pages) and evaluation process that can help narrow down contenders from pretenders. To complete the solution brief, offerors will be required to provide the following information:

- *Title page* that references the RPP and includes the Offeror's contact information
- *Executive summary* that provides a brief description of the methodology and technology the Offeror will employ, why it is relevant to the proposed objectives, and how the Offeror has completed similar work in the past.
- *Methodology/technology approach* that outlines the proposed methodology in sufficient detail to show a clear course of action as it relates to the topic area of interest.
- *Relevant experience* that identifies any work of a similar nature that could be used to gauge the effectiveness and worthiness of the technical or methodological approach.
- *Company viability* which provides a quick overview of the company or entity

Solution briefs will then be evaluated based upon the following four criteria and offerors will then be down-selected for participation in the Phase 2 Oral Presentations/Demonstrations:

- Feasibility of the proposed solution and its alignment with the RPP's topic area;
- Relevancy of the proposed methodology/technology/solution to the topic area with special interest toward any innovation or previously underutilized capabilities;
- Strength of the organization/team proposed to complete the work and its financial stability to potentially continue the maturation of the system beyond the scope of the I-PREDICT RPP; and
- Inclusion of nontraditional or small business participation or a 1/3 cost share.

Phase II Oral Presentations/Demonstrations

In Phase 2, it is envisioned that the Offeror(s) will provide a "pitch" of the proposed project during an in-person meeting with ONR. The pitch is intended to provide more details about the viability of the proposed work outlined in Phase 1. Offerors who are invited to give a Solution Brief Pitch are provided with the specific areas of interest to be included in the pitch at the end of Step 1 during the time of invitation to advance into Step 2. Offeror(s) will be asked the following information in their pitch:

- *Description*: The Offeror will provide a more robust description of their approach.
- *Progress*: The Offeror will describe the milestones that will be used to measure progress during the period of performance and describe the oversight managerial methods that will be employed to maintain a quality and timely performance.

- *Relevant experience*: The Offeror will convey details related to past performance(s) that demonstrate relevance to the scope of the proposed work and build confidence in the team's capabilities.
- *Effectiveness (opportunity and risk):* The Offeror will identify opportunities (e.g., reduction in cost or schedule and/or improvement in performance) and risks within each appropriate project Cost, Schedule, Performance measure of effectiveness. This should include a mitigation plan for each identified risk item.
- *Prototype*: A description of how this work effort will facilitate the development of the I-PREDICT prototype must be described.
- *Data rights assertions*: The Solution Brief will identify all proprietary and/or intellectual property involved in the efforts and any associated restrictions that may possibly affect the Government's use of the property in any way whatsoever.

Phase III Technical Development Team/Technical Committee Selection

It is envisioned that those offerors who are down-selected to participate in the TDT are referred to as finalists. These finalists are invited to attend a TDT meeting in person to help the Government scope out the technical requirements for the program in more detail to ensure that the project is organized to achieve its goals within the designated period of performance. These technical requirements will be worked into a Request for Project Proposals (RPP). Only members of the TDT will be invited to respond to the RPP. Finalists who are members of the TDT will be provided a participation stipend for their support in the TDT. The RPP to which the TDT members will respond includes the following components for evaluation:

- *Statement of work*: The Offeror is required to provide a detailed SOW. Based on the results of the Technical Evaluation, the Government reserves the right to negotiate and revise any or all parts of the SOW. Offerors will have the opportunity to concur with revised SOW and revise cost proposals as necessary.
- *Cost proposal submission*: Section I: Cost Proposal Narrative required. Separately, Section II: Cost Proposal is required.
- *Warranties and Representations*: If Nontraditional Defense Contractor participation is proposed, Warranties and Representations are required.
- *Royalty Payment Agreement or Additional Research Project Award Assessment*: Each Offeror is required to select either the MTEC Additional Assessment Fee or the Royalty Agreement (available on the MTEC members only website), not both, and submit a signed copy with the proposal.

It is envisioned that if a member of the TDT is not selected for funding, ONR may extend an invitation for them to become members of the TC. The advancement of non-awardee members of the TDT to become members of the TC is only by invitation. Invitations for non-awardees to participate on the TC may be renewed or rescinded on an annual basis at the discretion of ONR. Members of the TC who are Awardees will be provided funding commensurate with their final negotiated statement of work and accepted cost proposal. Members of the TC who are not Performers but who have been invited for TC participation only will receive an annual TC participation stipend.

Program Execution

The TC will operate with eight distinct roles for executing the program, some held by the Government, one with an FFRDC (MITRE), and several with industry/academia performers. The TC will be responsible for overseeing and executing the work performed under the program. In cases where there are new technical requirements, the TC may then again jointly prepare a SOW and offer it for proposals to the TC members. Once the TC is initially formed, all members will be asked to execute their roles based on and outlined in Figure 5 and the following descriptions:

- *Office of Naval Research Program Officer*: ONR is the funding agency managing the program. The ONR Program Officer will have ultimate decision authority for program goals, communication paths, responsibilities, program activities and scope, and delivery from all participants. The ONR Program Officer is supported by in-house staff and contractors.

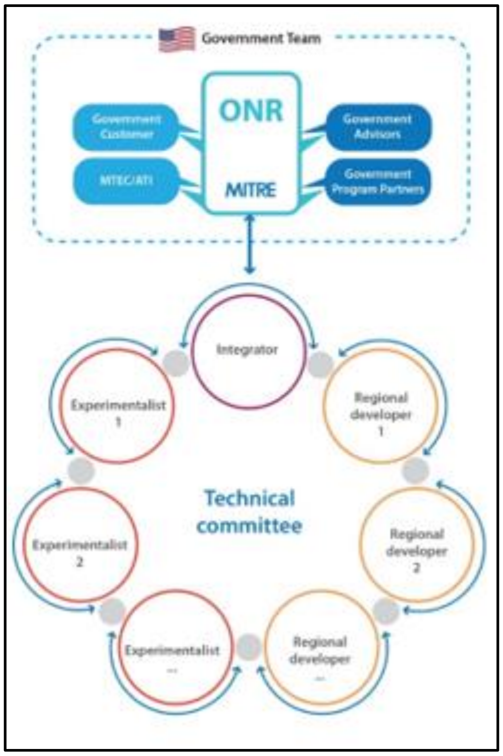


Figure 5. Technical Committee Structure

- *Government Program Partners*: One or more Government personnel from programs of record in the Navy or other service may be asked to provide direct program support including technical advice, use case development for program products, outreach to communities of interest (e.g., operational, medical, materiel, test and evaluation), interagency integration, advocacy, technology transition, and/or technology readiness assessments among potentially other program functions. Program support may include direct participation in internal ONR program meetings at the request of ONR. Government Program Partners may be consulted for their opinions on internal ONR program decisions. Government Program Partners will have no program decision authority.

- *Government Customers*: Internal funding for FNCs at ONR require one or more DoD organizations (hereafter, "Government Customers") to become signatories to a Technology Transition Agreement (TTA), so that FNC investments are measurably relevant to TTA signatory organization mission objectives. Government Customers provide vital information about their organization's mission objectives in the form of use cases for how I-PREDICT Program technologies might fulfill their organization's missions. Government Customers will have no program decision authority. Government Customers retain control of their funding, and TTA signatory authority. Government Customers will have a substantial influence over program priorities and execution.
- *Government Advisors*: One or more Government personnel from programs of record in the Navy or other service may provide advice to ONR on their organization's mission objectives and/or technical matters relevant to the I-PREDICT Program. Government Advisors will have no program decision authority.
- *MITRE*: As a DoD trusted agent, The MITRE Corporation will inform and advise ONR on technical matters and to mitigate programmatic and technical risk, to serve as a hub for communication among participants and stakeholders to develop technology acquisition strategies, and other activities as required by the Government. MITRE will have no program decision authority, except as delegated by the ONR Program Officer.
- *Government Team*: The "Government Team" is expected to consist of ONR, Government Program Partners, Government Customers, Government Advisors, and The MITRE Corporation as detailed above. This Team will form the nexus of decision making for the Government on all matters for program performance and delivery of technical products, in consultation with TC members and funded performers as detailed below. Ultimate decision authority rests with ONR.
- *Program performers and Technical Committee Members*: During various phases of the I-PREDICT project, funded performers from academia, industry, and/or Government may serve as performers and/or TDT members during the formulation of the funding vehicle and its goals. Once funded as performers and/or otherwise invited to serve as TC members, they will be expected to offer technical advice on program goals and scope, and to respond to ONR and the Government Team as described above. TC members will have no program decision authority.
- *Medical Technology Enterprise Consortium (MTEC)/Advanced Technology International (ATI):* In its capacity as Consortium Manager for MTEC, ATI will act as an administrative liaison between the ONR Program Officer/MITRE and offerors, TDT members, and TC members. In this role, ATI will publish ONR Program Officer–approved documentation related to I-PREDICT; communicate messages or sharing of information to offerors, TDT members, and TC members on behalf of the FFRDC and ONR; collect, organize, and share formal solicitation responses and inquiries from I-PREDICT participants with the ONR Program Officer; provide management and administration of funds dispersal to program performers upon approval from ONR; and provide management and administration of the base MTEC member agreement and

individual project performer agreements with input provided by ONR and MITRE where appropriate.

Conclusion

Accurate prediction of injuries and the resulting functional incapacitation under varying military hazard conditions would provide the ability to design safe equipment and behavioral practices, and to allow commanders to weigh operational risks during the planning and execution of missions and to allocate resources appropriate to those risks. The sooner these types of highly complex, innovative technologies can be transitioned to the field, the sooner warfighters can reap the benefits of this kind of cutting-edge research. However, designing a program to deliver a computational model that provides these capabilities is fraught with technical complexity, making the acquisition of such a model challenging. This paper provides a description of a decision framework that was developed for evaluating technical and acquisition options to meet project needs, building and evaluating potential project strategies, and the process for execution of the selected strategy. Additionally, this paper outlines the use of the OTA acquisition vehicle and MTEC along with a three-phase implementation strategy for award selection to MTEC members.

It is expected that the decision framework and implementation strategy developed may be used Navy-wide or across other military Services for any R&D program that requires acquisition flexibility coupled with highly collaborative technology development. The TC aspect of this process allows a way ahead to ensure that continued improvements and upgrades of the chosen solution can be transitioned to the fleet throughout the life cycle. Ultimately, the decision framework presented herein and its supporting processes may allow programs to benefit from a unique partnership with performers, while streamlining deployment and fielding, consequently yielding safer PPE, vehicles, weapons, and training regimens for the warfighter.

References

Acquisition guides. (n.d.). Retrieved March 14, 2018, from http://www.navair.navy.mil/nawctsd/Resources/Library/Acqguide/contract-numbers.htm

Arendt, M., Novak, R., Roe, S., & Jennings, L. (2018*). The MITRE challenge-based acquisition handbook* (PRS 17-4018). The MITRE Corporation.

Authority of the Department of Defense to Carry Out Certain Prototype Projects Act of 2017, 10 U.S.C. § 2371b.

Bianco, D. (n.d.). Research and development consortia. Retrieved March 14, 2018, from http://www.referenceforbusiness.com/encyclopedia/Pro-Res/Research-and-Development-Consortia.html

Broad Agency Announcements. (2017). Retrieved March 14, 2018, from https://www.arl.army.mil/www/default.cfm?page=8

DoD Open Systems Architecture—Data Rights Team. (2013). *Open systems architecture contract guidebook for program managers, v1.1*. Retrieved from http://bbp.dau.mil/docs/IP_Strategy_Brochure_FINAL_em.pdf

DoD Open Systems Architecture—Data Rights Team. (2014). *Intellectual property strategy*. Retrieved from http://bbp.dau.mil/docs/IP_Strategy_Brochure_FINAL_em.pdf

Eilenberger, J. M. (2016, January). *Other Transaction agreements*. Presented at the Chemical Biological Defense Acquisition Initiatives Forum, Falls Church, VA. Retrieved from http://www.ndia.org/-/media/sites/ndia/meetings-and-events/3187-

sullivan/divisions/chemical-biological-defense-acquisition-initiatives-forum/documents/16-acc-nj-ota.ashx?la=en

Federal Acquisition Regulation, 48 C.F.R. § 15.208 (2018).

Federal Acquisition Regulation, 48 C.F.R. § 15.306 (2018).

Federal Acquisition Regulation, 48 C.F.R. § 16.2 (2018).

Federal Acquisition Regulation, 48 C.F.R. § 16.3 (2018).

Federal Acquisition Regulation, 48 C.F.R. § 16.4 (2018).

Federal Acquisition Regulation, 48 C.F.R. § 16.6 (2018).

GSA. (2017a). Governmentwide acquisition contracts (GWACs). Retrieved March 14, 2018, from https://www.gsa.gov/portal/content/104874

GSA. (2017b). Indefinite delivery, indefinite quantity contracts. Retrieved March 14, 2018, from https://www.gsa.gov/portal/content/103926

GSA. (2018). About GSA schedules. Retrieved March 14, 2018, from https://www.gsa.gov/portal/category/100615

Intellectual property. (2017). In *Acquipedia online encyclopedia*. Retrieved March 14, 2018 from https://www.dau.mil/acquipedia/Pages/ArticleDetails.aspx?aid=7bfcfeee-b24b-4fdd-ad7b-046437729519

National Cooperative Research and Development Act, 15 U.S.C. § 4301-06.

Under Secretary of Defense for Acquisition, Technology, and Logistics. (2002). *"Other Transaction" (OT) guide for prototype projects.* Washington, DC: Author.

Bridging the Gap: Improving DoD-Backed Innovation Programs to Enhance the Adoption of Innovative Technology Throughout the Armed Services

Amanda Bresler—serves as Vice President of Business Development for PW Communications, Inc. (www.pwcommunications.com). She launched and manages the company's strategic initiative focused on democratizing the U.S. federal marketplace for innovative solutions providers. Prior to joining PW Communications, she served as COO for Maurice Cooper Brands. Bresler is passionately involved in numerous philanthropic causes and currently serves on the board of The Bresler Family Foundation and AlmaLinks. She is a member of the prestigious Milken Young Leaders Circle and Business Executives for National Security (www.BENS.org). She graduated cum laude from Georgetown University's McDonough School of Business. [abresler@pwcommunications.com]

Abstract

For over 60 years, Department of Defense (DoD)–backed innovation programs have played an outsized role in the narrative surrounding military innovation. While these programs provide valuable benefits, this paper specifically evaluates their effectiveness as a means of enhancing the adoption of innovative new technology throughout the armed forces. To assess how companies that participated in DoD-backed innovation programs performed in the defense sector subsequent to program completion, we compiled a data set of more than 1.29 million defense contract awards over seven years and analyzed the distribution of these awards across a data set of more than 8,000 DoD-backed innovation program award recipients. The results demonstrated that nearly half of participants achieved no meaningful growth in their defense business after program participation; and the small, innovative companies that did successfully bridge program participation into additional DoD business rarely contracted with customers outside of their initial branch sponsor. Through surveys and interviews of key stakeholders, we identified several causes for the low rate of adoption of participants' technology across the armed forces, and we present concrete recommendations for how the Department can address these problems to better leverage DoD-backed innovation programs as a means of enhancing force readiness.

Introduction

America emerged from World War II as the world's leading economic, political, and technological superpower, and this position remained largely uncontested for the duration of the 20th century. Post 9/11, however, the global landscape began to shift. Intensified competition with rival powers, including Russia and China, the advent of information warfare and a proliferation of threats, and the Global War on Terror and conflicts in the Middle East now pose acute challenges for both American hegemony as well as national security. Additionally, while in decades past Department of Defense (DoD) research often produced revolutionary technological breakthroughs for the civilian sector, commercial innovation now increasingly outpaces the DoD. The Department must therefore attempt to modernize as a whole, while it simultaneously competes to identify and integrate the most cutting-edge technological innovation. Furthermore, as elucidated by Defense Secretary James Mattis in the 2018 National Defense Strategy, "Success no longer goes to the country that develops a new technology first, but rather to the one that better integrates it and adapts its way of fighting" (DoD, 2018). To maintain a strategic overmatch demands mission planning and execution across the forces and requires unprecedented levels of Department-wide cooperation and communication. Just as innovation can serve as a force multiplier, it can also severely degrade military productivity and lethality if it is siloed within a single service branch or command structure. The DoD, therefore, faces a dual challenge today: it must not only identify and attract innovative and modernizing solutions providers, but also integrate revolutions in military technology across the forces as quickly and seamlessly as possible.

The DoD has widely acknowledged that the prowess and success of America's armed services demand ongoing, supported collaboration with private sector innovators, and increasingly so in light of 21st-century military and national security challenges. For more than 60 years, DoD-backed innovation programs have played an outsized role in the narrative surrounding military innovation. They consume billions of taxpayer dollars annually; enable thousands of disruptive technologies to enter the highly adaptive, risk-averse DoD ecosystem every year; and produce a network of companies with a rare combination of innovative prowess and DoD past performance.[1] Given that a competitive advantage in today's mission environment demands rapid, force-wide integration of innovative technologies, it is essential that these programs also connect the capabilities that they foster to as many prospective DoD customers as possible. However, while these programs implicitly and explicitly market themselves as "points of entry" for small, innovative companies into the overall DoD market, limited scholarship exists to evaluate how participant companies perform in the defense sector in the years following program completion, and if their capabilities are leveraged by the military at large. Our research sought to fill this gap; specifically, we were interested in understanding if and how DoD-backed innovation programs have evolved to ensure that the military remains flexible, agile, and advanced in an environment where rapid integration is essential.

DoD Innovation Programs

To evaluate DoD-backed innovation programs as a means of enhancing the adoption of new technology across the forces, it is important to first understand their history, how the largest are structured, and where they fit within the broader landscape of DoD innovation. The history of America's DoD-backed innovation programs begins in 1957, when the Soviet Union's surprise launch of Sputnik left the American people and its leaders fearful that the United States had lost its technological edge. President Eisenhower responded by creating the Advanced Research Projects Agency (later renamed the Defense Advanced Research Projects Agency, or DARPA) the following year. According to the DARPA website, it has a "singular and enduring mission: to make pivotal investments in breakthrough technologies for national security." Today, DARPA runs over 250 research and development (R&D) programs, all designed to further their core mission (DARPA, n.d.). DARPA is unique in the DoD innovation community as it exists as a stand-alone research agency that funds innovative research within industries and also conducts its own research. DARPA uses contracting authorities to solicit solutions directly and participates in the Small Business Innovation Research program (SBIR) and the Small Business Technology Transfer program (STTR) to fund research conducted by small businesses.

[1] In the Department of Defense's Fiscal Year 2019 Budget Request, $13.6 billion has been earmarked by the DoD for Science and Technology research in FY19. Over $286 billion has been requested for modernization efforts across the Department (Office of the Under Secretary of Defense, 2018).

The Small Business Administration (SBA) launched the SBIR program in 1977 to "support innovation through the investment of federal research funds in critical American priorities to build a strong national economy" (Small Business Association, n.d.-a). It established its identically-structured sister program, the STTR program, in 1992, designed to help further scientific research with potential for broader commercial benefit. SBIR/STTR programs run in three phases, all of which are restricted to small businesses.[2] In Phase I, companies establish the technical merit, feasibility, and commercial potential of their project over the course of one year in exchange for $150,000 in federal funding. For Phase II, participant companies continue to grow and develop their technology with additional funding not to exceed $1 million over a two-year period. A product with "commercial potential" may then proceed to Phase III. SBIR/STTR does not fund Phase III directly, but many federal agencies with SBIR/STTR authority provide follow-on contracts or funding to support this additional development. Phase III is intended to be the primary means of transitioning new technologies into the broader service branches or agencies that need them. The armed services and defense agencies each run their own SBIR/STTR programs and define the topics for which they are seeking small business applicants.

Based on the success of SBIR/STTR and the need to accelerate the fielding of innovative technology, the DoD created the Rapid Innovation Fund (RIF) in 2011, marketed on the DoD's Defense Innovation Marketplace site as "a collaborative vehicle for small businesses to provide the Department with innovative technologies that can be rapidly inserted into acquisition programs that meet specific defense needs." Much like SBIR/STTR, the armed services and various defense agencies run their own RIF programs and define their own project specifications. However, unlike SBIR/STTR, and despite its mission statement, RIF permits companies of any size to participate. The RIF process starts with a Broad Agency Announcement (BAA) and the request for companies to submit a white paper. The sponsoring agency reviews all submissions and invites shortlisted companies to submit a full-scale proposal. The company with the most competitive proposal receives the RIF award. One of the primary objectives of RIF is to better and more rapidly connect research and technologies developed in the SBIR/STTR program to a wider audience within the DoD. As such, from 2011 to 2015, RIF awarded more than half of its $1.4 billion in contract awards to companies that had previously participated in SBIR/STTR (Bujewski & Purdy, 2017).

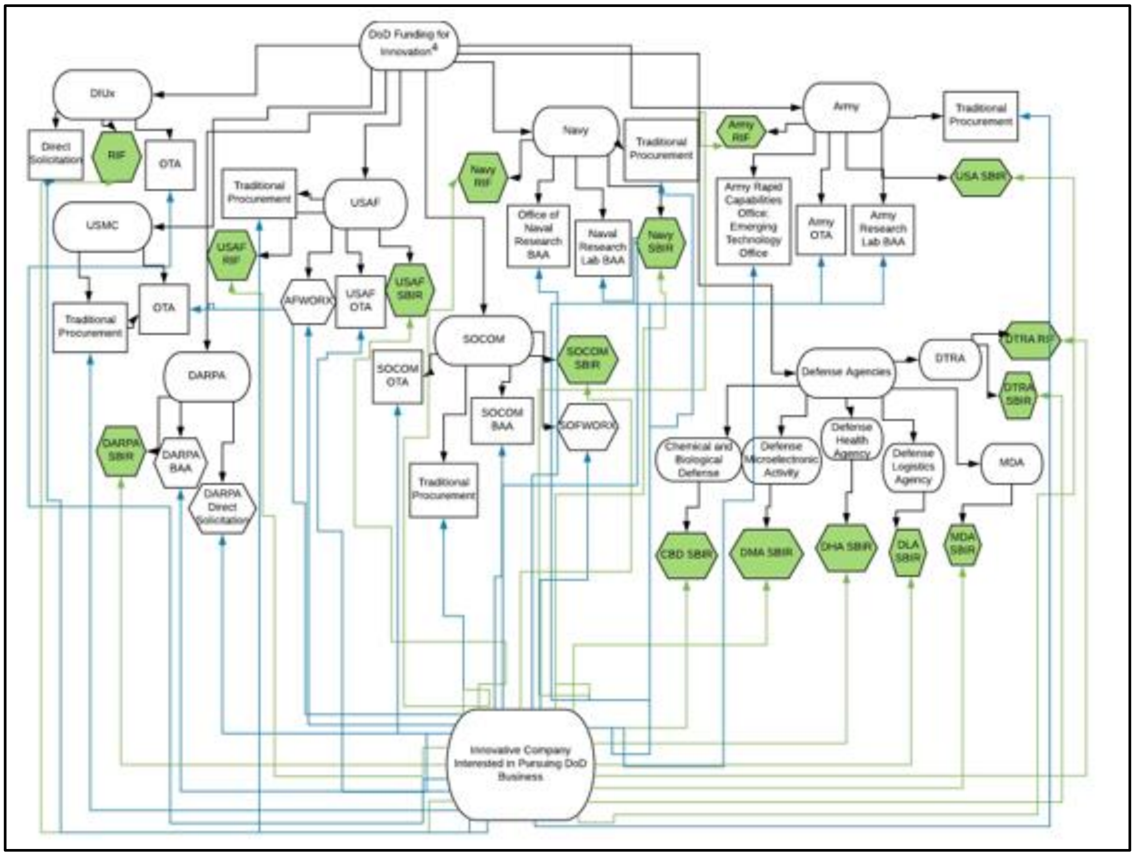
In 2015, Secretary of Defense Ash Carter established Defense Innovation Unit Experimental (DIUx) in support of the Third Offset Strategy initiated by Chuck Hagel in 2014. DIUx, like its predecessors, is designed to fund innovative companies with the purpose of solving national defense problems. It determines project specifications and areas of interest in concert with DoD entities. Companies of any size can respond to a DIUx solicitation by submitting a solution brief. According to DIUx's website, briefs are typically evaluated within 30 days, and shortlisted companies are then invited to submit a full-scale proposal and begin negotiations for a pilot contract. Unlike the other DoD innovation programs, however, DIUx utilizes "Other Transaction Authority" (OTA), a contract vehicle that streamlines the funding process and according to the DIUx website, enables them to fund projects in 90 days or fewer. According to U.S. Code 10 2371b, OTA funds include a designation for

[2] The SBIR size compliance guide defines a *small business* as "a business with 500 employees or fewer."

prototype projects deemed critical to enhancing the mission effectiveness of the military or to improve systems already in use by the armed forces. The armed services may contract up to $250 million annually for projects that meet OTA criteria, provided the primary awardee for a given project represents a "nontraditional defense contractor" or a small business. As such, OTA also affords DIUx greater latitude in allocating funding, making it more agile and more appealing to nontraditional solutions providers.

In addition to these five DoD-backed innovation programs, the DoD has approximately 50 additional external funding programs, 20 of which are designed to rapidly integrate technology (Small Business Technology Council, 2014). The DoD also continuously stands up new DoD-backed innovation programs and utilizes OTA funds independently to support internal innovation efforts.[3] This process compounds across the DoD, and innovative companies interested in pursuing the defense market—and the service branches and Departments that seek to collaborate with them—face a paralyzingly complex web of prospective routes and access points, as shown in Figure 1.

[3] Since January 2015 alone, the DoD has created DIUx, the Warfighting Lab Incentive Fund, AFWERX (an Air Force innovation initiative), SOFWERX (SOCOM's innovation initiative) among others (Gibbons-Neff, 2016; Kaplan, 2015; West, 2018).



Note. Programs reflected in our data set are highlighted in green.

Figure 1. Access Points for Innovative Companies into the DoD

This bloated, hyper-bureaucratic system also elucidates the Department's cripplingly risk-averse culture, one of the most acute obstacles thwarting DoD innovation efforts. Paralyzed by a "fear of failure," decision-makers are unwilling to experiment, collaborate, and take risks. This culture has resulted in countless "stove-piped" initiatives that are developed without cross-communication and operate in parallel. Given the Department's ongoing struggle to modernize and remain innovative, it should invest in building robust, long-term relationships with the innovative companies that do manage to navigate this web successfully and achieve proven DoD past performance. Consequently, rather than evaluating the effectiveness of innovation programs as a means of addressing singular, branch-specific requirements, our research sought to evaluate if and how the largest and most prominent DoD-backed innovation programs improve the rate of adoption of innovative technology force-wide.

Program Participants

In order to determine whether the companies fostered in DoD-backed innovation programs achieved force-wide adoption, we first looked at how those companies performed in the defense sector after program completion. Although there is no data available to determine the results of these programs in general, one indicator of a company's DoD performance is the number of defense contracts that it wins. Thus, we began by examining the number of defense contracts won by DoD-backed innovation program participants in the years following program participation. We focused our quantitative research on SBIR/STTR and RIF for several reasons: they are the largest of the DoD-backed innovation programs and the only hubs that publish complete lists of their program participants, which made it possible for us to gather a significant data set; and because all branches of the armed forces and all defense agencies participate in both SBIR/STTR and RIF, the data set is not only large but also comprehensive.

We first combined the publicly available lists of RIF project award recipients between 2011 and 2015 (103) and DoD-sponsored SBIR/STTR award recipients between 2013 and 2016 (8,158) from the SBA database. Both RIF and SBIR/STTR are rife with serial users, which means there were dramatically fewer unique companies in this ecosystem relative to total awards. For instance, from 2013–2016, the Physical Optics Corporation received 236 DoD sponsored SBIR/STTR awards, and Charles River Analytics received 129. Intelligent Automation received 138 SBIR awards from 2013–2016 and won seven RIF contracts from 2011–2015.[4] As such, we adjusted the 8,261 total awards to control for repeat usage and isolate unique companies, which resulted in a data set of "Program Participants" that contained 1,140 companies.

In order to understand how these 1,140 companies performed in the defense sector subsequent to their program participation, we scraped and filtered more than seven years' worth of publicly available defense contract award data from FBO.gov from January 1, 2011–January 15, 2018.[5] Because FBO.gov publishes only unclassified prime contract awards, our data did not include classified contracts or information about companies' performance as a subcontractor or teaming partner on DoD contracts. Nevertheless, the set contained more than 1.29 million defense contract awards, which we then filtered to isolate the defense contracts specifically awarded to the 1,140 companies in our Program Participant data set: a total of 13,449 defense contracts.

[4] While the SBA system is rife with flaws in how it reports and names companies, our data tool controlled for these errors to an extent by removing duplications and recognizing slight differences in company names to avoid errors in contract attribution (Cordell, 2018).
[5] These timeframes were selected to enable analyses of complete data sets. At the time this information was compiled, RIF awards were publicly available from program inception in 2011 through 2015. SBIR awards are all publicly available but the most recent, complete data is from 2013–2016.

Supplier Retention

The distribution of these 13,449 contracts across the 1,140 Program Participant companies in our data set produced striking results, as shown in Figure 2.

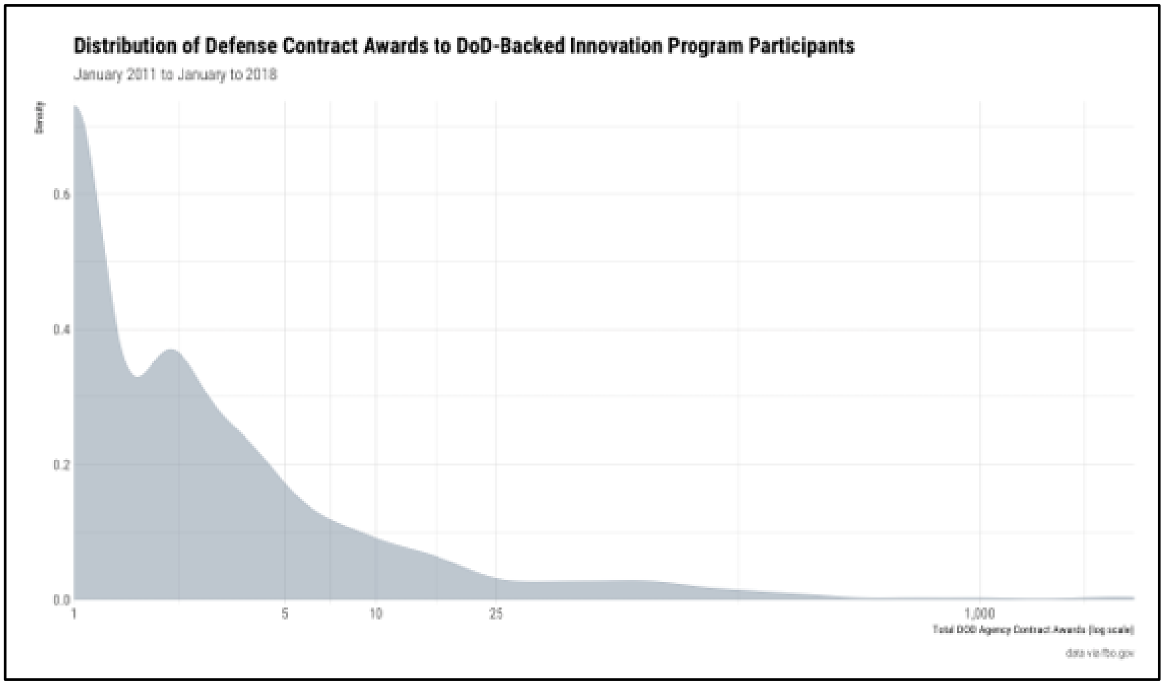


Figure 2. Distribution of DoD Contracts to Companies That Participated in a DoD-Sponsored SBIR/STTR Program or RIF Program (2013–2016 and 2011–2015, Respectively)

It is immediately apparent that a majority of participant companies won few if any follow-on defense contracts, as demonstrated by the concentration of companies on the left-tail of the distribution.

In fact, we determined that a startling 26% of the companies in our vast data set (~296 Program Participant companies) won zero defense contracts after completing their DoD-backed innovation programs.[6] Another 22% of companies in our set (~250 Program Participant companies) won only one single defense contract following their RIF/SBIR/STTR award. In other words, nearly half of the program graduates (48%) received fewer than 2% of the total 13,449 defense contract awards.

The format of publicly available data limited our ability to control for specific timeframes, as contract award data was fixed from January 1, 2011 to January 15, 2018 (per the parameters of the "scraper" we used). We assessed the entire data set of participant companies over this period of time; thus a company that completed SBIR/STTR or RIF in 2011 was evaluated over seven years, while a company that graduated in 2016

[6] ~296 companies exist as one contract in our data, but that contract is their award from SBIR/STTR or RIF.

was evaluated over two years. However, all companies in the data set had at least 13 months to win defense contracts, and given the size of our overall data sets and the overwhelming share of companies that failed to win DoD contracts following program participation, it is unlikely that the timeframe limitation significantly affected our results. Additionally, we recognize that small, innovative companies often lack the capacity and infrastructure to bid/perform as Prime contractors and instead subcontract or team with large Strategic Integrators (SIs) to expand their defense business. Unfortunately, we could not examine innovation hub participants' performance as subcontractors or teaming partners because no large, publicly available data sets exist containing that information. Presumably, a portion of program participants that won few or no defense contracts as a Prime did achieve some growth in their defense business indirectly. However, the significant and stark results of our data incontrovertibly reveal a trend and thus remain deeply troubling. Namely, that a significant share of participant companies went through the involved and drawn-out process of applying for an innovation hub, proved their relevance to a branch-specific requirement and received an award, and expended significant resources to develop the capabilities requested by their branch sponsor; and these efforts failed to bridge into any meaningful growth in their direct defense business. The implications are perhaps most concerning with respect to force readiness. The capabilities of these ~547 companies were honed by initial branch sponsors based on critical requirements for which private sector collaboration was deemed essential. For almost 300 of these companies, their capabilities were never procured directly by other defense customers, or from within the sponsor branch, or from other branches, and the other ~200 companies faced essentially the same fate.

The distribution of the 13,449 contracts across the 1,140 companies also made it clear that a small subset of companies won a disproportionate share of the remaining 98% of contracts in our data set. While RIF describes itself as a "vehicle for small businesses," the Participant Data data set includes names like 3M, BAE Systems, General Dynamics, Raytheon, and other experienced contractors. To better understand the effects of these outliers on our data, we isolated companies from our Participant Data data set that had won 50+ DoD contracts from January 1, 2011 to January 5, 2018. Forty companies in our data set fit those criteria.

These 40 companies, or approximately 3.5% of our total Program Participant data set, collectively received a staggering 80% of the defense contract awards in our overall defense contract award data set, as depicted in Figure 3.

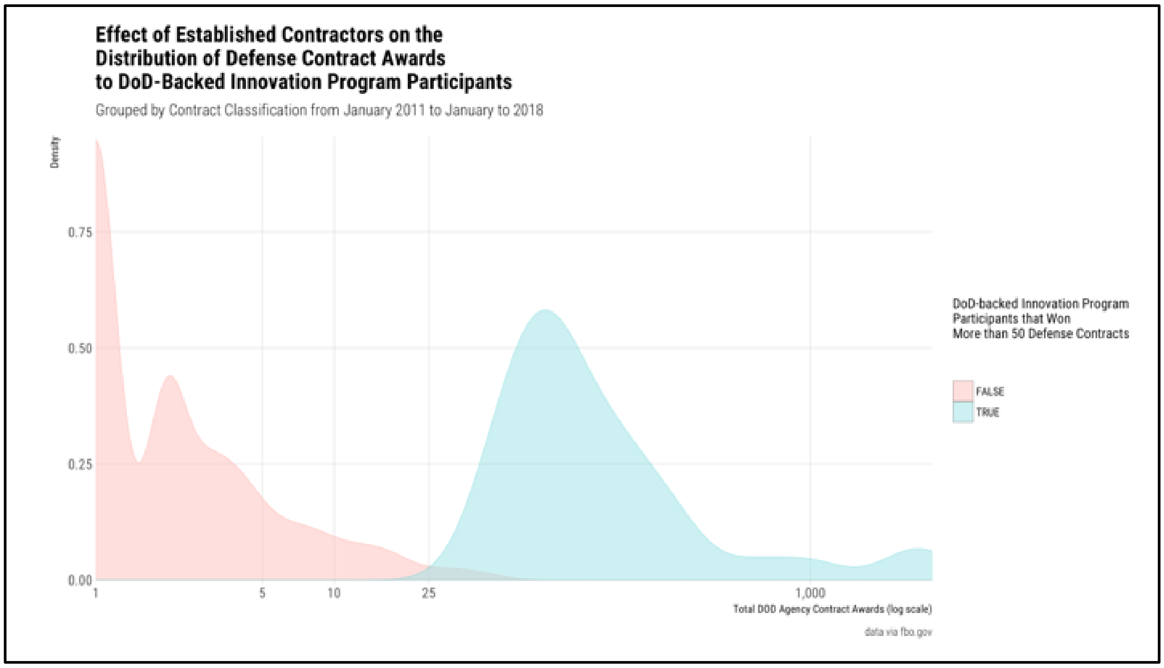


Figure 3. Distribution of DoD Contracts to DoD-Sponsored SBIR/STTR Participants and RIF Participants (2013–2016 and 2011–2015, Respectively) by Company Size

Removing these outlier companies from the data set, the percentage of the remaining 1,100 companies that won zero defense contracts after their SBIR/STTR or RIF participation also jumped by 10%, from 26% to 36%.

For the companies that stand to gain the most benefit from these programs, the reality—that the programs contributed to no meaningful growth in participants' direct defense business—is even more acute. Furthermore, these programs are designed to provide the forces with greater access to emerging technologies, whereas our forces are otherwise limited to the technologies that the established DoD contractors present to them. Our data suggests that these innovation hub programs have in fact become yet another channel for legacy contractors to gain DoD market share.

Customer Diversity

While our initial analyses demonstrated that innovation hubs fail to convert a large portion of participants into viable DoD suppliers, we also sought to examine, for emerging technology companies that did manage to survive the transition from DoD-backed innovation program into the defense market, how broadly their capabilities were integrated across the forces. To do so, we examined whether participants won contracts with defense customers other than their initial sponsor branch. Limitations to the functionality of our data analytics tools meant that we could not examine customer diversity for our entire data set, so we created a subset of our Participant Data to include companies that won between two and 12 defense contracts subsequent to program completion: 360 companies, or 32% of our Participant data set. As 48% of participants won zero or one defense contract after program completion and 3.5% of participants accounted for a startling 80% of all defense contract awards, the data left a narrow subset of companies that fell somewhere in the "middle"—that is, non-entrenched contractors that had successfully bridged the gap from program participation into a healthy, modest defense contracting business. By focusing on companies that won between two and 12 contracts, we were able to test customer diversity using a sizable data set from this narrow middle.

Our research found that, for a striking 76% of companies in this subset, all of their defense contract awards came from their initial sponsor branch.

In other words, 76% of these companies achieved no customer diversity, suggesting that even when the DoD manages to retain a supplier honed in an innovation hub, it largely fails to integrate the company's capabilities across the forces.

DIUx: Micro-Analysis

DIUx has garnered widespread attention as an especially promising approach to the DoD's innovation problem (Kaplan, 2016; Maucione, 2018; Williams, 2017). Its funding is slated to increase by more than 70% in Fiscal Year 2019, and Secretary Mattis recently lauded it, saying, "There is no doubt in my mind that DIUx will not only continue to exist, it will actually … grow in its influence and its impact on the Department of Defense" (Lanier-Chappellet, 2017). Although DIUx data was too limited to include in our broad quantitative analyses and despite the more acute impact of limitations on a small sample size, we felt it nevertheless important to determine if DIUx retains and disseminates its participants' capabilities more effectively than its predecessors.

We created a "DIUx Participants" data set using the company names marketed on the DIUx website as of Q4 2017 (31 companies). We then filtered our DoD contract data set to isolate DoD contracts awarded to these 31 companies from January 1, 2015 (the year DIUx was founded) through January 15, 2018: a total of 440 defense contracts. We immediately recognized that entrenched government contractors have utilized DIUx as well, which profoundly affected the results. Specifically, Rockwell Collins was one of the 31 companies in our "DIUx Participants" data set. The company is a legacy DoD contractor; thus unsurprisingly it won 408 of the 440 total defense contract awards from our FBO award data set. The distribution of the remaining 32 contracts across the 30 other DIUx participants in our data set reflected the same troubling trends we identified in our SBIR/STTR/RIF analyses:

Four DIUx participants won between three and six DoD contract awards after program completion, and the remaining 26 companies—or 81% of DIUx participants in our set—won zero or one DoD contract.

Despite its laudatory attention, DIUx, like its predecessors, does not appear to position private sector innovators for long-term success in the defense sector. Rather, DIUx engages with participants solely to address singular, branch-specific projects and does not disseminate supported capabilities to other prospective DoD customers thereafter.

Surveys Methods

While we concluded that DoD-backed innovation programs are not effective distribution channels for small, innovative companies into the broader defense market through our quantitative analyses, to better understand potential causes of these program failures, we conducted surveys and interviews with individuals from three key stakeholder groups:

- Representatives from DoD-backed innovation programs
- Representatives from companies that participated in a DoD-backed innovation program
- Members of the DoD community

We developed unique survey and interview questions for each of these stakeholder groups and contacted a total of 159 prospective respondents. While our quantitative analyses were limited to SBIR/STTR and RIF data and a small set of DIUx data, our qualitative research encompassed the broader landscape of DoD-backed innovation programs. We received a total of 20 responses: seven representatives who work for a DoD-backed innovation program, three of whom elected to participate in an in-depth follow-on interview; five representatives from participant companies, one of whom elected to participate in an in-depth follow-on interview; and 10 individuals from the general DoD community, three of whom elected to participate in an in-depth follow-on interview.

Survey & Interview Feedback

The anecdotal evidence consistently pointed to the same programmatic failures we identified in our quantitative analyses and most significantly, helped us better understand the reasons for these problems. The key issues revealed by our survey and interview data include the following:

- *DoD-backed innovation programs do not educate participants on how to succeed in the broader defense sector.*

  The majority of participants surveyed noted that their innovation program offered no instructions/education for how to identify or bid on government contracts after program completion. They also indicated that the program did not prepare them to support a broad base of DoD customers. Likewise, when program managers were asked how their program helps a participating company identify other relevant DoD requirements, it was clear that no formal, institutionalized process exists. For instance, one program manager explained, “We naturally become aware of [relevant opportunities] from our contractors, conferences, even our customers sometimes,” and another stated that they rely on “collaborative conversations” to find additional DoD opportunities for participants with promising capabilities. Similarly, when asked how they would improve the DoD innovation program, one respondent recommended, “Provide some sort of bridge to help small business survive the gap between the end of the ‘innovation contract’ and the follow on sustainment work.” Given that the innovation programs do not educate participants on how to find and bid on government contracts, it is therefore not surprising that, as indicated by our quantitative data, nearly half of program participants fail to win defense contracts after program participation.
- *DoD-backed innovation programs do not market participants’ capabilities to the broader armed services community.*

More than half of the program managers that provided feedback indicated that their programs have no formal process for circulating information about participants' capabilities to the broader armed services community. One of the most frequent comments from members of the DoD community was that they receive very few briefings on the projects their own branch funds and almost never receive information on the capabilities of companies funded by other branches. When members of the DoD do learn about the capabilities of companies that have participated in a DoD-backed innovation program, it is not because the programs are marketing them effectively. According to those surveyed, at best they might hear about a company through a random, one-off initiative like a "quarterly update" mailer that references an innovative technology, or "infrequent[ly] hosted industry days." Usually, they learn about the technology through a chance meeting with a colleague who is familiar with it. Our quantitative results are further explained by the fact that the innovation programs do little if any marketing of participants' capabilities to the broader DoD community. Simply put, a customer cannot buy something it does not know exists.

- *DoD-backed innovation programs do not track the performance of participant companies in the years following program completion.*

  Our qualitative research confirmed the troubling fact that DoD-backed innovation programs do not follow a systematic, consistent process for tracking the performance of participant companies after program completion. Many DoD-backed innovation programs do not track program participants at all; others may track some companies inconsistently or on an ad hoc basis. Without tracking participants, DoD innovation programs cannot discern how these companies fare in the defense sector or at large. They do not see changes in a company's performance year to year; they do not receive updates on a company's capabilities developments; and if a company changes its name, changes its leadership, or moves its headquarters, that information is not recorded in a central database. Perhaps most concerning, they do not monitor the long-term effectiveness of significant investments of public funds into DoD-backed innovation programs. The absence of a formal process for tracking the long-term performance of participants is the equivalent of a venture capital firm not tracking the performance of its portfolio companies. It elucidates these programs' egregious supplier retention problems and signals that these programs are not concerned with the long-term success of the companies that they fund.

- *A company's failure to win DoD contracts after program participation does not necessarily correlate to a lack of demand for that company's capabilities.*

  Our survey and interview feedback also demonstrated that, while in some cases innovation program participants may not achieve widespread adoption across the DoD due to the specificity of a sponsor project, a company's failure to win DoD contracts after program participation does not necessarily correlate to a lack of demand. For example, we interviewed the CEO of Monterey Technologies Incorporated (MTI), a company that develops mission planning software systems and has been an active defense contractor since 1984. MTI received one Navy-sponsored RIF award in 2012 and six Navy-sponsored SBIR awards between 2013 and 2016, and between January 1, 2011 and January 15, 2018, MTI won three defense contracts, all from Navy customers. While MTI has only won Navy contracts, there have been 144 DoD contracts awarded for "mission planning services" over the past five years alone, signaling Department-wide demand for their solution. The CEO shared that recently an Army Airborne officer serendipitously learned about MTI through a Navy contact. The Army had been handling mission planning by hand, and MTI had the potential to address this ongoing problem. He called MTI's CEO directly, and as a result of the ad hoc outreach they are now collaborating on a pilot project. This example demonstrates not only bona fide demand for MTI, but also the inadequacy of current communication channels across the services.

The interview also offered valuable insights relative to how small, innovative companies attempt to stay competitive in the defense sector. In addition to their own disparate business development efforts, MTI pays a business intelligence firm to compile a monthly report of relevant federal opportunities. MTI then undertakes an involved process to qualify relevant opportunities, many of which are designed for larger contractors. As such, after identifying a relevant opportunity, MTI must then develop a strategy to identify and engage with the appropriate teaming partners. Large integrators, from MTI's experience, do not give preference to companies that have participated in DoD-backed innovation programs. While their program participation has provided them with valuable funding and strategic customer opportunities within the Navy, SBIR/STTR and RIF have not offered them unique leverage in the defense market.

The experience of another company in our Participant data set, "Enomalies," tells a similar story. In 2016, Enomalies participated in Phase II of an SBIR project and, like many DoD-backed innovation program participants, has not won a single defense contract since. Enomalies specializes in advanced imaging and field-ready prototyping, and the Navy sponsored its SBIR project to further develop a Rapid Synthetic Environment Tool (Small Business Association, n.d.-b). The tool scans an area and quickly creates a 3D model that strategists and planners can walk through and interact with. Enomalies's tools also have the ability to rapidly scan objects and print prototypes on 3D printers (Enomalies, n.d.). Despite the fact that Enomalies has not won any follow-on defense contracts, since 2016 alone, there have been 26 defense contract awards for 3D printing systems and four for 3D scanning services. According to their website, Enomalies supports a broad range of commercial customers and appears to remain a viable company. Presumably, then, it is neither a lack of demand nor an inability to perform that has kept Enomalies from winning a defense contract since its SBIR award. Instead, it suggests that DoD innovation programs fail to adequately connect their participants to prospective DoD customers.

Further Analysis & Recommendations

The fact that DoD-backed innovation programs fail to market their participants to prospective DoD customers has myriad consequences. It results in lost revenue for the innovative companies who, as our data illustrated, do not become robust DoD suppliers. Perhaps more concerning, however, are the consequences for the strength and readiness of our forces at large. If the armed services are not made aware of the capabilities supported in DoD-backed innovation programs, they cannot adopt them—instead, they either do not modernize, or they conduct redundant market research. Furthermore, these programs introduce private sector innovators into the defense ecosystem, provide them with past performance, and then fail to nurture them as long-term suppliers. These results are costly, and they also damage the reputation of the Department: innovative suppliers undertake the complex, expensive, and time-consuming process of participating in a DoD-backed innovation program only to find that it does not contribute to meaningful growth in their defense business. They are expected to undertake full life-cycle sales processes for each individual DoD customer, in contrast to how large commercial customers operate (where new vendors typically undergo a vetting period and if successful, their product or service is distributed across the organization's broader portfolio). If these problems are not addressed, innovative companies will increasingly forgo public sector opportunities altogether, and innovation programs will both fail to attract top innovators and fail to realize their full potential as force multipliers. However, with the proper reforms, DoD-backed innovation programs do have the potential to drive vast improvements in the readiness of our armed forces. As stewards of significant tax dollars, the defense sector should focus on maximizing their investments in innovation and R&D over the long term. To do so first and foremost requires that DoD-backed innovation programs maintain consistent, clean, and accessible records about their projects and program participants.

As such, we first and foremost recommend creating an "innovators database"—a central, searchable database containing information about all DoD-backed innovation program participants.

The innovators database would be populated by innovation hub program managers and the companies themselves, and would contain company basics, information about the company's capabilities, and details about the projects that the company has supported for both government and commercial customers. A company would be required to maintain updated records in the database or face penalties. Likewise, program managers would be held accountable for maintaining these records for the five to 10 years following a company's program completion. The innovators database would enable DoD-backed innovation programs to easily share information with stakeholders across the armed forces on the capabilities of their participant companies. These stakeholders could access the database directly and search its rich pool of proven solutions providers to identify prospective suppliers. It would reduce redundancy in market research and requirements development and improve the likelihood of broader, more rapid integration of proven capabilities. Furthermore, the money saved by eliminating redundancy could be reallocated across the Department. A version of the innovators database would be made available to Prime Contractors and SIs as well, to help facilitate teaming arrangements.

For an entity as rigid and adaptive as the DoD, simply making more information available to key stakeholders will not drive change. The Department must encourage key stakeholders to better leverage DoD-backed innovation programs.

Specifically, we recommend that the DoD implement an incentive program that requires defense contracting entities and large defense contractors to allocate a set *percentage of business to "Proven Innovators," or companies that have graduated from a DoD-backed innovation program.*

This incentive program would be similar to existing set-aside programs designed to increase opportunities in the federal market for various historically disadvantaged groups. Proven Innovators would earn their "set-aside" status upon completion of their DoD-backed innovation program. Like the existing cadre of set-aside programs, contractors and contracting officers would be required to achieve minimum engagement levels with Proven Innovators and would be motivated to do so through tax incentives and other benefits. This system would force broader adoption of leading technologies across branches, as the different branches (and the contractors that serve them) would be required to stay abreast of various sponsors' projects by using the innovators database, and it would force collaboration and cross-communication in an otherwise siloed environment. This incentive program would also add tremendous value to participant companies, lending them a concrete advantage in the broader contracting environment. Participants would see an increase in their federal business overall and an increased willingness on the part of large integrators to support and engage with them.

To that end, in addition to these initiatives, DoD innovation hubs must also educate their participants on the fundamentals of the defense contracting sector—how to identify and bid on contracts, how to find and team with other firms, how to register for set-asides, etc.—in order to make them more competitive.

Innovation hubs could potentially partner with the SBA, which already offers similar training. The more benefits DoD-backed innovation programs can offer, the better positioned they are to attract and serve the most discerning, talented technology companies.

Conclusions and Future Work

While the aforementioned recommendations are crucial first steps toward improving the efficiency and functionality of DoD-backed innovation programs, further research is required to fully understand participants' experience in the defense sector after program completion. First and foremost, additional research is required to further analyze the large data sets we compiled. A more sophisticated analytics tool would enable us to draw additional and more extensive conclusions, such as if and how non-DoD agencies leverage the technologies fostered by DoD-backed innovation programs and the impact of the rampant serial usage on the effectiveness of these programs as a whole. Additional data and analyses are also needed to evaluate participants' performance as subcontractors and teaming partners on defense contracts. Relative to our recommendations, further research is also required to determine how best to structure the "Innovators Database" as well as the set-aside incentive program to enhance supplier retention and broad integration.

Once DoD-backed innovation programs are reformed and improved, additional research is needed to understand the most effective ways to market-to and attract the best and brightest innovators into these programs. Additional research is also needed to better understand opportunities for streamlining, merging, or eliminating redundant or ineffective entities throughout the vast, complex DoD innovation landscape, including the individual, specialized innovation programs within each branch and combat command. From the vantage points of both national security and efficiency, it is also critical that cutting-edge capabilities are integrated throughout the whole of the government. Therefore, further research is required to determine how to position participants in DoD-backed innovation programs for success as suppliers to both DoD and non-DoD customers.

It is essential for America's national security that the armed services have access to the best and brightest new technologies, and the continuous investment in DoD-backed innovation programs over the last 60+ years has resulted in a vast infrastructure of programs and access points for new, cutting-edge solutions. DoD-backed innovation programs provide both participants and the Department with valuable benefits. Participant companies benefit tremendously from the funding these programs provide, which allows them to grow their business, develop new research and prototyping, and commercialize. The programs also introduce participants to the various idiosyncrasies of the contracting space and to the unique challenges facing the armed forces. However, today's adversaries and threat environment demand unprecedented synchronicity and collaboration across the armed forces. It is, therefore, essential that all branches of the military adopt revolutions in technology as quickly and seamlessly as possible to ensure consistent standards in warfighting capabilities; to ensure fighters across the services can communicate and share information; and to ensure that the Department leverages its full potential and buying power to appeal to the small, private sector innovators it so desperately seeks to attract and retain. Unfortunately, DoD-backed innovation programs have failed to serve as viable entry-points for emerging capabilities into the broader defense sector and have failed to enhance the integration of these emerging capabilities across the forces. A superior military not only serves domestic national security interests, but also enables the United States to fulfill its role in safeguarding peace, prosperity, and freedom. As such, the DoD must stop at nothing to keep the forces agile, modern, and at the forefront of new technologies, and that includes adapting its existing resources—in this case, its vast network of innovation hubs—to be as effective as possible.

References

Bujewski, T., & Purdy, E. (2017). *Rapid Innovation Fund: Program overview* (PowerPoint slides). Retrieved from http://www.defenseinnovationmarketplace.mil/resources/rif/RIF_Overview(Feb2017).pdf

Cordell, C. (2018, February 2). SBA innovation program hobbled by database errors, GAO says. *FedScoop*. Retrieved from https://www.fedscoop.com/agency-database-errors-corrupt-progress-small-business-innovation/

Defense Advanced Research Projects Agency (DARPA). (n.d.). About DARPA. Retrieved from https://www.darpa.mil/about-us/about-darpa

DoD. (2018). S*ummary of the National Defense Strategy of the United States of America 2018*. Washington, DC: Author. Retrieved from https://www.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Enomalies. (n.d.). Products and services. Retrieved from http://enomalies.com/category/products-and-services/

Gibbons-Neff, T. (2016, May 25). Why the U.S. military turned a hipster tattoo parlor into a Special Operations lab. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/checkpoint/wp/2016/05/25/why-the-u-s-military-turned-a-hipster-tattoo-parlor-into-a-special-operations-lab/

Kaplan, F. (2016, December 19). Defense secretary tries to get around the Pentagon bureaucracy in a quest for innovation. *Technology Review*. Retrieved from https://www.technologyreview.com/s/603084/the-pentagons-innovation-experiment/

Lanier-Chappellet, J. (2017, August 11) Secretary Mattis ready to "enthusiastically embrace" DIUx. *FedScoop*. Retrieved from https://www.fedscoop.com/secretary-mattis-ready-enthusiastically-embrace-diux/

Maucione, S. (2017, October 19).OTA contracts are the new cool thing in DoD acquisition. *Federal News Radio*. Retrieved from https://federalnewsradio.com/acquisition/2017/10/ota-contracts-are-the-new-cool-thing-in-dod-acquisition/

Office of the Under Secretary of Defense. (2018). *Program acquisition cost by weapon system: United States Department of Defense Fiscal Year 2019 budget request.* Arlington, VA: Author. Retrieved from http://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2019/FY2019_Weapons.pdf

Small Business Association. (n.d.-a). About SBIR. Retrieved from https://www.sbir.gov/about/about-sbir

Small Business Association. (n.d.-b). Enomalies. Retrieved from https://www.sbir.gov/sbirsearch/detail/406391

Small Business Technology Council. (2014, July). *The Rapid Innovation Fund Program.* Retrieved from http://sbtc.org/wp-content/uploads/2014/07/SBTC-RIF-Report.pdf

West, H. (2018, January 12). AFWERX is smart risk for innovative solutions. Retrieved from http://www.af.mil/News/Article-Display/Article/1414478/afwerx-is-smart-risk-for-innovative-solutions/

Williams, L. C. (2017, December 7). DoD looks to DIUX for the future of acquisition. *Federal Computer Weekly*. Retrieved from https://fcw.com/articles/2017/12/07/diux-sasc-acquisition-future.aspx

Acknowledgments

The author would like to thank Alex Bresler for his assistance with the data analytics, Jillian Kutner for her significant contributions as the lead Research Assistant, and Phyllis Bresler and Blaise Durante for their feedback throughout the editing process. Additionally, the author would like to thank the individuals who offered valuable feedback via interviews and surveys.

Panel 10. Software Ecosystem: Sustaining Workforce Skillsets

Wednesday, May 9, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Captain Melvin Yokoyama, USN,** Commanding Officer, SPAWAR Systems Center Pacific ***Software Is Consuming DoD Total Ownership Cost*** Brad R. Naegle, LTC, U.S. Army (Ret.), Naval Postgraduate School ***Exploring the Department of Defense Software Factbook*** Christopher Miller, Carnegie Mellon University Forrest Shull, Carnegie Mellon University David Zubrow, Carnegie Mellon University ***DoD'S Software Sustainment Ecosystem: Needed Skill Sets and Gap Analysis*** Forrest Shull, Carnegie Mellon University Michael McLendon, Carnegie Mellon University Christopher Miller, Carnegie Mellon University

Captain Melvin Yokoyama, USN—assumed command of Space and Naval Warfare Systems Center Pacific (SSC Pacific) in May 2017.

A native of the Big Island of Hawaii, CAPT Yokoyama holds a Bachelor of Science degree from Jacksonville University and a Master of Science degree in Systems Engineering (Information Warfare and Electrical Systems Engineering subspecialties) from the Naval Postgraduate School.

A designated Naval Flight Officer in the Unrestricted Line (URL) community, CAPT Yokoyama is a member of the Defense Acquisition Corps and a Level II Joint-qualified Officer.

At-sea assignments include duties as Commanding Officer and Executive Officer of Tactical Air Control Squadron TWELVE (TACRON 12), Air Officer for the Bonhomme Richard Amphibious Ready Group, Officer-in-Charge Commander Task Group 57.1 (Manama, Bahrain and Al Udeid, Qatar), Officer-in-Charge Commander Task Group 72.5 (Misawa and Okinawa, Japan), EP-3E Operations Officer for Fleet Air Reconnaissance Squadron ONE (VQ-1), and ES-3A Electronic Warfare Mission Commander for Fleet Air Reconnaissance Squadron SIX (VQ-6) deployed onboard the USS *Enterprise* (CVN-65) and USS *John F. Kennedy* (CV-67).

Acquisition and shore assignments include duties as Principal Assistant Program Manager (PAPM) at PEO C4I Information Assurance and Cyber Security Program Office (PMW-130), Assistant Program Manager (APM) for SPAWAR's Data Center and Application Optimization (DCAO) program office, Chief of Staff for the Department of the Navy's Data Center Consolidation Task Force, Deputy Commander and Tomahawk Land Attack Missile (TLAM) C4I Officer for U.S. Strategic Command's (USSTRATCOM) Cruise Missile Support Activity Pacific, and as a graduate student in the Information Warfare curriculum at the Naval Postgraduate School.

In 2016, CAPT Yokoyama volunteered to deploy on a one-year Global War on Terrorism support assignment (GSA) to the U.S. Embassy, Baghdad, Iraq, where he served as the senior military advisor to the Iraqi Minister of Defense in support of Operation Inherent Resolve.

CAPT Yokoyama has received a number of personal awards but is most proud of his unit awards because of the teamwork they represent.

Software Is Consuming DoD Total Ownership Cost

Brad R. Naegle, LTC, U.S. Army (Ret.)—is a Senior Lecturer within the Graduate School of Business and Public Policy at the Naval Postgraduate School (NPS). He has served as the Academic Associate for Program Management Curricula at NPS and has managed Distance Learning graduate degree and certificate programs. He currently serves on the Navy's software community of practice. On active duty, he was the PM, 2½ Ton Extended Service Program and the Deputy PM for Light Tactical Vehicles. He holds a master's degree in Systems Acquisition Management from NPS and a BS from Weber State University in economics. He is a graduate of the Command and General Staff College and Combined Arms and Services Staff School. [bnaegle@nps.edu]

Abstract

Department of Defense (DoD) software-intensive systems and the software content in other systems will continue to grow and may dominate total ownership costs (TOC) in the future. These costs are exacerbated by the fact that, in addition to contracted development costs, the bulk of software sustainment costs are also contracted. All of these factors indicate that DoD system software will continue to be a very expensive portion of TOC.

The software engineering environment remains immature, with few, if any, industry-wide standards for software development or sustainment. The Defense Acquisition System (DAS) is significantly dependent on mature engineering.

System software size and complexity are key indicators of both development costs and sustainment costs, so initial estimates are critical for predicting and controlling TOC. Unfortunately, the software size estimating processes require a significant amount of detailed understanding of the requirements and design that is typically not available when operating the DAS without supplementary analyses, tools, and techniques. Available parametric estimating tools require much of the same detailed information and are still too inaccurate to be relied upon. Similarly, understanding the potential software complexity requires in-depth understanding of the requirements and architectural design.

It is clear that the DoD must conduct much more thorough requirements analyses, provide significantly more detailed operational context, and drive the software architectural design well beyond the work breakdown structure (WBS) functional design typically provided. To accomplish this, the DAS must be supplemented with tools, techniques, and analyses that are currently not present.

System Software Development and Sustainment Environmental Challenges

While many of the TOC initiatives apply equally to hardware-oriented systems and software-oriented systems, there are some significant differences in both the software development and sustainment environments that need to be considered to gain better software-TOC performance. Understanding these differences in environments will help managers at all levels better manage the acquisition management system and provide the warfighter with systems that are easier and cheaper to sustain.

The Software Engineering Environment (Naegle, 2015)

The software engineering environment is not mature, especially when compared to hardware-centric engineering environments. Dr. Philippe Kruchten (2005) of the University of British Columbia remarks, "We haven't found the fundamental laws of software that would play the role that the fundamental laws of physics play for other engineering disciplines" (p. 17). Software engineering is significantly unbounded because there are no physical laws that help define environments. There is significant evidence for software engineering immaturity, and it is nearly impossible to find widely accepted, industry-wide development standards, protocols, architectures, or formats. There is no dominant programming language, design and development process, standard architectures, or software engineering tools, which means that reusable modules and components rapidly become obsolete. All of these combine to make it nearly impossible to institute a widely accepted software reuse repository. Without significant software architecture and code reuse in developing software-intensive weapon systems, each development process essentially starts from scratch. This fact is one of the main reasons that the Technology Readiness Assessment (TRA) and the software Technology Readiness Levels (TRLs) are ineffective in predicting software development risk (Naegle & Petross, 2007).

The software engineering state-of-the-practice currently is wholly dependent on the requirements and operational environment cues that are passed to the software development team. From the requirements, a software architecture is designed, and the requirements “flow down” through that architecture to the individual modules and computer software units that are to be constructed. The software build focuses on the requirements that flowed down to that level and the integration required for functionality. The standards, protocols, formats, languages, and tools used for the build will likely be unique to the contractor developing the software, and will most certainly not be universally accepted or recognized across the software industry.

The software architectural design is the basis for all of the current and future system performance, including TOC performance, that the system will achieve, and the current state-of-the-practice in software engineering has each project design a unique architecture. Like hardware, the software design will significantly impact system attributes that are important to the warfighter, including TOC-oriented elements of maintainability, upgradability, interoperability, reliability, safety, and security. Most hardware-oriented engineering environments address these critical areas through widely accepted industry standards. For example, all DoD ground combat vehicles use a 24 volt, direct current, negative ground electrical system. Any current or future subsystem requiring vehicle power will automatically be designed to operate using those industry-wide electrical power standards.

The software engineering environment is in stark contrast to even our most advanced hardware-centric engineering environments. For example, in the automotive engineering field, a design that provides for easy replacement of wear-out items such as tires, filters, belts, and batteries obviously provides sustainability performance that is absolutely required. This engineering maturity helps account for derived and implied requirements not explicitly stated in the performance specification. Most performance specifications do not explicitly address this capability because they would be automatically considered by any competent provider within the mature automotive engineering environment. A mature engineering environment includes design elements and industry-wide standards, processes, materials, and techniques to which we have grown to expect. A significant problem will exist if we expect the software engineering environment to perform the same way as other, more mature engineering fields (Naegle & Petross, 2007).

As the example above illustrates, many system TOC elements are often standardized across hardware-oriented engineering environments due to the maturity of the sector’s engineering maturity. Without the engineering maturity, software sustainability performance and expectations must be specified as part of the requirements generation process. The capabilities-based user requirements and performance-based acquisition requirements are specifically not designed to provide that level of specificity.

The Software Engineering Environment Challenge

The DoD's acquisition management system is designed to garner innovation from the commercial marketplace by leveraging the mature engineering environments present in most disciplines. The DoD develops its requirements beginning with the capabilities-based language provided by the users, then translating them into performance-based language for the RFP. This requirements generation system is purposely designed to allow the maximum contractor flexibility in satisfying the warfighter's needs.

Within the immature software engineering environment, this requirements generation process creates an opportunity for significant misinterpretation, and derived and implied requirements that are not addressed, all resulting in requirements creep that fuels cost increases and schedule slippage. Unlike mature hardware-oriented engineering environments, where the widely accepted industry standards will be employed whether or not they are specified, with software, you get what you specify and very little else (Naegle, 2015, p. 13).

Addressing the Challenge

There are several necessary steps to effectively address the immature software engineering environment challenge:

1. The acquisition community must understand that the software engineering environment is different, and not mature. This must be an essential part of Knowledge Point 1 and of the Navy gate reviews 1 through 5, detailed earlier. The BBP memoranda help support this step by its direction to "improve the professionalism of the total acquisition workforce."
2. The acquisition community must take active steps to compensate for the software immature engineering environment.
    a. Requirements. Fully develop all requirements so that derived and implied requirements are specified. Sustainment performance including maintainability, upgradability, interoperability, reliability, and safety/security must be specified to improve TOC attributes.
    b. Operational context. Provide context for the requirements beyond what is provided in the typical OMS/MP. Software engineers need to understand how the system will be used and maintained, how it will be modified and interfaced in the future, which features are critical and which are non-critical enhancers, and how the user expects the system to operate under stressful conditions at the limits of the operational envelope. All of this required information is not available from any other source, and certainly not available in the software engineering environment.
3. The acquisition community must drive and monitor the software architectural design process to a much greater extent than what is needed for hardware-centric system. This is an essential function to reach Knowledge Point 2, and you literally could not achieve Knowledge Point 2 without the ability to drive the software architectural design. This would also be an essential function to effectively pass through the Navy gate reviews 4 through 6.

Estimating Software Size and Cost

Estimating the software size is essential to estimating development and sustainment costs. Unfortunately, estimating size is difficult for any software-intensive effort, and nearly impossible for unprecedented development efforts, including many DoD weapon systems. The DoD often seeks cutting-edge technologies pursuing dominant capabilities, driving the need for developing unprecedented software development.

The Estimating Software Size and Cost Challenge

Estimating software size, especially for a cutting-edge weapon system, is challenging, at best. It is essential for understanding both software developmental and sustainment costs, so is critical to understanding TOC.

> Software Size Estimating is an important activity in software engineering that is used to estimate the size of an application or component in order to be able to implement other program management activities such as cost estimation or schedule progress. The software engineer is responsible for generating independent estimates of the software size throughout the life cycle. These estimates are sometimes expressed as Software Lines of Code (SLOC), Function Points (FP), or Equivalent Software Lines of Code (ESLOC). An effective software estimate provides the information needed to design a workable Software Development Plan (SDP). This estimate is also input to the Cost Analysis Requirements Description (CARD) process. ("Software Management," n.d., p. 1)

The U.S. Air Force has published a guide for weapon system software development management and describes the software estimating challenge as follows:

> Weapon system acquisition programs routinely aim to develop and deliver unprecedented warfighting capability. This unprecedented capability is often realized by developing complex, SIS [software intensive system] or integrating existing systems and subsystems with other equally complex systems in new ways. Since acquisition programs are planned and estimated when only top-level performance requirements are available, it is extremely difficult to develop high confidence estimates and align expectations early in the program life cycle. Such early estimates are relatively subjective, involve numerous assumptions, and are almost always optimistic since the engineering activities that result in a complete understanding of the work to be accomplished have not been completed. This complete understanding typically does not mature until well into the design phase, and when it does, it usually confirms that initial estimates were optimistic, key assumptions (such as significant reuse) cannot be achieved, more work than planned needs to be done, and the amount of software that has to be developed and/or integrated is growing. (SecAF, 2008, p. 7)

Both the AcqNotes website and the Air Force guidebook offer some guidance in estimating the amount of software that needs to be developed, which is not the only factor in the development cost, but certainly one of the most important.

The AcqNotes website recommends the following:

There are various ways available to the software engineer to develop a size estimate. It is recommended that multiple techniques be used and the results combined to produce the final size estimate. Methods that can be used of estimating size are:

- **Comparable to existing programs**: Compare the proposed functionality and other similarities to existing programs. If the proposed program has 20% more functionality than one program and 15% less than another, a fairly accurate estimate can be achieved using the actual sizes from the existing programs.
- **Historical data**: Within a program, historical data of previous developments (estimates and actual) may exist. Since many of the parameters are usually the same (developer team, environment, platform, etc.) this is a good method to compare previous software builds and the proposed code. The more data that is used will increase the accuracy.
- **Contractor estimate**: It is generally true the contractor has written software similar previously. They often maintain a database of past efforts (estimates and actual) and can produce a very accurate estimate. Since the contractor and the Government have different objectives, their estimate should never be relied on solely.
- **Expert judgment (Delphi technique)**: Engineers that have domain experience and knowledge can often accurately estimate the software size. Without extensive experience however, expert judgment is seldom more accurate than guessing.
- **Level of effort or schedule**: This method does not really estimate the size to be developed, but rather defines the most that could be developed given unchangeable level of effort or schedule constraints. The software engineer uses productivity rates, integration time and software defect data from recently delivered programs to define the maximum size that could be developed. ("Software Management," n.d., p. 1)

The Air Force guidebook also has recommended considerations for estimating software size:

> The software estimating process consists of a series of activities that include estimating size of the software to be developed, modified, or reused; applying estimating models and techniques; and analyzing, crosschecking, and reporting the results. The following steps should be considered as part of any software estimating process:
>
> - Develop a notional architecture for the system, and identify program requirements likely to be satisfied by software.
> - Identify potential COTS, GOTS, and other sources of NDI software.
> - Identify existing software that will be modified, including the size of the overall software as well as the size of the expected modifications.
> - Identify software that will be newly developed for this program to provide functionality not available from existing software, or to adapt/integrate all the necessary software components.
> - Obtain software size information for all software elements, where size is carefully defined and measured in one of the two standard software size measures: non-comment source lines of code (SLOC) or function points.
> - Assess the uncertainty in the new and modified software sizes, based on historical data (if available) and engineering judgment.

- Assess the uncertainty associated with the reusability of existing software (COTS, GOTS, and NDI) in the context of the program (see section 3.2.4). Estimate the trade studies, familiarization, and the integration and testing efforts required to accommodate the unmodified reused code.
- Account for software complexity and the proposed development approach/processes, and assess any overlaps in software builds.
- Be realistic about expected software productivity and any assumption of significantly higher than historical productivity due to applying the best people, improved/more efficient processes, or new and improved development tools. Past performance, where actual size, cost, and same program or a very analogous program, should be heavily weighted. It is rare to have the A-team people for a long-duration embedded system development, and new processes and tools often fall short of expectations.
- Apply growth factors to new/modified and reuse software, based on past experience and the level of uncertainty.
- Account for all remaining uncertainties as estimate risks (see section 3.2.2).
- Ensure the estimate includes software support to systems engineering, system and sub-system requirements definition, configuration management, quality assurance, program management, system integration, and system test as appropriate.
- Address the software development life-cycle from software requirements analysis through software-related system integration and testing. The chosen modeling/estimation approach may not address the entire software effort since some commercial parametric models focus on the period starting with the baseline set of software requirements and ending with a fully integrated and tested subsystem/functional software product ready for software/hardware integration and test. Estimate and include any additional effort required to develop, allocate, and analyze the subsystem and software requirements; perform software to hardware (subsystem) integration and test; and perform system integration and test.
- Crosscheck estimate results with other methods such as other models, expert advice, rules of thumb, and historical productivity.
- Improve the estimate over time. (SecAF, 2008, pp. 27–28)

Both the AcqNotes and U.S. Air Force size estimating guidance suggest using multiple methodologies to form a more informed estimate of the likely software size of a developmental system. Nearly all of the guidance is dependent on an excellent understanding of the system requirements and operational context.

One common method to estimate the software size on a new developmental program is to use the analogy method, that is, to compare the new system to a similar system that was recently developed, assuming that the software will be similar in overall size. The following is the first bullet in the AcqNotes software estimating guidance detailed previously in this section. It seems a logical approach, but has not proven particularly accurate in recent history:

The premise is that the existing system's architecture, complexity, and functions are similar enough to fairly accurately predict the software development resources required for the new system. Unfortunately, this technique has proven to be ineffective as evidenced by the F-22 Raptor development and the follow-on F-35 Joint Strike Fighter (JSF) effort. The two high-performance, supersonic aircraft have overlapping missions, are significantly similar, and are both developed by the same contractor. The F-22 would seem to be a very good predictor of the F-35 software development effort with the SwTRL [Software Technology Readiness Level] model, but it clearly was not:

> The lines of code necessary for the JSF's capabilities have now grown to over 24 million—9.5 million on board the aircraft. By comparison, JSF has about 3 times more on-board software lines of code than the F-22A Raptor and 6 times more than the F/A-18 E/F Super Hornet. This has added work and increased the overall complexity of the effort. The software on-board the aircraft and needed for operations has grown 37 percent since the critical design review in 2005. … Almost half of the on-board software has yet to complete integration and test—typically the most challenging phase of software development. (GAO, 2012, p. 11)

The report goes on to state that typical software size growth in DoD systems development ranges from 30% to 100%.

> JSF design changes were originally supposed to taper off and be completed by January 2014. Actual design changes through September 2011 failed to taper off and continue at a significantly high rate. The projections in the GAO (2012) report indicated that the revised design change projections would continue and actually grow in number, until January 2019 (p. 16). Given this level of redesign, the software and system complexity growth are likely to continue. (Naegle, 2015)

The second bullet guidance from AcqNotes indicates that the use of historical data may be useful in estimating a new system's software size. This is particularly challenging for the DoD as the new weapon systems often have capabilities or features that are unprecedented (cutting-edge technologies). Certainly, there will be many subsystems in which historical data may be a good predictor for software size in existing, identical, or similar subsystems. However, the analogy method uses the historical data of a similar system as a surrogate for actual historical data, but suffers the challenges detailed previously.

The third AcqNotes bullet is "contractor estimates for software size." The problem with contractor estimates is that the size estimate is needed far before a development contractor would be involved in the process. Of course, market research contractors could be used to garner "contractor estimates," but this would require two essential preconditions. First, the market research contractor would need an extraordinary amount of requirements, operational context, and design detail on the proposed system to be able to provide to the marketplace to garner reasonably accurate software size estimates. Second, the market research would be conducted with industry members who can only respond to the information provided, so the estimates are only as accurate as the requirements-oriented information provided. In addition, the surveyed companies may be unwilling to provide much detail about their estimate as it could provide competitors with valuable competitive information.

The expert judgement, or Delphi Method (AcqNotes bullet 4), depends on the level of expertise of the engineers providing the estimate and their total understanding of the system to be developed. The DoD may gain access to expert software engineers that are inside the Government or through contracting for such expertise, but the level of understanding is dependent on the requirements generations system and the operational context provided.

There are also numerous parametric models, like Barry Boehm's Constructive Cost Model (COCOMO), that may be used in an attempt to estimate effort and cost (USC, 2002). COCOMO, like other estimating models, requires a software size estimate to be used. One of the inputs to the model is the Annual Change Traffic (ACT), or the percentage of the software that needs to be accessed for sustainability purposes. Obviously, the model would need to know the software size to perform the percentage calculations.

Because of all of the variables that are needed for the models, they can be quite misleading. For example, the University of Southern California (USC) used the models and then compared actual results to those estimated. They found that COCOMO "demonstrates an accuracy of within 20% of actuals 46% of the time for effort, and within 20% of actuals 48% of the time for a nonincremental development schedule" (USC, 2002). They found that, with more initial data input, the model accuracy improved to 30% of actuals 75% of the time. Boehm himself stated that "a software cost estimation model is doing well if it can estimate software development costs within 20% of the actual costs, 70% of the time, and on its home turf (that is, within the class of projects to which it is calibrated)" (SecAF, 2008, p. 21).

Obviously, using the results of parametric models alone would not result in the accurate estimates required by the DoD. The BBP memoranda specify "would cost" and "should cost" estimates that the models simply could not accurately produce. The software development cost and schedule estimate would necessarily need to be sufficiently accurate to avoid a Nunn-McCurdy violation in a software-intensive system development program.

Addressing the Challenge

Obviously, a fairly accurate software size estimate is necessary to predict both developmental and sustainment costs on a new system, and it is clear that obtaining an accurate size estimate is significantly challenging. The necessary precursor to software estimation is described earlier in this paper as compensating for the immature software engineering environment. Without more clearly defined requirements and operational context, accurately estimating software size is nearly impossible.

As suggested in both the AcqNotes and U.S. Air Force software estimating guidelines, a multi-faceted approach is needed. To be successful, each approach must be completed with significant discipline and rigorous systems analysis that goes beyond the current practices. If successful, the software size estimate will help predict both developmental and sustainment software costs.

Software Sustainability Architecture

A system's architecture and sustainability performance are strongly linked. Much of the design priority has been delegated to the contractor as the requirements language is capabilities-based on the user side and performance-based on the program management side. The DoD is responsible for driving the architectural design through the performance-based specification language, which requires a very in-depth understanding and development of the requirements.

The Software Architecture Challenge

Driving the software architectural design towards improved system TOC performance has numerous and complex challenges. The DoD requirements generation process is designed around the premise that the commercial marketplace has solutions for achieving the system performance specified by the DoD. This philosophy came from the acquisition reforms of the '90s, when systems were much more hardware-oriented, and the associated engineering environments were mature. As the DoD has moved to software-oriented systems, the philosophy did not change, even though the software engineering environment is not mature. This has created a significant mismatch in what the DoD communicates and what it expects to be delivered. Much of the mismatch can be linked to the software engineering immaturity:

> The lack of software engineering maturity impacts both requirements development and design of the architecture. To compensate for the relative immaturity of the software engineering environment, the DOD must conduct significantly more in-depth requirements analysis and provide potential software developers detailed performance specifications in all areas of software performance and sustainability. This is a significantly different mind-set than the hardware-dominated systems acquisition of the past.
>
> In addition to the performance requirements, software architectures must be similarly shaped to include system attributes expected by the warfighter. Many DOD user representatives and acquisition professionals have grown accustom to the engineering maturity levels offered by the hardware-oriented systems that dominated past acquisitions. Providing the system requirements in the same fashion may not drive the architecture for needed attributes. As demonstrated by the F-35 JSF redesign problems, changing software architectures during the development cycle will likely be costly in terms of schedule and funding. (Naegle, 2014, p. 14)

The DoD also provides the top levels of the work breakdown structure (WBS) to provide cues to the necessary design structures, but like the requirements generation process, the communication through the WBS is often too vague or lacking in necessary detail for the software engineers to understand important aspects of the design.

> Contracts resulting from proposals that are based on underdeveloped, vague, or missing requirements typically result in catastrophic cost and schedule growth as the true demands of the software development effort are discovered only after contract award. (Naegle, 2015, p. 8)

The design metrics are very important to ensure that the software architecture is meeting the warfighter needs and expectations for the new system, including the TOC performance. Too often, this process serves to identify missing requirements or clarify vague requirements, causing significant requirements creep impacting the cost and schedule.

Addressing the Challenge

The requirements generation process, the Operational Mode Summary/Mission Profile (OMS/MP), the WBS, and the resulting performance specification and Government-specified functional architecture (top levels of the WBS) must drive the software engineer to develop the detailed system architecture to the total needs of the warfighter. The software engineering environment will not compensate for vague or missing requirements, and there are virtually no industry-wide standards for sustainability.

Processes to both drive the software architecture and monitor the design activities are unlike the contractor's hardware architecture activities and significantly more critical. Fifty percent or more of the software effort is expended in requirements and architectural design, which is far greater than typical hardware-oriented systems. This means that half or more of the software development resources have been used by the Preliminary Design Review (PDR), which occurs quite early in the developmental process. Requirements creep and software changes after the PDR are significantly disruptive to the design process and are costly in both funding and schedule. In addition, changes occurring after the design is complete are typically accommodated through the use of software patches. While these patches may function adequately, they typically weaken the software structure and add difficulty to the sustainment effort as they add lines of code, are not generally well documented, and add complexity to problem analyses in the deployed system.

Software Sustainment Activities

The Post Deployment Software Support (PDSS) structure—maintainers, software engineering tools, documentation, licenses, and so forth—must all be funded and in place at the initial deployment as software maintenance will likely be required immediately due to the complexity. Most of the DoD software sustainment effort is accomplished through Contracted Logistics Support (CLS) strategies, so the support contracts are critical to system deployment.

As with hardware-oriented systems, the software sustainability performance is significantly defined by the system architecture. The software engineering immaturity means that there are no industry-wide standards for software sustainability, so the DoD must drive the desired sustainability performance into the software design.

The two major components that help determine a system's software sustainment cost are software size (SLOC count) and complexity. Many of the effort estimating tools need the software size to estimate the number of software maintainers that need to be dedicated to the sustainment effort. Complexity factors are then added into the calculations.

The Software Sustainment Challenge

The DoD system acquisition process is driven through the performance-based specifications, program WBS functional architectural cues, and high-level OMS/MP and, therefore, relies heavily on the contractor's expertise backed by the industry's mature engineering environments. This process is not adequate for driving the software architecture to a sustainable design as the immature software engineering environment has no industry-wide standards for sustainability, so software sustainability performance must be totally driven through the DoD front-end processes.

Unlike even the most sophisticated hardware system, the software maintainers must have the same skill sets as the design engineers, and so the DoD is typically contracting for software engineers to maintain the software. The software sustainment cost factors include maintainers, software tools, license fees, and associated contract costs for most DoD systems. While the non-personnel costs can be considerable, the cost of the maintainers is usually the largest part of the sustainment cost because the DoD is typically contracting for software engineers to maintain the software components.

The events driving the need for software maintenance are not always within the control of the system's PM. As the DoD continues to network platforms into Systems of Systems (SoSs), each platform is subject to the network's complexities and interoperability requirements.

Addressing the Challenge

The solutions for addressing the software sustainability challenge are rooted in solving the other issues presented in this section, as they all tend to build on one another. The DoD needs to recognize that the software engineering environment is immature, significantly different than the hardware-oriented engineering environments. That immaturity renders much of the DoD front-end processes ineffective for software-intensive systems, so active steps augmenting the standard acquisition processes must be taken to compensate.

TOC performance is being influenced by the ever increasing software functionality of DoD systems, so improving TOC performance means effectively addressing software development and sustainability costs. The software costs and performance are dependent on how effective the acquisition front-end processes address them, and the standard DoD acquisition management system appears to be insufficient for the software components.

The software TOC issues presented, and their underlying causes, call for supplementary Systems Engineering Process (SEP) tools, techniques, and analyses to be applied to the DoD acquisition process. The following sections describe recommended tools, techniques, and analyses that would help address the issues presented. All of these are designed to work within the Defense Acquisition System (DAS).

Driving the Software Requirements and Architectures for System Supportability

While the tools and techniques described in this section were designed for the software components, they would be just as effective for any non-software component because they are Systems Engineering (SE) oriented. The SEP focus used does not attempt to separate software from other components, so all system components would benefit from using these tools and techniques.

Software Supportability Analysis

As with hardware system components, software supportability attributes must be designed into the system architecture. Many hardware-oriented engineering fields are now quite mature, so that a number of supportability attributes would be automatically included in any competent design, even if they were not specified by the user community. For example, the state of maturity for the automotive engineering field means that, in any automotive-related program, there would be supportability designs allowing for routine maintenance of system filters, lubricants, tires, brakes, batteries, and other normal wear-out items. There are few, if any, corresponding supportability design attributes that would be automatically included in even the best software construct. Virtually all of the software supportability attributes required must be explicitly specified because they would not likely be included in the design architecture without clearly stated requirements. With software, you get what you specify and very little else. So how does one ensure that required software supportability attributes are not overlooked?

Logistics Supportability Analysis (LSA), performed extremely early, is one of the keys for developing the system supportability attributes needed and expected by the warfighter. The F/A 18 Super Hornet aircraft was designed for higher reliability and improved ease of maintenance compared to its predecessors (“F/A 18,” n.d.) because of warfighter needs for generating combat power in the form of aircraft sorties available. The LSA performed on the F/A 18 determined that a design fostering higher reliability and faster maintenance turnaround time (the engines are attached to the airframe at 10 locations and can be changed in about 20 minutes by a four-man team) would result in more aircraft being available to the commander when needed. The concept for software LSA is no different, but implementing sound supportability analyses on the software components has been spotty, at best, and completely lacking, at worst.

To assist in effective software LSA, a focus on these elements is key: Maintainability, Upgradeability, Interoperability/Interfaces, Reliability, and Safety & Security—MUIRS.

Maintainability

The amount of elapsed time between initial fielding and the first required software maintenance action can probably be measured in hours, not days. The effectiveness and efficiency of these required maintenance actions is dependent on several factors, but the software architecture that was developed from the performance specifications provided is critical. The DoD must influence the software architecture through the performance specification process to minimize the cost and time required to perform essential maintenance tasks.

Maintenance is one area in which software is fundamentally different from hardware. Software is one of the very few components in which we know that the fielded product has shortcomings, and we field it anyway. There are a number of reasons why this happens; for instance, there is typically not enough time, funding, or resources to find and correct every error, glitch, or bug, and not all of these are worth the effort of correcting. Knowing this, there must be a sound plan and resources immediately available to quickly correct those shortcomings that do surface during testing and especially those that arise during warfighting operations. Even when the system software is operating well, changes and upgrades in other interfaced hardware and software systems will drive some sort of software maintenance action to the system software. In other words, there will be a continuous need for software maintenance in the planned complex SoS architecture envisioned for net-centric warfare.

Because the frequency of required software maintenance actions is going to be much higher than in other systems, the cost to perform these tasks is likely to be higher as well. One of the reasons for this is that software is not maintained by ”maintainers,” as are most hardware systems, but is maintained by the same type of people that originally developed it—software engineers. These engineers will be needed immediately upon fielding, and a number will be needed throughout the lifespan of the system to perform maintenance, add capabilities, and upgrade the system. There are several models available to estimate the number of software engineers that will be needed for support; planning for funding these resources must begin very early in the process. Because the DoD has a very limited capability for supporting software internally, early software support is typically provided by the original developer and is included in the RFP and proposal for inclusion into the contract or as a follow-on Contractor Logistics Support (CLS) contract.

Upgradeability

A net-centric environment composed of numerous systems developed in an evolutionary acquisition model will create an environment of almost continuous change as each system upgrades its capabilities over time. System software will have to accommodate the changes and will have to, in turn, be upgraded to leverage the consistently added capabilities. The software architecture design will play a major role in how effective and efficient capabilities upgrades are implemented, so communicating the known, anticipated, and likely system upgrades will impact how the software developer designs the software for known and unknown upgrades.

Trying to anticipate upgrade requirements for long-lived systems is extremely challenging to materiel developers, but is well worth their effort. Unanticipated software changes in the operational support phase cost 50 to 200 times the cost in early design, so any software designed to accommodate an upgrade that is never realized costs virtually nothing when compared to changing software later for a capability that could have been anticipated. For example, the Army Tactical Missile System (ATACMS) Unitary was a requirement to modify the missile from warhead air delivery to surface detonation—that is, flying the warhead to the ground. The contract award for the modification was $119 million. The warhead was not new technology, nor particularly challenging to integrate with the missile body. The vast majority of this cost was to reengineer the software to guide the missile to the surface. Had there been an upgrade requirement for this type of mission in the original performance specification, this original cost (including potential upgrades, even if there were 10 other upgrade requirements that were never applied) would have been a fraction of this modification cost.

Interfaces/Interoperability

OA design focuses on the strict control of interfaces to ensure the maximum flexibility in adding or changing system modules, whether they are hardware or software in nature. This presupposes that the system modules are known—which seems logical, as most hardware modules are well-defined and bounded by both physics and mature engineering standards. In sharp contrast to hardware, software modularity is not bounded by physics, and there are very few software industry standards for the modular architecture in software components. This is yet another area in which the software developer needs much more information about operational, maintenance, reliability, safety, and security performance requirements, as well as current, planned, and potential system upgrades. These requirements, once well-defined and clearly communicated, will drive the developer to design a software modular architecture supporting OA performance goals. For example, if a system uses a Global Positioning System (GPS) signal, it is likely that the GPS will change over the life of the system. Knowing this, the software developer creates a corresponding discrete software module that is much easier and less expensive to interface, change, and upgrade as the GPS system does so.

With the system software modular architecture developed, the focus returns to the interfaces between hardware and software modules, as well as to the external interfaces needed for the desired interoperability of the net-centric force. Software is, of course, one of the essential enablers for interoperability and provides a powerful tool for interfacing systems, including systems that were not designed to work together. Software performing the function of “middleware” allows legacy and other dissimilar systems to interoperate. Obviously, this interoperation provides a significant advantage, but it comes with a cost in the form of maintainability, resources, and system complexity. As software interfaces with other components and actually performs the interface function, controlling it and ensuring the interfaces provide the desired OA capability becomes a major software-management and software-discipline challenge.

One method being employed by the DoD attempts to control the critical interfaces through a set of parameters or protocols rather than through active management of the network and network environment. This method falls short on several levels. It fails to understand and control the effects of aggregating all of the systems in a net-centric scheme. For instance, each individual system may meet all protocols for bandwidth, but when all systems are engaged on the network, all bandwidth requirements are aggregated on the network—overloading the total bandwidth available for all systems.

> While these standards may present a step in the right direction, they are limited in the extent to which they facilitate interoperability. At best, they define a minimal infrastructure that consists of products and other standards

> on which systems can be based. They do not define the common message semantics, operational protocols, and system execution scenarios that are needed for interoperation. They should not be considered system architectures. For example, the C4ISR domain-specific information (within the JTA) identifies acceptable standards for fiber channels and radio transmission interfaces, but does not specify the common semantics of messages to be communicated between C4ISR systems, nor does it define an architecture for a specific C4ISR system or set of systems. (Morris et al., 2004, p. 38)

Clearly, understanding and controlling the interfaces is critical for effective interoperation at both the system and SoS levels. The individual PM must actively manage all systems' interfaces impacting OA performance, and a network PM must do the same for the critical network interfaces. Due to this necessity of constant management, a parameters-and-protocols approach to net-centric OA performance is unlikely to produce the capabilities and functionality expected by the warfighter.

Understanding the software interfaces begins with the software architecture; controlling the interfaces is a unique challenge encompassing the need to integrate legacy and dissimilar systems and the lack of software interface standards within the existing software engineering environment. As stated earlier, the architecture needs to be driven through detailed performance specifications, which will help define the interfaces to be controlled. An effective method for controlling the interfaces is to intensely manage a well-defined Interface Control Document (ICD), which should be a Contract Data Requirements List (CDRL) deliverable on any software-intensive or networked system.

Reliability

While the need for highly reliable weapon systems is obvious, the impact on total system reliability of integrating complex software components is not so obvious. Typically, as system complexity increases, maintaining system reliability becomes more of a challenge. Add the complexity of effectively networking an SoS (all of which are individually complex) to a critical warfighting capability that is constantly evolving over time, and reliability becomes daunting.

Once again, the software developer must have an understanding of reliability requirements before crafting the software architecture and developing the software applications. Highly reliable systems often require redundant capability, and this holds true for software components as well. In addition, software problems tend to propagate, resulting in a degradation of system reliability over time. For example, a Malaysian Airlines Boeing 777 suffered several flight control problems resulting in the following: a near-stall situation, contradicting instrument indications, false warnings, and difficulty controlling the aircraft in both autopilot and manual flight modes. The problems were traced to software in an air data inertial reference unit that was feeding erroneous data to the aircraft's primary flight computer (PFC), which is used in both autopilot and manual flight modes. The PFC continued to try to correct for the erroneous data received, adjusting flight control surfaces in all modes of flight, displaying indications that the aircraft was approaching stall speed and overspeed limits simultaneously, and causing wind shear alarms to sound close to landing (Dornheim, 2005, p. 46). It is critical for system reliability that the software developers understand how outputs from software applications are used by interfaced systems so that appropriate reliability safeguards can be engineered into the developed software.

Software that freezes or shuts down the system when an anomaly occurs is certainly not reliable nor acceptable for critical weapon systems, yet these characteristics are prevalent in commercially based software systems. Mission reliability is a function of the aggregation of the system's subcomponent reliability, so every software subcomponent is contributing to or detracting from that reliability. The complexity of software makes understanding all failure modes nearly impossible, but there are many techniques that software developers can employ when designing the architecture and engineering the applications to improve the software component reliability. Once requirements are clearly communicated to the developers, the software can be engineered with redundancy or "safe mode" capabilities to vastly improve mission reliability when anomalies occur. The key is identifying the reliability requirements and making them clear to the software developers.

Safety & Security

Very few software applications have the required safety margins associated with critical weapon systems used by warfighters in combat situations—where they are depending on these margins for their survival. Typically, the software developers have only a vague idea of what their software is doing and how critical that function is to the warfighter employing the weapon system. Safety performance must be communicated to the software developers from the beginning of development so they understand the link between software functionality and systems safety. For example, suppose a smart munition senses that it does not have control of a critical directional component, and it calculates that it cannot hit the intended target. The next set of instructions the software provides to the malfunctioning system may well be critical to the safety of friendly troops, so software developers must have the necessary understanding of operational safety to decide how to code the software for what will happen next.

Software safety is clearly linked with reliability since software that is more reliable is inherently safer. It is critical that the software developer understands how the warfighter expects the software to operate in abnormal situations, in degraded modes, and when inputs are outside of expected values. Much commercially based software simply ceases to function under these conditions or gives error messages that supersede whatever function was being performed, none of which are acceptable in combat operations.

With software performing so many critical functions, there is little doubt that software applications are a prime target for anyone opposing U.S. and Allied forces. Critical weapon system and networking software must be resistant to hacking, spoofing, mimicking, and all other manners of attack. There must be capabilities for isolating attacks and portions of networks that have been compromised without losing the ability to continue operations in critical combat situations. The software developer must know that all of these capabilities are essential before he or she constructs software architectures and software programs, as this knowledge will be very influential for the software design and application development. The Software Engineering Institute's *Quality Attribute Workshop* states, "As an example, consider security. It is difficult, maybe even impossible, to add effective security to a system as an afterthought. Components as well as communication mechanisms and paths must be designed or selected early in the lifecycle to satisfy security requirements" (Barbacci et al., 2003, p. 2).

Interoperability challenges are increased when the SoS has the type of security requirements needed by the DoD. Legacy systems and existing security protocols will likely need to be considered before other security architecture can be effectively designed. OA capabilities will be hampered by the critical need for security; both must be carefully balanced to optimize system performance and security. This balance of OA and security must be managed by the DoD and not the software developer.

Physical security schemes and operating procedures will also have an impact on the software architecture. For example, many communication security (COMSEC) devices need only routine security until the keys, usually software programs, are applied; then, much more stringent security procedures are implemented. Knowledge of this security feature would be a key requirement of the developer; he or she must understand how and when the critical software pieces are uploaded to the COMSEC device. The same holds true for weapon systems that upload sensitive mission data just prior to launch.

Residual software on equipment or munitions that could fall into enemy hands presents another type of security challenge that needs to be addressed during the application development. For example, the ATACMS missile air-delivers some of its warheads, leaving the missile body to freefall to the surface. It is very conceivable that the body could be intact and, of course, unsecured. If critical mission software was still within the body and found by enemy forces, valuable information might be gleaned from knowing how the system finds its targets. The Government would certainly want the developer to design the applications in a way that would make anything recovered useless to the enemy, but this is a capability that is not intuitive to the software developers (Naegle, 2006, pp. 17–25).

Effective Software Development Tools Supporting System TOC Analyses

Software Engineering Institute's Quality Attribute Workshop

The Quality Attribute Workshop (QAW) is designed to help identify a complete (or as complete as possible) inventory of system software requirements through analysis of system quality attributes. One of the intents is to develop the derived and implied requirements from the user-stated requirements, which is a necessary step when user-stated requirements are provided in terms of capabilities needed as prescribed by the Joint Capabilities Integration Development System (JCIDS) process. A system's TOC, and those elements that contribute to TOC, are system quality attributes. Although obviously important to the warfighter, the associated operations and support, training/education, and facility costs are rarely addressed in much detail and need to be derived from stated requirements or augmented with implied requirements through the QAW process, or something similar.

The QAW helps provide a facilitating framework and process designed to more fully develop the derived and implied requirements that are critical to clearly communicate to potential contractors and software developers. Including a robust LSA process using the MUIRS focus elements, described previously, within the QAW process will likely significantly improve requirements analysis for those associated TOC elements and vastly improve the accuracy of system TOC projections. While improving the system requirements development, QAW is designed to work with another SEI process called the Architectural Tradeoff Analysis Methodology^{SM} (ATAM^{SM}) to further improve the understanding of the system for potential contractors and software developers.

SEI's Architectural Tradeoff Analysis Methodology^{SM}

The Software Engineering Institute's Architectural Tradeoff Analysis Methodology^{SM} (ATAM^{SM}) is an architectural analysis tool designed to evaluate design decisions based on the quality attribute requirements of the system being developed. The methodology is a process for determining whether the quality attributes, including TOC attributes, are achievable by the architecture as it has been conceived before enormous resources have been committed to that design. One of the main goals is to gain insight into how the quality attributes trade off against each other (Kazman, Klein, & Clements, 2000, p. 1).

Within the Systems Engineering Process (SEP), the ATAM^{SM} provides the critical requirements loop process, tracing each requirement or quality attribute to corresponding functions reflected in the software architectural design. Whether ATAM^{SM} or another analysis technique is used, this critical SEP process must be performed to ensure that functional- or object-oriented designs meet all stated, derived, and implied warfighter requirements. In complex systems development such as weapon systems, half or more than half of the total software development effort will be expended in the architectural design process. Therefore, the DoD PMs must ensure that the design is addressing requirements in context and that the resulting architecture has a high probability of producing the warfighters' JCIDS stated, derived, or implied requirements.

The ATAM^{SM} focuses on quality attribute requirements, so it is critical to have precise characterizations for each. To characterize a quality attribute, the following questions must be answered:

- What are the stimuli to which the architecture must respond?
- What is the measurable or observable manifestation of the quality attribute by which its achievement is judged?
- What are the key architectural decisions that impact achieving the attribute requirement? (Kazman et al., 2000, p. 5)

The ATAM^{SM} scenarios are a key to providing the necessary information to answer the first two questions, driving the software engineer to design the architecture to answer the third. This is a critical point at which all of the MUIRS elements need to be considered and appropriate scenarios developed.

The ATAM^{SM} uses three types of scenarios: *Use-case scenarios* involve typical uses of the system to help understand quality attributes in the operational context; *growth scenarios* involve anticipated design requirements, including upgrades, added interfaces supporting SoS development, and other maturity needs; and *exploratory scenarios* involve extreme conditions and system stressors, including Failure Modes and Effects Criticality Analysis (FMECA) scenarios (Kazman et al., 2000, pp. 13–15). As depicted in Figure 1, the scenarios build on the basis provided in the JCIDS documents and requirements developed through the QAW process. These processes lend themselves to development in an Integrated Product Team (IPT) environment led by the user/combat developer and including all of the system's stakeholders. The IPT products will include a set of scenarios, prioritized by the needs of the warfighter for system capability. The prioritization process provides a basis for architecture trade-off analyses. When fully developed and prioritized, the scenarios provide a more complete understanding of requirements and quality attributes in context with the operation and support (including all of the MUIRS elements) of the system over its life cycle. A more complete understanding of the system's TOC elements should emerge from this type of analysis.

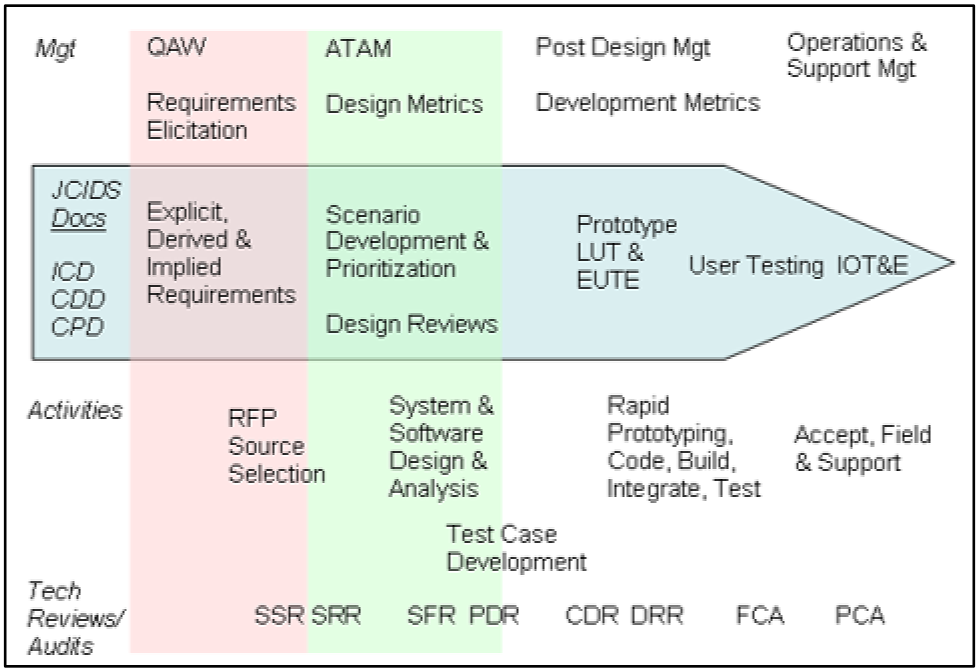


Figure 1. QAW & ATAM^{SM} Integration Into Software Life-Cycle Management

Just as the QAW process provides a methodology supporting RFP, source-selection activities, and the Software Specification and System Requirements Reviews (SSR and SRR), the ATAM^{SM} provides a methodology supporting design analyses, test program activities, and the System Functional and Preliminary Design Reviews (SFR and PDR). The QAW and ATAM^{SM} methodologies are probably not the only effective methods supporting software development efforts, but they fit particularly well with the DoD's goals, models, and SEP emphasis. The user/combat developer (blue arrow block in Figure 4) is kept actively involved throughout the development process—providing key insights the software developer needs to successfully develop warfighter capabilities in a sustainable design for long-term effectiveness and suitability. The system development activities are conducted with superior understanding and clarity, reducing scrap and rework, and saving cost and schedule. The technical reviews and audits (part of the DoD overarching SEP) are supported with methodologies that enhance both the visibility of the necessary development work as well as the progress toward completing it.

One of the main goals in analyzing the scenarios is to discover key architectural decision points that pose risks for meeting quality requirements. Sensitivity points are determined, such as real-time latency performance shortfalls in target tracking. Trade-off points are also examined so that TOC impacts resulting from proposed trade-offs can be analyzed. The Software Engineering Institute explained, “Trade-off points are the most critical decisions that one can make in an architecture, which is why we focus on them so carefully” (Kazman et al., 2000, p. 23).

The ATAM^{SM} provides an analysis methodology that complements and enhances many of the key DoD acquisition processes. It provides the requirements loop analysis in the SEP, extends the user/stakeholder JCIDS involvement through scenario development, provides informed architectural trade-off analyses, and vastly improves the software developer's understanding of the quality requirements in context. Architectural risk is significantly reduced, and the software architecture presented at the Preliminary Design

Review (PDR) is likely to have a much higher probability of meeting the warfighters' need for capability, including TOC elements.

Together, the QAW and ATAM^SM provide effective tools for addressing problem areas common in many DoD software-intensive system developments: missing or vaguely articulated performance requirements, significantly underestimated software development efforts (resulting in severely underestimated schedules and budgets), and poor communication between the software developer and the Government (both user and PM). Both tools provide frameworks for more detailed requirements development and more effective communication, but they are just tools—by themselves, they will not replace the need for sound planning, management techniques, and effort. Both QAW and ATAM^SM provide methodologies for executing SEP Requirements Analysis and Requirements Loop functions, effective architectural design transition from user to developer, and SEP design loop and verification loop functions within the test-case development.

A significant product resulting from the ATAM^SM is the development of test cases correlating to the use case, growth, and exploratory scenarios developed and prioritized. Figure 2 depicts the progression from user-stated capability requirements in the JCIDS documents to the ATAM^SM scenario development, and finally to the corresponding test cases developed. The linkage to the user requirements defined in the JCIDS documents is very strong as those documents drive the development of the three types of scenarios, and, in turn, the scenarios drive the development of the use cases. The prioritization of the scenarios from user-stated Key Performance Parameters (KPPs), Critical Operational Issues (COIs), and FMECA analysis flows to the test cases, helping to create a system test program designed to focus on effectiveness and suitability tests—culminating in the system Operational Test and Evaluation (OT&E). FMECA is one of the focus areas that will have a dynamic impact on TOC analysis because it will help identify software components that need higher reliability and back-up capability. The MUIRS focus helps ensure that TOC elements are addressed in design and test.

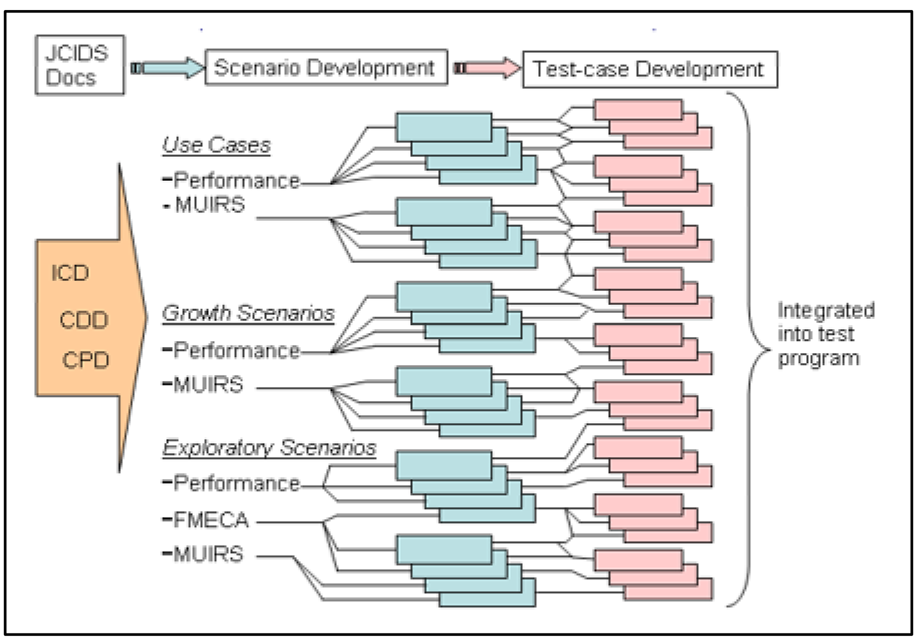


Figure 2. Progression From User-Stated Capability Requirements Through Scenario Development to Test-Case Development

Capabilities-Based ATAM^{SM} Scenario Development

The traceability from user-stated requirements through scenario development to test-case development provides a powerful communication and assessment methodology. The growth scenarios and resulting test cases are particularly suited for addressing and evaluating TOC design requirements because the system evolves over its life cycle, which is often overlooked in current system development efforts.

The software developer's understanding of the eventual performance required in order to be considered successful guides the design of the architecture and every step of the software development, coding, and testing through to the Full Operational Capability (FOC) delivery and OT&E. Coding and early testing of software units and configuration items is much more purposeful due to this level of understanding. The MUIRS and FMECA focus will help the design process for better TOC performance.

The resulting test program is very comprehensive as each prioritized scenario requires testing or other verification methodologies to demonstrate how the software performs in each related scenario and satisfies the quality attributes borne of the user requirements. The testing supports the SEP design loop by verifying that the software performs the functions allocated to it and, in aggregate, performs the verification loop process by demonstrating that the final product produces the capability identified in the user requirements through operational testing.

Both QAW and ATAM^{SM} require the capturing of essential data supporting decision-making and documenting decisions made. These databases would be best used in a collaborative IT system, as described in the next section.

Conclusions and Recommendations: Major Thrusts to Control Software Component TOC

Conclusions

DoD software-intensive systems and the software content in other systems will continue to grow and may dominate the TOC costs in the future. These costs are exacerbated by the fact that, in addition to contracted development costs, the bulk of the software sustainment costs are also contracted. In addition, the skill sets needed for software sustainment are the same as for software development, so the DoD is contracting for software engineers to perform maintenance functions. All of these factors indicate that DoD system software will continue to be a very expensive portion of TOC.

The software engineering environment remains immature, with few, if any, industry-wide standards for software development or sustainment. The Defense Acquisition System (DAS) is significantly dependent on mature engineering environments to compensate for the gaps and interpretation requirements presented with the performance-based specifications, vague Operational Mission Summary/Mission Profiles, and high-level work breakdown structures (WBSs) that the DoD provides during the request for proposal (RFP) process.

The developer software engineers will consume 50% or more of their contract resources analyzing requirements and developing the architectural design. This effort is expended before the Preliminary Design Review (PDR) and requirement additions (requirements creep), or changes beyond that point have disastrous effects on the software design and can even cause a complete redesign at extreme cost in funding and schedule.

The system software size and complexity are key indicators of both the development costs and the sustainment costs, so the initial estimates are critical for predicting and controlling TOC. Unfortunately, the software size estimating processes require a significant amount of detailed understanding of the requirements and design that is typically not available when operating the DAS without supplementary analyses, tools, and techniques. Available parametric estimating tools require much of the same detailed information and are still too inaccurate to be relied upon. Similarly, understanding the potential software complexity requires in-depth understanding of the requirements and architectural design.

It is clear that the DoD must conduct much more thorough requirements analyses, provide significantly more detailed operational context, and drive the software architectural design well beyond the WBS functional design typically provided. To accomplish this, the DAS must be supplemented with tools, techniques, and analyses that are currently not present.

Recommendations

Program managers for software-intensive systems must supplement the DAS processes to

- compensate for the immature software engineering environment
- gain sufficient detailed information to perform reasonable software size and complexity estimates critical to understanding and managing system TOC
- complete the inventory of derived and implied requirements, including the often neglected sustainability requirements, before the RFP is issued
- provide more detailed system operational context, beyond what exists in most OMS/MP documents
- obtain more realistic contractor proposals in terms of cost and schedule associated with the software development and sustainment
- drive the software architecture for a more sustainable, less complex design
- monitor the software design process (metrics) to ensure the effort is progressing towards an effective, supportable, and testable design supporting the warfighter

The tools, techniques, and analyses presented in this research are designed to accomplish the tasks outlined above, and are compatible with the Systems Engineering Process (SEP) supporting the DAS. They also are designed to work together in a synergistic method to improve the software-intensive system development and sustainment performance influencing system TOC. They are certainly not the only tools, techniques, and analyses available to improve the process, and others may be as effective, as long as they can address the bulleted items above.

The maintainability, upgradability, interoperability, reliability, and safety/security (MUIRS) analysis technique is designed to help identify derived and implied requirements that need to be more fully articulated to ensure that the software engineer adequately considers these critical system attributes. These were selected because they are often missing from the user's capability-based requirements documents and the resulting performance specification, yet they are critical for the warfighter and are significant TOC drivers.

The Quality Attribute Workshop (QAW) is a technique to help more fully detail all requirements, including derived and implied. It is often used with the system WBS to more fully develop the desired functional design, especially when combined with the MUIRS analyses.

The Architectural Tradeoff Analysis Methodology[SM] (ATAM[SM]) is designed to be used with the QAW and provides detailed operational context through the scenario development, providing critical design cues to the software development engineers. The scenarios include Use Cases (how the system will be used and maintained if fielded today), Growth Cases (how the system will likely change over its life cycle, including future networking), and Exploratory Scenarios (how the system is to operate under unusual or stressful conditions). This research recommends including the MUIRS analyses in the ATAM, as well as Failure Modes and Effects Criticality Analyses (FMECA) to identify critical functionality requirements.

Combined, the tools, techniques, and analyses provide a much improved understanding of the system and identify critical attributes that the software developers need to know to design an effective and supportable design. These tools help compensate for the immature software engineering environment, provide more detailed information needed to perform size and complexity estimates, and provide detailed operational context needed for proper software architectural design. They help produce superior RFPs and garner more realistic contractor proposals. They provide processes for monitoring critical software design activities and full test matrix crosswalks. All of these enhancements will help more accurately estimate and manage software TOC attributes.

References

Barbacci, M., Ellison, R., Lattanze, A., Stafford, J., Weinstock, C., & Wood, W. (2003, August). *Quality attribute workshops (QAWs)* (3rd ed.; CMU/SEI-2003-TR-016). Pittsburgh, PA: Carnegie Mellon University, Software Engineering Institute.

Dornheim, M. A. (2005, September). *A wild ride. Aviation Week & Space Technology, 163*, 46.

F/A 18. (n.d.). In *Wikipedia*. Retrieved from http://www.wikipedia.org/wiki/McDonnell_Douglas_F/A-18_Hornet

GAO. (2012, March 20). *Joint Strike Fighter: Restructuring added resources and reduced risk, but concurrency is still a major concern* (GAO-12-525T). Retrieved from http://www.gao.gov

Kazman, R., Klein, M., & Clements, P. (2000, August). *ATAM[SM]: Method for architecture evaluation* (CMU/SEI-2000-TR-004). Pittsburgh, PA: Carnegie Mellon University, Software Engineering Institute.

Kruchten, P. (2005, March/April). Software design in a postmodern era. *IEEE Software, 18*(2), 17.

Morris, E., Levine, L., Meyers, C., Place, P., & Plakosh, D. (2004, April). *System of systems interoperability (SOSI): Final report.* Pittsburg, PA: Carnegie Mellon University, Software Engineering Institute.

Naegle, B. R. (2006, September). *Developing software requirements supporting open architecture performance goals in critical DoD system-of-systems* (NPS-AM-06-035). Monterey, CA: Naval Postgraduate School.

Naegle, B. R. (2015, February 4). *Gaining control and predictability of software-intensive systems development and sustainment* (NPS-AM-14-194). Monterey, CA: Naval Postgraduate School.

Naegle, B. R., & Petross, D. (2007, October). *Developing software requirements supporting open architecture performance goals in critical DoD system-of-systems* (NPS-AM-07-104). Monterey, CA: Naval Postgraduate School.

Secretary of the Air Force (SecAF). (2008, August). *Weapon system software management guidebook.* Washington, DC: Author.

Software management: Software size estimate. (n.d.). In AcqNotes. Retrieved March 9, 2017, from http://www.acqnotes.com/acqnote/careerfields/software-size-estimate

University of Southern California. (2002, September). COCOMO. Retrieved March 9, 2017, from http://sunset.usc.edu/cse/pub/research/COCOMOII/cocomo_main.html

Exploring the Department of Defense Software Factbook

Christopher Miller—is a Senior Researcher at the Software Engineering Institute (SEI). Dr. Miller's expertise is in software metrics, measurement, and estimation of software intensive systems. His quantitative analysis background is focused on life cycle cost estimation, evaluating project feasibility measures, establishing performance measurements, and providing analysis for the optimization of systems. Dr. Miller served as the Chair of the International Council on Systems Engineering (INCOSE) Measurement Working Group (MWG) for seven years and is a certified trainer for Practical Software and Systems Measurement (PSM). Dr. Miller earned a Masters of Engineering Management and PhD in Systems Engineering at the George Washington University and is currently a member of the adjunct faculty in the School of Engineering and Applied Science (SEAS). [clmiller@sei.cmu.edu]

Forrest Shull—is Assistant Director for Empirical Research at Carnegie Mellon University's Software Engineering Institute. His role is to lead work with the U.S. Department of Defense, other government agencies, national labs, industry, and academic institutions to advance the use of empirically grounded information in software engineering, cybersecurity, and emerging technologies. He has been a lead researcher on projects for the U.S. Department of Defense, NASA's Office of Safety and Mission Assurance, the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation, and commercial companies. He serves on the IEEE Computer Society Board of Governors and Executive Committee. [fjshull@sei.cmu.edu]

David Zubrow—is an Associate Director for the Software Solutions Division and Manager of the Software Engineering Measurement and Analysis Initiative. He has been employed at the Software Engineering Institute (SEI) at Carnegie Mellon University since 1992. For much of his 26 years at the SEI, he has developed and transitioned quality and process improvement principles and practices directly through the CMMI and applications of statistical analysis. More recently, his research and transition projects have involved using machine learning methods in conjunction with software engineering data to model program status assessments and risks as well as predicting defects. [dz@sei.cmu.edu]

Abstract

The Carnegie Mellon Software Engineering Institute (SEI) conducted an analysis of software engineering data owned and maintained by the Department of Defense (DoD) to produce high-level, DoD-wide heuristics and domain-specific benchmark data. This work yielded basic facts about software projects, such as averages, ranges, and heuristics for requirements, size, effort, and duration. Factual, quantitatively-derived statements were reported to provide users with easily digestible benchmarks.

Findings were also presented by system type, or super domain. The analysis in this area focused on identifying the most and least expensive projects and the best and worst projects within three super domains: real time, engineering, and automated information systems. It also provided insight into the differences between system domains and contained domain-specific heuristics.

Finally, correlations were explored among requirements, size, duration, and effort and the strongest models for predicting change were described. The goal of this work was to determine how well the data could be used to answer common questions related to planning or replanning software projects. The paper provides a high-level overview of the SEI's research and primary findings.

Introduction

In 2015 the SEI undertook an analysis of the most extensive collection of software engineering data owned and maintained by its primary sponsor, the Department of Defense (DoD). The resulting *Department of Defense Software Factbook* provides an analysis of the software resources data report (SRDR), a contract data deliverable that applies to major contracts and subcontracts for programs with software development elements that include a projected software effort greater than $20 million.[1]

The SRDR is a contract data deliverable that formalized the reporting of software metrics data and is the primary source of data on software projects and their performance. The SRDR reports are provided at the project level or subsystem level, not at the DoD Acquisition Program level. **It is important to note that when the analysis refer to a "project" in this report, a project is synonymous with a software build, increment, or release**. In many cases, several projects (i.e., data points) would contribute to the overall scope and make up of Acquisition Program (i.e., an entire weapon system).

This work builds on a field of research begun in the 1970s into how to estimate the cost of software development. An entire industry focused on parametric software estimation has grown since that time, and at the core of this industry is a fundamental assumption that the cost of developing software can be estimated based on an accurate estimate of the size of the software product to be developed. This concept might be more accurately described as an assumed empirical relationship between cost and software size.

A new SRDR Data Item Description (DID), DI-MGMT-82035A, with updated formats for software development and maintenance was approved for release in November 2017. This new DID replaces the 2016 version of the DID which superseded the 2011 Initial and Final SRDR DIDs. The SRDR DID remains at $20 million or over for all new contracts, and over $1 million per year maintenance efforts. Key parameters related to software cost include functional size (in requirements), physical size (in equivalent source lines of code), effort hours, and duration of software projects. In most DoD environments, size is measured by requirements and the final physical size of the software product, commonly measured in source lines of code (SLOC). The amount of effort required to deliver the software can be estimated if you know the size. Similarly, duration (or schedule) can be derived from the effort.

The majority of the SRDR data used in this analysis is based on the final report that contains data about actual results. Projects used for this analysis had to include the following information: size data (functional and product), effort data, and schedule data. Our analysis included 287 projects from the product-event final report data and 181 pairs of initial and final case data (to compare estimated versus actual performance).

[1] For a more detailed description of programs types that require the use of the SRDR, see the *Department of Defense Software Factbook*, or CSDR Requirements, OSD Defense Cost and Resource Center, http://dcarc.cape.osd.mil/CSDR/CSDROverview.aspx#Introduction.

Functional Size (Requirements)

Functional size represents the overall magnitude of the software capabilities without regard to the final solution. The benefit of using functional measures is their availability early in the software development lifecycle. In the DoD acquisition community, requirements are rigorously derived and used as the contractual basis for acquiring systems. Therefore requirements and requirements documents are produced as part of the system acquisition life cycle and are readily available for the extraction of the number of requirements.

The drawback of using functional measures is that the requirement does not consistently correlate to a unit of effort (i.e., not all requirements take the same amount of effort to satisfy). Using the total number of requirements to represent size is useful, but trying to attach a unit cost (i.e., the cost per requirement) is not advised.

In general, software project data tends to be skewed. So making a transformation to get it into a normal (Gaussian) distribution is usually necessary. This was necessary for the SRDR requirements data. Since it was quite skewed, with the bulk of the data between 102 (~100) and 1110 (~1100) requirements, it needed to be transformed. Once the data was normalized using a natural log transformation, the median is e6.04, or 420 requirements with a mean of 368 requirements. Both are much closer to the raw data median of 399 than the raw data mean of 1118 requirements (see Figure 1).

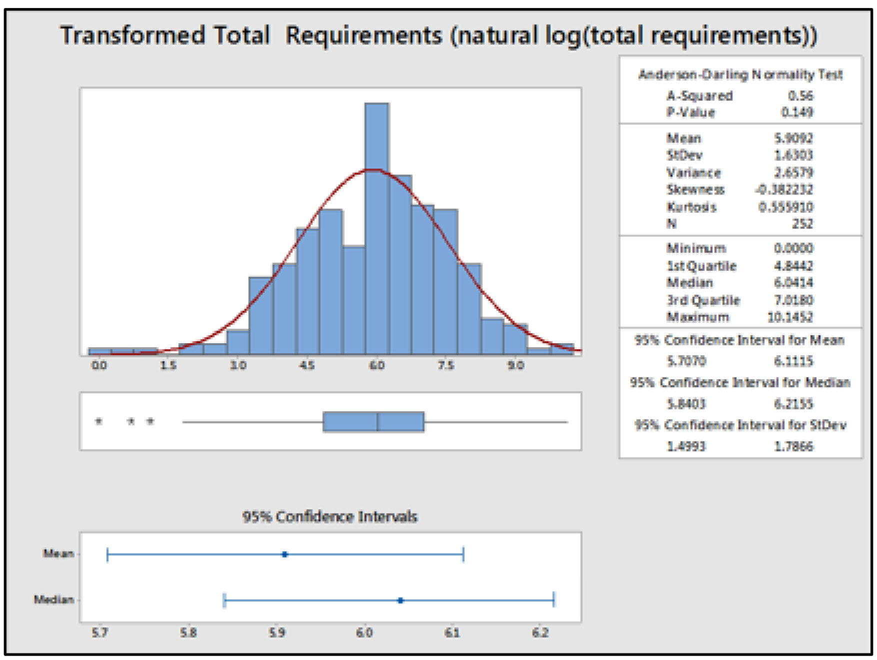


Figure 1. Functional Size, Normalized

Requirements data analyzed by super domain are presented in Figure 2. As is in shown on the top of the figure, to the left of the line is the 25th percentile value. This indicates that 25% of the projects have less than 100 requirements. Similarly, on the right the 75th percentile value indicates that 25% of the projects have more than 1100 requirements. Note that 50% of the projects have between 100 and 1100 requirements, with relatively more toward the lower end and a median or typical view of 400. The additional lines in the figure can be similarly interpreted. Similar figures are provided throughout this paper showing the 25th percentile, median, and 75th percentiles. An easy heuristic for the average functional size of a DoD software project is **400 requirements**.

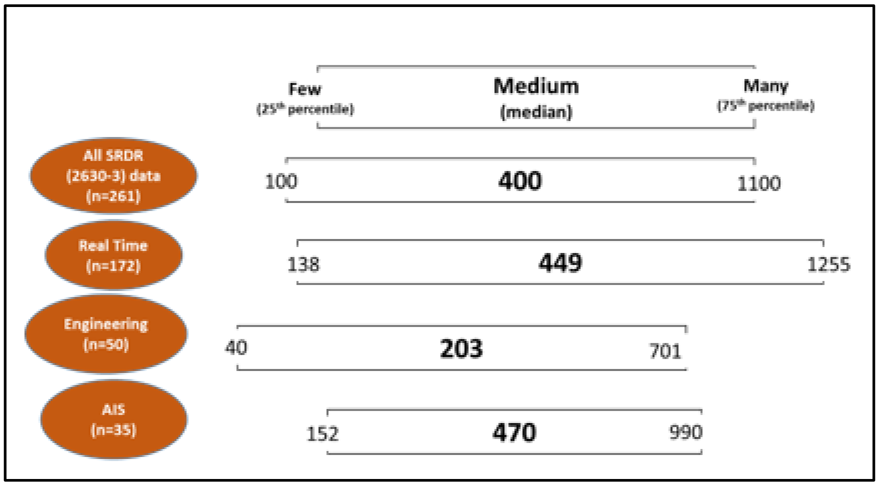


Figure 2. Requirements Data by Super Domain

Product Size (ESLOC)

Another common measure of interest is product size, which is often measured in source lines of code (SLOC). A key issue in using SLOC as a measure of work effort and duration is the difference in work required to incorporate software from different sources, including new code, modified code (changed in some way to make it suitable), reused code (used without changes), and auto-generated code (created from a tool and used without change).

Each of these sources requires a different amount of work effort to incorporate into a software product. The challenge is in coming up with a single measure that includes all of the code sources. The approach taken here is to normalize all code sources to the *equivalent* of a new line of code. This is done by taking a portion of the measures for modified, reused, and auto-generated code. The portioning is based on the percentage of modification to the code based on changes to the design, code and unit test, and integration and test documents. This approach is adopted from the COCOMO II Software Cost Estimation Model (Boehm et al., 2000, p. 22).

Equivalent source lines of code (ESLOC), then, is the homogeneous sum of the different code sources. The portion of each code source is determined using a formula called an Adaptation Adjustment Factor (AAF):

$$AAF = (0.4 \times \%DM) + (0.3 \times \%CM) + (0.3 \times \%IM)$$

Where

%DM: Percentage Design Modified

%CM: Percentage Code and Unit Test Modified

%IM: Percentage Integration and Test Modified

Using a different set of percentages for the different code sources, ESLOC is expressed as

ESLOC = New SLOC +

(AAFM x Modified SLOC) +

(AAFR x Reused SLOC) +

(AAFAG x Auto-Generated SLOC)

New code does not require any adaptation parameters, since nothing has been modified.

Auto-generated code does not require the DM or CM adaptation parameters. However, it does require testing, IM. If auto-generated code does require modification, then it becomes modified code, and the adaptation factors for modified code apply.

Equivalent source lines of code (ESLOC) normalize all code sources to the equivalent of a new line of code by computing a portion of the physical measures for modified, reused, and auto-generated code. Figure 3 shows the ESLOC data normalized using a natural log transformation. ESLOC by super domain is presented in Figure 4. An easy heuristic to use for average project size is **around 40,000 ESLOC** for all projects.

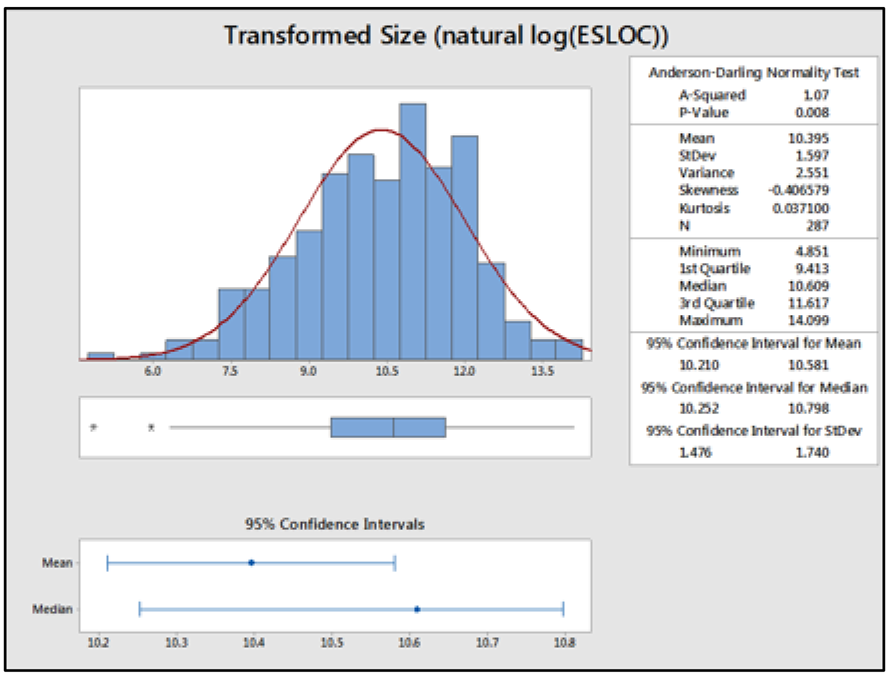


Figure 3. Product Size in ESLOC, Normalized

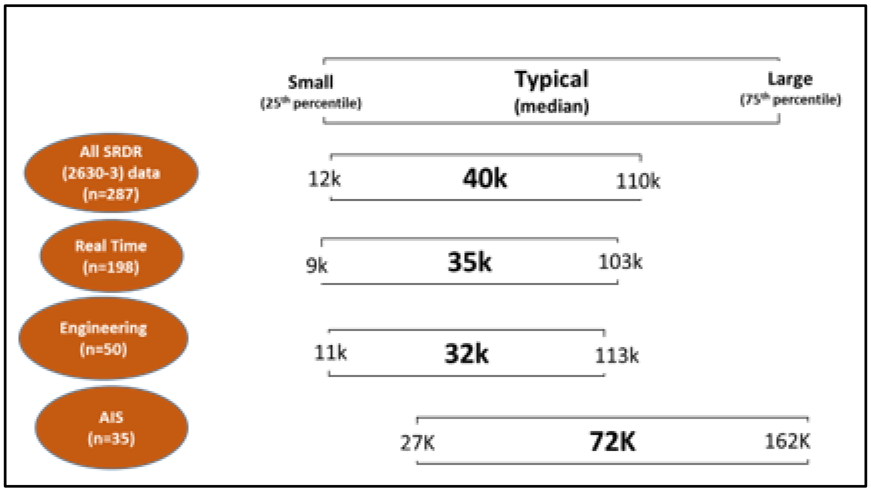


Figure 4. ESLOC by Super Domain

Effort

The amount of effort used to create software is the major driver of the cost of the development; the effort estimate in dollars provides the largest element in the cost estimate for software. Effort is usually collected in hours. For simplification purposes many estimation tools and equations use person months. When comparing effort data, ensure that the same conversion rate is used across the data set (i.e., the number of hours in a person month and/or number of hours in a full time equivalent). As detailed in Appendix G: Burden Labor Rate, it is assumed here that there are 152 hours in a labor month and 1824 hours per full-time equivalent (FTE), based on an annual labor rate of $150,000.

Figure 5 shows the effort data normalized. The effort hour data analyzed by super domain are presented in Figure 6. An easy heuristic to use for average project effort is around **40,000 hours, 263 person months, or 22 FTEs** for a DoD software project.

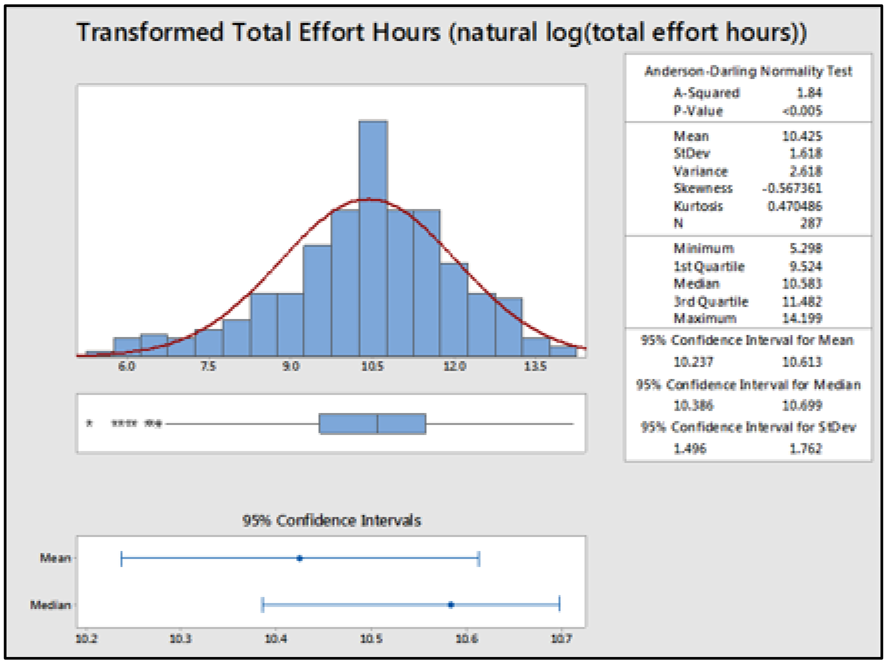


Figure 5. Effort, Normalized

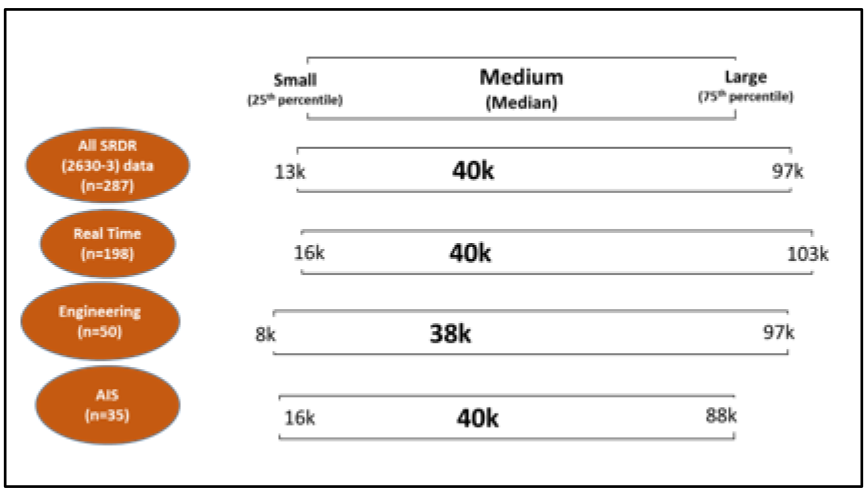


Figure 6. Effort Hours by Super Domain

Duration

Duration is a measure of the calendar time it takes to complete the software project. Many factors affect duration, including staffing profile, schedule constraints, and release plan. No adjustments are made for these factors in the data reported in this section.

Figure 7 shows the duration data normalized. The data indicate that the majority of projects take between 2 ½ to 3 years. An easy heuristic to use for the duration of an average DoD software project is **approximately 3 years**. Duration data analyzed by super domain is presented in Figure 8.

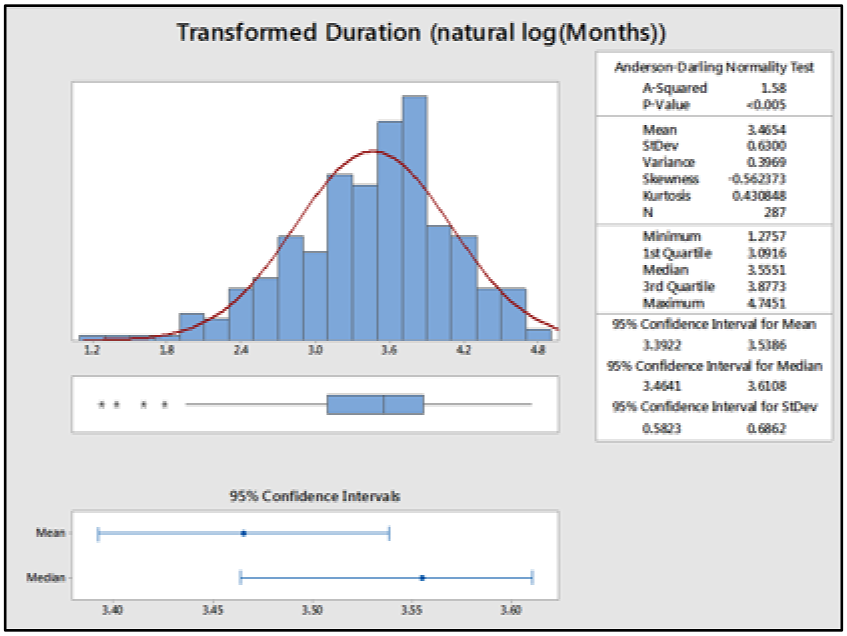


Figure 7. Duration, Normalized

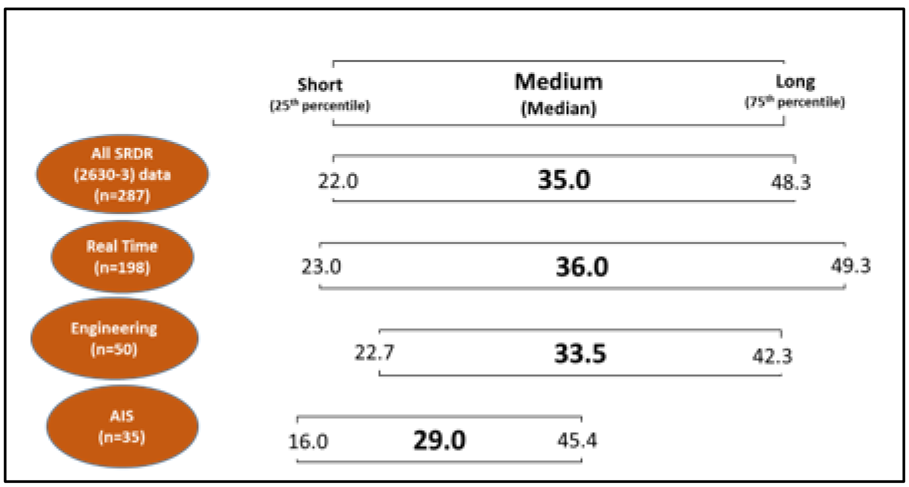


Figure 8. Duration Data by Super Domain

Team Size (People)

The size of the development team reported here is based on measures of project effort and duration. The effort for a project is reported in labor hours. Labor hours are converted to person months of effort (based on 152 hours/month) and divided by months of project duration. This derives the average level of project staffing or full time equivalent (FTE).

Figure 9 shows a histogram of the data in natural log scale. It shows that most teams have 20 or fewer people. Recall that SRDRs are required for contracts over $20 million. These contracts have multiple product events resulting in seemingly small team sizes which, in fact, are due to low levels of effort over relatively long durations.

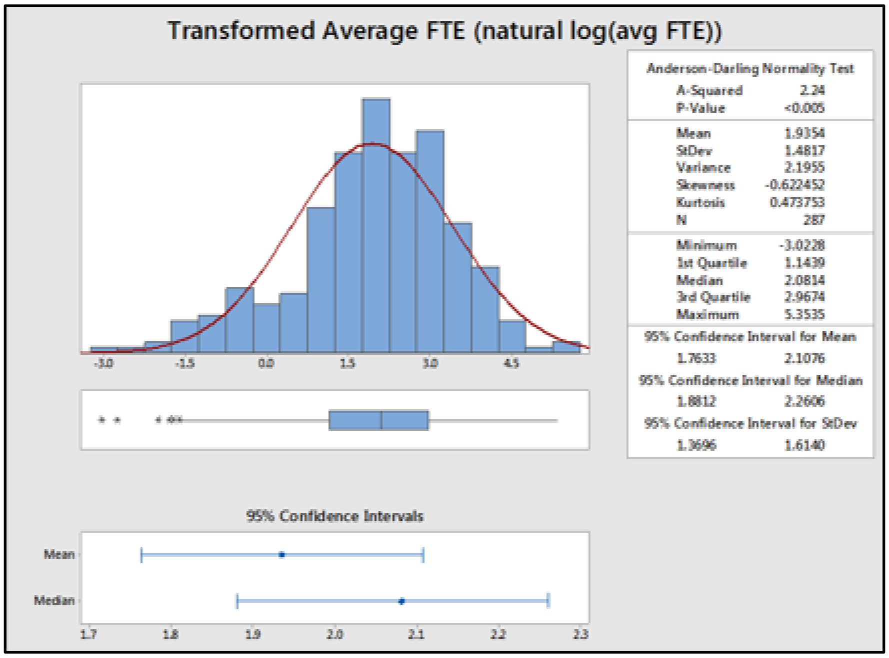


Figure 9. Time Size, Normalized

Figure 10 shows the data divided into three groups: small-, medium-, and large-team-size projects. The groups are based on a cumulative percentage divided into thirds. Small teams make up the lower third, medium size teams are in the middle third, and large teams make up the upper third. Based on the groupings the team sizes are as follows:

- small-size teams: < 5 average staff
- medium-size teams: 5–14 average staff
- large-size teams: > 14 average staff

Medium and large team sizes are used in the effort/schedule tradeoff analysis.

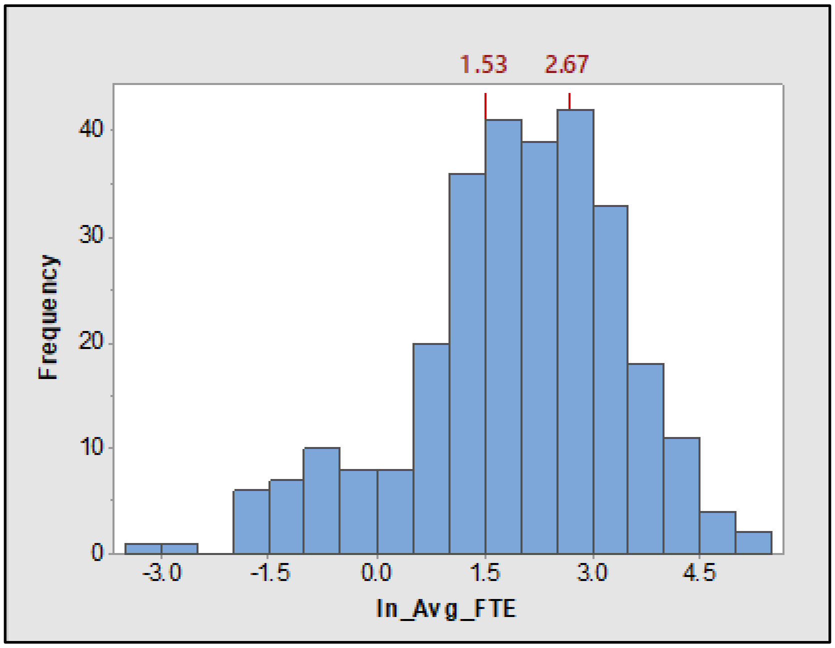


Figure 10. Team Size Distribution

Team size data analyzed by super domain is presented in Figure 11.

Figure 11. Team Size Data by Super Domain

Productivity

Productivity (also referred to as efficiency) is the amount of product produced for an amount of resource. For software, productivity is commonly measured by size (ESLOC) divided by effort hours.

Productivity in general is considered very competition sensitive and therefore rarely shared publicly by the private sector. Since the SRDR data set is owned and maintained by the government and the individual data provider's productivity is protected, this compilation of data provides a rarely available insight into software productivity across the industrial base.

Figure 12 shows the productivity data after normalization. For practical purposes, the data shows a 1:1 ESLOC: hour ratio. Productivity data analyzed by super domain is presented in Figure 13.

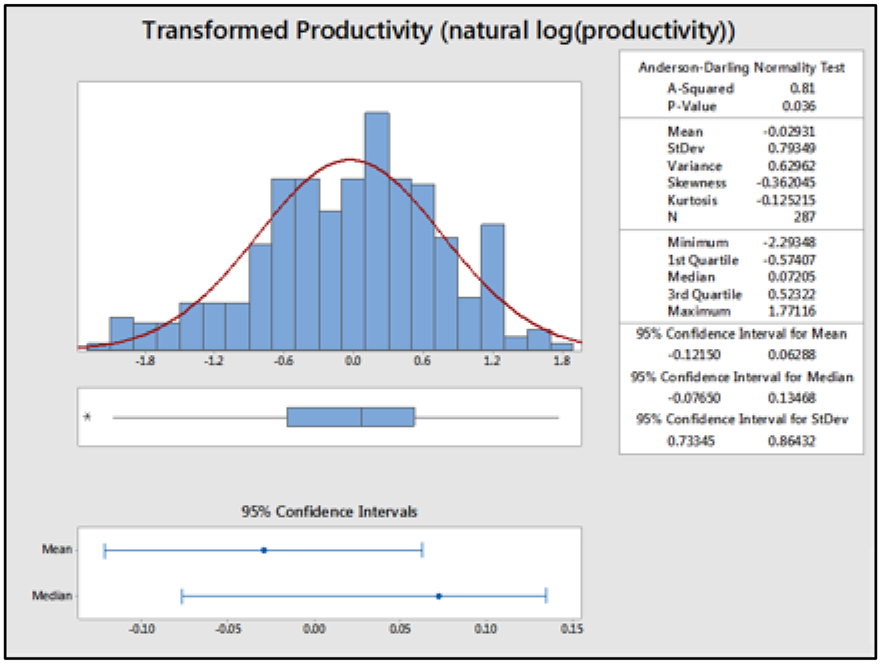


Figure 12. Productivity, Normalized

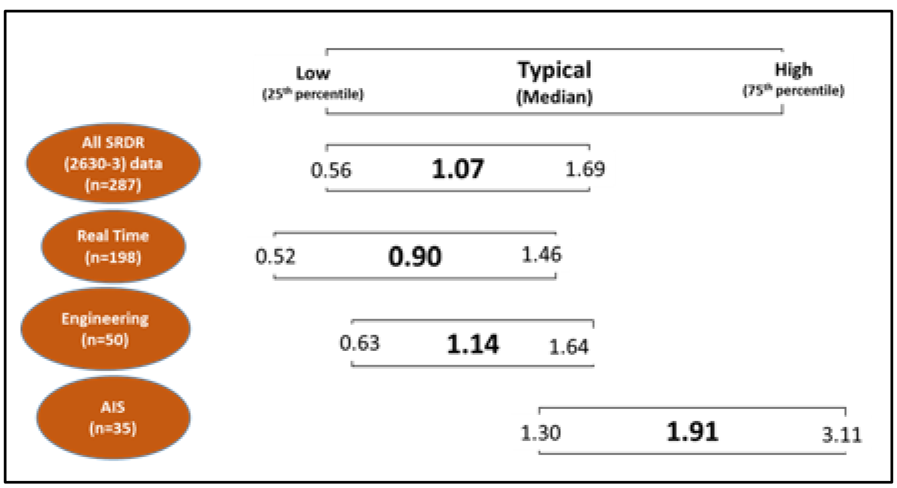


Figure 13. Productivity by Super Domain

Profiles of Typical Projects

Integrating the analysis results of the individual parameters provides a general software project profile. This section contains the profiles for a generic DoD software project, as well as profiles for RT, ENG, and AIS projects.

As a reminder, the SRDR reports are provided at the project level or subsystem level, not at the DoD Acquisition Program level. **It is important to note that when the analysis refer to a "project" in this report, a project is synonymous with a software**

build, increment, or release. In many cases, several projects (i.e., data points) would contribute to the overall scope and make up of Acquisition Program (i.e., an entire weapon system).

Snapshot of a Typical DoD Software Project

Figure 14 provides a snapshot of the overall dataset, showing the size and scope of a typical DoD software project. Keep in mind SRDR data points are typically submitted by subsystem or potential increment; these numbers do not represent an entire DoD program of record.

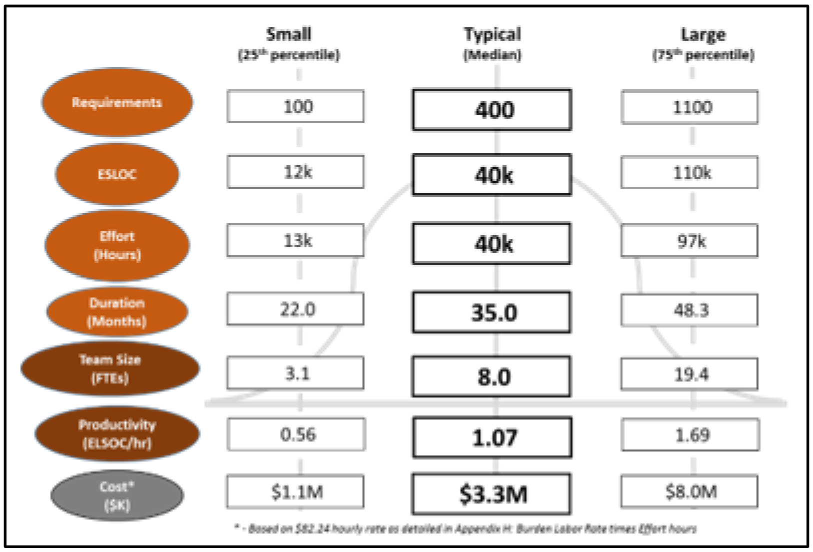


Figure 14. Parameters of DoD Software Projects

This data can be used to answer general questions about DoD software projects. For example,

- Question: What is the typical (average) size of a software delivery?
  Answer: 40 KESLOC
- Question: How long does an increment take?
  Answer: 35 months (~3 years)
- Question: How many FTEs does a typical software project require?
  Answer: 8 FTEs; some large projects may require upwards of 20 FTEs.
- Question: In general how much does a software project cost?
  Answer: Software projects tend to range between $1 million and $8 million; without knowing any details about what type of software or its composition, a generic DoD project costs a little over $3 million.

The percentile numbers help convey the variation in the data. These data can be utilized by oversight offices when assessing overall program feasibility. A project plan that contains parameter values outside the 25th and 75th percentile range signifies a situation that is not common and might require additional scrutiny. In this case, the oversight office would want to ask for more information about the engineering and technical rationale to justify this plan.

Given the mix of system domains, language types, environments, platforms, functionality, and associated quality/performance parameters, these rules of thumb may not provide a lot of value to project managers estimating their software efforts. To get the information useful to them, they would need to isolate like projects in the dataset and generate a parameter profile that best represents the system they are developing. In this vein, the following sections provide heuristics by super domains.

Snapshot of Real-Time Software Projects

RT software is typically the most complex and intricate type of software. It tends to be embedded in the system architecture and contributes to the success or failure of key performance parameters of the system. Given the level of rigor this type of software requires, the variations between the RT super domain parameters in Figure 15 are not surprising. Of the 287 data points analyzed, 198 were classified as real time.

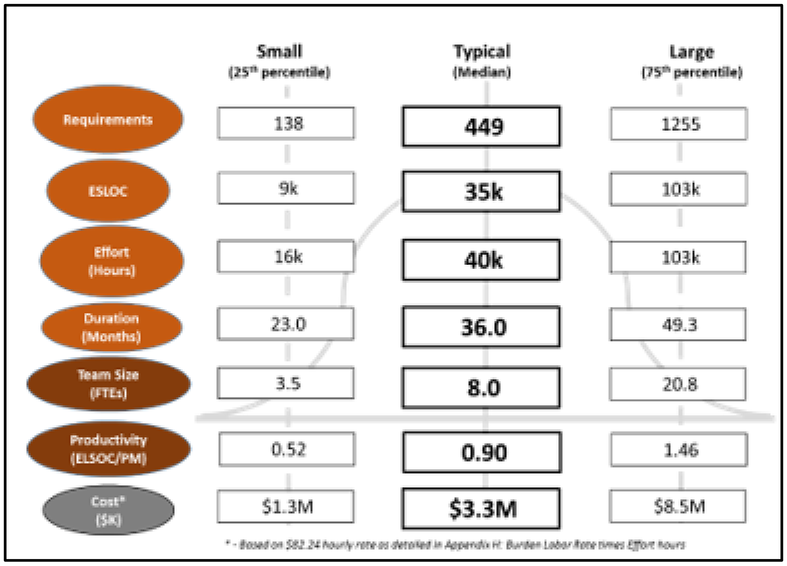


Figure 15. Parameters of Real-Time Software Projects

It is logical that increased system complexity would require a more detailed articulation of the requirements, resulting in a higher requirements count and lower productivity in comparison to the overall data set. This can also be seen in the slightly higher effort hour percentile values.

Snapshot of Engineering Software Projects

ENG super domain software is of medium complexity. It requires engineering external system interfaces, high reliability (but not life-critical) requirements, and often involves coupling of modified software. Examples of software domains in this super-domain are: mission processing, executive, automation and process control, scientific systems, and telecommunications.

Figure 16 shows the key software parameters for the 50 ENG super domain data points in the 287 data set.

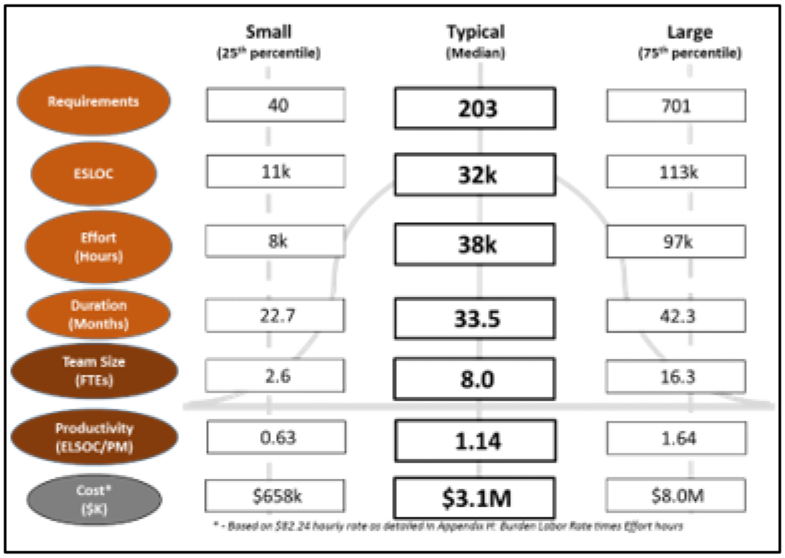


Figure 16. Parameters of Engineering Software Projects

In comparison to RT systems, ENG systems tend to state their requirements at a slightly higher level. For example, a typical requirement may be, "System X shall interface with System Y." In this case there are several nuances to meeting this requirement. This can be seen by comparing the requirements parameters, ESLOC, and effort parameters of the RT data to the ENG data.

Snapshot of Automated Information System Software Projects

AIS software automates information processing. These applications allow the designated authority to exercise control over the accomplishment of the mission. Humans manage a dynamic situation and respond to user input in real time to facilitate coordination and cooperation. Examples of software domains in this super-domain include intelligence and information systems, software services, and software applications.

Figure 17 shows the key software parameters for the 35 AIS super domain data points in the 287 data set.

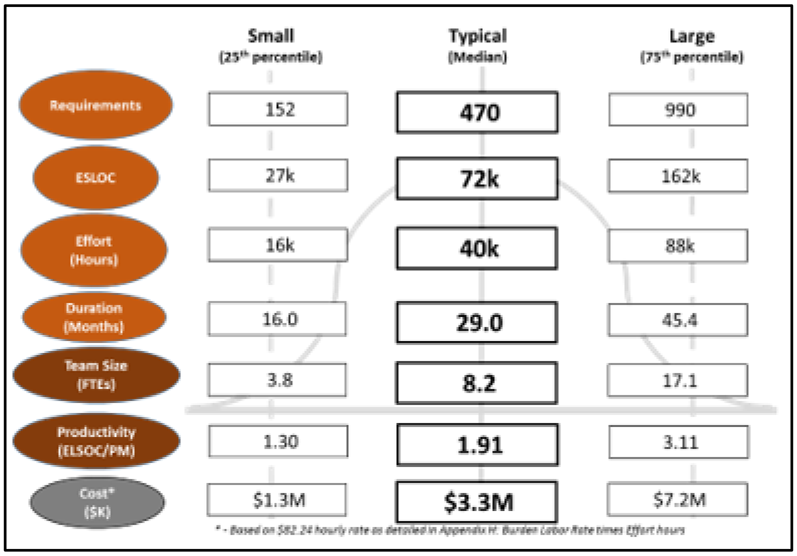


Figure 17. Parameters of Automated Information System Software Projects

The size and productivity parameters vary the most from the overall super domain parameters. Based on the way AIS are developed (i.e., adaptation of existing COTS/GOTS applications), the increase in comparison to the other super domains is not surprising.

Portfolio Performance: Common Questions

This section explores the findings by super domain to answer some common questions about software types.

Most and Least Expensive Software

What are the most and least expensive software types to develop?

Our analysis is based on the rationale that some types of software are more difficult to develop than other types and therefore require more effort to develop. The level of difficulty can be caused by factors such as execution timing constraints, interoperability requirements, commercial-off-the-shelf (COTS) software product incorporation, algorithmic complexity, communication complexity, data-bandwidth requirements, and security requirements. To account for the dissimilarities in project difficulty, projects are segregated into three super domains.

The analysis proceeds by introducing two concepts: unit cost and production rate.

- Unit cost is the cost of producing a unit of software with some amount of effort. In this case, the unit of software is thousands of equivalent source lines of code (KESLOC). The effort is reported in labor hours, which can be translated into cost using an average labor rate.
- Production rate is the rate at which a unit of software is delivered over a period of time. The unit of software is a KESLOC and the time is days of project duration.
- Cost is derived by applying a burdened labor rate to the number of labor hours worked in a day. Hours per day are determined by dividing total hours

by the duration (total days). For example, if a real-time project required 1,007 total hours and 25 days, the labor hours expended in a day is 40.3 (implying several people were working on the project).

The analysis then normalizes the unit cost with the production rate, creating a high-level comparison. This is done because some projects may choose to employ more staff to increase their production rate and deliver the software sooner or vice versa. The resulting effort per day is then multiplied by an average burden labor rate to derive cost.

Unit Cost

With an average project size of 40,000 ESLOC, each of the three groups are analyzed separately. Trends for each group were created based on a natural log-transformation of the data. This transformation made it easier to see the relationships between the three groups for an average project size of 40,000 ESLOC.

The difference in unit costs between the three groups is shown in Table 1. Real-time software shows that for small amounts of size, a large amount of effort is required. Automated information system software data shows the opposite: for large amounts of size, a small amount of effort is required.

Table 1. Unit Costs for Different Domains

Domain	Hours / KESLOC
Real-Time Software	1,070
Engineering Software	936
Automated Information System Software	578

Production Rate

The production rate data analysis focused on the relationships between size and duration for the three super domains. The analysis revealed much greater variability than the unit cost plot. This indicates a very weak systematic relationship between size and duration. The dispersion of the data is attributed to other factors that influence the size-duration relationship (e.g., different levels of staffing on similar size projects can impact duration). This is an area for further research.

For an average-size project, the production rate (how long it takes to deliver a unit of software) is shown in Table 2.

Table 2. Production Rate for Different Domains

Domain	Days / KESLOC
Real-Time Software	25
Engineering Software	26
Automated Information System Software	20

Cost Comparison

When unit cost is divided by production rate, the average number of hours each month is determined. Using an average burden labor rate, the normalized monthly cost for each group is shown in Table 3. The hours/day indicate that more than one person is working per day.

Table 3. Costs for Different Domains

Domain	Hrs / Day	Cost / Day
Real-Time Software	40.4	$3,324
Engineering Software	35.4	$2,912
Automated Information System Software	29.1	$2,393

Real-time software is the most expensive to develop and automated information system software is the least expensive. RT software costs 14% more to develop than ENG software and 39% more than AIS software.

Cost Heuristics

Units for cost vary based on the office reporting them and the types of decisions that are being made. Engineering organizations often prefer to discuss things in technical units (e.g., requirement and SLOC) and effort (e.g., hours or person months, months). Cost offices tend to communicate in terms of dollars and fiscal years. Table 4 is a translation table that shows the same unit cost, production rate, and cost data expressed in different units.

Table 4. Unit Cost and Productivity

Project Size (40 KESLOC)	Unit Cost	Production Rate		
Domain	Hours / KESLOC	Days / KESLOC	Hrs / Day	Cost / Day
Real-Time Software	1,007	25	40.4	$3,324
Engineering Software	936	26	35.4	$2,912
AIS Software	578	20	29.1	$2,393

Project Size (40 KESLOC)	Productivity				
Domain	ESLOC / Hour	ESLOC / Day	People (FTEs)	Cost Month	Cost per Year
Real-Time Software	0.99	40	5.1	$99,720	$1,196,640
Engineering Software	1.07	38	4.4	$87,360	$1,048,320
AIS Software	1.73	50	3.6	$71,790	$861,480

Table 4 provides the unit cost (hours/KESLOC) and its inverse, productivity (ESLOC/hour). Depending on the type of information needed, one of the metrics may be preferred over the other. Alternatively, production rate is a metric that can be expressed in terms of units of product produced in a period of time (days/KESLOC) or units of time to produce a single product (ESLOC/day). It also provides monthly and annual costs by domain. The cost by year represents the annual costs for an average project for a full calendar year. This number doesn't help an engineering organization determine the total cost of a particular project, but it is a useful metric to technical managers when they are required to submit an annual budget.

Best-in-Class/Worst-in-Class

What differences are there between best-in-class and worst-in-class software projects?

This analysis examines the best- and worst-in-class projects within each of the three super-domains discussed in the previous section. To assess differences between projects, the three derived metrics explained in the previous section are used: unit cost, production rate, and cost.

Analysis Approach

An average size project within each super domain is used to derive unit cost, production rate, and cost. A ±1 standard error (SE) about the unit cost and production rate trend lines were used to identify best- and worst-in-class projects.

The definition of best-in-class and worst-in-class projects were developed as follows:

- Best-in-class projects: at or below the −1 SE value are projects that used less effort or less time to finish than an average project. This boundary represents the worst of the best-in-class projects—performance may actually be better.
- Worst-in-class projects: at or above the +1 SE value are projects that used more effort or more time to finish than an average project. This boundary represents the best of the worst-in-class projects—performance may actually be worse.

Real-Time (RT) Software

Unit Cost

The average-size RT project (34,000 ESLOC for the RT domain) expends 39,664 labor hours of effort. Best-in-class projects expend 18,361 labor hours and worst-in-class projects expend 85,687 labor hours, a 10-fold increase. The difference between a best- or worst-in-class project from the average project is 21,304 labor hours. It is important to understand the context of the labor-hour differences in conjunction with project duration.

Production Rate

The average-size project delivers a product in 997 days (32.8 months). A best-in-class project delivers a product in 538 days (17.7 months). A worst-in-class project delivers a product in 1,848 days (60.8 months).

Cost

Table 5 summarizes the differences in unit cost and production rate between best-, average-, and worst-in-class RT projects. An average RT size project of 34,000 ESLOC was used to determine effort and schedule. Best-in-class RT projects are 2 times more efficient than average projects and 4.7 times more efficient than worst-in-class projects. Best-in-class projects are 1.8 times faster than an average projects and 3.4 times faster than a worst-in-class project. As mentioned earlier, the noted results for the best-in-class are the lowest reported numbers in the best-in-class bracket. Conversely, the reported results for worst-in-class are the highest reported numbers in the worst-in-class bracket.

Table 5. Real-Time Software Best and Worst Summary

Metric	Best-in-Class	Average	Worst-in-Class
Effort (Labor Hours)	18,361	39,664	85,687
Schedule (Days)	538	997	1,848
Cost (per Day)	$2,805	$3,271	$3,813
Total Cost ($M)	$1.510	$3.262	$7.047

Using a burden labor rate of $150,000 per year, the best-in-class project saves $1.752 million dollars over an average project and $5.537 million over a worst-in-class project.

Engineering (ENG) Software

Unit Cost

There are 50 projects in the ENG super-domain. The average-size project (32,000 ESLOC for the ENG domain) expends 30,780 labor hours of effort. The best-in-class expends 14,468 labor hours and the worst-in-class expends 65,485, a 4.5 increase times the amount of best in class. The difference between a best- and worst-in-class project from the average project is 16,312 hours.

Production Rate

The best-in-class project delivers a software product in 640 days (21 months), an average project in 1,031 days (33.9 months), and a worst-in-class project in 1,659 days (54.6 months).

Cost

Table 6 summarizes the differences in unit cost and production rate between best, average, and worst-in-class ENG projects. An average ENG size project of 32,000 ESLOC was used to determine effort and schedule. The best-in-class ENG projects are 2.3 times more efficient than average projects and 5.3 times more efficient than worst-in-class projects. The best-in-class project is 1.6 times faster than an average project and 2.6 times faster than a worst-in-class project.

Table 6. Best and Worst Summary of Engineering Software

Metric	Best-in-Class	Average	Worst-in-Class
Effort (Labor Hours)	14,468	30,780	65,485
Schedule (Days)	640	1,031	1659
Cost (per Day)	$1,859	$2,456	$3,246
Total Cost ($M)	$1.190	$2.531	$5.385

Best-in-class projects save $1.341 million dollars over average projects and $4.195 million dollars over a worst-in-class project.

Automated Information System (AIS)

Unit Cost

Using an average-size project of 72,000 ESLOC, best-in-class, average, and worst-in-class projects expended an average of 22,400, 39,114, and 68,297 labor hours of effort, respectively. There is a three-fold increase in effort between best and worst-in-class. The difference between a best or worst-in-class project and the average project is 16,713 labor hours.

Production Rate

The best-in-class average-size project delivers a product in 445 days (14.6 months). The average project delivers a product in 880 days (29 months). The worst-in-class a project delivers product in 1,743 days (57.3 months).

Cost

Table 7 summarizes the differences in unit cost and production rate between best, average, and worst-in-class projects. An average AIS size project of 72,000 ESLOC was used to determine effort and schedule. That makes best-in-class projects 1.7 times more efficient than average projects and 3 times more efficient than a worst-in-class projects. Best-in-class projects are 2 times faster than average projects and 4 times faster than worst-in-class projects.

Best-in-class projects save $1.375 million over average projects and $3.774 million over worst-in-class projects.

Table 7. Best and Worst Summary of AIS Software

Metric	Best-in-Class	Average	Worst-in-Class
Effort (Labor Hours)	22,400 (% of avg.)	39,114	68,297 (% of avg.)
Schedule (Days)	445	880	1,743
Cost (per Day)	$4,144	$3,654	$3,223
Total Cost ($M)	$1.842	$3.217	$5.616

Project Planning, Trade-Offs and Risk

As part of our analysis we also explored correlations among requirements, size, duration, and effort. The goal of this work was to determine how well the data could be used to answer common questions related to planning or replanning software projects, such as “How much growth should we plan for?” and “How well can initial estimates be used to predict final outcomes?”

The Department of Defense Software Factbook provides a more extensive description of our work in this area, while this paper provides a brief overview of the strongest models we found to predict growth in requirements, ESLOC, schedule, and effort from the initial estimates. Each of the models can be used to construct predicted growth intervals for any given initial estimate, although we caution against using the model outside the bounds indicated by the 5th and 95th percentiles for each variable.

Estimation Relationships

Among the many factors and models for estimating effort, the SRDR data allows us to investigate the relationship between requirements and the size of the effort and then the relationship between the estimated size and the estimated effort as well as the final effort. A simple look at the correlations among requirements, size, duration, and effort found that the only actionable correlation was between size and effort.

Predicting Actual Total Effort by Estimated ESLOC

The following model shows that an initial estimate of ESLOC can also be used to predict the total actual effort. Although the model is only moderately strong, it is presented here in case an initial estimate of effort is not available, but an estimate of size (ESLOC) is available.

n = 163

Regression Equation:

$$\ln \text{Total Hours\_Actual} = 2.031 + 0.8259 \ln \text{ESLOC\_Estimated}$$

which translates to:

$$Actual\ Total\ Hours = 7.614 * (Estimated\ ESLOC)^{.83}$$

The table shows the predictions have a “sweet spot” that is +/− 10% in the range from 75KESLOL to 200 KESLOC. The model accounts for over 67% of the variance. Below are the predicted (forecast) values and prediction ranges for a set of new given inputs, followed by a graphic showing the actual data fitted to the model along with the associated prediction intervals. Predicted values show an underestimate of the initial by 158% at the low end (500 ESLOC) but an overestimate of −22% at the high end (500K ESLOC). This indicated that the model is reasonably good fit to the data.

Table 8. Prediction Values for Actual Total Hours (Effort) Using ESLOC

Initial ESLOC Estimate	Forecast Total Hours	Percent Difference From Estimate	Prediction Interval—Total Hours	
			Lower 95%	Upper 95%
500	1,291	158%	264	6,305
750	1,805	141%	372	8,747
1,000	2,289	129%	475	11,040
2,500	4,879	95%	1,024	23,235
5,000	8,648	73%	1,828	40,911
7,500	12,088	61%	2,562	57,025
10,000	15,330	53%	3,255	72,213
25,000	32,675	31%	6,949	153,635
50,000	57,921	16%	12,300	272,755
75,000	80,961	8%	17,158	382,026
100,000	102,674	3%	21,717	485,437
150,000	143,515	−4%	30,249	680,898
200,000	182,006	−9%	38,248	866,094
300,000	254,403	−15%	53,200	1,216,562
400,000	322,634	−19%	67,199	1,549,009
500,000	387,926	−22%	80,526	1,868,786

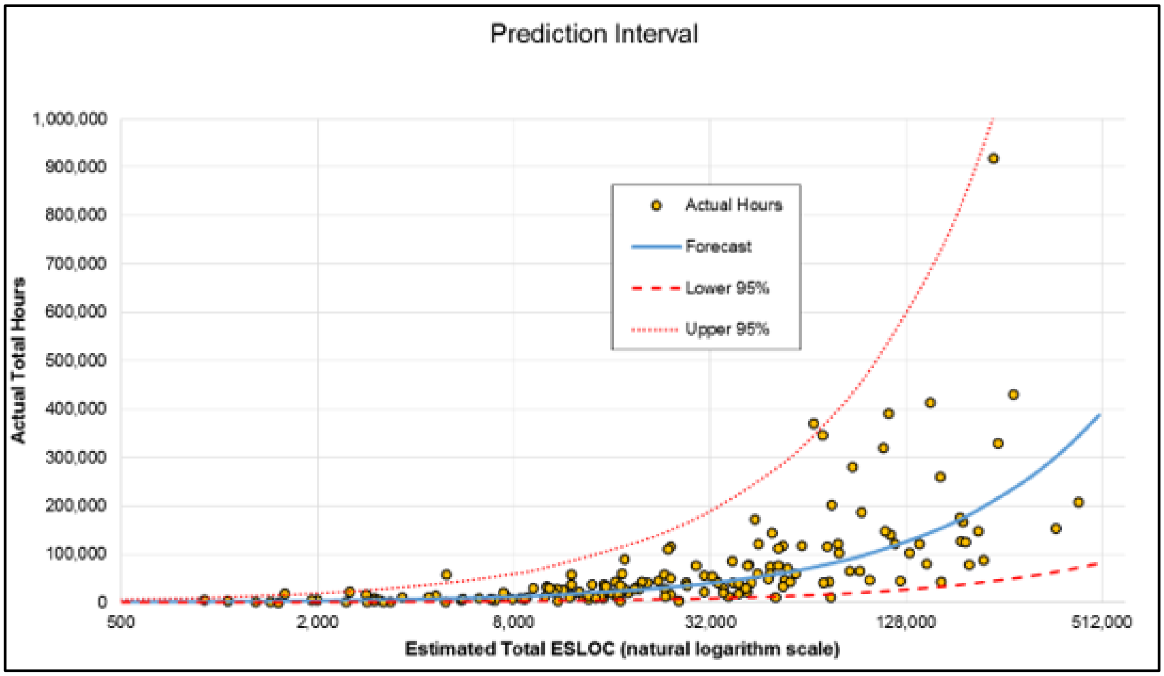


Figure 18. +/−10% is the Range for 75,000 to 200,000 Initial ESLOC Estimates With +/−10%

Software Growth—Predicting Outcomes

To determine if final outcomes can be predicted from initial estimates we examined the project performance as represented by 181 paired initial and final contractor submissions. As such, we measured the difference between the initial estimates and the actual outcomes.

Based on historical SRDR data transformed to natural logarithms, we determined that we can predict (with a known degree of certainty) the expected outcomes for software size, schedule, and effort. The models presented enable predictions of final outcomes based on initial estimates. Each of the models can be used to construct outcome prediction intervals for any given initial value, although we caution against using the model outside the bounds indicated by the 5th and 95th percentiles for each variable.

While the full report describes the data and statistical analyses in more detail, we provide here an overview of the strongest models to emerge from this analysis:

Requirements	$(r^2 = .936)\ Actual\ Total\ Reqts = 1.2838 * (Estimated\ Total\ Reqts)^{.9456}$
ESLOC	$(r^2 = .849)\ Actual\ Total\ ESLOC = 2.0157 * (Estimated\ ESLOC)^{.964}$
Schedule	$(r^2 = .776)\ Actual\ Total\ Duration = 2.3054 * (Estimated\ Total\ Duration)^{.7878}$
Effort	$(r^2 = .898)\ Actual\ Total\ Hours = 3.3128(Estimated\ Total\ Hours)^{.9097}$

Predicting productivity is less strong unless we separate the underestimated cases from the overestimated cases, which then yield very strong models (r^2 equals .886 and .758, respectively). This indicates that if the productivity could be assessed during the development effort, then the actual outcome could be more accurately predicted. If we also account for the type of super domain, these models increase in strength.

Schedule duration can also be separately predicted for the three services. We show that total effort hours can also be predicted by using the initial estimate for ESLOC, although the fit is not as strong (r^2 = .674) as using the initial estimate for hours. We also show how the prediction interval becomes tighter when the confidence level for the prediction is reduced.

Perhaps the most useful takeaway from this analysis are the prediction tables. The tables provide the predicted value along with the prediction interval at a 95% confidence level. These can be used in the absence of any available estimates, or as a sanity check against estimates coming from other sources. New values can easily produce a ballpark forecast by interpolation or the actual equation can be used for calculation. The data sets we used are also available for distribution which enable users to reproduce the models with their own statistical software.

As mentioned earlier, no further adjustments were made in case selection once the data were trimmed. Undoubtedly, the models could be improved (and the predictive intervals narrowed) with substantive knowledge concerning the behavior of outliers which could provide meaningful reasons for their exclusion from a model. Also, any additional data supplied during the interim of the project—which is under consideration by the DoD—could further calibrate and improve a model's fit. This would be especially useful in the productivity models where the best fits were determined by whether the original submission over- or underestimated the productivity. A midcourse determination of productivity would then indicate which sub-model was appropriate to estimate the final productivity for the project.

Conclusion

The analyses conducted by the SEI shows that the cost of software development varies depending on several factors. The class or super-domain of software makes a difference in the cost of software. Different super domains have different levels of difficulty that cause more effort to be expended on more difficult software. On an average-size project, AIS software costs $31,350 a month and RT software costs $101,250 a month—more than three times as much.

The time to develop software also drives cost. Based on an average-size project, shorter duration projects cost disproportionately more than longer duration projects. It was shown that team size is clearly NOT determined solely by the size of the software to be built. The performance of a project also drives cost. The analysis looked at best-, average-, and worst-performing projects within each super-domain.

Perhaps the most valuable contribution of this study is the ability to provide, for the first time, guidance to decision makers about projects that is based on empirical analysis. Table 9 summarizes the basic benchmarks that can now be used throughout the DoD.

Table 9. Basic Benchmarks for DoD Software Projects

DoD Software Projects: Basic Benchmarks	*Small Projects* *(25th percentile)*	***Average/Typical***	
Requirements: What is the functional size of a DoD software project?	100 requirements	**400 requirements**	1100 requirements
ESLOC: How many lines of code do DoD software projects contain?	12,000 lines of code	**40,000 lines of code**	110,000 lines of code
Effort: How many hours of work does it take to complete DoD software projects?	13,000 hours	**40,000 hours**	97,000 hours
Duration: How long do DoD software projects last?	22 months	**35 months**	48.3 months
Team size: How many people work on DoD software project teams?	3.1 FTEs	**8 FTEs**	19.4 FTEs
Productivity: How many lines of code per hour do DoD software projects produce?	0.56 ESLOC per hour	**1.07 ESLOC per hour**	1.69 ESLOC per hour
Cost: How much do DoD software projects cost?*	$1.1 M	**$3.3 M**	$8 M

**Based on an $82.24 hourly rate*

While this information is valuable, there was not enough background data on projects to investigate some important issues, such as why best and worst projects perform differently. This leads to the next steps. There is an effort to link the project data back to source documents and other data to provide the capability to investigate the data more fully. There is a lot of data and source material, and some progress has been made to date with a lot more to do. While more analysis will be done, we would like to hear from you. What are the important questions that need answers? For comments and suggestions, please contact: fact-book@sei.cmu.edu.

Reference

Boehm, B., Abts, C., Brown, W., Chulani, S., Clark, B., Horowitz, E., Madachy, R., Reifer, R., & Steece, B. (2000). *Software cost estimation with COCOMO II.* Prentice Hall.

Disclaimer and Distribution Statement

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

DoD’S Software Sustainment Ecosystem: Needed Skill Sets and Gap Analysis

Forrest Shull—is Assistant Director for Empirical Research at Carnegie Mellon University’s Software Engineering Institute. His role is to lead work with the U.S. Department of Defense, other government agencies, national labs, industry, and academic institutions to advance the use of empirically grounded information in software engineering, cybersecurity, and emerging technologies. He has been a lead researcher on projects for the U.S. Department of Defense, NASA’s Office of Safety and Mission Assurance, the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation, and commercial companies. Dr. Shull serves on the IEEE Computer Society Board of Governors and Executive Committee. [fjshull@sei.cmu.edu]

Michael McLendon—is an Associate Director at Carnegie Mellon University’s Software Engineering Institute. Previously, he served as Senior Advisor in the Office of the Assistant Secretary of Defense for Systems Engineering. He was also a principal in the Office of the Assistant Secretary of Defense for Program Analysis & Evaluation and also in the Office of the Under Secretary of Defense for Policy. He was a professor at the Defense Systems Management College and a career Air Force officer, serving in a range of leadership and management positions in system and technology development and acquisition as well as the federal level and the private sector. [mmclendon@sei.cmu.edu]

Christopher Miller—is a Senior Researcher at Carnegie Mellon University’s Software Engineering Institute. Dr. Miller’s expertise is in software metrics, measurement, and estimation of software intensive systems. His quantitative analysis background is focused on lifecycle cost estimation, evaluating project feasibility measures, establishing performance measurements, and providing analysis for the optimization of systems. Dr. Miller served as the Chair of the International Council on Systems Engineering (INCOSE) Measurement Working Group (MWG) for seven years and is a certified trainer for Practical Software and Systems Measurement (PSM). He earned a Master’s of Engineering Management and PhD in Systems Engineering at the George Washington University and is currently a member of the adjunct faculty in the School of Engineering and Applied Science (SEAS). [clmiller@sei.cmu.edu]

Introduction

Software is the foundational building material for the engineering of the Department of Defense (DoD) systems—the principal means for delivering almost 100% of the integrated functionality of kinetic weapon systems. Software is also the means for creating warfighter competitive advantage in today’s net-centric warfare environment, where the flow of information in real time is critical to the execution of the DoD’s mission across all domains. There is no plateau in sight for the advancement of software technology and its extensive use by the DoD in new systems, as well as to enhance the capabilities of legacy systems and extend their operational value far beyond their designed service life.

To maintain its competitive edge, it is imperative that the DoD have the capability and capacity to affordably acquire and sustain software-reliant systems to continually operate and achieve mission success in a dynamic threat, cyber, and net-centric environment. However, the DoD is strategically challenged to produce high-quality software more affordably and efficiently across the system lifecycle, as noted by the Defense Science Board (2000) and others (National Research Council [NRC], 2010a). The acquisition and sustainment of software, particularly for distributed real-time and embedded systems, remains high risk and more problematic as individual system and system-of-systems complexity continues to grow.

As long recognized, successful acquisition of software-intensive systems by the DoD is driven to a significant degree by the competencies of the DoD’s organic software engineering workforce in applying evidence-based knowledge and practice throughout a system's lifecycle (Software Engineering Institute [SEI], 1998). In a prior study, we emphasized the need to better address software sustainment issues, particularly by engaging appropriate software expertise at the right points early in the acquisition lifecycle, when critical engineering decisions are made (Shull & McLendon, 2017). This means that the DoD organic software engineering sustainment community must be an active participant early in the requirements and engineering process and that the product support manager in acquisition programs must be knowledgeable and proactive in representing software sustainment equities.

Achieving this early engagement to influence software design-for-sustainability requires DoD organic software engineering staff who are not just knowledgeable about software but also "street-smart" in system acquisition. Many DoD sustainment programs with which we have interacted over the last few years have been separately developing software workforce competencies to enable effective engagement early in the acquisition process.

In this paper, we synthesize what we have learned to date regarding an initial model to assist the DoD in thinking about DoD software engineering competencies. We emphasize that this is an initial model recognizing that defining workforce competencies is complex and also dynamic given the nature of software and system sustainment policies.

Research Goal, Scope, and Methodology

Our research goal was to characterize the state of the practice regarding the DoD's software sustainment workforce with respect to the range of roles and related skills required, from which an initial model of the relevant competencies could be created. We also captured some of the recurring challenges related to workforce issues and the role of contractor versus the DoD organic software sustainment workforce in addressing those challenges.

This work was conducted in the context of ongoing work focused on software sustainment in weapons systems; therefore, the direct applicability of our results are limited to that domain. Software in this domain can typically be characterized as embedded software (i.e., software that interacts with physical components to provide functionality for the overall weapons system). Acquisition of embedded software presents some of the most technically difficult and resource-intensive software engineering challenges because of tightly coupled interfaces, integration with unique hardware, real-time requirements, and very high reliability and assurance needs due to life-critical and mission-critical demands. However, the DoD has a substantial amount of software across many other domains: business systems; mission support systems (e.g., test equipment, mission planning, engineering models, and simulations); mission-critical, non-embedded systems; and modeling and simulation, among others. While our initial results can be used to understand the software workforce issues for other types of software-intensive systems, a more detailed description of how to tailor results for those domains is a subject of future work.

Our team leveraged multiple streams of data and information for this study.

- Literature Search—The body of knowledge related to software engineering is extensive. However, there has been limited systematic study focused on DoD software sustainment; therefore, there is no organized set of literature and ongoing study or research agenda to create and refresh a software sustainment body of knowledge
- SEI DoD Engagements—The SEI has been actively engaged with the military services for three decades to provide technical expertise to enhance organizational capabilities (processes, practices, and competencies) for

software engineering across the lifecycle and solve technical challenges for specific weapon system and information system programs.

- Interviews With Key Leaders—The SEI complemented its research with information from meetings with key leaders across all four Services, including (1) those in the Senior Executive Service (SES), (2) senior managers and staff in the Office of the Secretary of Defense (OSD), and (3) those from industry. This study was conducted at the unclassified level, and our interviews with DoD sustainment staff were conducted under the conditions of non-attribution to enable an open exchange of perspectives with senior leaders, managers, and staff engaged in software sustainment.

Context: Workforce Issues in the Software Sustainment Ecosystem

Results from our prior research indicate that the DoD's software sustainment infrastructure is best described and understood as an *ecosystem* composed of interrelated elements. We found over and over that the factors that drive software sustainment are highly interrelated. For example, it is difficult to discuss the workforce needed to perform necessary sustainment activities without first understanding the business model in terms of public-private partnerships, which activities can be done by contractors, and which activities need to remain in the organic DoD workforce. Decisions about the nature and types of these business models may also be influenced by the degree to which the government has provisioned for and exercised its technical data rights for a given program at the time of developing an acquisition strategy and contract. These decisions have implications for the scope of the software sustainment system. Because of the high degree of connectivity that exists among the drivers and factors, we use the metaphor of an "ecosystem" to describe the interdependencies among these elements; decisions made at any point are affected by and affect whole series of other decisions.

Based on our research, we created a framework that describes the software sustainment ecosystem, depicted in Figure 1. We abstracted the issues raised in our discussions with DoD sustainment stakeholders into *six demand drivers* and 10 *ecosystem elements*, which were described more fully in a prior paper (Shull & McLendon, 2017).

The six demand drivers, shown in the outer ring of Figure 1, represent requirements that are generated by changes in the weapon system's mission profile, funding availability, evolution of the underlying technologies, and so forth. These drivers capture the fact that DoD systems exist in an environment that is highly dynamic, where there is a need to respond to constantly changing threats and mission needs. This dynamism drives many of the system changes that must be made during software sustainment. For many of these changes, the most cost-effective way of implementing the new capability relies on the unique flexibility of software.

The 10 ecosystem elements, shown as interconnected "bubbles" in Figure 1, are the tightly interconnected factors that sustainment organizations must manage to effectively and continuously engineer the software. The drivers and elements of this ecosystem represent a virtual spider web of linkages and relationships.

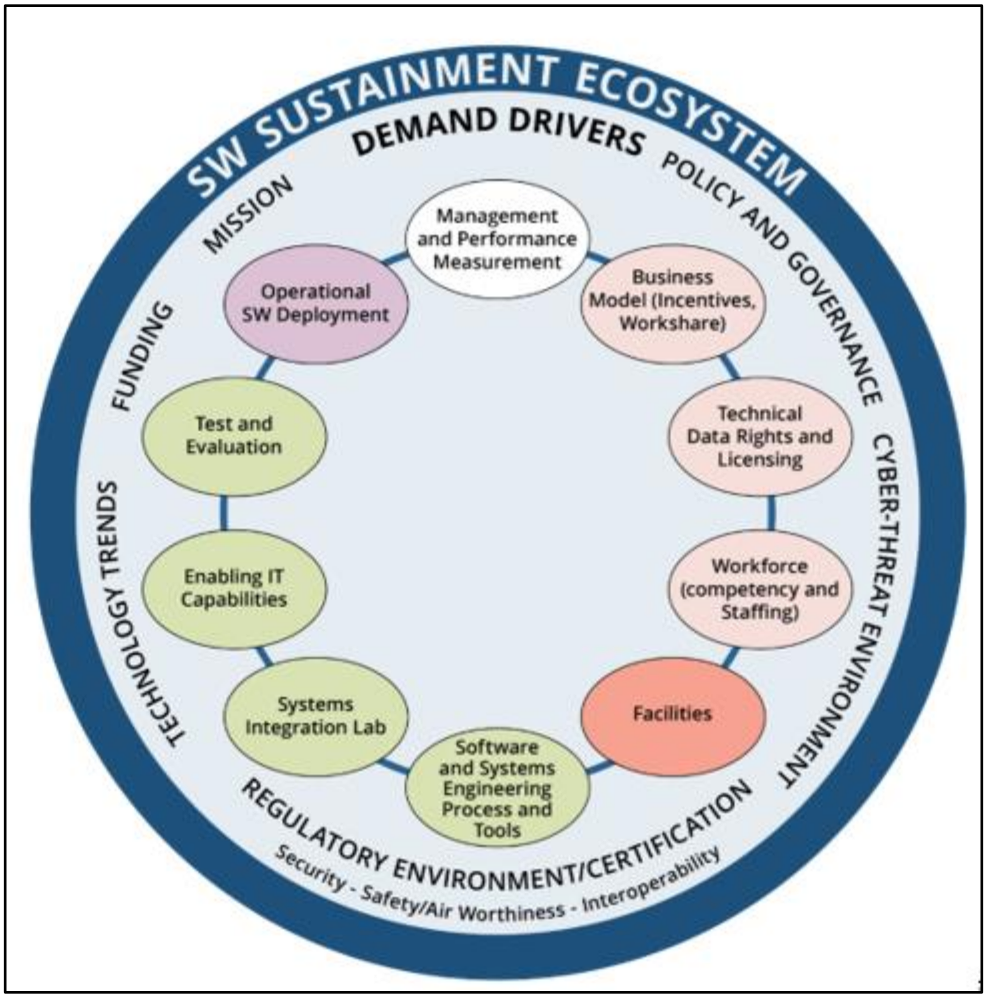


Figure 1. The DoD Software Sustainment Ecosystem Framework

Among these ecosystem elements, the three **knowledge and expertise** elements (shown in light orange in Figure 1) are most closely related to workforce skill sets. These elements include the factors that describe how the necessary skill sets are brought to bear for sustainment activities, and how the government grows its organic workforce and accesses necessary technical information—perhaps with some level of interaction with the private sector—to deliver and deploy the capabilities that must go to the warfighter. They consist of the following:

- **Workforce (Competency and Staffing)—**The means of accessing a sufficient workforce with appropriate skill sets, as well as a balance of organic and non-organic staff
- **Business Model (Incentives, Workshare)—**The strategic decision regarding which parts of the work will be done by the organic workforce and which by contractors, and how the overall work is managed both technically and contractually
- **Technical Data Rights and Licensing—**The tactical decisions governing what technical information is necessary to be accessed by the organic workforce, and the mechanisms by which they have access and the ability to maintain their working knowledge

Historical Context of DoD's Software Workforce

Several DoD studies dating back to 1982 have raised concerns about the technical competencies and size of the DoD's software workforce.[1] The key threads from previous relevant studies are summarized below. We argue that the DoD today remains challenged to find the highly skilled systems and software engineers needed now and in the future.

As early as 1993, the DoD Acquisition Management Board identified the need to review the DoD's software acquisition management education and training curricula, which was the first attempt by the DoD to establish a set of software acquisition management key competencies for the acquisition workforce. At that time, no existing DoD workforce functional management group was responsible for identifying the software competencies needed in the workforce. The board asserted that no new career field was needed for software acquisition managers and also made a key assumption. The board assumed that some personnel in acquisition programs clearly require more knowledge of software development and acquisition management than others, and that within each program there was an experienced individual fulfilling that role.

By 2001, the same concerns regarding the software competencies of the DoD acquisition work-force surfaced again. The DoD Software Intensive Systems Group, then in the Office of the Deputy Under Secretary of Defense for Science & Technology, conducted a software education and training survey of the acquisition workforce (SEI, 2017a). The findings from this study led to three specific recommendations: (1) Institute mandatory software-intensive systems training for the workforce, (2) develop a graduate-level program for software systems development and acquisition, and (3) require ACAT 1 programs to identify a chief software/systems architect.

A year later in 2002, Congress mandated in the NDAA for fiscal year 2003 that the DoD establish a program to improve the software acquisition processes. Subsequently, each Service established a strategic software-improvement program (Army 2002, Air Force 2004, and Navy 2006). These Service initiatives have continued at some level; however, with the sun setting of the Software Intensive Systems Group at the OSD level, the enterprise focus on software has waned.

In 2006, the DoD once again recognized that the sound application of modern software technologies and the use of sound software engineering practices over the acquisition lifecycle was critical to program execution given the increasing reliance on software in DoD systems. As a result, the DoD sponsored an industrial base study (Chao, 2006) to assess the nation's software workforce. The study concluded that while the nation's overall supply of software engineers may be adequate for the near term, there was a significant shortfall in the number of "top tier" software managers, architects, and domain experts. More importantly, the study estimated that perhaps as few as 500 engineers had the skills to develop the DoD's complex, software-intensive systems. This study did not specifically address the DoD's demand for government software engineers in acquisition

---

[1] Defense Joint Service Task Force Report on DoD Software Action Plans, 1982; Report of the Defense Board Report on Military Software, July 1987; Adapting Software Development Poli-cies to Modern Technology, Air Force Studies Board, National Research Council, 1989; DoD Information Systems Workforce Education, Training, and Career Development, Executive Re-sources Task Force Report, October 1992; DoD Software Master Plan, DAB S&T Committee, February 1990.

and sustainment, but as recognized by the DoD, it reported that these "shortfalls in top tier talent are evident there as well." The National Defense Industrial Association echoed similar concerns about the need for software knowledge and skills in its 2008 Report on Systemic Root Cause Analysis of Program Failures.[2]

During this same period, the Navy started the Software Process Improvement Initiative (SPII), which identified issues preventing software-intensive projects from meeting schedule, cost, and performance goals. This initiative highlighted the lack of adequately educated and trained software acquisition professionals and systems engineers. Subsequently, the Navy SPII Human Resources Focus Team recommended that the DoD use the findings from the Navy's report as a baseline to analyze the software competencies of the acquisition workforce.

As the result of this focus on the software workforce, the OSD issued guidance to create the Software Acquisition Training & Education Working Group (SATEWG) with a charter to affirm required software competencies, identify gaps in Defense Acquisition Workforce Improvement Act (DAWIA) career fields, and develop a plan to address those gaps. This group was composed of representatives from the Services, the OSD, and other organizations, including the SEI. The group developed a software competency framework that identified four key knowledge areas and 29 competencies that could inform the different acquisition workforce managers about the software competencies to be integrated into their existing career field competency models (Lucero, 2010). There has been no follow-on effort to evaluate the progress of the SATEWG or its outcomes.

More recently, each Service, as well as at the software sustainment organizational level, has evolved its own approach or model for identifying software competencies for its workforce. For example, one model in the DoD is based on establishing a software competency based organization with skilled people, processes, tools, mission facilities, and core technologies to support program teams and other customer needs. This model is enabled by a standard skills package that provides a description of the unique competency skills and associated criteria necessary for individuals to gain certification at one of three levels. In this competency-based organizational model, software staff are assigned as appropriate to serve as members of a program office functional engineering team or a software sustainment organization. As of 2018, the DoD Information Technology (IT) Functional Integrated Product Team (FIPT) has responsibility for DoD software workforce competencies.

This historical context highlights two key points. First, the DoD has long recognized the challenges of addressing the technical competencies and size of the software workforce across the lifecycle. However, there is limited evidence these different efforts had any lasting impact or resulted in meaningful outcomes. Second, this history clearly indicates that acquiring software human capital and equipping that workforce with the necessary competencies is a persistent and dynamic challenge that demands a continuous enterprise strategy. Our engagements with the DoD's software-sustainment organization clearly demonstrate the strategic and practical challenges in dealing with these issues.

---

[2] https://www.ndia.org/-/media/sites/ndia/divisions/systems-engineering/ndiasrcareportfina18dec2008.ashx?la=en

Challenges in the State of the Practice

A key challenge today is that there is no visibility into the number of personnel in the DoD's software sustainment organic workforce, and even less insight into the skills and background that they bring to their job.

Figure 2 shows 2016 data (the most recent we could find) that identifies the DoD's engineering workforce by occupational code and size as well as the number of personnel who are considered to be part of each acquisition engineering career field. (Those personnel engaged specifically in software sustainment represent a subset of these numbers.)

From this view, the software workforce in the DoD, whether in acquisition programs or organic sustainment organizations, cannot be accounted for at the DoD or Service enterprise levels. It is also not clear if all engineers across the DoD engaged in software sustainment are known. Based on our interviews, the principal engineering occupational codes most represented in the software sustainment workforce are 0855, 0801, 1550, 0854, and 0850. However, a list of all government personnel in these codes does not capture all government personnel performing software engineering duties.

DoD-Wide

Engineering (Non-Construction)

Civilian Occupational Series & Description	Count
0801 - GENERAL ENG	16,493
0802 - ENG TECHNICAL	10,691
0806 - MATERIALS ENG	840
0818 - ENG DRAFTING	2
0830 - MECHANICAL ENG	11,129
0840 - NUCLEAR ENG	2,430
0850 - ELECTRICAL ENG	3,458
0854 - COMPUTER ENG	3,486
0855 - ELECTRONICS ENG	16,413
0856 - ELECTRONICS TECHNICAL	5,574
0858 - BIOENG AND BIOMEDICAL ENG	99
0861 - AEROSPACE ENG	4,287
0871 - NAVAL ARCHITECTURE	866
0873 - MARINE SURVEY TECHNICAL	118
0881 - PETROLEUM ENG	1
0893 - CHEMICAL ENG	807
0895 - INDUSTRIAL ENG TECHNICAL	993
0896 - INDUSTRIAL ENG	1,044
0899 - ENG AND ARCH STUDENT TRAINEE	1,197
GRAND TOTAL	79,928

Engineering (Non-Construction)

Component	Count
Army	19,025
DoN	41,687
Air Force	15,381
4th Estate	3,835
GRAND TOTAL	79,928

Defense Acquisition Workforce

Acquisition Engineering Career Field

Civilian Occupational Series & Description	Count
0855 - ELECTRONICS ENG	10,730
0801 - GENERAL ENG	7,869
0830 - MECHANICAL ENG	5,976
1550 - COMPUTER SCIENTIST	3,496
0861 - AEROSPACE ENG	2,907
0854 - COMPUTER ENG	2,523
0850 - ELECTRICAL ENG	1,423
1515 - OPS RESEARCH ANALYST	659
1310 - PHYSICIST	553
0893 - CHEMICAL ENG	499
OTHER (INCLUDING ACTIVE DUTY MILITARY)	4,690
GRAND TOTAL	41,325

Acquisition Engineering Career Field

Component	Military	Civilian	Total Count
Army	0	9,063	9,063
DoN	224	21,019	21,243
Air Force	1,403	7,546	8,949
4th Estate	0	2,070	2,070
GRAND TOTAL	1,627	39,698	41,325

Data Sources:
Engineering (Non-Construction) – Defense Civilian Personnel Data System, 31 March 2016
Acquisition Engineering Career Field – USD(AT&L) Defense Acquisition Workforce Data Mart, 31 March 2016

Figure 2. DoD Engineering Workforce

It is clear that the DoD's software-sustainment organizations place a high priority on software human capital across the cycle of recruitment, retention, and training in what is a highly competitive market. Since the DoD organizations engaged in software work are distributed across the country, geographic constraints (i.e., locations in parts of the country without substantial commercial software industry) can be challenging. However, there are numerous examples of innovative DoD approaches to building relationships with colleges and universities to enhance recruiting and provide for continuing education.

Our engagements with software-sustainment stakeholders identified several challenges these organizations face in their ability to hire, develop, and retain the skills of this critical organic workforce. The most relevant of these challenges are highlighted below.

- Each software-sustainment center is challenged by the fact that there can be great diversity in knowledge and understanding about software engineering among new hires. Such new hires come from different engineering disciplines and educational programs, where there may be limited emphasis on software engineering. As a result, there are often significant gaps in knowledge and practical skills. These gaps must be filled to enable a new hire to be productive in an organization that has standards, practices, and a defined software engineering lifecycle development process. During our study, we noted and confirmed with software engineering managers that entry-level engineers do not appear to have been exposed to secure coding practices or vulnerable code detection as part of their education or previous on-the-job experience. Further, these entry-level engineers tended to know about vulnerabilities in code but were not aware of how these vulnerabilities could be exploited, the impact of that exploitation, or how to detect and fix these vulnerabilities.

  There is a larger issue at hand regarding the diversity of knowledge and understanding of new hires coming into the DoD software-sustainment workforce. Cyber-physical systems (including weapons systems) pose exceptional technical challenges to systems and software engineering practitioners. Many software engineering academic programs emerged out of computer science- and math-focused programs, requiring little in the way of classical engineering courses in physics, electronics, chemistry, and mechanics. Systems engineering academic programs grew out of the classical engineering programs, requiring little in the way of software engineering competencies. As a result, systems and software engineering graduates are well prepared to work on computational systems, but fewer graduates are well prepared to work on cyber-physical systems (e.g., weapon systems). Further complicating this challenge is that those trained in computer science and those trained in software engineering have different skill sets.

  The root cause of this issue is beyond the ability of the DoD's organic software engineering and sustainment organizations to solve. However, this issue highlights the significant challenge faced by software sustainment centers in filling the gap in terms of the practical competencies required for continuous engineering of software intensive systems. How to best and affordably fill this workforce competency gap is a DoD and Service enterprise challenge.

- The number of different software languages and versions of those languages used in the DoD's legacy systems is staggering and creates significant challenges in achieving and sustaining critical competencies. Within this diversity of programming languages, the use of the Ada language presents a critical challenge to the software-sustainment centers because Ada still represents a significant portion of the software code base in use within the DoD. This is problematic for the DoD across the workforce since Ada is no longer commonly taught or supported outside of legacy DoD applications. Thus, there is no college or university pipeline for training and education or ongoing development of tools. Further, the DoD is no longer the driver of the technology marketplace as it was decades ago, so it has limited ability to influence the selection of programming language within the supply chain.

The graph in Figure 3 highlights this point. This data is taken from a snapshot of DoD software data that has been collected from certain programs and analyzed by the SEI (2017b). By far, the C family of programming languages dominates in terms of the software in DoD systems. (The C family includes C, ANSI C, C++, C#, C/Assembly, and C# Net languages.) Ada represents another substantial subset and, therefore, will continue to be a challenge.

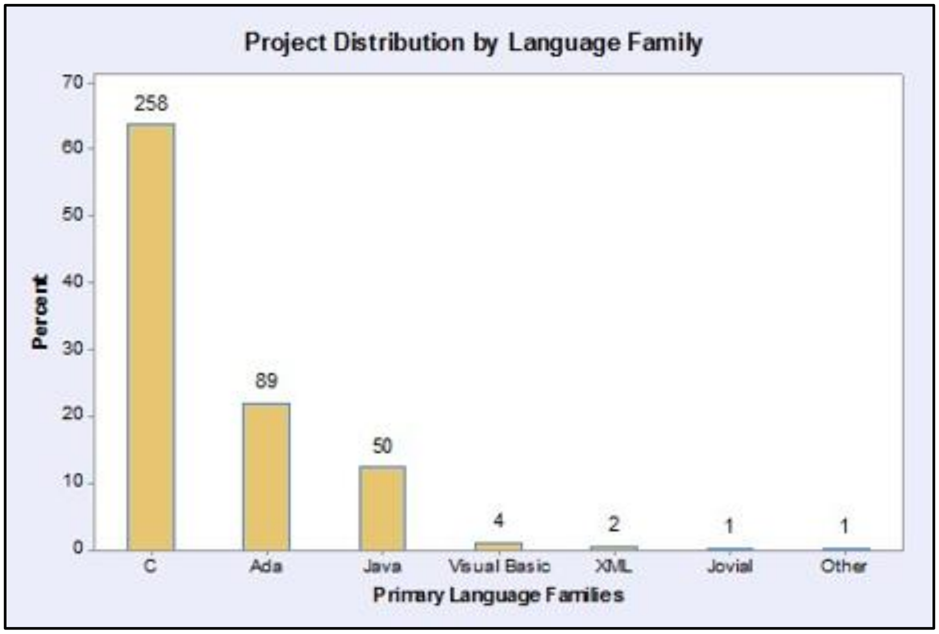


Figure 3. Distribution of Software Languages

- Managing the dynamics of an aging workforce in the DoD is not unique to the software-sustainment workforce. However, our view is that this fact-of-life issue for the sustainment community may be exacerbated by several factors. The DoD must continually make adjustments to accommodate a decline in the acquisition of new systems while the service life of legacy systems is extended in the force structure. This, in turn, creates a demand for the workforce to be continually refreshed in legacy system technical knowledge and skills. The software sustainment environment is inherently dynamic due to technology and mission demands, the evolving nature of program-by-program decisions regarding public-private partnerships, and cyber demands. Finally, there is lead time associated with acquiring and training software human capital, which means it is critical to plan for the organic software-sustainment workload as early as possible in the acquisition process.

  A challenge we heard repeatedly from all levels of organizational leadership and management was the slowness of the government hiring process. Analysis of the talent acquisition process was beyond the scope of this study, but those we interviewed cited many examples of this problem. In a competitive marketplace, interested recruits are unlikely to stay available for the weeks or months required for the DoD human resources process to complete. In our view, this behavior is out of sync with the tenets of the DoD's

Better Buying Power policies, which emphasize imperatives such as greater productivity and efficiency.

A Model for Analysis of Inherently Governmental Responsibilities and Needed Skills

To facilitate DoD programs thinking through their software workforce issues, we propose that the DoD first focus on identifying critical organic engineering competencies needed for specific programs, domains, and technologies. The DoD should use this list of competencies to identify gaps in the current organic workforce and ensure that a pipeline of talent is constantly being recruited. What we mean by *workforce pipeline* is a mechanism by which junior personnel are explicitly mentored by senior personnel to avoid the risks that skill gaps open up when personnel leave the project for any reason. An important part of this analysis is understanding what skills must be established and maintained in a government software engineering organization for it to perform inherently organic software engineering functions. A key element in this analysis is determining the software engineering functions that can be appropriately performed by a supplemental workforce.

This proposed approach is based on recognition that software sustainment is an *engineering activity*, a very different model from hardware sustainment. In many instances, software changes directly enable weapon system capabilities and/or overall system performance in support of maintaining national security. Therefore, there must be certain functions that are inherently governmental (i.e., required for the government to perform) to understand the technical baseline and then exercise technical authority to make appropriate engineering decisions. This implies the need for a model describing the engineering functions, capabilities, and competencies needed to perform sustainment and continually refresh the software technical baseline.

As an initial model, Figure 4 depicts one view of key weapons systems depot-level software sustainment functions, which the SEI study team developed. The software engineering functions in the center of Figure 4 represent the basic processes that create software. These activities are based on current software engineering standards, which were derived from ISO/IEC 2167a. The functions on the perimeter of Figure 4 represent activities related to the ecosystem elements and other functions the SEI team identified during Phase I and Phase II activities and site visits of DoD maintenance facilities.

Not all the functions depicted in this model are performed by every software sustainment organization. These functions can be decomposed and analyzed to identify associated technical task requirements. These requirements can then be analyzed to determine the competencies (skills, knowledge, and experience) necessary to execute those tasks. Following this logic leads to defining a baseline set functions, tasks, and competencies.

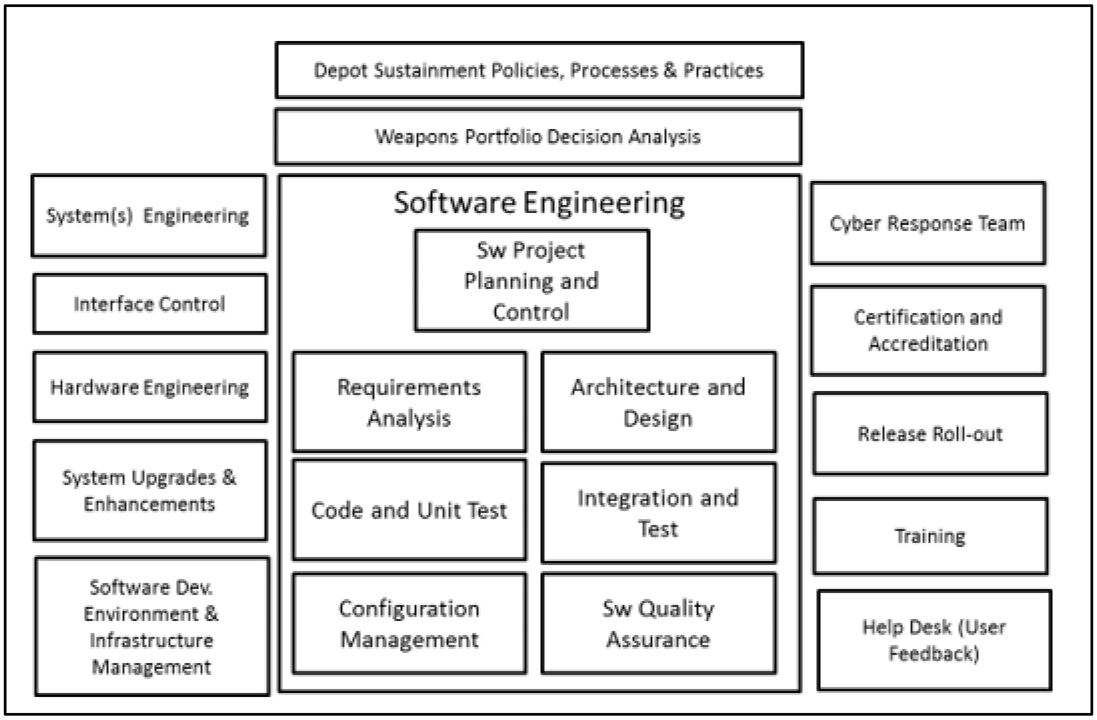


Figure 4. Weapons Systems Depot-Level Software Sustainment Functions

Applying this function-based model requires knowledge and understanding of the technical baseline of the specific system. Each function can be decomposed into a more detailed description of the work content involved, and specific and relevant work flows, processes, and practice. These functions and the details of each function may vary in scope and degree depending on specific functionality and system domains.

It is important to understand that a critical consideration in applying this model is identifying which software-sustainment functions are most critical from a national security perspective. Since software engineering functions involve changing the software baseline, a working knowledge and capability to perform these functions is probably high on the list of functions to staff organically.

The activities on the left side of Figure 4 represent system and technology influences on the operational software baseline. For weapon systems that experience a frequent rate of change driven by threats and technology imperatives, the DoD may elect to employ an organic workforce to ensure overall weapon system readiness. For systems that have stable interfaces and infrequent refresh cycles, the DoD may elect to employ a mix of organic and contracted staff based on best value considerations. In these situations, the government will likely want to establish a minimal set of competencies in these functions so it can maintain technical authority in making decisions and provide some level of organic software sustainment to mitigate changes in the software sustainment supply chain.

Another critical consideration in applying this model relates to addressing engineered-in-security, mitigating vulnerabilities via rapid deployment of security patches, and accommodation of the rapid pace of technological change. During our study, we observed that the DoD's software-sustainment community is acutely aware of the need to enhance software workforce competencies for software assurance.

For each function, careful attention must be paid to ensuring that the capabilities that are inherently governmental functions are maintained in the organic workforce and understanding which ones are candidates to be outsourced to a supplemental workforce. The inherently governmental functions then must be managed accordingly. The requisite education, skills, domain knowledge and system experience must be documented, and a deliberate process to hire, train, retain, and grow organic personnel must be put in place. This approach ensures that the government is always capable of (1) understanding the technical baseline for the system and (2) making appropriate, long-term decisions about engineering alternatives.

A key factor is applying this model is judging the appropriate DoD program official regarding the choice of using organic government software engineering capabilities, relying on defense industrial base capabilities, or selecting some mix of these capabilities to execute the sustainment mission. These choices are driven by a number of considerations, such as national security, affordability, and what the DoD believes is in its best interest in the long term.

Summary

A key takeaway message is that software sustainment is an engineering task. Almost any non-trivial change to the software requires analyzing the change and the current system, tracing the impact of the change on the existing requirements and design, and developing a new solution. For these reasons and others, policies and practices that are based on hardware sustainment, which can be treated as a discrete series of activities intended to restore form, fit, and function, do not apply well to software nor to understanding the requirements for the software workforce.

Executing the DoD software engineering sustainment mission demands a keen focus on defining and continually refreshing the DoD’s software-engineering competencies to drive workforce development and organization performance.

In this paper, we highlighted an initial functional model for thinking about software engineering competencies. We recognize that defining these competencies is a complex task that is a continuous activity that must be aligned with the nature and pace of technology change. Defining these competencies nests within the total workforce acquisition, development, and management strategy and plan.

References

Chao, P. (2006, October). *An assessment of the national security software industrial base.* Center for Strategic and International Studies. Retrieved from https://www.csis.org/analysis/assessment-national-security-software-industrial-base

Defense Science Board. (2000, November). *Report of the Defense Science Board Task Force on defense software.* Washington, DC: Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA385923

Level of Repair Analysis (LORA). (2017, March 15). In *Acquipedia*. Retrieved from Defense Acquisition University website: https://shortcut.dau.mil/acq/lora

Lucero, D. S. (2010, May/June). Influencing software competencies across the DoD acquisition workforce. *CrossTalk.* Retrieved from http://static1.1.sqspcdn.com/static/f/702523/9185115/1288380566797/201005-Lucero.pdf?token=eT0HIzSVufpRpLIqCts6pKvgnvk%3D

National Research Council. (2010a). *Critical code: Software producibility for defense.* The National Academies Press. Retrieved from https://doi.org/10.17226/12979

National Research Council. (2010b). *Achieving effective acquisition of information technology in the Department of Defense.* The National Academies Press. Retrieved from https://doi.org/10.17226/12823

Software Engineering Institute. (1998, June). *Review of DoD policy impacts on the Software Acquisition Management Education Program.*

Software Engineering Institute. (2017a, February). *DoD Software Sustainment Study Phase I: DoD's software sustainment ecosystem.*

Software Engineering Institute. (2017b, April). *Department of Defense software factbook.* Retrieved from https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=502651

Shull, F., & McLendon, M. (2017). *The policies and economics of software sustainment: DoD's software sustainment ecosystem*. Presented at the Acquisition Research Symposium, Monterey, CA.

Acknowledgments

During the course of this study, SEI researchers interacted with a range of senior leaders, key managers, and staff across the DoD's software-sustainment organizations. Universally, the Services and those we engaged were generally responsive to our data and information requests; they candidly shared their experiences and perspectives on this critical and complex issue. We are exceedingly grateful for the many thoughtful conversations and insights we received in the course of this work.

Disclaimer and Distribution Statement

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

Panel 11. Sustaining the Defense Industrial Base

Wednesday, May 9, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Nancy Spruill,** Director, Acquisition Resources & Analysis, OUSD (AT&L) ***Evaluating Consolidation and the Threat of Monopolies Within Industrial Sectors*** Andrew Hunter, Center for Strategic and International Studies Gregory Sanders, Center for Strategic and International Studies Zach Huitink, Syracuse University ***Trends in the DoD Industrial Base*** Nancy Moore, RAND Corporation ***Contractual Flow-Down Clauses: Deterrence to Non-Traditional Defense Contractors From Doing Business With DoD*** Mark F. Kaye, Institute for Defense Analyses Daniel L. Cuda, Institute for Defense Analyses Kevin Y. Wu, Institute for Defense Analyses

Nancy Spruill—is the Director for the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, Acquisition Resources and Analysis (ARA). Dr. Spruill integrates the diverse aspects of defense acquisition into a balanced and coherent program that supports the national strategy and makes the most effective use of resources provided. The Director, ARA also serves as the Executive Secretary to the Defense Acquisition Board; oversees the Defense Acquisition Executive System (DAES); manages AT&L's participation in the Planning, Programming, Budgeting, and Execution System (PPBE); and is responsible for the timely and accurate submission to Congress of Selected Acquisition Reports and Unit Cost Reports for Major Defense Acquisition Programs.

Evaluating Consolidation and the Threat of Monopolies Within Industrial Sectors

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a Senior Executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee. Hunter holds an MA degree in applied economics from Johns Hopkins University and a BA in social studies from Harvard University.

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds an MA in international studies from the University of Denver, and a BA in government and politics and BS in computer science from the University of Maryland.

Zach Huitink—is a Postdoctoral Fellow at Syracuse University and an externally affiliated researcher working with the Defense-Industrial Initiatives Group at CSIS. His research and analyses focus on national security, government contracting, and public-private partnerships. Huitink holds a PhD from the Department of Public Administration and International Affairs at Syracuse University's Maxwell School, an MPP with a concentration in international public policy from the Martin School at the University of Kentucky, and an honors BA in economics and business administration from Coe College.

Abstract

Economics scholars and policy-makers in recent years have rung alarm bells about the increasing threat of consolidation and concentration within industrial sectors. This paper examines the importance of industrial consolidation in two ways: first, as a direct relationship between concentration and performance outcomes; and second, as an indirect relationship, where concentration influences performance through reduced competition for government contract business. The paper finds that both increasing consolidation and decreasing competition are associated with an increase in contract cost ceiling breaches but also lower rates of termination. Subsequent stages of research will examine the interrelation of consolidation and competition.

Introduction

Project Motivation—Monopoly, Consolidation, and Implications for Performance

In recent years, economists, policy-makers, and other observers have expressed growing concerns over industrial concentration and the threat of monopolies in the U.S. economy.[1] Data on revenue concentration, for example, show that the largest firms in a number of U.S. industries are accruing an increasing percentage of their respective industry's market share. The 50-firm concentration ratio (CR_{50})—which measures the proportion of an industry's revenue accruing to its 50 largest firms—has grown by 10% or

[1] For a recent summary and synthesis of current views regarding industrial consolidation, monopoly, and their implications for policy, see Shapiro (n.d.).

more over the last 15 years (1997–2012, based on the latest available information) in industries ranging from transportation and warehousing to retail trade to finance and insurance (White House Council of Economic Advisers, 2016). For example, in the case of finance and insurance, the latest available data (as of 2012) shows the 50 largest firms account for nearly half (48.5%) of all revenue in the industry. This figure is even higher elsewhere. In utilities, for instance, the CR_{50} stands at 69.1% (White House Council of Economic Advisers, 2016).

These trends may reflect an actual decline in competition, but it is important to note they could also stem from superior economic performance among firms that may have driven their competitors out of the market. Moreover, production in many industries (like utilities) is subject to at least some degree of economies of scale—where per unit costs fall as production increases, and an industry's total output can be produced more efficiently by fewer, rather than more, firms—making those industries more concentrated to begin with. Finally, while the data reflect what is happening nationally, the actual effects of concentration tend to play out on a lower geographical scale (such that the issue is not strictly one of growing concentration nation-wide but one that affects regional and local markets in particular). Acknowledging these caveats (and their implications for whether increasing concentration warrants one or another type of public policy response), the increasingly concentrated nature of many industries in the U.S. remains a noteworthy economic development.

Concerns over industrial concentration and potential monopolies also extend to the U.S. defense industry. Maintaining a vibrant, dynamic defense industrial base with vendors that compete vigorously to win contracts and provide the government with products and services is critical to U.S. national security. Indeed, while historically the government has relied on mobilizing a mix of federally-funded arsenals and civilian contractors during wartime to meet its military needs, following WWII, these needs have been met principally by a permanent private defense establishment.

This research project seeks to evaluate the urgency of these concerns by examining the connection between industrial consolidation and contract outcomes. It examines the relationship in two ways: first, directly—through the influence of consolidation in the contract's sector on performance, and second—indirectly, through the effects of competition on contract performance. Future stages of this paper will combine both of these examinations into a single model.

Literature Review

As the primary buyer of the defense industry's goods and services, the U.S. government can play a significant role in shaping the industry's size, composition, and economic viability. As a result, the defense industrial organization has evolved (at least in part) in accordance with military spending. Since WWII, the defense budget has cycled between a series of peaks and troughs, generating significant expansions in industrial capacity followed by more modest declines. This pattern resulted in a particularly acute case of capacity overhang following the end of the Cold War, because during the war, contractors had invested heavily in plants, equipment, and other assets that were no longer needed following the war's end (and the subsequent drop in defense expenditures). To eliminate inefficiencies stemming from excess capacity, the Department of Defense (DoD) explicitly encouraged its contractors to merge, and offered to share in savings generated from consolidations. Merger activity in the defense industry increased dramatically. Between 1993 and 2000, the number of major prime contractors fell from 50 to six (Gansler, 2011). However, it is still an open question whether and to what extent these mergers actually generated savings—or even stemmed as much from the DoD's pro-consolidation policy and post–Cold War budget cuts as they did from economy-wide trends that also drove mergers in non-defense industries.[2]

Defense budgets reversed following 9/11, and grew at rapid double-digit rates for nearly a decade. However, spending reductions mandated by the Budget Control Act (BCA) of 2011 and the cuts to Overseas Contingency Operations (OCO) funding around that time—policies collectively referred to as "the drawdown"—have significantly impacted the defense industry. Across individual product and service platforms, a recent analysis showed declines in defense contract obligations from 16% for Ships and Submarines to as high as 56% for Land Vehicles (McCormick, Hunter, & Sanders, 2017). Declines in other portfolios varied, according to the analysis, from 19% for Aircraft, to 20% for Ordinance and Missiles, to 32% for Space Systems (McCormick et al., 2017). Obligations for products, services, and R&D activities not falling under one of these specific platform categories fell by 30%, 28%, and 19% respectively (McCormick et al., 2017). Within product, service, and R&D categories, the analysis showed shares of obligations going to small businesses tended to grow or remain steady, but tended to fall for the Big 5 (Lockheed Martin, Boeing, Raytheon, Northrop Grumman, and General Dynamics) and especially for large- and medium-size vendors (McCormick et al., 2017). Across categories and vendor sizes, the analysis found that the number of vendors receiving prime contracts from the Department of Defense (DoD) dropped in all by 17,000, or nearly 20% over the drawdown period (McCormick et al., 2017).

Whether these vendors fully exited the defense marketplace or remained (e.g., as subcontractors) cannot be definitively established. Nonetheless, existing evidence suggests the U.S. defense industry is in the process of another significant episode of transformation, and officials from both the previous and current administrations have signaled worries over the industry's health and competitiveness. As far back as 2011, Ash Carter, then Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) and later Secretary of Defense, stressed the importance of avoiding excessive consolidation among large prime contractors (Weisberger, 2015). His successor as USD(AT&L), Frank Kendall, took the same view, calling Lockheed Martin's proposed and subsequently executed acquisition of rotary-wing aircraft manufacturer Sikorsky "the most significant change to the defense industry since the general consolidation that followed the Cold War" (Weisberger, 2015). Kendall warned more generally that continued consolidation, particularly of large prime contractors, could diminish competition and the number of suppliers available to the military, erect barriers to entry, and hinder innovation that is key to sustaining U.S. technological superiority (Weisberger, 2015). Around the same time, the U.S. Department of Justice (DoJ) and the Federal Trade Commission (FTC) issued a joint statement saying "many sectors of the defense industry are already highly concentrated [and others] appear to be on a similar trajectory" and reiterated their commitment to take action against mergers that would dampen innovation and competitive forces (DoJ & FTC, n.d.). More recently, under Executive Order 13806 (2017), President Trump directed a sweeping review of the industrial base with the aim of determining if its broad composition, capacity, and resiliency can meet a growing set of security threats through having a robust base of capable

[2] For a review of competing explanations of post–Cold War U.S. defense industry consolidation, see Brady (2009).

suppliers. Questions of industrial concentration and monopoly power, as well as their implications for competition and performance, relate importantly to these issues.

Industrial Concentration—Definition and Measurement

Industrial concentration refers to the degree to which a smaller versus a larger number of firms account for production or other measures of market share (e.g., revenue) in some part of the economy.

Taking this idea as a point of departure, a large discourse in the literature has developed around alternative approaches to measuring concentration in practice.[3] One approach is to use concentration ratios, which add shares (whether of production, revenue, or some other activity) of a pre-determined number of firms in a particular market. Commonly used numbers include the top 4, 8, 20, or 50 firms in the market of interest. These ratios are relatively simple to calculate and, compared to other metrics—such as the Herfindahl–Hirschman Index (HHI)—do not impose as large a challenge with data collection because they do not require data on the shares of every firm in the relevant market place. By contrast, to calculate the standard HHI requires data on the shares of every firm in the relevant market place and entails squaring each individual share before adding them (so as to weight the index more strongly toward larger companies). The upsides of this approach include counting shares of every applicable firm and weighting firms with larger shares more heavily in the calculation. Whereas concentration ratios are expressed in percentage terms (with a 100% maximum), the HHI varies between a minimum of 0 and a maximum of 10,000 (where one firm accounts for 100% of the market and $100^2 = 10{,}000$). For purposes of evaluating mergers and their antitrust implications, the DoJ deems HH indices of 2,500 or higher to be significantly concentrated.[4]

As noted, both concentration ratios and the HHI continue to be used in practice, with the choice of one versus the other depending principally upon data availability and the objectives of the analysis. Common challenges that must be addressed in using either measure include identifying and collecting reliable data on market shares and, more fundamentally, defining the scope of the marketplace in which concentration will be analyzed. As noted above, concentration metrics are often calculated and presented on nation-wide basis, whereas evaluating the implications of concentration for competition, consumer welfare, and public policy often requires examining trends at a less aggregated level. Moreover, concentration measures can be sensitive to the specificity with which products are defined and categorized. With all else equal, defining a particular class of product more broadly—and thereby including more firms—will tend to reduce concentration levels, whereas a more precise definition will raise them.

Causes of Industrial Concentration and Monopolies

Variation in levels of industrial concentration—from very low to monopoly levels where one firm accounts for all of an industry's production, revenue, sales, or other economic activity—stems from several sources. Differences across industries or within a given industry over time may reflect an underlying decline in competition and attendant increases in market power for leading firms—a common interpretation of recent trends in the

---

[3] See Curry and George (1983) for a commonly-cited review of the literature.
[4] See DoJ (2015).

U.S.[5]—although one of at least four other forces may also be at play (and, depending on which, may suggest alternate explanations for changes in concentration levels).

First, higher industrial concentration may stem from *economies of scale,* a technological feature of production that leads per unit production costs to fall as output rises. The upshot of this dynamic is that an industry's aggregate output can be most efficiently produced by a smaller, rather than a larger, number of firms (Carleton & Perloff, 2015). Accordingly, in a case like this, the industry actually operates most efficiently and can charge lower prices for its output with less as opposed to more firms in operation. The number of firms may fall due to some firms exiting the marketplace, or through mergers and acquisitions. In extreme cases, economies of scale are so high as to make it most efficient for a single firm to produce all of an industry's output, a situation referred to as a *natural monopoly*. Unlike (as discussed below) situations where monopoly power derives from purposefully erected barriers to entry (e.g., government conferring operating privileges exclusively to a single company), natural monopolies arise due to the underlying technology for production of a good or service.[6] Commonly-cited natural monopolies include utilities, where entry of additional firms would entail highly inefficient (and arguably infeasible) recreation of distribution infrastructure like pipes or power lines that one firm has already incurred the costs to build (Kunneke, 1999).

Second, and similarly, production may be subject to *learning curves*, where (however high or low scale economies may be) per unit costs fall as firms discover more efficient ways to produce output. According to learning curve theory, through repeated production, firms accumulate knowledge and experience that can be used for purposes of process improvement, efficiency enhancements, and lower per-unit pricing, which may make them more competitive relative to their peers and lead them to capture higher market share (Brady & Greenfield, 2009). Manufacture of large capital assets like ships, planes, or construction equipment, which may initially entail high costs for design and early unit production but entail lower costs as production expands, are often suggested to benefit from the learning curve dynamic.

Third, firms may create barriers to entry or force competitors out through strategic behavior like predatory pricing, hostile takeovers, or alternative forms of vertical acquisition where an incumbent firm acquires lower-level suppliers (thus eliminating potential sources of productive inputs that new entrants need in order to operate). Incumbent firms may act alone to create entry barriers, or they might potentially collude with one another for this purpose. A commonly cited example of collusion to prevent competition involves incumbent firms dividing up customers in lieu of vying with each other to capture as much business as possible. The firms may divide up sales territories, for example, and work together to prevent competitors from entering. Such conduct has been suspected or documented to have happened in industries as diverse as health insurance and chemicals.[7]

Finally, in some instances governments purposefully erect structural barriers to entry that may limit competition that is otherwise likely to arise (e.g., in cases where scale economies do not operate at high levels and concentrate production in a few firms).

---

[5] See Shapiro (n.d.).
[6] For an early overview of natural monopoly, see Posner (1969).
[7] For further explanation and specific examples, see FTC (n.d.).

Governments may create entry barriers through extending protections for intellectual property and innovation (e.g., through patents), through establishing legal and regulatory requirements that must be fulfilled in order to do business in a particular area, or by granting only one or a few firms permission to do a form of business (thereby foreclosing competitors from entering the market). Sufficiently high entry barriers can create monopolies in cases where the underlying technology of production implies strong efficiency gains from having one or only a few producers. Taxis are an often-cited example of a monopoly that city governments have created through regulations, such as requiring the purchase of a medallion to drive a cab.

The monopsony of the market can also be a barrier. The defense industry sells its products principally to a single buyer: the U.S. government (from which decisions about policy, budgets, and procurement priorities can significantly impact defense industry structure). In addition, concentration in different sectors of the defense industry may stem at least partially from underlying scale economies, learning curve dynamics, and government-imposed regulations, which are often cited as a barrier to further entry by commercial firms. Scale economies and learning curves are fundamental to the production of large, complex assets such as fighter jets and ships, leading to high concentration in these sectors (U.S. aircraft carriers, for example, are built exclusively in one shipyard, operated by Newport News Shipbuilding). And, in both of these sectors (and all others from which government purchases products, services, and R&D support) rules and regulations that firms must adhere to for purposes of bidding on contracts and winning business may constitute a substantial barrier to further competition—particularly for nontraditional firms that could be significant sources of innovation. Experiments with alternative acquisition models and partnerships, such as the Defense Innovation Unit Experimental (DIU_x)—the DoD's Silicon Valley–based unit focused on identifying and acquiring cutting-edge commercial technology solutions for the U.S. military, are ongoing, but large-scale entry of commercial players into the defense marketplace (and attendant growth in competition) remains to be seen.

Concentration, Competition, and Performance

To the extent it stems from factors such as reductions in competition and barriers to entry (whatever their source), rather than economies of scale, learning effects, or other forces that reflect firms actively searching for ways to enhance efficiency, industrial concentration is concerning because it can reduce economic welfare and generate market power firms that may use to extract rents in the form of higher prices to consumers (Carleton & Perloff, 2015).

Empirically, there is a large and now decades-old body of evidence relating increasing concentration to elevated prices and profits for firms.[8] Whether these relationships reflect firms exercising market power to charge excessively high prices and make additional profits is less clear, however. Some research, for example, attributes the observed link between concentration and profits to efficiency gains stemming from learning and harnessing scale economies. These arguments suggest that efficiency-enhancing concentration generates reductions in both prices and costs, but greater reductions in the latter than the former (leading, on average, to higher observed profitability as price-cost

[8] Literature reviews date back as far as the 1970s, with one review (Weiss, 1974) cataloging the results of 40 preexisting studies.

differentials grow; Peltzman, 1977). This finding is supported by other research demonstrating that, after controlling for firm size, the relationship between concentration and profitability is less strong—suggesting profit growth comes from efficiencies brought about by increasing the scale of production, of which increased concentration is just a byproduct (Brozen, 1982). More recent research comes to the opposite conclusion, finding robust connections between growing concentration, profits from both ongoing business as well as from mergers and acquisitions, and higher stock prices. Rather than reflecting operational efficiency and declining costs, however, this analysis suggests higher profitability is a function of increased market power (Grullon, Larkin, & Michaely, 2017).

Compared to research on relationships between concentration, competition, and firm- performance outcomes like profitability, there has been less research conducted on the implications of concentration for other measures of performance. While, as noted, higher profitability from increasing concentration may reflect stronger operational efficiency, there are other possible sources that do not imply better performance. As a result, this still leaves open the problem of explicitly examining links between concentration and firm performance along nonfinancial dimensions.

Moreover, compared to research on the private sector, very little work has been done to examine the implications of industrial concentration for government, specifically in the context of procurement and contracting. Competition is deemed a fundamental source of value in public procurement and is argued to provide higher quality products at lower prices, along with ancillary benefits such as accountability, fraud prevention, and better stewardship of taxpayer resources (Manuel, 2011). In buying simple goods and services for which many suppliers already exist, the benefits of competition can be powerful. For more complex products—whether inputs into government's provision of public services (e.g., fighter jets for national defense) or public services delivered by nongovernmental actors (e.g., social services provided by a nonprofit organization)—markets may be thinner and competition less viable (Kettl, 1993). However, in these cases too, the focus has been on examining the relationships between the quality of products and services on the one hand and competition on the other. Moreover, this work has often been done in the context of one or a few different product types.

Research that independently (or through competition as a mediating channel) explores the link between program level outcomes and concentration, competition, and contractor performance appears to be mostly absent from the existing literature and would be considerable value-added to the literature. In particular, there's an absence of work that uses large amounts of data to look across numerous product and service categories. Example studies explicitly assessing the link between industrial concentration and performance outcomes in the U.S. defense arena appear to be very few. One example is an analysis finding a positive relationship between concentration and firm profitability in the aerospace industry (Davis, 2006). Another analysis, more closely related to the research presented in this paper, finds evidence that some defense industry mergers generated cost savings in Major Defense Acquisition Programs (MDAPs) but also found that mergers do not categorically generate program-level savings (Hoff, 2007). Unlike the present study, however, this analysis is focused on financial dimensions of performance at the program level. This study extends the literature by looking at both financial and nonfinancial dimensions of performance and considers outcomes at the contract, rather than the program, level.

Conceptual Framework and Hypotheses

This paper posits and tests a conceptual argument linking industrial concentration and contract performance in two ways: first, as a direct relationship between concentration and performance outcomes; and second, as an indirect relationship, where concentration influences performance through reducing competition for government contract business. Specified in this manner, the argument broadens the approach to observing the relationship between concentration and contract performance, accounting for multiple ways the two variables may be connected.

Industrial Concentration and Contract Performance—Direct Relationship

The most straightforward way that industrial concentration impacts different markers of contract performance is through a direct relationship between the two variables. That is, changes in the level of industrial concentration are associated with an observable variation in alternative performance benchmarks, including (as considered in this paper) terminations and breaches of cost ceilings.

While arguments about concentration and contract performance may suggest the two are negatively related, with higher concentration leading to poorer performance, these arguments usually imply the presence of a mediating variable. Competition, as discussed previously, is one such variable. Economies of scale is another, which is often cited when arguing that concentration and performance may instead be positively related. In this case, rather than decreasing competition (and the attendant accumulation of market power a vendor may wield over the government), increasing concentration leads to positive performance, as it reflects efficiency gains from one or more vendors consolidating to operate at a larger scale of production.

Arguments that do not imply or explicitly reference a mediating variable—but instead posit a direct concentration–performance link—are agnostic with respect to whether growing concentration levels foster better or worse performance. For hypothesis testing purposes, the study team therefore does not suggest the direct relationship between concentration and contract performance is positive or negative. Instead, we simply hypothesize that the former may have a direct influence on the latter:

$\underline{H_1}$: Industrial concentration leads to changes in contract performance.

Industrial Concentration and Contract Performance—Mediating Role of Competition

While concentration and contract performance may be directly related, one common argument is that higher concentration negatively impacts performance by hindering competition that would otherwise act to discipline incumbent vendors. With all else equal, greater competition gives the government greater control in their relationship with vendors, providing them with multiple options while forcing vendors to perform well as they are considered more replaceable.

Through reducing the number of vendors from which government can select for awarding a contract, the argument is that concentration effectively reduces competitive forces. In addition, this would reduce the incumbent vendor’s incentive to perform effectively, as the prospect of being replaced is now lower. The incumbent may therefore be less motivated to innovate, control costs, or otherwise ensure its product meets or exceeds the government’s requirements. Consequently, the risk of termination or a cost ceiling breach may be elevated.

This line of reasoning points to two hypotheses. First, the logic that concentration as influencing performance through a competition channel implies a relationship between concentration and competition per se. Put simply, as concentration increases, competition decreases. Second, it implies a link between competition and performance outcomes, where reduced competition makes poorer performance more likely. In other words,

> $\underline{H_2}$: Increasing (decreasing) industrial concentration leads to decreasing (increasing) competition.
>
> $\underline{H_3}$: Decreasing (increasing) competition makes poor contract performance more (less) likely.

Data and Methods

Data Sources and Structure

Data Sources

The study team's primary source of data for this study is the Federal Procurement Data System (FPDS), which tracks all prime federal contract transactions worth $3,500 or more, conducted by most U.S. government department and agencies.[9] CSIS has created its own copy of this database, using data downloaded from USAspending.gov and supplemented at times with the FPDS-NG ad hoc search webtool.[10] During the period of this study, USAspending.gov underwent a major update that CSIS is still incorporating into the study team's analysis.

Data Structure

The unit of analysis for the dataset is prime contracts and task orders. Each contract entry has a unique procurement identifier and each task order entry has a unique combination of a parent award identifier and procurement identifier. The dataset includes all completed DoD contracts and task orders initiated between fiscal years 2007 and 2015 that were completed by the end of fiscal year 2015.[11] For task orders, the dates of inclusion and completion are based on each specific task order, not the date of the larger parent. The data set contains over 8.8 million entries, of which 12.6% were removed due to missing data, primarily with reference to undefinitized contract awards. These removed entries accounted for about 13.5% of obligations in the original dataset. For computational efficiency purposes, the study team has limited the analysis to a random sample of 250,000 to 1,000,000 contracts and task orders from the filtered dataset.

The study team has created the contract dataset from FPDS, which expands and updates a dataset used in previous CSIS reports on fixed-price (Hunter et al., 2015) and crisis contracting (Sanders & Hunter, 2017). To create this dataset, the study team decided how to handle contradictory information within the same field and how to consolidate large numbers of categories in the raw data to the more manageable number used in the regression to mitigate contradictions and to emphasize information available at the time a

---

[9] Prominent exceptions include classified contracts, which excludes the entirety of the CIA and some DoD contracts, most prominently in the U.S. Air Force. Other parts of the government are not required to report, such as the Defense Commissary Agency or the U.S. Postal Service.

[10] The study team is exploring additional sources, such as economics statistics broken down by NAICS category report by the U.S. Census and Bureau of Labor Statistics. However, they have not been incorporated into the dataset at this stage of the project.

[11] Completion is measured by having surpassed the current completion date of the contract or task order by at least one year or by contract closeout or a partial or complete contract termination.

contract is awarded. As a general principle, the most weight is given to a contract or task orders' initial unmodified transaction. The primary addition to the datasets used in previous reports relates to the North American Industrial Classification System (NAICS). First, the study team calculated the top 6-digit NAICS codes for each contract in the dataset. Second, the study team added a measure for consolidation, calculated at the NAICS sectoral-level.[12]

In addition to the contract dataset, the measures of consolidation also relied on past and updated work by the study team to consolidate large vendors who may be represented by multiple DUNSnumbers, the primary unique identifier for vendors within FPDS. The study team uses an obligation-weighted approach to choose identifiers for manual classification that have received more than $1 billion in obligations from 2000 to 2017 or $250 million in any year, in constant 2017 dollars. Those identifiers which the study team has not manually classified are instead handled via parent codes provided by the database. One disadvantage to this approach is that merger and acquisition activity is sometimes backdated to years before the merger occurred. However, the value weighted approach applied by the study team is appropriate for the consolidation measures described in the literature review because the largest firms in a sector are disproportionately important to calculating the HHI.

Measures of Dependent and Independent Variables

This section introduces the variables used in our regression model. For consistency and ease of data replication, the shortened name of the variable is included in parentheses after the full name. This shorthand name is also used in the definition of the equation and the results.[13]

Dependent Variables

Partial or Complete Terminations (b_Term) measures whether contracts and task orders experience a partial or complete termination, which yields a value of 1, while contracts with no terminations are given the value 0 for this variable. FPDS does not differentiate between complete and partial terminations, so this can include both a cancelled program and a contract that was completed after being initially protested and reassigned. 1.2% of contracts and task orders have experienced at least one partial or complete termination, and those records account for about 5.6% of obligations in the dataset.

Ceiling Breaches (b_CRai) tracks whether the contract had to be changed in a means that risked significant cost increases. To measure this, the study team observes transactions that are contract change orders and considers a ceiling breach to have occurred (assigning a value of 1) if any of these modifications also increased the contract or task order's cost ceiling, and assigning 0 otherwise. While only 1.2% of contracts and task orders have experienced a ceiling breach, the total obligations of those entries account for

[12] CSIS has made this dataset publicly available through our github repository (https://github.com/CSISdefense/Vendor/) to other researchers to be used with attribution.
[13] Some of the variables were transformed from categorical variables to the mathematical formats used in the dataset. For example, Term has a value of "Terminated" or "Not Terminated" while b_Term has a value of 0 or 1. Different prefixes are used depending on data type: "b_" refers to binary variables, "n_" refers to numerical variables, "l_" refers to variables that have undergone a logarithmic transformation, and "c_" labels to variables that were center (and thus "cl_" is a centered logarithmically transformed variable).

over 21% of obligations in the dataset. In addition, a slim fraction of terminations overlaps with ceiling breaches, despite both accounting for a similar percentage of contracts and task orders.

Study Independent Variables

Study Variables

Effective Competition (n_EffComp) is a numerical variable with three values:

- 1 for contracts competed with multiple offers (54% of contracts and task orders).
- 0.5 for contracts competed receiving only 1 offer (13% of contracts and task orders).
- 0 for non-competed contracts (33% of contracts and task orders).

The term "effective competition" is used by the DoD when monitoring their own competition rates.[14] The study team draws on multiple variables in FPDS to make this determination, with some contracts and task orders relying on the extent of the competitive field and others relying on the fair opportunity field. The study team considered other variations on the measure for competition, including the possibility of increasing gradations for competition with 2+ offers, before settling on effective competition as the best measure. However, effective competition has limitations, one being that the number of offers for competitive contracts and task orders is not always reported. Effective competition information is missing for 1.8% of contracts and task orders, and less than 0.8% of obligated dollars in the dataset.

Herfindahl-Hirschman Index (c_HHI_lag1) is a measure of consolidation in the defense industrial base. It is broken down into sectors as defined by six-digit NAICS codes. As described in the literature review, the HHI is calculated by squaring the market share of each participant in a sector. For the purposes of this study, market share refers to the percentage of prime obligations within a given fiscal year, which has the notable drawback of not capturing subcontracting activity. In the dataset this measure is lagged by one year, so for a contract signed in 2009, the consolidation measure of industry in 2008 is used. The variable is centered,[15] by subtracting its mean (2,056) and dividing by its standard deviation (1,867). The mean of this variable is roughly in the center of the DoJ's moderately consolidated category, which ranges from an HHI of 1500 to 2500. Data is missing for less than 0.1% of records and obligated dollars.

---

[14] See, for example, DoD (2015).

[15] Centering a variable is a way of making sure the different variables in a regression model are operating on the same scale, which makes it easier to compare coefficients across different variables. Mathematically to center x means c_x = (x –average of x) / (standard deviation of x).

Other Independent Variables

Initial Contract Scope

Initial Cost Ceiling (cl_Ceil) is the natural log of the initial contract cost ceiling as reported by the base and all options field, in then-year dollars.[16] The variable is centered, by subtracting its mean (9.12) and dividing by its standard deviation (2.26). Values of -1; 0; 1; and 2 correspond to $952; $9,121; $87,359; and $836,709 dollars respectively. Data is missing for less than 0.24% of contracts and transactions, which accounts for just over 0.3% of obligated dollars in the dataset.

Initial Duration (cl_Days) is the natural log of the initial maximum duration of the contract in days. The maximum duration is determined by comparing the contract effective date to the current completion date. The variable is centered, by subtracting its mean (3.05) and dividing by its standard deviation (1.92). Values of -1, 0, 1, and 2 correspond to 3.1 days, 21 days, 143 days, and 974 days respectively. Data is missing for just under 1% of contracts and transactions, which represents a miniscule percent of dataset obligations.

Contract Vehicle

Contracts and task orders come in a variety of types, some of which are simple purchase orders, others are complex but single use contract awards, and yet others are task orders that are a specific instance of an overarching indirect delivery vehicle. These types are explained below and help define the nature of the contractor/customer relationship.[17] The dataset uses dummy variables for four different types of indirect delivery vehicles:

- **SIDC** is 1 if the vehicle is a single-award indefinite delivery contract and 0 otherwise. These contracts may be initially awarded via competition, but afterwards are only used for task orders to a single vendor. They constitute over 58% of all contracts and task orders.
- **MIDC** is 1 if the vehicle is a multiple-award indefinite delivery contract and 0 otherwise. These vehicles have a pool of potential vendors that can receive task orders and make up 3.6% of contracts and task orders.
- **FSSGWAC** is 1 if the vehicle is a Federal Supply Schedule or Government-Wide Acquisition Contract and 0 otherwise. These two consistently multiple-award indirect delivery vehicles constitute 5.3% of task orders and contracts.
- **BPABOA** is 1 if the vehicle is a Blank Purchase Agreement or Basic Ordering Agreement and 0 otherwise. These indirect vehicles can be either single-award or multi-award, but taken together, only constitute 1.8% of task orders and contracts.
- The remaining types, definitive contracts and purchase orders, are intentionally left out.

[16] Constant dollars are not to allow for comparability between the contract ceiling and contract's actual expenditures in multiyear contracts. The base and all options ceiling of the contract is in nominal dollars but does not break out the cost ceiling for each individual year of a contract's life. As a result, the ceiling in constant dollars could be approximated, for example, by assuming that the ceiling will be split evenly over the life of a contract but cannot be calculated with any certainty.

[17] For more detail on contract vehicle types, see the glossary at USAspending.gov.

The remaining 31% of contracts and task orders are contract awards and purchase orders with no parent contract. This is the baseline for the regression model, true when all four dummy variables are zero. Vehicle classifications are missing for less than 0.1% of contracts and task orders and for a similarly small percentage of dataset obligations.

Contract Pricing

Fixed-Price (n_Fixed) is a numeric variable based on contract pricing. It has a value of 0 for cost-based (3.5% of records), 0.5 for "combination or other" (0.1% of records), and 1 for any fixed price (97% of records). While the overwhelming percentage of contracts and task orders are fixed price, nearly 29% of obligations go to cost-based contracts. Slightly more than 0.1% of contracts and task orders are unlabeled, along with a miniscule proportion of obligations. The study team is experimenting with including fee type as well but has not been able to replicate results by other researchers on the benefits of incentive fee contracts. This may be explained by the rarity of that fee type and the range of potential confounds. The study team intends to return to this question in later stages of the research.

Undefinitized Contract Action (b_UCA) is a binary variable with a value of 1 for contracts and task orders that begin as letter contracts or undefinitized contract awards (UCA) and a value of 0 otherwise. They account for a miniscule proportion (less than 0.01%) of contracts and task orders and only 3.4% of obligations, but do significantly correlate with a greater risk of terminations and ceiling breaches. Unfortunately, due to a reporting error in recent years on the now retired version of USAspending.gov, UCA classification is missing for nearly 10% of records and over 12% of obligations in the dataset. Nonetheless, the predictive power of this variable is sufficient, and, therefore, it is still included in the study.

Contract Location

Any International (b_Intl) is a binary variable with a value of 1 for contracts and task orders with any transactions performed internationally and a value of 0 otherwise. Nearly 10.5% of contracts and task orders had an international component as well as nearly 15% of obligations. Only a miniscule portion of records were unlabeled.

Contract Industrial Sector

NAICS represents the top North American Industrial Classification Code of each contract and is measured by obligated amount. This paper uses a multilevel model that allows for setting a different intercept for each industrial sector, discussed in greater detail in the next section. Due to computational limitations, the level of detail varies between models and is shown by the number at the end of the code (e.g., NAICS6 is the full six-digit NAICS code while NAICS2 is the minimal two-digit version). Less than 0.1% of contracts and task orders have no NAICS labeling whatsoever.

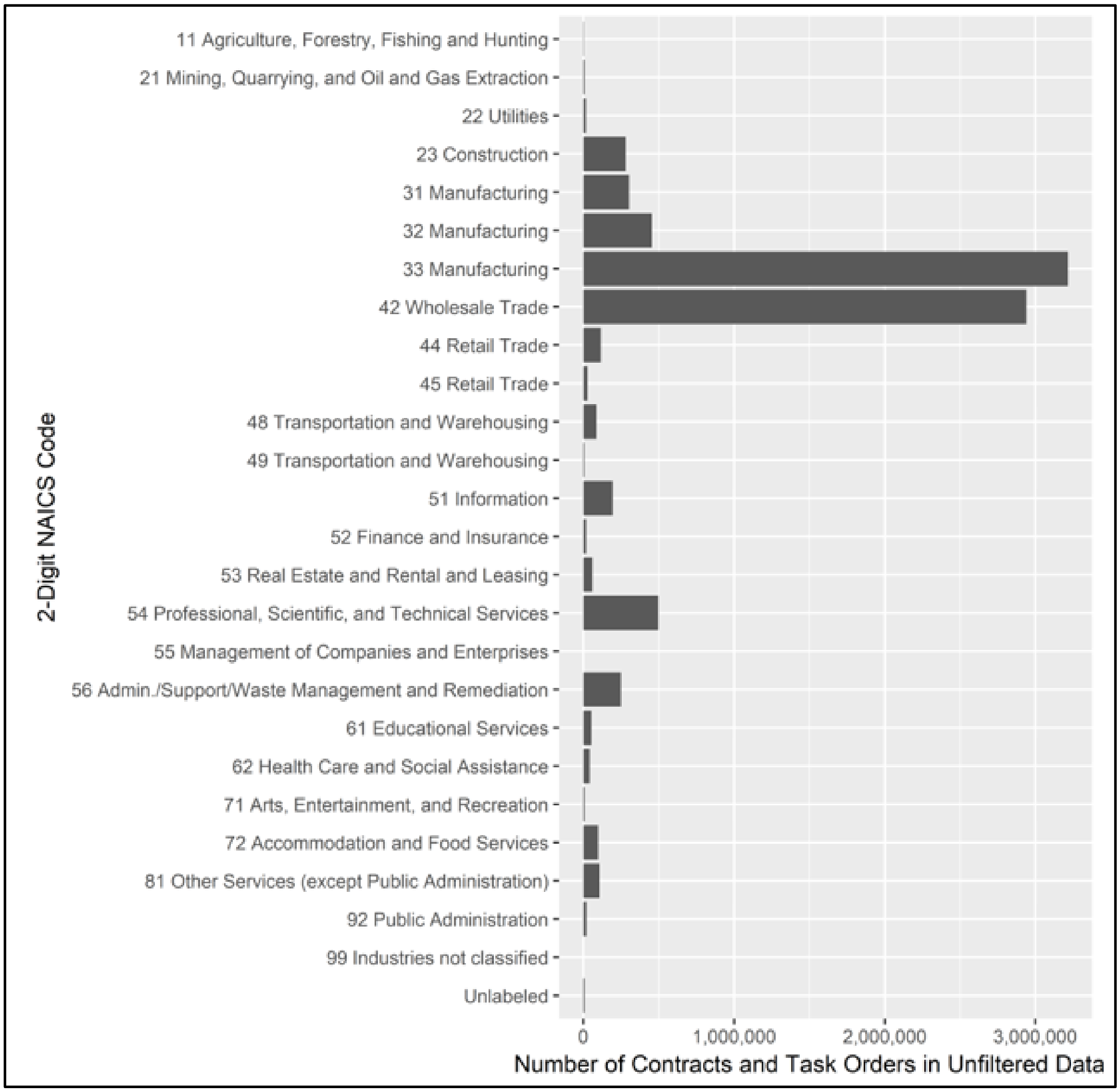


Figure 1. Distribution of Contract Obligations by NAICS 2-Digit Code

As shown in Figure 1, the distribution of DoD contract obligations is focused in a subset of the 24 NAICS 2-digit codes. Manufacturing (31–33) particularly stands out, as that category (Manufacturing), like Transportation and Warehousing (48–49), as well as Retail Trade (44–45), spills over into multiple 2-digit codes. In dollar terms, Professional, Scientific, and Technical Services (54) and Construction (23) are the second and third most prevalent industrial sectors; however, they are less significant in terms of the number of contracts and task orders because those sectors have higher value contracts. At the other end of the scale, Wholesale Trade (42) has lower obligations contracts, with less dollars obligated in that sector than either Construction or Professional, Scientific, and Technical Services.

Empirical Approach

At this stage of the project, the study team has four working models evaluating all combinations of the study and mediating variables: competition and consolidation respectively, with the two contract outcome variables, terminations and ceiling breaches. These initial models allow the study team to study H_1 and H_3, but leave H_2, the connection between consolidation and competition, to a later stage of the research.

Choice of Econometric Model

The study team uses a maximum likelihood logit analysis to analyze both termination and ceiling breaches. Logit is suited to dependent variables which can be true or false, 1 or 0, but not values outside of that range. This approach does not allow for evaluation of the size of a ceiling breach or variations of partial or complete terminations. However, less than 5% of contracts or task orders ever experience ceiling breaches or termination, therefore, the study team is only focusing on when these events occur and not differences between these cases.

In addition, when examining competition, the study team employs multilevel modeling techniques to capture the differences in expected outcomes between industrial sectors as categorized by NAICS codes. Each contract is assigned to a 2-digit NAICS sector based on the NAICS code that received the most overall obligations over the contract's lifespan. The equations below use a varying intercept model, which is to say that each of the 24 2-digit NAICS codes has a constant term added to the equation based on the termination or ceiling breach rate within that sector. Multilevel modeling techniques are a means to balance between two extremes when considering how to combine data from different groups. The first technique is complete pooling, which means there would be no varying intercept and no differentiation based on a contract's NAICS sector. The second technique is no pooling, which means there is a separate model for each NAICS sector. Multilevel modeling uses "soft constraints," which are covered in more detail in the next section.

Presentation of Estimating Equation

For competition as a mediating variable when estimating the probability of termination, the study team used the following model (subscript i refers to the individual contract or task order, while subscript j refers to the NAICS sector):

Consolidation Equations

$$
\begin{aligned}
Probability\ of\ Termination\ (y_i = 1)\\
&= Logit^{-1}\big(\alpha + \alpha_{k[i]}^{NAICS} + \beta_1 c\_HHI\_lag1_{\ i} + \beta_2 cl\_Ceil + \beta_3 cl\_Days_i + \beta_4 SIDV_i \\
&+ \beta_5 MIDV_i + \beta_6 FSS\text{-}GWAC_i + \beta_7 BPA\text{-}BOA_i + \beta_8 n\_Fixed_i + \beta_9 b\_UCA_i \\
&+ \beta_{10} c\_HHI\_lag1_i \cdot SIDV_i + \beta_{12}\, c\_HHI\_lag1_i \cdot MIDV_i \\
&+ \beta_{13} c\_HHI\_lag1 \cdot FSS\text{-}GWAC_i + \beta_{14}\, c\_HHI\_lag1_i \cdot BPA\text{-}BOA_i \\
&+ \beta_{15} c\_HHI\_lag1_i \cdot b\_UCA_i + \epsilon_i\big), for\ i = 1\ to\ 1{,}000{,}000
\end{aligned}
$$

$$\alpha_j^{NAICS2} \sim N(\mu_{\alpha,}, \sigma_a^2), for\ j = 1\ to\ 24$$

The second half of the equation merits additional explanation. α_j^{NAICS2} refers to the intercept, which in this and the subsequent equations will vary for each of the 24 2-digit NAICS codes. Gelman and Hill (2007) explain the concept in their introductory textbook:

> In the multilevel model, a "soft constraint" is applied to the [α_j^{NAICS2}]'s: they are assigned a probability distribution [see above], with their mean $\mu_{\alpha,}$ and standard deviation σ_a^2 estimated from the data. The distribution has the effect of pulling the estimates of [α_j^{NAICS2}] toward the mean level $\mu_{\alpha,}$ but not all the way. (p. 257)

$$
\begin{aligned}
&Probability\ of\ Ceiling\ Breach\ (y_i = 1\ ) \\
&\quad = Logit^{-1}\big(\alpha + \alpha_{k[i]}^{NAICS} + \beta_1 \text{c\_HHI\_lag1}_i + \beta_2 cl\_Ceil + \beta_3 cl\_Days_i + \beta_4 SIDV_i \\
&\quad + \beta_5 MIDV_i + \beta_6 FSS\text{-}GWAC_i + \beta_7 BPA\text{-}BOA_i + \beta_8 n\_Fixed_i + \beta_9 b\_UCA_i \\
&\quad + \beta_{10} c\_HHI\_lag1_i \cdot cl\_Ceil_i + \beta_{11}\, b\_Intl + \beta_{10} cl\_Ceil_i \cdot SIDV_i \\
&\quad + \beta_{12}\, cl\_Ceil_i \cdot MIDV_i + \beta_{13} cl\_Ceil_i \cdot FSS\text{-}GWAC_i + \beta_{14}\, cl\_Ceil_i \cdot BPA\text{-}BOA_i \\
&\quad + \beta_{15} cl\_Ceil_i \cdot b\_UCA_i + \beta_{16} c\_HHI\_lag1_i \cdot b\_UCA_i + \epsilon_i\big), \\
&\quad for\ i = 1\ to\ 1{,}000{,}000
\end{aligned}
$$

$$\alpha_j^{NAICS2} \sim N(\mu_{\alpha,}, \sigma_a^2), for\ j = 1\ to\ 24$$

Competition Equations

$$
\begin{aligned}
&Probability\ of\ Termination\ (y_i = 1\ ) \\
&\quad = Logit^{-1}\big(\alpha_0 + \alpha_{j[i]}^{NAICS2} + \beta_1 b\_Comp\ _i + \beta_2 cl\_Ceil\ _i + \beta_3 cl\_Days_i + \beta_4 SIDV_i \\
&\quad + \beta_5 MIDV_i + \beta_6 FSS\text{-}GWAC_i + \beta_7 BPABOA_i + \beta_8 n\_Fixed_i + \beta_9 b\_UCA_i \\
&\quad + \beta_{10} SIDV_i \cdot Comp_i + \beta_{11}\, MIDV_i \cdot Comp_i + \beta_{12} FSS\text{-}GWAC_i \cdot Comp_i \\
&\quad + \beta_{13}\, BPABOA_i \cdot n\_Comp_i + \beta_{14} b\_UCA_i \cdot n\_Comp_i + \epsilon_i\big), \\
&\quad for\ i = 1\ to\ 1{,}000{,}000
\end{aligned}
$$

$$\alpha_j^{NAICS2} \sim N(\mu_{\alpha,}, \sigma_a^2), \qquad for\ j = 1\ to\ 24$$

$$
\begin{aligned}
&Probability\ of\ Ceiling\ Breach\ (y_i = 1\ ) \\
&\quad = Logit^{-1}\big(\alpha + \alpha_{k[i]}^{NAICS} + \beta_1 n\_Comp\ _i + \beta_2 cl\_Ceil + \beta_3 cl\_Days_i + \beta_4 SIDV_i \\
&\quad + \beta_5 MIDV_i + \beta_6 FSS\text{-}GWAC_i + \beta_7 BPA\text{-}BOA_i + \beta_8 n\_Fixed_i + \beta_9 b\_UCA_i \\
&\quad + \beta_{10} n\_Comp \cdot cl\_Ceil + \beta_{11} b\_intl_i + \beta_{12} cl\_Ceil_i \cdot SIDV_i + \beta_{13}\, cl\_Ceil_i \cdot MIDV_i \\
&\quad + \beta_{14} cl\_Ceil_i \cdot FSS\text{-}GWAC_i + \beta_{15}\, cl\_Ceil_i \cdot BPA\text{-}BOA_i + \beta_{16} cl\_Ceil_i \cdot b\_UCA_i \\
&\quad + \beta_{17} n\_Comp_i \cdot b\_UCA_i + \epsilon_i\big), for\ i = 1\ to\ 1{,}000{,}000
\end{aligned}
$$

$$\alpha_j^{NAICS2} \sim N(\mu_{\alpha,}, \sigma_a^2), for\ j = 1\ to\ 24$$

Results and Discussion

Consolidation and Performance—Direct Relationship

In keeping with H_1, consolidation significantly correlated with both outcome measures, which supports the hypothesis on the importance of industrial consolidation. Surprisingly, as the c_HHI_lag1 row in Table 1 shows, the relationships of consolidation to the two dependent variables are opposite. Supporting perceptions of the risk of industrial consolidation, more consolidation is associated with a greater prevalence of ceiling breaches. In addition, the increased likelihood of cost escalation also undercuts the explanation that the lower associated rate of terminations may simply be the result of consolidated sectors having superior economics of scale or efficiencies.

Table 1. Logit Model Results for Consolidation

	Complete and Partial Termination	Ceiling Breach
(Intercept)	-5.50 (0.16)*	-4.75 (0.21)*
c_HHI_lag1	-0.15 (0.02)*	0.28 (0.01)*
cl_Ceil	-0.02 (0.01)	0.64 (0.02)*
cl_Days	0.67 (0.02)*	0.19 (0.01)*
SIDV	-1.04 (0.03)*	-0.07 (0.03)*
MIDV	-0.22 (0.05)*	0.37 (0.05)*
FSSGWAC	-0.28 (0.05)*	0.16 (0.06)*
BPABOA	-0.45 (0.08)*	-0.01 (0.08)
n_Fixed	1.02 (0.09)*	0.30 (0.04)*
b_UCA	1.64 (0.07)*	2.01 (0.07)*
c_HHI_lag1:SIDV	-0.50 (0.04)*	
c_HHI_lag1:MIDV	0.18 (0.05)*	
c_HHI_lag1:FSSGWAC	0.21 (0.05)*	
c_HHI_lag1:BPABOA	-0.02 (0.11)	
c_HHI_lag1:b_UCA	0.37 (0.09)*	0.37 (0.07)*
b_Intl		-0.27 (0.03)*
c_HHI_lag1:cl_Ceil		-0.17 (0.01)*
cl_Ceil:SIDV		-0.14 (0.02)*
cl_Ceil:MIDV		-0.24 (0.03)*
cl_Ceil:FSSGWAC		0.04 (0.04)
cl_Ceil:BPABOA		-0.32 (0.08)*
cl_Ceil:b_UCA		-0.39 (0.05)*
AIC	112213.41	105872.70
BIC	112402.46	106097.20
Log Likelihood	-56090.70	-52917.35
Num. obs.	1000000	1000000
Num. groups: NAICS2	24	24
Var: NAICS2 (Intercept)	0.32	0.98

*$p < 0.05$

Statistical models

The interactions may merit further exploration in the future. In consolidated sectors, use of single-award IDCs are significantly correlated with a lower probability of terminations which may reflect an institutionalized partnership between government and industry. On the other hand, UCA contracts in a consolidated sector appear to magnify the already significant correlation with both terminations and ceiling breaches.

While the relationship is statistically significant, the coefficient for the HHI is not notably impressive. The range of industry that the DoJ considers to be moderately consolidated ranges from 1500 to 2500 on the HHI index or from 0.7 to 1.24 on the centered version of the variable used for this study. While there is significant variation between sectors, few individual sectors are likely to change that rate of consolidation so much that they move by 1 point on this centered scale which corresponds to 1867 points on the HHI. The correlation with lower probability of ceiling breaches is also strongest for smaller contracts.

Role of Competition as a Mediating Variable

As shown in Table 2, competition correlates with terminations and ceiling breaches in the same direction as consolidation.

Table 2. Logit Model Results for Competition

	Complete or Partial Termination	Ceiling Breach
(Intercept)	-6.16 (0.10)*	-5.38 (0.05)*
n_Comp	0.58 (0.03)*	-0.25 (0.03)*
cl_Ceil	-0.04 (0.01)*	0.87 (0.02)*
cl_Days	0.83 (0.01)*	0.47 (0.01)*
SIDV	-0.58 (0.04)*	0.19 (0.03)*
MIDV	0.01 (0.07)	1.11 (0.05)*
FSSGWAC	-0.57 (0.07)*	0.24 (0.05)*
BPABOA	0.00 (0.09)	0.45 (0.08)*
n_Fixed	1.42 (0.09)*	0.38 (0.04)*
b_UCA	1.81 (0.06)*	1.94 (0.07)*
n_Comp:SIDV	-0.74 (0.05)*	
n_Comp:MIDV	-0.37 (0.09)*	
n_Comp:FSSGWAC	0.20 (0.09)*	
n_Comp:BPABOA	-0.74 (0.15)*	
n_Comp:b_UCA	-1.96 (0.27)*	-0.35 (0.14)*
b_Intl		0.13 (0.03)*
n_Comp:cl_Ceil		0.24 (0.02)*
cl_Ceil:SIDV		-0.33 (0.02)*
cl_Ceil:MIDV		-0.53 (0.03)*
cl_Ceil:FSSGWAC		-0.18 (0.04)*
cl_Ceil:BPABOA		-0.64 (0.08)*
cl_Ceil:b_UCA		-0.64 (0.06)*
AIC	115293.63	115669.09
BIC	115482.68	115893.58
Log Likelihood	-57630.81	-57815.54
Num. obs.	1000000	1000000
Num. groups: NAICS2	24	24
Var: NAICS2 (Intercept)	0.00	0.00

*$p < 0.05$

Statistical models

However, unlike consolidation, the literature suggests that competition should have a positive effect of performance. H_3 is only borne out in part. Competition is associated with a lower probability of ceiling breaches, in keeping with the hypothesis, but it is also correlated with a higher probability of partial or complete terminations. Notably, the coefficient for terminations is more than twice that of the coefficient for ceiling breaches.

The correlation for more competition and a greater risk of termination, as with consolidation, has multiple straightforward explanations. Alternately, in the absence of any competition, the government may be locked-in to a given vendor. This explanation would assume that the cost and effort of termination is more appealing if another vendor waits in the wings. Another possible implication is that competition sometimes allows technically unqualified vendors or price-to-win bids to emerge victorious even if they are ultimately unable to deliver on the terms agreed. Finally bid-protests are one source of partial or complete terminations that are exclusive to competed contracts.

However, the interactions of competition with terminations complicate the story. When it comes to contract vehicles, single-award IDCs, blanket purchase agreements, basic ordering agreements, and, to a lesser extent, multiple-award IDCs are all less likely to be terminated when competed. Even more striking, the coefficient of the interaction of competition and UCAs is large enough to cancel out the greater probability of termination associated with that contracting method.

The support for H_3 comes from ceiling breaches, where competition is correlated with a lower probability of change orders raising the cost ceiling. One caveat to this finding, as with consolidation, is that the correlation of competition with a lower probability of ceiling breaches is strongest for smaller contracts and fades away as ceilings grow larger.

Other Noteworthy Results

Contract or task order vehicles and pricing wield a significant influence over contract outcomes. Fixed-price contracts were somewhat correlated with ceiling breaches but strongly correlated with terminations. The former result merits closer scrutiny as it runs against past findings by CSIS and outside research. Likewise, at this stage the study team was unable to replicate research on that and found that incentive fees are linked with lower costs, although ceiling breaches only capture the cost growth part of that equation. Finally, as covered under both consolidation and competition, UCAs have significant negative correlations with both terminations and ceiling breaches, justifying their classification as a high risk contract type.

Concluding Thoughts and Next Steps

The results at this stage of the research do support the idea that even in the sometime monopsony of DoD acquisition, competition and lower rates of consolidation do correlate with a lower risk of cost escalation. At the same time, the findings regarding terminations emphasize the complex interactions of acquisition policy decisions and the risk of unexpected results. On that same note, the significant explanatory power of contract vehicles and their varying situational relevance suggest that the choice of vehicles should perhaps be given additional attention as a factor influencing contract outcomes. As ever, these findings reinforce the judgment and human capital needed for successful acquisition policy and the absence of one-size fits all solutions, even for foundational approaches such as competition.

The next step for the study team will be to examine the direct relationship between consolidation and competition as well as to create models of contract performance that include both consolidation and competition. In addition, the study team will iterate the existing models by considering refinements of existing inputs as well as new inputs such as sector-level economic data for the defense-industrial base and for the U.S. economy as a whole.

References

Brady, R. R., & Greenfield, V. A. (2009). Competing explanations of U.S. defense industry consolidation in the 1990s and their policy implications. *Contemporary Economic Policy, 28*(2).

Brozen, Y. (1982). *Concentration, mergers, and public policy*. New York, NY: Macmillan.

Carleton, D. W., & Perloff, J. M. (2015). *Modern industrial organization*. New York, NY: Pearson.

Curry, L. B., & George, K. D. (1983). Industrial concentration: A survey. *Journal of Industrial Economics, 31*(3).

Davis, J. B. (2006). *The impact of the defense industry consolidation on the aerospace industry.* Washington, DC: Industrial College of the Armed Forces.

Department of Defense (DoD). (2015). *Department of Defense competition report for FY 2014*. Retrieved from https://www.acq.osd.mil/dpap/cpic/cp/docs/DoD_FY_2014_Competition_Report.pdf

Department of Justice (DoJ). (2015, July 29). Herfindahl-Hirschman index. Retrieved from https://www.justice.gov/atr/herfindahl-hirschman-index

Department of Justice (DoJ), & Federal Trade Commission (FTC). (n.d.). *Joint statement of the Department of Justice and the Federal Trade Commission on preserving competition in the defense industry*. Retrieved from https://www.ftc.gov/system/files/documents/public_statements/944493/160412doj-ftc-defense-statement.pdf

Exec. Order No. 13806 (2017). Retrieved from

https://www.whitehouse.gov/presidential-actions/presidential-executive-order-assessing-strengthening-manufacturing-defense-industrial-base-supply-chain-resiliency-united-states/

FTC. (n.d.). Market division or customer allocation. Retrieved from https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/dealings-competitors/market-division-or

Gansler, J. (2011). *Democracy's arsenal: Creating a twenty-first century defense industry.* Cambridge, MA: MIT Press.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

Grullon, G., Larkin, Y., & Michaely, R. (2017). *Are US industries becoming more concentrated?* [Working paper].

Hoff, R. V. (2007). *Analysis of defense industry consolidation effects on program acquisition costs. Monterey*, CA: Naval Postgraduate School.

Hunter, A., Sanders, G., Meitiv, A. L., McCormick, R., McQuade, M., & Nzeribe, G. (2015). *Avoiding terminations, single-offer competition, and costly changes with fixed-price contracts* (CSIS-CM-15-119). Retrieved from https://calhoun.nps.edu/bitstream/handle/10945/53598/CSIS-CM-15-119.pdf?sequence=1

Kettl, D. (1993). *Sharing power: Public governance and private markets*. Washington, DC: Brookings Institution.

Kunneke, R. W. (1999). Electricity networks: How "natural" is the monopoly? *Utilities Policy*, 8.

Manuel, K. (2011). *Competition in federal contracting: An overview of the legal requirements.* Washington, DC: Congressional Research Service.

McCormick, R., Hunter, A. P., & Sanders, G. (2017). *Measuring the impact of sequestration and the drawdown on the defense industrial base*. Washington, DC: Center for Strategic and International Studies. Retrieved from https://csis-prod.s3.amazonaws.com/s3fs-public/publication/180111_McCormick_ImpactOfSequestration_Web.pdf?A10C65W9Qkx07VaJqYcJguCH.7EL3O7W

Peltzman, S. (1977). The gains and losses from industrial concentration. *Journal of Law and Economics*, 20.

Posner, R. A. (1969). Natural monopoly and its regulation. *Stanford Law Review*, 21.

Sanders, G., & Hunter, A. (2017). *Overseas contingency operations contracts after Iraq: Enabling financial management research and transparency through contract labeling.* Retrieved from https://www.researchsymposium.com/conf/app/researchsymposium/unsecured/file/145/SYM-AM-17-051-005_Sanders.pdf

Shapiro, C. (n.d.). Antitrust in a time of populism. Manuscript in preparation.

Weisberger, M. (2015). Lockheed-Sikorsky deal stokes fears about industry consolidation. Retrieved from http://www.defenseone.com/business/2015/09/sikorsky-lockheed-deal-stokes-fears-about-industry-consolidation/122445/

Weiss, L. (1974). The concentration-profits relationship and antitrust. In H. J. Goldschmid et al. (Eds.), *Industrial concentration: The new learning*. New York, NY: Little Brown & Co.

White House Council of Economic Advisers. (2016, April). *Benefits of competition and indicators of market power*. Retrieved from https://obamawhitehouse.archives.gov/sites/default/files/page/files/20160414_cea_competition_issue_brief.pdf

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Contractual Flow-Down Clauses: Deterrence to Non-Traditional Defense Contractors From Doing Business With DoD

Mark F. Kaye—received a BA from Dickinson College, JD from Georgetown University Law School, and MBA from The University of Chicago. [mkaye@ida.org]

Daniel L. Cuda—received a BS from the U.S. Air Force Academy, MS from the Air Force Institute of Technology, MA from Georgetown University, and PhD from George Mason University. [dcuda@ida.org]

Kevin Y. Wu—received a BA from Middlebury College and is currently working on an MPS at the University of Maryland. [kywu@ida.org]

Abstract

Section 887 of the National Defense Authorization Act (NDAA) for Fiscal Year 2017 directed the Secretary of Defense to conduct a study of contractual flow-down clauses related to major defense acquisition programs (MDAPs) and subcontractors. The MDAP data sample indicates no clear misapplication of flow-down clauses from prime contractors to subcontractors; however, the significant number of flow-down clauses reflects the voluminous nature of the Federal Acquisition Regulation (FAR) and Defense Federal Acquisition Regulation Supplement (DFARS). The existing and growing burden of regulatory compliance will reduce private sector participation in MDAPs, and the possibility exists that in the future, the Department of Defense may not have ready access to advanced technologies and capabilities for MDAPs that reside in the private sector due in part to this regulatory burden.

Introduction

Project Background

The Congress perceives that regulatory burden may deter many types of innovative firms from doing business with the Department of Defense (DoD) and that prime contractors exacerbate the reluctance of firms to engage in defense business by overzealously extending regulatory requirements to subcontractors.

In this context, the Congress, in the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2017 (Public Law 114-328), Section 887(a), required the Secretary of Defense to “conduct a review of contractual flow-down provisions related to major defense acquisition programs on contractors and suppliers, including … nontraditional defense contractors.”

As directed by Section 887(b), the DoD asked the Institute for Defense Analyses (IDA) to conduct the required review, specifically to

- Determine if there are instances in which flow-down clauses in contracts between DoD and prime contractors have been misapplied to the prime’s subcontractors, through sampling of a limited number of major defense acquisition programs (MDAPs). The focus will be on clear misapplications and not administrative or extraneous misapplications, due to schedule and resource limits.
- Conduct literature reviews and interviews with governmental and commercial sector personnel in order to help determine the effect, if any, of Federal

Acquisition Regulation (FAR)/Defense Federal Acquisition Regulation Supplement (DFARS) flow-down provisions on DoD MDAPs in terms of access to advanced research and technology capabilities available in the private sector.

Defining FARS, DFARS, and Contract Flow-Downs

The FAR and its specified contract clauses control and shape most acquisitions by U.S. government executive branch agencies. Broadly stated, the FAR is a publicly accessible set of rules that controls almost all U.S. government contracting with the global commercial economy. It is a codification of general and permanent rules or regulations published by executive branch departments and agencies in the Code of Federal Regulations (CFR). The FAR represents Parts 1 through 53 of Title 48 of the CFR.[1]

Most federal agencies—including the DoD—have their own unique supplements to the FAR. The DFARS implements FAR policies and procedures and supplements the FAR to meet DoD-specific needs (Manuel, Halchin, & Lunder, 2015). The DFARS contains authorized deviations from the FAR as well as requirements of law, DoD-wide policies, and so forth. The DFARS needs to be read in conjunction with the FAR, as the FAR is the primary set of acquisition rules.[2]

When federal agencies undertake acquisitions, the FAR provides the basis for contract clauses that form the legally enforceable agreement between the U.S. government and a private contractor. In turn, when this same private contractor contracts with another firm to help execute the government's contract, this second firm becomes a U.S. government subcontractor. The first firm is identified here as the prime contractor. The prime contractor, in their contract with the subcontractor, will "flow-down" a number of the original contract clauses from the government contract. These are referred to hereafter as *flow-down clauses*. In general, our research seeks to understand the impact of these government flow-down clauses on actual and potential government subcontractors.

Analytical Approach

The NDAA reviews language focused on subcontractors and their contributions to the overall excellence of DoD military technical systems. These contributions are overwhelmingly regulated by the contractual relationship between the prime and subcontractor but are nevertheless shaped by FAR/DFARS flow-down clauses. Our analytical approach to this general issue and the specific areas of emphasis described by the NDAA began with first understanding the dynamic nature of the FAR/DFARS themselves. These regulations are living documents updated monthly. We established the December 2016 version of the FAR/DFARS as our baseline. Given their importance to commercial industry, contracting departments within commercial defense companies specialize in FAR/DFARS interpretation and application to DoD contracts. In turn, the FAR/DFARS have generated a broad set of written literature. The research team first examined this literature and then proceeded to interview government and commercial FAR/DFARS specialists with an emphasis on commercial primes and subcontractors. These

---

[1] The FAR is available at https://www.acquisition.gov/browsefar.
[2] The DFARS is available http://www.acq.osd.mil/dpap/dars/dfarspgi/current/. The DFARS is often accompanied by PGI memos.

interviews sampled current FAR/DFARS application and practice, solicited commentary on administrative burdens, and worked to understand subcontractor participation in DoD projects that addressed wider congressional concerns. Finally, the IDA team obtained copies of specific contracts between prime contractors and the government and then obtained related subcontracts to sample the practice of flow-down clauses over a range of MDAP types.

Source Data

We used three sources of data and information for this paper. First, we examined previous academic and commercial trade literature in and around the general subject area. We used a variety of subscription databases and publicly available information. Such information was instructive for understanding legal and business implications of FAR/DFARS flow-down clauses. Overall, the literature specifically addressing flow-down clauses is a relatively limited subset of the wider literature of FAR/DFARS and government contracting. Second, we employed qualitative research methods to supplement our literature review by interviewing industry and government representatives to gain practical insights. Last, we obtained copies of prime contracts from the DoD for a small number of MDAPs and then contacted the prime contractors to obtain a small sample of subcontracts pertaining to each prime MDAP contract. Each subcontract required non-disclosure agreements (NDAs) with the prime.

Flow-Down Clauses

FAR/DFARS flow-down clauses derive from the situation that most of the cost of an MDAP is driven by subsystems and components provided by subcontractors[3] to the prime. It is not unusual for the prime to only account for 10–20% of the total MDAP cost. Subcontractors—inclusive of basic parts and material suppliers that feed the upper tiers of the supply chain—account for the bulk of the defense industrial base (Gansler, 2011, pp. 132–133). MDAP prime contractors are often viewed as integrators of systems and parts and not necessarily as weapons manufacturers (Gansler, 2011; GAO, 2010).[4]

To produce and deliver an MDAP to the DoD, prime contractors enter into contracts with an array of firms. The vast majority of MDAP prime contractors are well-known major defense firms. The first tier of subcontractors are a diverse group of firms—it is not uncommon for some of the key first-tier subcontractors to be other large defense or industrial manufacturers. Below the first tier subcontractors are additional tiers that provide sub-components, parts, commodity type items (e.g., fasteners), and various basic materials. The subcontractors will often have contractual arrangements among themselves (e.g., a first tier provider of a major subsystem will obtain parts from a second tier and, in turn, the second tier parts firm will obtain raw materials from a third-tier firm).

The government nominally has no role in these subcontracts. By well-established legal precedents, the U.S. government has no “privity of contract”[5] with any of the two

---

[3] *Subcontractor* is generally defined as any firm that supplies materials or provides services for a prime contractor or for a higher tier subcontractor (FAR 44.101).
[4] The following are estimates on percentage of costs driven by subcontractors of all tiers: (1) ships: 82–88%, (2) missiles: 70–80%, and (3) fighter aircraft: 80%.
[5] The U.S. government is not a direct party to the subcontract.

parties in a subcontract; the subcontract is viewed in the courts as exclusively between the commercial prime contractor and their subcontractor. In practice, subcontracts[6] under a DoD prime contract are "hybrid" contractual documents (Feldman, 2009).[7] They are combinations of clauses from the Uniform Commercial Code (UCC) as enacted by the various states (e.g., California or Connecticut), industry-preferred commercial clauses, and governmental clauses. Although the U.S. government disclaims any direct responsibility or liability for subcontracts, it does maintain a measure of control by both maintaining the right to approve subcontracts themselves and to direct the introduction of the flow-down clauses.[8] The clauses themselves are contained in FAR Part 52[9] and DFARS Part 252.[10]

The most common means of government control over subcontracts is through its mandatory flow-down contract clauses. These are clauses the government insists will be placed in almost all subcontracts with only limited exceptions.[11] These clauses are written and mandated to protect the government's rights and interests within the overall project. They can serve to promote formal national security interests such as preventing espionage and business fraud, but also to shape procurement practices and promote socio-economic policies (Feldman, 2009, p. 254). Congressional interest has been piqued by the flow-down of product-unrelated government policy items such as the following:

- **52.237-11 Accepting and Dispensing of $1 Coin (Sept. 2008).** All business operations conducted under this contract that involve coins or currency, including vending machines, shall be fully capable of accepting $1 coins in connection with such operations.
- **52.223-18 Encouraging Contractor Policies to Ban Text Messaging While Driving (Aug 2011).** The Contractor is encouraged to adopt and enforce policies that ban text messaging while driving.
- **52.204-10 Reporting Executive Compensation and First-Tier Subcontract Awards (Oct 2016).** First-tier subcontract information … with a value of $30,000 or more, the Contractor shall report the information at http://www.fsrs.gov for that first-tier subcontract.

The legal literature and commentary does make a distinction between mandatory flow-down clauses and non-mandatory flow-down clauses. While this is a relevant legal

---

[6] *Subcontract* in this context is related to purchase of good or services substantially or directly related to the performance of the MDAP prime contract.
[7] The primary body of governing law for subcontracts is commercial contract law—generally expressed in the individual State UCC.
[8] For example, subcontracts identified by the contracting officer as of a critical nature or high value, in order to protect the government's interests (FAR 44.201).
[9] FAR Part 52: Solicitation Provisions and Contract Clauses. This is a matrix of provisions and clauses with reference to the particular CFR Title 48 section that prescribes a particular provision or clause.
[10] DFARS Part 252: Solicitation Provisions and Contract Clauses.
[11] In theory, FAR/DFARS flow-downs can extend to an infinite tier of subcontractors. The limited exceptions are contained in the FAR/DFARS themselves.

distinction in drafting subcontracts, for the purposes of this research, it is less relevant.[12] The mandatory flow-down clauses are certainly non-negotiable and must be flowed down.[13] However, some of the non-mandatory clauses can be as important in terms of promoting government policy,[14] ensuring that the subcontractor will provide adequate support or cooperation to enable the prime contractor to meet its contractual obligations, or in protecting the prime's financial position vis-à-vis subcontractors from unilateral government action.[15] The architecture of clauses in a subcontract is a result of the prime contractor's contracting personnel, industry practice, and government mandates and emphasis, along with the negotiating positions of the prospective subcontractor.

Among congressional concerns, there are additional suggestions of a growing quantity of FAR/DFARS clauses increasing the size of flow-down contracts. As determined on the basis of clause change date, Figure 1 and Figure 2 show the latest issue dates for 586 current FAR clauses and 377 DFARS clauses, respectively. The data are derived from a study performed by Richard Ginman (2016) and used with his permission. Figure 1 shows a large number of clauses revised in 1984 due to the evolution of the FAR system in the early 1980s from the older Armed Forces Procurement Regulations. Since that time, generally between one and 30 FAR clauses have been issued, modified, or amended annually. The exception was a post-1984 high point in 2014 with approximately 50 clauses either changed or newly issued. The trend of issues from 2014 to 2016 appears to be departing from the historical norm. This period is too short to make firm conclusions, but is suggestive of concern expressed in the NDAA.

[12] The literature and legal professional materials also note that clauses may be flowed down by reference (e.g., FAR part number), full text of the FAR/DFARS clause, and as substantially the same as the FAR/DFARS clause. While important for legal drafting, this is not relevant to this research.
[13] Some examples of mandatory clauses that could be included in a subcontract are FAR 52.203-7: Anti-kickback Procedures, and FAR 52.215-23: Limitations on Pass-Through Charges.
[14] E.g., FAR 52.225-1: Buy American.
[15] E.g., FAR 52.249-2: Termination for Convenience of the Government. If the prime does not include this clause in its subcontract, it will not have a right to terminate the subcontract.

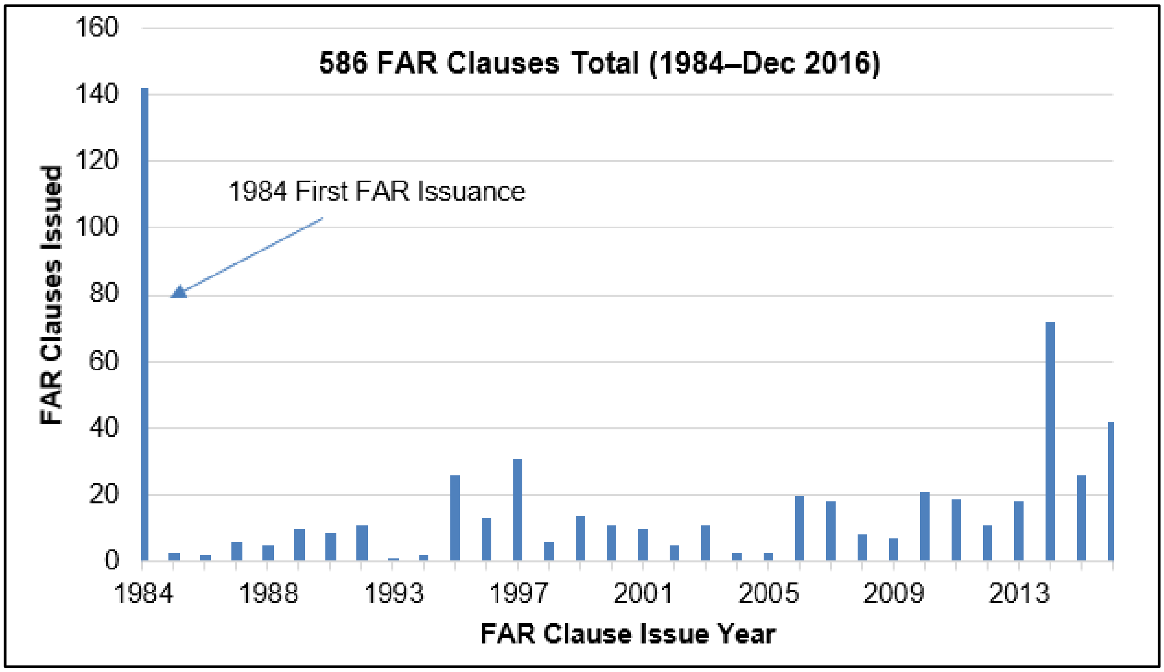


Figure 1. Basic FAR Clause Issue Data
(Ginman, 2016)

DFARS issue or re-issue dates shown in Figure 2 suggest extensive efforts to revise its clauses since approximately 2008. The chart displays the changing number of clauses by year—the data do not distinguish between new issues and re-issues. The chart shows an overwhelming spike of issues around 1991. These issues were connected with the July 1989 Defense Management Report, which “concluded that much of the stifling burden of Department of Defense (DoD) regulatory guidance, including the Defense Federal Acquisition Regulation Supplement (DFARS) was self-imposed” (DoD, 1991). After this activity, between 1991 and 2008, there were occasional spikes of activity with otherwise relatively small numbers of annual revisions. However, after 2008, the rate of issue or revision increased notably, and in the period since 2012, each year exceeds all previous issue years except for 1991.

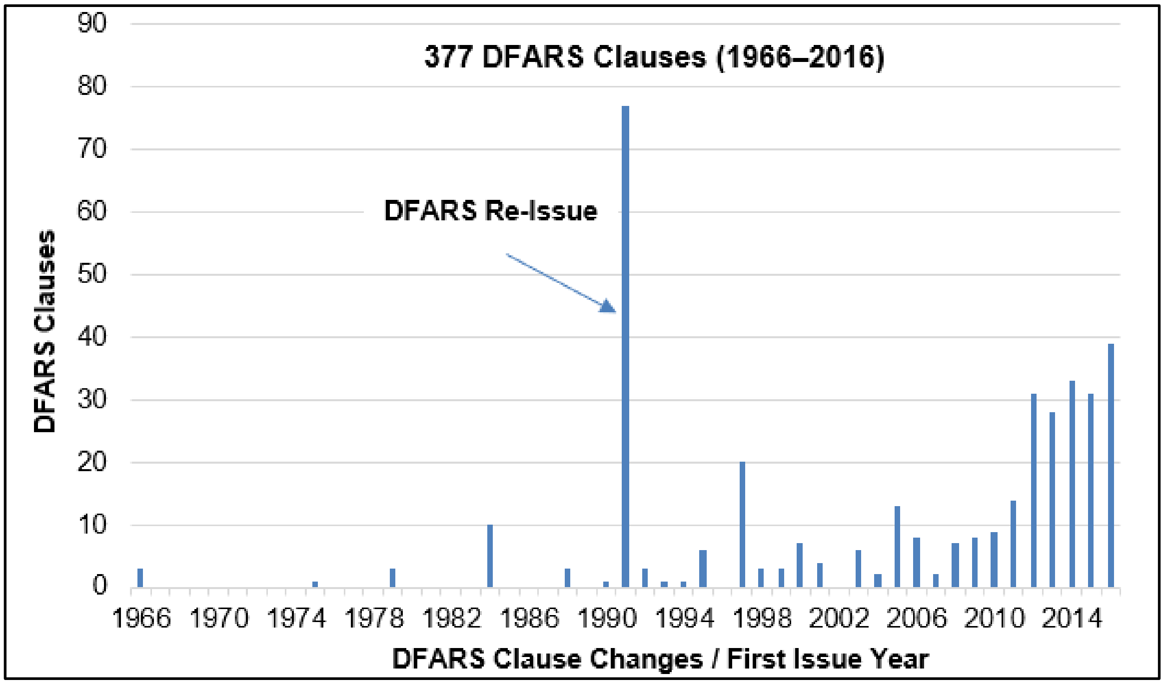


Figure 2. Basic DFARS Clause Issue Data
(Ginman, 2016)

Statutes are a rich source of new regulations. NDAA Title VIII is the ordinary source for statutory changes to the DoD acquisition and contracting regulations and the CFR. Final publication in the CFR represents a formal change to the FAR/DFARS. In addition to NDAA Title VIII, specific legislation can also lead to FAR/DFARS changes. Figure 3 also counts the number of sections in specific legislation such as the 2009 Weapon Systems Acquisition Reform Act (WSARA), the Services Acquisition Reform Act (SARA), and the Acquisition Improvement and Accountability Act of 2007. The assumption here is that the greater number of sections of either the annual NDAA Title VIII or specific legislation, the more likely that a change to the FAR/DFARS is to occur. Figure 3 shows the trend line for the total number of sections.

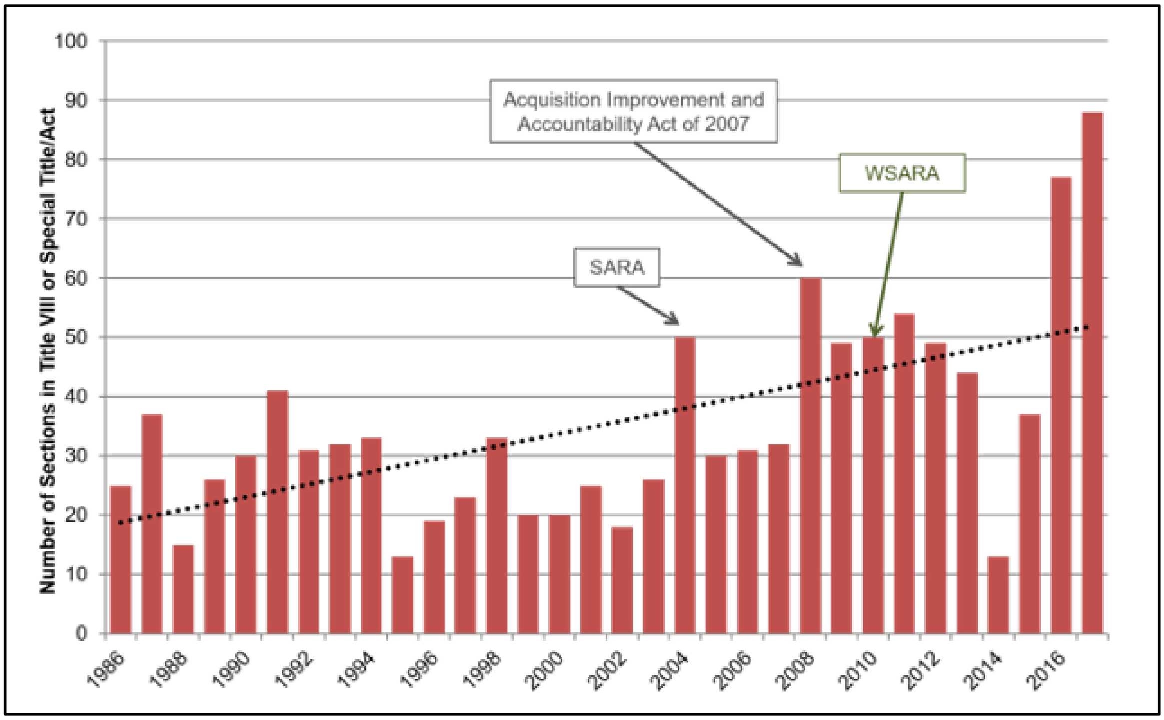


Figure 3. NDAA and Policy Direction Sections
(NDAAs FY 1986–FY 2017)

Review and Analysis

Potential Misapplication of MDAP Flow-Down Clauses

A nearly universal commercial attitude toward government contracts and their associated FAR/DFARS clauses is one of burden. They are perceived to impose administrative burden on both primes and subcontractors, and set government and the DoD apart from other commercial partners. Yet the FAR/DFARS exist to fulfill legitimate government purposes. Beyond this, the NDAA expresses the concern that some FAR/DFARS clauses are inappropriately applied to subcontractors—that there is a clear misapplication of mandatory and customary clauses by prime contractors to their subcontractors. The working hypothesis suggested by the NDAA is that primes are reacting to U.S. government contracting officers who are using standard sets of clauses (1) disconnected from the factual situation of the MDAP or (2) to simply avoid risk by adding extraneous clauses; in turn, prime contracting personnel are seeking to eliminate or minimize risk exposure by flowing down the proverbial “kitchen sink” of clauses.

To address this hypothesis, the research team worked with a set of major defense contractors with whom the DoD has contracted to develop and produce a diverse set of MDAPs—Raytheon, General Atomic, BAE, General Electric, and General Dynamics.[16]

The research team requested from the participating contractors that they allow us to obtain copies of their first-tier subcontracts.[17] These subcontracts were limited to one major program per contractor. Proprietary data issues were addressed by non-disclosure agreements with all firms. Discussions with the primes revealed that all five use standardized contracts for dealing with their first-tier subcontractors; therefore, the same set of FAR/DFARS clauses and other commercial terms were flowed-down to all first-tier subcontractors.

Table 1 shows FAR flow-down clauses in each subcontract in the wider context of the total FAR clauses in the prime contract (first row). Total subcontract clauses by MDAP contract are shown in the second row, together with the percentage of this number relative to the number in the first row. This total number of subcontract clauses is broken down between the number of clauses that are actual flow-downs from the prime contract (third row) and those typically flowed from prime contract Terms and Conditions (T&C) unrelated to the U.S. government–associated flow-down clauses (fourth row).

Table 1. FAR Prime Contract Clause Flow-Downs

Statistic	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total Number of FAR Clauses in Prime Contract	**123**	**73**	**140**	**76**	**99**
Total Number of FAR Clauses in Subcontract (Percent of Prime Contract Clauses)	**97 (79%)**	**110 (151%)**	**81 (58%)**	**78 (103%)**	**83 (84%)**
Specific FAR Clauses in DoD Prime Contract Flowed Down to Subcontractor	70	43	51	39	80
Additional FAR clauses from Prime T&C to Subcontractor	27	67	30	39	3

For example, under MDAP 1, a total of 123 FAR clauses were present in the prime contract, 70 of which then became clauses identifiable in the prime's subcontracts. In addition to these 70, 27 FAR clauses were also present in the subcontracts—additional FAR clauses from the prime contract's standard T&C.[18] Together, these two sources generated a total of 97 FAR clauses identifiable in the subcontract. This total of 97 represented 79% of the total FAR clauses contained in the prime contract.

[16] The largest U.S. defense firm, the Lockheed Martin Corporation (LMC), declined to participate in the research. LMC indicated they would participate in the research if their MDAP contracts were modified to cover the increased scope of work entailed by this effort; however, the responsible command with budget authority declined. To our knowledge, the prime contractors who did participate did not make similar requests.
[17] Obtained from DAMIR.
[18] The clauses in the prime's Terms and Conditions were direct references to FAR/DFARS clauses.

The construction of T&C varies by firm and their internal contracting practices, but the prime's T&C can be a significant source of subcontract FAR clauses.[19] MDAP 2 and its unidentified prime contractor placed a significant number of clauses in its subcontract and then added more from T&C—as a result, for MDAP 2, the total number of clauses in its subcontract represented 151% of the total number of clauses in its own prime contract. T&C generated more clauses than the original contract itself.

Considering the sample size and research limitations, it is difficult to categorize the variation in quantity of FAR clauses flowed down to subcontractors. Discussions with government and private sector contracting personnel suggest the primary drivers are (1) prime contractor corporate legal and risk management policies; (2) the need for contract standardization, since primes often have thousands of contracts; and (3) the specialization of firms in the defense industry (especially the prime to first-tier sample) that have financial, accounting, legal, and other compliance systems in place to manage U.S. government/DoD contracting requirements, regardless of whether the firm is a prime contractor or subcontractor—and therefore are accustomed to a significant number of FAR/DFARS clauses and generally accept them.

A legal principle called the Christian Doctrine (*G.L. Christian & Assocs. v. United States,* 1963) was also noted by some contracting experts. A Federal Court of Appeals ruled that a mandatory contract clause that conveys a deeply ingrained strand of public procurement policy is considered to be included even if it is not in the agreement between the U.S. government and the contractor. The same experts noted that, in addition to the reasons cited in the preceding paragraph, the primes may add flow-down clauses to their T&C in view of the Christian Doctrine. For example, if the U.S. government or DoD contracting officer unintentionally omitted a mandatory clause, the prime will still be bound by the omitted FAR/DFARS clause. Thus, as corporate protection vis-à-vis its subcontractors, the prime will include all mandatory clauses in its T&C. Although this subject matter is outside the scope of this research, this legal doctrine—established via federal court ruling—should be acknowledged when examining flow-down clauses.

Once the research team had defined the set of contracts/subcontracts and their respective clauses, we analyzed the nature of the clauses themselves to ascertain if any of the FAR flow-down clauses were misapplied by the primes to their subcontractors. As noted earlier, the focus was on clear misapplications and not administrative error or extraneous contracting oversights. An administrative error was judged by evaluating the clause in light of the overall context of the contract and the likely work it generated to achieve compliance. If an included clause was irrelevant to the contract and it required no work to achieve compliance, it was evaluated as an error. Alternately, if review of the contract clause found no government interest in its actual application and its compliance generated work on behalf of the contractor, the clause was judged a "clear misapplication."

To assist in this analysis, we used Ginman's (2016) categorization of FAR and DFARS clauses as *mandatory, optional,* or *neither mandatory nor customary*. Mandatory clauses are those required to be flowed-down to subcontractors. Optional clauses may also be referred to as non-mandatory (i.e., not specifically required by the FAR/DFARS)—the

[19] We did not address Uniform Commercial Code (UCC) and other commercial terms as they are outside the scope of this research effort.

legal literature often refers to such clauses as *discretionary* or *recommended*. The practicalities of contract and financial management, adherence to U.S. laws and regulations, and need for the prime contractors to protect their legal, financial, and reputational interests deem large numbers of these so-called non-mandatory clauses to be effectively required. In this paper, we refer to Ginman's (2016) optional clauses as *customary* because the prime contractors require these clauses to produce the MDAP while adhering to U.S. government laws and regulations in a manner that protects the prime contractors' legitimate interests. The third category—*neither mandatory nor customary*—consists of clauses deemed not required to be flowed-down to subcontractors.

You may recall that in Table 1, the third row indicates the number of DoD-Prime FAR contract clauses that were flowed down to the subcontractors (e.g., for MDAP 1, there were 70 such clauses). In Table 2, we take those clauses and apply Ginman's (2016) analytical tool to break down the DoD-Prime FAR clauses that were flowed down to subcontractors into our three categories (mandatory, customary, and neither mandatory nor customary). We also include another row entitled "Solicitation Provisions"[20] and conclude with a summary row entitled "Misapplications."

Table 2. FAR DoD-Prime Clauses Flowed Down From Prime to Subcontractor

	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total	**70**	**43**	**51**	**39**	**80**
Mandatory	42 (60%)	25 (58%)	30 (59%)	25 (64%)	33 (41%)
Customary	22 (31%)	16 (37%)	13 (25%)	11 (28%)	30 (38%)
Neither Mandatory nor Customary	6 (9%)	2 (5%)	7 (14%)	3 (8%)	17 (21%)
Solicitation Provisions	0	0	1 (2%)	0	0
Misapplications*	**4 (6%)**	**0**	**1 (2%)**	**0**	**0**

* Flow-downs that appear to be administrative error or contract drafting oversight, not necessarily a clear misapplication

In analyzing if there were clear misapplications, we examined all of the clauses that are not categorized as mandatory to determine if they appeared to be reasonable or within the realm of expected FAR clauses when considering the subject matter of the MDAP. Due to research constraints, we were not able to analyze each FAR clause in exacting detail. Such an intensive analysis may be extremely challenging, if not impractical, as we did not have insight into the prime-subcontractor negotiations, past histories between the firms, MDAP program specifics, prime corporate policies on legal and risk management, and a host of other variables. Therefore, our approach was one of general reasonableness within the general construct of the prime MDAP contract type, size of the contract, and recognition

[20] We included this row in the analysis because we came across a number of solicitation provisions that were apparently unintentionally flowed down from the prime to the subcontractor.

that the subcontractors are first-tier and almost certainly have ample negotiating abilities and resources.[21]

As can be seen in Table 2, only two of the five MDAP subcontracts examined were assessed as having misapplications; however, even these are not clear misapplications. MDAP 1 contained the most clauses assessed as misapplication (four, or 6%) and MDAP 3 had one solicitation provision. We considered the bulk of neutral FAR clauses—neither mandatory nor customary—as benign, and therefore not clear misapplications according to the reasonableness standard described above.[22]

We next reviewed those FAR clauses incorporated by reference or verbatim[23] in the primes' T&C[24] that are accepted by the subcontractors. We found the primes' T&C to be an unexpected source of additional FAR clauses that were included in the prime-to-subcontractor agreements.

As shown in Table 3, the misapplication totals in the T&C are larger than those in the contracts, using the same methodology described above. Clause misapplication was as high as 19% in MDAP 1, but assessed as below 10% in MDAPs 2, 3, and 4, and zero for MDAP 5. Once again, pursuant to a reasonableness standard (and considering the other factors noted), we did not observe clear misapplications of flow-down clauses. The misapplications that appeared were mainly nuisance clauses; that is, primarily solicitation provisions that were apparently inadvertently captured in standardized agreements.

---

[21] At least two of the subcontractors are major U.S. defense contractors (and primes on other DoD MDAPs) and others are publicly listed manufacturing firms with global operations.
[22] We identified a small number of clauses that were not solicitation provisions for MDAP 1 that seemed inapplicable; however, the appearance was more of administrative error or extraneous clauses of no importance. For example, FAR 52.237-02 (Protection of Government Buildings, Equipment and Vegetation) seems not applicable in view of the nature of the contract; and 52.242-03 (Penalties for Unallowable Costs) deals with indirect rates and is not considered a flow-down clause.
[23] An indication of the specialization in DoD business or strong familiarity in contracting with the U.S. government is that incorporation of FAR clauses in the subcontracts was by reference. That is, only the FAR clause and title were listed, without clause language or reference to FAR clause prescription contained elsewhere in the FAR.
[24] Primes' T&C refer to the obligations accepted by the parties via private contractual arrangement. As noted earlier, most commercial contracts are governed by the UCC. The contracts we examined are a hybrid of UCC and FAR/DFARS.

Table 3. FAR Clauses Flowed Down From Prime to Subcontractor via Primes' T&C

	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total	**27**	**67**	**30**	**39**	**3**
Mandatory	17 (63%)	22 (33%)	17 (57%)	26 (67%)	0 (0%)
Customary	4 (15%)	30 (45%)	10 (33%)	7 (18%)	3 (100%)
Neither Mandatory nor Customary	1 (4%)	9 (13%)	1 (3%)	3 (8%)	0
Solicitation Provisions	5 (19%)	6 (9%)	2 (7%)	3 (8%)	0
Misapplications*	**5 (19%)**	**6 (9%)**	**2 (7%)**	**3 (8%)**	**0**

* Flow-downs that appear to be administrative error or contract drafting oversight, not necessarily a clear misapplication

For both sets of FAR flow-down clauses examined—those from the DoD that the prime passed to the subcontractors and those that the primes appropriated from the FAR and flowed down—we did not observe clear misapplication of FAR clauses. Since the parties to the subcontract agreements are veteran DoD contractors, we can assume that they are sophisticated in such matters, regularly deal with each other concerning MDAPs, and therefore have refined their legal processes to reflect the realities of DoD contracting.

We continued our examination of flow-down clauses by analyzing DFARS clauses, using the same methodology and analytical approach as with the FAR clauses. The results were comparable to the FAR flow-downs. Flow-down of prime contract clauses were sometimes less than the total clauses in the prime contract, but, again, flow-down of T&C clauses proved highly variable. As shown in Table 4, the prime contractors for MDAP 2 and MDAP 4 made extensive use of the T&C section of their contracts to generate subcontractor clauses.

Table 4. DFARS Prime Contract Clause Flow-Downs

	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total Number of DFARS Clauses in Prime Contract	**93**	**35**	**78**	**41**	**56**
Total Number of DFARS Clauses in Subcontract (Percent of Prime Contract)	**88 (95%)**	**74 (211%)**	**55 (71%)**	**64 (156%)**	**46 (82%)**
Clauses in DoD-Prime Contract Flowed Down to Subcontractor	60	26	38	18	41
Additional DFARS Clauses from Prime T&C to Subcontractor	28	48	17	46	5

We scored DFARS clause misapplications with results similar to those obtained for FAR clauses. Misapplications were present, yet were of small numbers, and we did not observe clear misapplications. However, another interpretation of these data is worth addressing—whether the intent of misapplication scoring is to help understand the discretionary dimension of subcontracting; that is, to understand if prime contractors might be burdening subcontractors with flow-downs either to reduce their own contract risk with the government or from a misunderstanding of government intent. This is a broader interpretation of “misapplication”; the data can be interpreted to mean that prime contractors used their discretion to make misapplications, but that this discretion included clauses that were judged to be neither mandatory nor customary. This broader interpretation suggests that an element of independent choice exists among prime contractors regarding their imposition of subcontract clauses in their subordinate contracts. In evaluating these additional clauses, all appear appropriate and justified, but this wider interpretation does help scope the dimension of the broader problem to include prime contractor discretion.

Table 5 summarizes DFARS clauses from the DoD-Prime MDAP contracts that were flowed down to the first-tier subcontractors. The pattern and results mirror those of the FAR clause flow-downs shown in Table 2 (i.e., no clear misapplications and solicitation provisions being the prime generator of apparent administrative error).

Table 5. DFARS DoD-Prime Clauses Flowed Down From Prime to Subcontractor

	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total	**60**	**26**	**38**	**19**	**41**
Mandatory	34 (57%)	14 (54%)	20 (53%)	9 (47%)	19 (46%)
Customary	16 (27%)	6 (23%)	13 (34%)	5 (26%)	13 (32%)
Neither Mandatory nor Customary	10 (17%)	4 (15%)	3 (8%)	4 (21%)	8 (20%)
Solicitation Provisions	0	2 (8%)	2 (5%)	1 (5%)	1 (2%)
Misapplications*	**4 (7%)**	**2 (8%)**	**2 (5%)**	**1 (5%)**	**1 (2%)**

* Flow-downs that appear to be administrative error or contract drafting oversight, not necessarily a clear misapplication

As with the FAR clauses, the primes also apparently used their discretion and included DFARS clauses that are neither mandatory nor customary; however, utilizing the reasonableness standard, we did not observe clear misapplications. The clauses appear harmless and may even facilitate contract management.

Last, the DFARS clauses included in the primes’ T&C—but not in the DoD-Prime MDAP contract—corresponded to what we observed in the FAR clauses appropriated by primes for incorporation into their T&C (shown in Table 3). The results are shown in Table 6. We observed no clear misapplications in these data per our analysis.

Table 6. DFARS Clauses Flowed Down From Prime to Subcontractor via Primes’ T&C

	MDAP 1	MDAP 2	MDAP 3	MDAP 4	MDAP 5
Total	**28**	**48**	**17**	**46**	**5**
Mandatory	17 (61%)	22 (46%)	11 (65%)	26 (57%)	2 (40%)
Customary	6 (21%)	18 (38%)	3 (18%)	14 (30%)	2 (40%)
Neither Mandatory nor Customary	1 (4%)	4 (8%)	0	1 (2%)	1 (20%)
Solicitation Provisions	4 (14%)	4 (8%)	3 (18%)	5 (11%)	0
Misapplications*	**6 (21%)**	**4 (8%)**	**3 (18%)**	**5 (11%)**	**0**

* Flow-downs that appear to be administrative error or contract drafting oversight, not necessarily a clear misapplication

Overall, we found that prime contractor misapplication of FAR/DFARS clauses occurs, but we did not observe clear misapplications of flow-down clauses. Quantitatively, the noted misapplications appear small within this limited sample of first-tier subcontracts. Interactions with contracting departments of prime contractors during our research left an impression of highly skilled and experienced personnel who manage large volumes of contract work via standardization. We conclude that first-tier subcontractors experienced in defense work are likely to be equally expert in FAR/DFARS provision. Often prime contractors and their contracting organizations act as subcontractors themselves and carry this expertise into this role. As discussed elsewhere in the paper, the choice to become a government contractor appears to be a choice not taken lightly without some knowledge of the skills necessary to be a successful participant.

Impact of Flow-Down Clauses on Participation of Subcontractors on DoD Contracts and DoD's Access to Advanced Research and Technology Capabilities

The 2017 NDAA appears to suggest that subcontractor participation rates could vary inversely with the regulatory burden[25] of flow-down clauses. In turn, the DoD's ability to access advanced research and technology capabilities available in the private sector (non-traditional DoD contractors) could be stymied.

The wider issue evident in the NDAA is the concern that the technology of future military advantage is increasingly emanating from commercial industry sectors wider than the ordinary defense industrial base. Specifically, the concern is that a new set of military advantages is developing beyond the research and development (R&D) directed by the U.S. government and DoD. In its review of the 2017 NDAA, the Senate Armed Services Committee (SASC) stated, "DOD has struggled to tap into the technological advancements that are increasingly being driven not by large defense contractors, but by commercial technology firms that generally choose not to do business with the DOD." The SASC stated it believes these firms have been generally deterred by the unique demands of the defense acquisition system, including acting as subcontractors, in light of FAR/DFARS flow-down provisions (U.S. Senate Armed Services Committee, 2017).

In general, administrative burden represents a cost of doing business with the DoD and the wider U.S. government. Ordinarily, cost is associated with a contractor or vendor; it is an attribute of the product to be purchased by the government. In contrast, FAR/DFARS provisions represent conditions or information the government expects from its contractors in addition to the production of the contracted service or good. These requirements are not useless—they assist the DoD in its responsibilities to the U.S. taxpayer, and they seek to avoid fraud, malfeasance, and waste. However, they do require the prime federal contractor to achieve certifications and share information in a way that distinguishes these contracts from almost any other type of commercial exchange. These provisions clearly represent an additional cost to both the prime providers of federally procured products and—in the case of flow-down provisions—to the prime's subcontractors.

To do business directly or indirectly with the DoD and the U.S. government, contractors must often must make a conscious addition to their overhead costs to administer

---

[25] *Regulatory burden* in this context means cost and time spent on compliance with FAR/DFARS. This is inclusive of tracking, training personnel, maintaining databases, ensuring subcontractors/vendors are in compliance, etc.

the FAR/DFARS required provisions: U.S. government and DoD contracting expertise must be recruited, internal accounting practices must be modified, and hiring and reporting provisions must be implemented. This essentially becomes an element of overhead if the business owner plans to systematically seek DoD and U.S. government contracts now and in the future—in effect, it becomes a part of the cost basis for products produced for the prime contractor.

This requirement for firms to configure themselves for DoD and government contracting effectively creates the situation feared by the NDAA: Firms often must make a strategic decision to become a DoD subcontractor. It is not a casual decision. The requirements are unusual and distinctive from ordinary commercial contracting. Essentially, FAR/DFARS flow-down requirements serve to make the DoD and the U.S. government a unique customer for all but the simplest commercial items. The imposed costs codified in the FAR/DFARS provisions can effectively raise costs for government purchases. These additional imposed costs may be charged back to the government in the form of higher prices, particularly if the firm is not required by the circumstances of the contract to trim its prices. In effect, the additional costs of FAR/DFARS flow-down provisions may frequently be paid for by the government itself in terms of higher prices.[26]

Our independent research was recently validated in a Government Accountability Office (GAO) report concerning DoD acquisitions and challenges in attracting non-traditional firms (i.e., firms that do not traditionally do business with the DoD either directly or indirectly as subcontractors) to do business with the DoD (GAO, 2017). This GAO report is noteworthy in that the same major deterrence factors—in addition to regulatory burden and compliance—are universally expressed by government and private sector members of the U.S. defense industrial base. The report noted the variety of factors affecting participation—with FAR/DFARS clauses being one (GAO, 2017).[27]

The issue of DoD access to advanced research and technology capabilities is closely related to subcontractor participation rates in DoD contracts. The underlying presumption in the 2017 NDAA is that non-traditional contractors possess advanced technologies and capabilities that are not readily accessible in the present DoD universe of prime contractors and subcontractors—and these advanced technologies are much needed by the DoD.

One point not directly discussed earlier is the role of prime contractors in recruiting specific types of subcontractors. Since prime contractors are in active competition for large DoD contracts, they are often more likely to recognize problems faster than the DoD does. The current business model of a prime contractor is to form teams in order to deliver solutions that meet DoD needs with the prime as the integrator of systems and technologies.

---

[26] Interviews with firms that do business with the DoD and non-DoD government entities suggest some costs cannot be passed on to the commercial sector; thus, profit margins may be negatively affected.

[27] Other major factors identified by the GAO and uncovered by IDA analysts in preparing this paper were (1) the complexity of the DoD's acquisition process, (2) an unstable budget environment, (3) a lengthy contracting timeline, (4) an inexperienced DoD contracting workforce, and (5) intellectual property rights concerns. The cost of regulatory compliance was also addressed as part of U.S. government/DoD contract clauses.

A successful MDAP bidder will be held responsible for failing to deliver technological solutions for key weapon systems.

We discussed this perspective with a limited set of prime contractor personnel, and identified a general sense that the current subcontracting regimes allow sufficient flexibility to recruit necessary subcontractor technology through a variety of contractual and business arrangements. However, these contracting delivery teams often specialize in DoD work and may not be achieving the most advanced or cost-effective technological solutions—that is, they may be re-inventing technologies and capabilities that could be obtained in the commercial market.

In contrast, we received comments from DoD's Defense Innovation Unit (Experimental; DIUx)[28] that many technology firms in the commercial sector are either unaware of DoD needs or are aware of the potential for business with the DoD but refuse to contend with the obstacles identified (see footnote 27). These comments echo the fears of the Congress; the widely reported 2014 withdrawal of the Google-owned Schaft Robot from a Defense Advanced Research Projects Agency (DARPA) competition helps to support these concerns (Templeton, 2014; Hoffman, 2014). Google is reported to avoid military contracts by corporate policy. The Schaft robot was a leading design in the DARPA competition when Google purchased its parent company and then partially withdrew from the competition.

Another important economic perspective on this issue is the possibility that the DoD is no longer the dominant driving force in many of these areas of new technologies (GAO, 2017).[29] Even if the DoD is allotted more funds for R&D, due to the many obstacles to doing business with the DoD, it is not clear that certain non-traditional contractors will participate.

Government regulatory burden constitutes an additional potential cost placed on government contractors and thus demotes the DoD in contractor preferences as a customer. If these companies can identify sufficient customers among many to satisfy their total profit expectations, they will choose the customers yielding the greatest profit margin and serve others as an afterthought. In turn, competitors of these first-tier companies will pivot to service these second-tier customers. There is a sense within the NDAA that the DoD has slipped into this category for certain types of technology: that DoD technical demands at best parallel other customers such as retail consumers and may even lag this standard, thus creating a rational condition of the DoD as a second-best customer. This is a speculative hypothesis subject to future examination.

In interviewing DoD personnel, the current perception is that the barriers to doing business with the DoD—including flow-down clauses—have not yet barred access to advanced technologies and capabilities. However, there are apparent costs and delays. We were informed of instances in which DoD R&D personnel seeking access to certain

[28] Defense Innovation Unit (Experimental): https://www.diux.mil/. DIUx was established to build relationships with commercial firms that have not previously done business with the DoD in order to obtain advanced technologies that may meet DoD needs. The goal is to avoid the current acquisition cycle and rapidly prototype concepts and deliver them to the Services.

[29] DoD spending on R&D was relatively flat for the 1987–2013 period, and private sector R&D spending skyrocketed. In 2013, DoD spent $75 billion, while the aggregate spending of U.S. firms on R&D was $341 billion.

technologies had to contract via a third party (e.g., a university or not-for-profit) to obtain the technology, as certain firms do not wish to deal with U.S. government/DoD regulatory burdens. The cost then increases, as the third party takes a portion of the DoD R&D funds[30] as part of the arrangement and the DoD faces delays in receiving the technology. We understand similar situations arise when MDAP primes that need to obtain certain technologies experience additional costs and delays.[31]

In summary on this issue, we conclude that FAR/DFARS flow-downs do have an impact on firm participation in DoD contracts; however, the flow-down clauses have not yet created access problems with regard to advanced technology and capabilities. Nonetheless, flow-down clauses are yet another barrier (and an additional cost) to doing business with the DoD and tend to isolate the DoD from technologies in the commercial sector—in particular with the growth of the FAR/DFARS over the years. The situation is not expected to improve at its current trajectory and could plausibly hinder access to advanced technologies unless reforms and improvements are implemented. We deem primary research into this area to be a very high priority.

Conclusion and Recommendations

Subcontractor Participation and DoD Access to Advanced Technologies for MDAPs

Cases exist in which commercial firms have strategically chosen not to pursue DoD business. Our research suggests that the inherent regulatory burden of FAR/DFARS flow-down clauses is one of several factors influencing subcontractor participation—and eventually DoD access to advanced research and technology. Other factors include profit potential, market size, funding stability, intellectual property, and the complexity and length of the DoD acquisition process. It is also clear that the DoD is no longer the driving force for some advanced technologies associated with relatively lucrative commercial markets. Some innovative firms have chosen to pursue advanced technology and profits in these markets rather than engage in the complexities and lesser profits of DoD business—either directly as a prime or as a subcontractor. Removing unnecessary regulatory burdens—including flow-down clauses—can certainly assist in improving the DoD acquisition process, but further research is necessary.

Misapplication of Flow-Down Clauses

In order to quantify the impact of FAR/DFARS flow-down clauses, we compared five prime MDAP contracts with five corresponding subcontracts. In this sample, we found no widespread practice of burdensome flow-down misapplication. However, we did find administrative errors, measured at 1–10% of the total number of flow-down clauses. Aside from possible misapplication and error, the primary driver of flow-down clauses appears to be the ever-expanding size of the FAR/DFARS, along with prime contractors' attempts to manage their large number of subcontracts through rote standardization. As part of this larger issue, a certain number of flow-downs appear driven by defensive risk management on the part of the DoD and its prime contractors.

---

[30] The IDA team was told that fee percentages range from 5–20%.

[31] Although outside of the scope of this paper, we understand that DoD prime contractors will use various legal mechanisms such as licensing agreements to obtain technologies. There are costs and delays associated with such actions.

Recommendations

Based upon our analysis, we recommend the following:

- Conduct primary research on non-participating firms that possess technologies of interest to the DoD to understand incentives/disincentives, and propose legal and regulatory changes that may encourage their participation.
- Learn from Defense Innovation Unit (Experimental; DIUx) experiences—including statutory and regulatory changes to incorporate insights.
- Per Executive Order (EO) 13777, cull FAR/DFARS of regulations that do not directly affect the quality and performance of the acquired product in order to reduce the volume of regulations and flow-downs.
- Analyze regulations in order to quantify costs to assist in reduction of FAR/DFARS clauses.
- Restrict new regulations to those that can accelerate weapons development and production and achieve cost efficiencies.

Summary

In summary, all of the above issues are intertwined and directly affect the cost, technical abilities, and scheduled deployment of DoD MDAPs or other end products. They should be analyzed and recommendations should be implemented as part of a wider DoD acquisition reform effort.

References

Defense Innovation Unit (Experimental; DIUx). (n.d.). *Commercial solutions opening (CSO)* [White paper]. Washington, DC: Author. Retrieved from https://diux.mil/download/datasets/736/DIUx-Commercial-Solutions-Opening-White-Paper.pdf

Defense Innovation Unit (Experimental; DIUx). (2016). *Commercial solutions opening: How-to guide*. Silicon Valley, CA/Boston, MA: Author. Retrieved from https://diux.mil/download/datasets/740/CSOhowtoguide.pdf

Defense Procurement Acquisition Policy (DPAP). (2018, January 31*). Defense Federal Acquisition Regulation Supplement (DFARS) and Procedures, guidance, and information* (PGI). Retrieved from http://www.acq.osd.mil/dpap/dars/dfarspgi/current/

DoD. (1991). Acquisition regulations; Defense Federal Acquisition Regulation Supplement—Final rules and interim rules, 56 Fed. Reg. 36,280 (proposed July 31, 1991). Retrieved from https://cdn.loc.gov/service/ll/fedreg/fr056/fr056147/fr056147.pdf

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018).

Feldman, S. W. (2009) *Government contracts guidebook* (4th ed., 2008–2009). Toronto, Canada: Thomson Reuters.

G. L. Christian & Assocs. v. United States, 312 F.2d 418 (Ct. Cl. 1963).

Gansler, J. S. (2011). *Democracy's arsenal: Creating a twenty-first-century defense industry.* Cambridge, MA: MIT Press.

GAO. (2010). Defense acquisitions: *Additional guidance needed to improve visibility into the structure and management of major weapon system subcontracts* (GAO-11-61R; Draft). Washington, DC: Author. Retrieved from https://www.gao.gov/assets/690/686012.pdf

GAO. (2017). *Military acquisitions: DOD is taking steps to address challenges faced by certain companies* (GAO-17-644). Washington, DC: Author. Retrieved from http://www.gao.gov/assets/690/686012.pdf

Ginman, R. (2016). *A study of the applicability of Federal Acquisition Regulation (FAR) clauses to subcontracts under prime defense and NASA contracts*. Arlington, VA: National Defense Industrial Association. Retrieved from http://www.ndia.org/issues/acquisition-reform/order-ndia-far-flowdown-book

Hoffman, M. (2014). *Google rejects military funding in robotics.* Retrieved from https://www.defensetech.org/2014/03/25/google-rejects-military-funding-in-robotics/

Manuel, K. M., Halchin, L. E., Lunder, E. K., & Christensen, M. D. (2015). *The Federal Acquisition Regulation (FAR): Answers to frequently asked questions* (R42826). Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/misc/R42826.pdf

National Defense Authorization Acts FY 1986–FY 2017, Title VIII and legislative sections.

Templeton, G. (2014). *Google finally proves it won't pursue military contracts, pulls leading robot from DARPA competition*. Retrieved from https://www.extremetech.com/extreme/185570-google-finally-proves-it-wont-pursue-military-contacts-pulls-leading-robot-from-darpa-competition

U.S. Senate Armed Services Committee. (2017). Reforming the Defense Acquisition System. In *NDAA FY17 Bill Summary*. Retrieved from https://www.armed-services.senate.gov/imo/media/doc/FY17%20NDAA%20Bill%20Summary.pdf

Acknowledgments

The research for this paper was funded by the Office of the Secretary of Defense. We thank them for their support.

Panel 12. Cost and Pricing Considerations in Defense Procurement

Wednesday, May 9, 2018	
1:45 p.m. – 3:15 p.m.	**Chair: Elliot Branch,** Deputy Assistant Secretary of the Navy, Acquisition and Procurement ***Comparing Online B2G Marketplaces: Purchasing Agent Preferences and Price Differentials*** Capt Holland Canter, USAF, Air Force Installation Contracting Agency Capt Tabitha Gomez, USAF Lt Col Karen Landale, USAF, Air Force Installation Contracting Agency Maj William Muir, USAF, Naval Postgraduate School ***Analysis of Contract Prices: Comparing Department of Defense With Local Governments*** Latika Hartmann, Naval Postgraduate School Lt Col Karen Landale, USAF, Air Force Installation Contracting Agency Rene Rendon, Naval Postgraduate School

Elliot Branch—is the Deputy Assistant Secretary of the Navy (Acquisition and Procurement) in the Office of the Assistant Secretary of the Navy (Research, Development, and Acquisition). He is the senior career civilian responsible for acquisition and contracting policy that governs the operation of the Navy's worldwide, multibillion-dollar acquisition system. Branch is the principal civilian advisor to the Navy Acquisition Executive for acquisition and procurement matters. He serves as the Department of the Navy's Competition Advocate General, and he is the leader of the Navy's contracting, purchasing, and government property communities.

Prior to joining the Navy Acquisition Executive's staff, Branch was the first civilian Director of Contracts at the Naval Sea Systems Command. In that role, he led one of the largest and most complex procurement organizations in the federal government. As the senior civilian for contracting at NAVSEA, Branch was responsible for the contractual oversight of the nation's most complex shipbuilding and weapons systems procurement programs. His duties involved the obligation and expenditure of approximately $25 billion annually.

He is a member of the Senior Executive Service (SES). Members of the SES serve in the key positions just below the top presidential appointees. They are the major link between these appointees and the rest of the federal work force. SES members operate and oversee nearly every government activity in approximately 75 agencies.

Branch spent time in the private sector, where he specialized in acquisition and project management education, training, and consulting for the federal workforce and its associated contractors. In this role, he was responsible for the design, development, delivery, and maintenance of a wide variety of course materials on subjects ranging from project management to contract law. Branch's clients included Computer Sciences Corporation, QSS Group, BAE Systems, the Pension Benefit Guaranty Corporation, and the Departments of Defense, Energy, Justice, and State.

Prior to that, he served as the Chief Procurement Officer for the government of the District of Columbia, where he was the agency head responsible for procurement operations and policy, and for formulating legislative proposals for local and congressional consideration. Branch led a staff of over

200 employees that supported over 40 city agencies, administered a $14 million annual operating budget, and oversaw the placement of $1.5 billion annually in city contracts.

Before joining the District of Columbia's government, Branch held various positions in the SES with the Department of the Navy (DoN). In 1993, he became a member of the SES as the Director of the Shipbuilding Contracts Division at NAVSEA. He next served as Executive Director of Acquisition and Business Management for the DoN, responsible for policy and oversight of contract operations throughout the entire Navy. While in this position, he also served as Project Executive Officer of Acquisition Related Business Systems. In this role, he was responsible for the formulation and execution of a multi-year effort transforming the Navy's acquisition system from a paper-based system into one that made use of electronic technologies and methods. In this role, Branch was directly responsible for a portfolio of projects worth more than $200 million.

Branch graduated with a Bachelor of Science degree in Economics from the University of Pennsylvania Wharton School and completed the Executive Program at the University of Virginia Darden School. He has received the Navy Distinguished Civilian Service Medal, the David Packard Excellence in Acquisition Award, two Presidential Rank Awards for Meritorious Executive, the Vice Presidential Hammer Award for Reinventing Government, and the 2012 Samuel J. Heyman Service to America Medal for Management Excellence.

Comparing Online B2G Marketplaces: Purchasing Agent Preferences and Price Differentials

Capt Holland Canter[1], USAF—is a Contracting Officer at the Air Force Installation Contracting Agency, Wright-Patterson Air Force Base, Dayton, OH. [holland.canter@us.af.mil]

Capt Tabitha Gomez, USAF—is a Contract Manager of the Range and Network Division, Los Angeles Air Force Base, Los Angeles, CA. [tabitha.gomez@us.af.mil]

Lt Col Karen Landale, USAF—is the Commander of the 773d Enterprise Sourcing Squadron, Joint Base San Antonio-Lackland, San Antonio, TX. [karen.landale@us.af.mil]

Maj William Muir, USAF—is an Assistant Professor of Acquisition Management in the Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. [wamuir1@nps.edu]

Abstract

Language within the 2018 National Defense Authorization Act (NDAA) seeks to improve the federal acquisition of commercial products through agency use of commercial e-commerce portals, requiring the General Services Administration (GSA) to establish a program and enter into contracts with commercial portal providers. While the GSA has long supported agencies through its own business-to-government (B2G) portal, *Advantage*, little is currently known about how commercial portals and their associated business-to-business (B2B) or B2G marketplaces may be able to support the needs of federal agencies and their personnel who acquire commercial products. Accordingly, our research seeks to identify comparative advantages and disadvantages of the federal government's largest online B2G marketplace, *GSA Advantage*, with a leading private-sector B2B marketplace, Amazon's *Amazon Business*. We focus on the benefits and limitations of each platform for government purchase cardholders, comparing prices, shipping costs, shipping time, ease of use, and customer satisfaction, while considering future improvement initiatives. Our findings highlight several benefits, limitations, and risks of each platform for repetitive, purchase card–based transactions.

Introduction

Section 846 of the Fiscal Year (FY) 2018 National Defense Authorization Act (NDAA), *Procurement Through Commercial E-Commerce Portals*, requires the General Services Administration (GSA) to contract with private-sector marketplaces to satisfy government demand for commercial off-the-shelf (COTS) products. The intent of this program is to improve alignment between public-sector practices for the acquisition of COTS products and those of the private sector (GSA, 2018), while providing for enhanced competition, faster purchasing, improved government insight into the supply market, and reasonableness in prices paid by the government (Pub. L. No. 115-91). Prior to its passage, the provision gained the title "The Amazon Amendment" (Miller, 2017) within the popular press, a label indicative of recent efforts by proprietors of private-sector online marketplaces such as the Amazon.com corporation to further capture federal demand within the nearly

[1] Corresponding author

$53 billion spent annually by federal agencies on commercial items (The Coalition for Government Procurement, 2017). Many of these online marketplaces, including Amazon's *Amazon Business* B2B marketplace, are well-suited for use by industrial purchasers because of the business-oriented functionality offered, such as online requests for proposals, reverse auctioning, graduating pricing, and access to a wide range of goods and services. These marketplaces may also be able to readily adapt to idiosyncratic regulatory requirements faced by federal purchasers (e.g., Buy American Act, Javits–Wagner–O'Day Act, vendor exclusion) as well as needs and objectives specific to federal organizations, such as those related to access, transparency, supply chain security, and socioeconomic participation.

One area of federal spend that is particularly attractive to these marketplace proprietors is the nearly $19 billion in purchases that occur annually for commercial items under the Government Purchase Card (GPC) program (GAO, 2016). In recent years, agency GPC purchase requirements have increasingly been made through GSA's own *GSA Advantage* platform and under GSA Multiple Award Schedule (MAS) contracts (where annual obligations total more than $30 billion; GAO, 2015). The *GSA Advantage* purchasing platform has been designed around the unique requirements that exist for a B2G marketplace, and the underlying schedule contracts provide terms and conditions that protect the interests of federal buyers as well as those of the contracted, private-sector vendors. The platform also provides access to several strategic sourcing contract vehicles such as those under the Federal Strategic Sourcing Initiative's programs for office products and maintenance, repair, and overhaul supplies. Despite the numerous benefits, *GSA Advantage* and the MAS contracts continue to lose regulatory ground. For instance, prior to 2014 (FAC 2005-72-1), federal supply schedules were prioritized ahead of commercial sources in the Federal Acquisition Regulation (FAR), but are now placed in a "non-mandatory" status, on-par with commercial sources. Section 846 of the 2018 NDAA may further shift a large portion of the remaining spend away from *Advantage* and toward private-sector sellers via online marketplaces. Thus, it is important to gain a better understanding into differences (and commonalities) that exist between government-operated and industry-operated marketplaces and how federal purchasers view these platforms. To do so, we examine two marketplaces, *GSA Advantage* and *Amazon Business*, and focus our research on use under the federal GPC Program given its gravity and importance to both sectors. This report provides a brief overview of our research and findings.

As with any research, there were limitations. Due to time constraints, we limited our research to only Air Force historical GPC data and surveyed only Air Force members affiliated with the GPC. While the results contained in this report are specific to the Air Force, we have no reason to believe they are not generalizable to the entire federal government. Also, due to time constraints, we were limited in the number of exact item price comparisons that we could perform. We focused on comparing prices of 60 commercially available items (i.e., not military specific, which would bias results toward *GSA Advantage*) most frequently purchased by the Air Force in FY 2015. Further, while supply chain issues and legal concerns are relevant and prevalent, these types of risks are outside the scope of this study and were not addressed. Examples of these risks include brand protection, supply chain integrity, counterfeit items, product tampering, cardholder and supplier security, and Berry Amendment concerns.

Research Methodology

We adopt a multi-method approach to gain an understanding of *GSA Advantage's* and *Amazon Business's* relative positions in the overall 'market of online marketplaces' for the government buyer. First, we gathered qualitative information from interviews with management of *GSA Advantage*. Next, we corroborate and extend findings from these interviews using survey data collected from 428 Air Force members affiliated with the GPC (e.g., cardholders, approving officials). Lastly, we tested for price differentials between *GSA Advantage* and *Amazon Business* using a market basket of products developed from Air Force spend data.

Interviews

We sought to interview leaders from both the GSA and Amazon. We developed similar interview questions for both platforms. As *GSA Advantage* is listed in the FAR as a supply source and *Amazon Business* is not, we tailored our questions for each entity. The interview questions asked about current goals for the respective marketplace, customer service, policies, and continuous improvement processes. The questions also explored small business processes and the potential to achieve greater insight into federal GPC transactions. Interview requests were forwarded, along with interview questions, to each company's point of contact. While we were unable to arrange for an interview with Amazon, GSA agreed to participate.

Survey

We surveyed Air Force members affiliated with the GPC to better understand current platform use (for GSA and Amazon.com platforms) and trends as well as individual preferences and demographics. For each platform that the respondent had experience with, we obtained respondent ratings on product search, pricing, shipping, and return policies. Respondents were also asked to rate their level of satisfaction with the platform and indicate preference, if any, between platforms. A pre-test was used to further develop and refine the survey instrument with a pool of GPC subject matter experts. To ensure a sufficient level of response (i.e., to obtain a suitable level of statistical power for subsequent analysis), we coordinated distribution of the survey through the Air Force Installation Contracting Agency and to Air Force Level-Three Agency/Organization Program Coordinators (A/OPCs), who distributed the survey to base-level A/OPCs. Installation-level A/OPCs forwarded the survey to individual cardholders within their area of responsibility. The survey was deployed in a single wave, and 428 responses were received from a pool of 24,610 potential respondents, representing 1.74% of the total cardholder population.

Comparative Price Differentials

A price comparison was conducted between *GSA Advantage* and *Amazon Business* using a market basket approach. The market basket was developed using GPC data from FY 2015, consisting of 1,048,575 line items, obtained from the Air Force Installation Contracting Agency. Top categories of spend were (1) computers, computer peripheral equipment and software; (2) medical, dental, ophthalmic, hospital equipment and supplies; (3) industrial supplies; (4) stationary, office supplies, printing and writing paper; and (5) business services. Unfortunately, product and service identifiers were sparse and inconsistent within the dataset. A text frequency analysis was performed to develop the market basket, identifying the 60 most commonly purchased items based on textual information and percentage of spend on that item within the overall dataset. We excluded items that were commercial yet specific enough to the military that the items would not be commonly sold in a private-sector marketplace from the basket.

After identifying the 60-item market basket, we proceeded to collect item-level prices from each platform. Due to the abbreviated nature of the text descriptions in the GPC dataset, matching exact items on both platforms proved difficult. We found it most efficient to search from *GSA Advantage* first, and then match the item on *Amazon Business*. However, regardless of which platform we first searched for an item, the nature of the search results did not change. After matching the exact item, we documented purchase data from the five lowest cost vendors of that item. However, not every item could be matched to five vendors on both platforms. In those cases, we collected data from all available vendors. We captured the following details from each platform: (1) item description; (2) manufacturer part number; (3) vendor name; (4) price; (5) socioeconomic designation, if any; (6) vendor rating, if any; (7) quantity discount; if any; (8) shipping days; (9) pack-size; (10) product origin; (11) fulfillment source; (12) minimum purchase requirement, if any. These item details were compared across marketplaces.

Research Results

Interview

Our interview with the GSA occurred in August 2017. The interviewee worked for the organization for 15 years and was a self-described "technologist by heart." The interviewee is now a division director for 27 systems and 70 brick-and-mortar stores across the United States. The interviewee provided insight into *GSA Advantage* during two one-hour interview sessions, which occurred via teleconference. After the interviews, the interviewee also emailed several written responses. We summarize this interview next.

Current State and Goals

The interview began with a discussion on the current state and goals of *GSA Advantage*. The interviewee stated that the primary goal for *GSA Advantage* is to "provide a government marketplace that is compliant with federal, military, and state and local government rules and regulations to deliver quality products and services to government buyers and to promote fair and equal competition between suppliers." The interviewee explained that the GSA was under new leadership and explained that the new leadership is "setting the new bar or resetting the new baseline as to where they want to take their business and how commodities and services will play a role in that."

The interviewee was aware of current legislation involving transformation of federal purchasing, to include the (then-proposed) "Amazon Amendment." The interviewee stated, "GSA has been performing their own study and analysis of (the) government marketplace." The interviewee explained that the GSA has been going through system consolidation and streamlining its processes. The interviewee stated that several modernization tracks for *GSA Advantage* have been created. The modernization effort includes tracks such as "sign on, registration, user management, and the shopping cart experience, and all the capabilities around that."

The interviewee also explained that *GSA Advantage* was just one of many systems managed by the Federal Acquisition Service organization. The interviewee stated, "GSA provides a vast array of offerings and many diverse methods and technologies for acquiring these offerings, passing the savings, knowledge, and compliance onto all of government." The interviewee spoke highly of GSA eBuy, which is another system of capabilities under the *GSA Advantage* umbrella. eBuy draws in about $11 billion in awards each year. It allows users to build Requests for Quotes and Requests for Proposals, and connects with the vendors who hold GSA contracts.

Shadow of E-Commerce

Currently, the GSA is facing challenges in the shadow of e-commerce. The interviewee stated, "We are not private industry. We will never be Amazon." The interviewee emphasized that the GSA was a government organization. The interviewee said,

> GSA's purpose is to provide as much current information on catalogs and contracts to assist consumers and suppliers to do market research, not just for price comparison, but also to identify and support socioeconomic programs, environmentally friendly products, and mandatory or preferred sources of supply for the government.

The interviewee stated, "The biggest issues and challenges with meeting and exceeding customer expectations would be policy and compliance within the government and existing terms and conditions in the contracts." The rules and regulations that the GSA is bound to creates an atmosphere where the GSA is unable to provide the level of customer service available on commercial platforms. The interviewee said, "For example, *Advantage* cannot provide vendor ratings, and is very limited as to what products can be promoted on the site. The system follows the terms and conditions stated in the contract, which limits capabilities for upselling, and influencing a purchase." The GSA believes vendor ratings promote one vendor over the other, which is not allowed due to government policy and rules and regulations. The interviewee did say that they would love for the company to provide vendor ratings in the future.

Small Business Goals

The GSA has aggressive small business goals and assists ordering activities in achieving or exceeding their goals. The interviewee stated that within the GSA, "Approximately 80 percent of all GSA contractors are small businesses." The interviewee explained that all socioeconomic items and services are identified on the site in a way that is clearly visible to buyers, and all transactions are captured.

Minimum Order Requirements

While the interviewee explained that *GSA Advantage* was geared "towards the smaller commodity buys," GPC holders have voiced concerns regarding minimum order requirements. The interviewee agreed the search results within *GSA Advantage* are skewed because of the minimum order requirements—minimum order requirements are not taken into account when displaying what appears to be the lowest priced item. The interviewee agreed that disparity makes it difficult to accurately compare prices.

The interviewee stated that *GSA Advantage* is currently developing a prototype to make the user interface filter search results and incorporate minimum order requirements into the displayed filtered search results. The interviewee also emphasized that all the minimum order requirements and price discounts are per the terms and conditions of the contract previously established under the MAS program. The interviewee said, "*Advantage* shows what a vendor provides based on the terms and conditions of their contract." The interviewee explained that to achieve a lower price per unit, the vendors claim that they must have a minimum dollar amount to break even. However, the interviewee stated that you could find the items at a lower cost per unit on websites like Amazon. The interviewee stated in the future, vendors should provide the government with wholesale prices, which would provide lower prices.

Level Three Data/Transactional Level Data

Level three data and transactional level data, which require the capture of specific line item data (e.g., merchant name, address, invoice number, and other line item details), are necessary for agencies to accurately understand and assess their GPC transactions. The interviewee stated that the GSA captures everything and produces an analytics report. When asked if *GSA Advantage* required participating vendors to provide level three data, the interviewee stated, "Yes, all of this data comes in during contract or catalog submission. It is then matched when items are purchased on the site, so all level three transactional data is captured."

While the GSA provides level three data, vendors are not required to provide transactional level data; these vendors frequently claim it is too expensive. However, we learned that the catalog data is not linked to vendor data, so it does not provide the complete picture of each transaction.

Survey

The survey began by collecting sample data. The 428 GPC holder respondents were nearly evenly split on gender, with 57% responding male and 41% responding female (2% declined to answer). Of the total, 35% were in the 51–60 age group, 21% were in the 41–50 age group, 19% were in the 31–40 age group, 12% were in the 18–30 age group, and 13% over 61 years of age. We also collected data on grade (rank) of the respondents. Thirty-five percent of the respondents were civilian employees in the grade of GS-7 to GS-9; 20% were civilian GS-12 to GS-13; 20% were civilian GS-10 to GS-11; 11% were military E-5 to E6; 6% were military E-7 to E-9; 3% were military E-1 to E-4; 3% were military O-1 to O-3; 1% was O-4 to O-6; and 1% was GS 14+. Regarding experience in the GPC program, 46% of the respondents had over five years of experience, 20% had one to two years of experience, 18% had three to five years of experience, and 16% had less than one year of experience.

GSA Advantage

The survey asked cardholders about their experience with the two platforms. Of the 91% of respondents who had used a GPC to purchase from *GSA Advantage*, 42% had purchased from *GSA Advantage* more than 10 times, 31% had purchased from *GSA Advantage* two to five times, 21% had purchased from *GSA Advantage* six to 10 times, and 6% had only purchased from *GSA Advantage* once. Compared to other online ordering platforms, 46% of respondents said that *GSA Advantage's* website was more difficult to use, 44% said it was similar to use, and 10% said it was easier to use. Regarding search, 48% of respondents said that *GSA Advantage's* search engine results page was less comprehensive than other online ordering websites, while 44% said it was similar, and 8% said it was more comprehensive. A majority (54%) of respondents stated that *GSA Advantage's* shipping policies were similar to other online ordering websites. When asked about return policies, 26% of respondents stated that *GSA Advantage's* return policies were similar to other online ordering websites; however, 58% had never attempted to return a product purchased through *GSA Advantage*. Regarding pricing, 51% of respondents stated that *GSA Advantage* was more expensive than other online ordering websites, 39% said prices were similar, and 10% said *GSA Advantage* was less expensive. Forty-six percent stated that finding the lowest price on *GSA Advantage* was similar to other online ordering websites, 41% stated that finding the lowest price was more difficult, and 13% said that it was easier. A majority (62%) of respondents never sought additional discounts or rebates when purchasing from *GSA Advantage*. However, when respondents asked for a discount, 72% stated they "sometimes" receive it, 16% said they "never" receive it, and 12% said they received it "most of the time." Finally, when asked to rate their level of customer satisfaction with *GSA Advantage*, the participants' ratings varied widely between very dissatisfied and somewhat satisfied; 28% of the respondents were somewhat satisfied, 25% were neither satisfied nor dissatisfied, 21% were somewhat dissatisfied, 13% were very satisfied, and 13% were very dissatisfied.

Amazon Business

Of the 428 respondents, only 77 (18%) had conducted a GPC transaction on *Amazon Business*. Of those that had placed a purchase using the platform, 45% had purchased from *Amazon Business* two to five times, 26% had purchased more than 10 times, 20% had purchased six to 10 times, and 9% had purchased only once. Compared to other online ordering platforms, 68% of respondents said that *Amazon Business's* website was easier to use, 30% said it was similar to use, and 2% said it was more difficult. Regarding search, 58% of respondents said that *Amazon Business's* search engine results page was more comprehensive than other online ordering websites, 37% said it was similar, and 5% said it was less comprehensive. A majority of respondents (57%) said *Amazon Business's* shipping policies were better, while 41% said they were similar, and 2% said they were worse. *Amazon Business's* return policies were rated as better than other online ordering websites by 34% of respondents, 21% said they were similar, and 5% said they were worse; however, a full 40% had not completed a return. No respondents felt that *Amazon Business's* prices were higher than other online marketplaces; 55% of respondents stated that, in their experience, *Amazon Business* was less expensive, while 42% said the platform's prices were similar to those offered on other online platforms. A majority of respondents (57%) stated that finding the lowest price on *Amazon Business* was easier than it was on other online ordering websites, while 42% stated that it was similar to other online ordering websites, and 1% said it was more difficult. When asked if they sought additional discounts or rebates from vendors on *Amazon Business*, 76% of respondents responded that they had never sought additional discounts or rebates when placing purchases on the platform, 14% responded they had sometimes asked for an additional discount, 6% responded that they often ask for a discount, and 4% responded that they always ask for a discount. However, when respondents asked for a discount, 57% stated they "sometimes" receive it, 27% said they receive it "most of the time," and 16% said they "never" receive it. Finally, when asked to rate their level of customer satisfaction with *Amazon Business*, 61% of respondents were very satisfied, 20% said somewhat satisfied, 15% were neither satisfied nor dissatisfied, 2% were somewhat dissatisfied, and 2% were very dissatisfied.

Amazon.com

While industrial purchasing differs in several ways from personal shopping, we asked respondents to provide information on their personal interactions with Amazon.com's retail platform (Amazon) as these personal experiences may frame expectations and preferences in a professional context. The vast majority of respondents (394 respondents; 92%) had purchased from Amazon in a personal capacity. Of these, 58% had purchased from Amazon more than 10 times, 22% purchased two to five times, 16% purchased six to 10 times, and 4% had purchased only once. Compared to other online ordering platforms, 70% of respondents said that Amazon was easier to use, 29% said it was similar to use, and 1% said it was more difficult to use. Regarding search, 60% of respondents said that Amazon's search engine results page was more comprehensive than other online ordering websites, 36% said it was similar, and 4% said it was less comprehensive. A majority (64%) of 394 respondents who had purchased from Amazon stated that shipping policies were better than other online ordering websites, while 35% said policies were similar, and 1% said policies were worse. We also asked about experiences with returning products to Amazon; 51% stated that Amazon's policies for returns were better than other online ordering websites, while 23% said policies were similar, 2% said policies were worse, and 24% had never returned a product purchased from Amazon. When asked about product pricing, 64% of respondents stated that prices on Amazon were lower than other online ordering websites, while 35% responded that prices were similar, and 1% responded that prices on Amazon were more expensive than comparable sites. Regarding ease of locating lowest prices for products, 60% of respondents stated that finding the lowest price on Amazon was easier than other online ordering websites, 37% stated that it was similar to other online ordering websites, and 3% said it was more difficult. We also asked about propensity to seek discounts; 67% of respondents never sought additional discounts or rebates when purchasing from Amazon, 22% sometimes sought additional discounts, 6% always sought additional discounts, and 5% often sought additional discounts. However, when respondents asked for a discount, 70% stated they "sometimes" receive it, 23% receive the discount "most of the time," and 7% "never" receive it. Finally, when asked to rate their level of customer satisfaction with Amazon, a majority (66%) of respondents were very satisfied, 23% were somewhat satisfied, 8% were neither satisfied nor dissatisfied, and 2% were very dissatisfied.

Online Product Reviews

The volume and valence of online customer reviews (i.e., product reviews, vendor ratings) influence buyer expectations and buyer preferences (Wu et al., 2015) as well as retailer sales (Floyd et al., 2014). In B2B settings, online customer reviews also influence the behavior of industrial purchasers (i.e., within online B2B marketplaces), where purchasers frequently reconcile internally-generated reviews with external review information (Steward, Narus, & Roehm, 2017). However, little is known about how public-sector purchasers view online customer reviews and how they incorporate these reviews into their industrial purchasing decisions. To investigate these issues, we asked respondents a series of questions about (1) online vendor ratings and (2) online product reviews. When asked about online vendor ratings, a majority (84%) of respondents indicated that vendor ratings are important, and within that group, another 84% said they factored these ratings into their purchasing decisions. Respondents also indicated that online product reviews were important. Of the total respondents, 92% stated that product reviews were important to them, and of these, 91% responded that they made purchasing decisions based on product reviews.

Marketplace Preference

We asked respondents if they would prefer to place GPC purchases at *GSA Advantage* or at *Amazon Business*, given the choice. More than three-quarters (78%) of respondents indicated that they would prefer *Amazon Business*. To better understand this result, we examined the effects of influential factors on GPC holders' online platform preference within a generalized linear model using a logit link function, a logistic regression. Online platform preference was a binary response measured by self-report on a cardholder's preference to order from Amazon (Amazon.com or *Amazon Business*) instead of *GSA Advantage*, given the opportunity. Thus, we subset our sample to cardholders who are current users of *GSA Advantage*—those who reported placing at least one purchase annually through *GSA Advantage*—and who have experience placing GPC purchases through either Amazon.com or *Amazon Business*. This filtering procedure resulted in a sample of 360 respondents. We are unable to conclude, based on the results of chi-square testing of distributions from cardholder demographics, that respondents from this subset otherwise differ significantly from those in the larger random sample.

Regressors in the model accounted for cardholders' perceptions of *GSA Advantage's* price and quality competitiveness. Price competitiveness was based on a comparative price assessment against other online marketplaces. Quality competitiveness captured the following dimensions of site quality: (1) overall ease of site use; (2) ease of locating lowest item pricing; (3) comprehensiveness of site search and (4) adequacy of logistics (shipping and returns) policies. For all regressors, competitiveness was measured as a comparative assessment of *GSA Advantage* against other online marketplaces. Given Amazon.com's prominence as an online marketplace and likelihood for cardholders to anchor their comparisons against an Amazon marketplace, we reduce our exposure to multicollinearity by including in our model only competitiveness assessments of *GSA Advantage*. All regressors were measured using single-item, categorical scales.

To control for potential confounding effects, we also included several covariates in the model. These factors included respondent (1) gender, (2) age, (3) years of experience as a cardholder, (4) frequency of *GSA Advantage* use, (5) propensity to request price discounts, and (6) overall customer satisfaction with *GSA Advantage*. Customer satisfaction was assessed on a five-point Likert-type scale and was treated in the model as a continuous variable. Similarly, age intervals were treated as continuous. All other covariates were included in the model as categorical factors.

Model estimation was performed in R (R Core Team, 2017) using maximum likelihood estimation. The model, Model 1, offered improved fit to the data over a null model and correctly predicted the preference for 88.33% of cardholders (32 of 65 who prefer *GSA Advantage* and 286 of 295 who prefer Amazon). In an effort to produce a parsimonious model of cardholder preference, we utilized iterative backward selection (see Table 1) to identify potential factors for exclusion. Factor contribution to model fit was assessed by chi-square change and Akaike Information Criterion (AIC).

Table 1. Single-Term Deletions for the Full and Parsimonious Model

Variable	*df*	Model 1 Dev.	Model 1 LR	Model 2 Dev.	Model 2 LR
Null		207.63		223.58	
Cardholder Gender	2	208.42	0.78		
Cardholder Age	1	211.40	3.77†	229.70	6.13*
Cardholder GPC Experience	3	209.71	2.08		
Frequency of GSA Advantage Use	3	211.30	3.67		
Customer Sat. w/ GSA Advantage	1	214.63	7.00**	234.18	10.60**
Website Ease of Use	2	217.64	10.01**	241.15	17.58**
Search Comprehensiveness	2	210.47	2.84		
Shipping Policy Adequacy	2	210.11	2.48		
Returns Policy Adequacy	3	214.75	7.12†	230.41	6.83†
Pricing Competitiveness	2	215.74	8.11*	234.76	11.19**
Ease of Locating Lowest Price	2	210.31	2.68		
Propensity to Seek Discounts	3	208.87	2.24		

Note: $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .10$; AIC = 261.63 (Model 1), 243.58 (Model 2).

Based on these assessments, gender, GPC program experience, the frequency of *GSA Advantage* use, propensity to seek discounts, search comprehensiveness, shipping policy, and ease of locating lowest price were excluded from the model. The removal of these factors did not result in a significant reduction to model fit. The parsimonious model, Model 2, correctly predicted preference for 85% of cardholders (24 of 65 who prefer *GSA Advantage* and 282 of 295 who prefer Amazon). Beta coefficients in the table represent the estimated (conditional) change in log-odds of a cardholder preferring Amazon over *GSA Advantage* when a regressor is changed by one unit. Exponentiated coefficients are presented within the text. Logistic regression results are listed in Table 2.

Table 2. Logistic Regression Results

Variable	OR[§]	β	SE	Pr(> |z|)
(Intercept)	4.90	1.59	.43	<.01**
Cardholder Age[‡]	.97	−.03	.01	.02*
Customer Sat. w/ GSA Advantage[‡]	.53	−.64	.20	<.01**
Website Ease of Use				
Easier	.22	−1.50	.46	<.01**
Similar		*Referent Group*		
More Difficult	2.77	1.02	.53	.05[†]
Returns Policy Adequacy				
Better	.08	−2.57	1.23	.04*
Not Sure	.83	−.19	.39	.64
Similar		*Referent Group*		
Worse	1.50	.40	.69	.56
Pricing Competitiveness				
Less Expensive	.61	−.49	.50	.33
Similar		*Referent Group*		
More Expensive	2.89	1.06	.40	<.01**

Note: $^{**}p < .01$; $^{*}p < .05$; $^{\dagger}p < .10$; [‡]mean-centered; [§]OR (Odds Ratio) = e^{β}.

Cardholder Age and Customer Satisfaction

A cardholder's odds of preferring *Amazon Business* to *GSA Advantage* decrease by 3.41% for each additional year group (β = −.03, se = .01, p = .02). Similarly, a cardholder's odds of preferring Amazon decrease by 47.32% with each one-unit increase in their self-reported level of satisfaction with *GSA Advantage* (β = −.64, se = .20, p < .01).

Website Ease of Use

For the categorical regressor, *website ease of use*, we selected "similar" as our referent category. When cardholders perceive *GSA Advantage* to be easier to use (comparatively to other online ordering sites), their odds of preferring Amazon to *GSA Advantage* decrease by 77.67% (β = −1.50, se = .46, p < .01). Alternatively, when cardholders perceive GSA Advantage to be more difficult to use, their odds of preferring Amazon increase by 177.39%. However, this difference is borderline in statistical significance (β = 1.02, se = .53, p = .05)

Return Policy Adequacy

For *return policy adequacy*, we again selected "similar" as our referent category. Cardholder odds of preferring Amazon over *GSA Advantage* only differ (from the referent category) for those cardholders who perceive *GSA Advantage's* return policies to be better in comparison to policies of other online order sites. For these cardholders, the odds of preferring Amazon decrease by 92.35% (β = −2.57, se=1.23, p = .04).

Price Competitiveness

For *price competitiveness*, we again used "similar" as our referent category. Our data does not suggest that cardholders who view *GSA Advantage's* pricing as being less expensive (in comparison to other online ordering sites) are more or less likely to prefer Amazon to *GSA Advantage* (β = −0.49, se = .50, p = .33) than cardholders who feel that *GSA Advantage's* pricing is similar to other online ordering sites. However, when cardholders view *GSA Advantage's* pricing as being more expensive, their odds of preferring Amazon increase by 188.64% (β = 1.06, se = .40, p < .01).

Between-Marketplace Price Comparison

We found that, of the vendors that offered the 60 compared items, prices on *GSA Advantage* were lower than *Amazon Business* 80% of the time (241 times out of 300). However, every *GSA Advantage* item had a minimum order requirement. In contrast, *Amazon Business* did not have a minimum order requirement for any of the items that we examined. The tables in this section show different observations about the data. For each observation, we display a subset of the 60 items that were compared, abbreviating the full results for brevity.

Lowest and Highest Prices

Table 3 compares the prices of the items between each platform. The table highlights which platform offered the lowest price and which platform had the highest price for each item. At times, *GSA Advantage* offered the lowest and highest price for the same item; other times, *Amazon Business* offered the lowest and highest price for the same item.

Table 3. Lowest and Highest Prices

Item	Item Description	Average Price	
		GSA	AB
1	Smead Mead Heavyweight 2-Pocket Portfolio	$18.30	$27.56
2	Boise Polaris Premium Multipurpose Paper	$39.25	$47.58
3	7510012360059 Document Protector	$5.32	$7.33
4	Skilcraft Gregg Ruled Steno Book	$8.29	$22.14
5	7530 Notebook, Steno	$16.49	$26.81
6	Double Pocket Portfolio, Letter Size, Dk Blue	$12.42	$13.56
7	Mechanix Wear MP3-F55-010 TAA Compliant	$40.27	$85.63
8	Energizer Industrial Alkaline Batteries, AA	$3.31	$11.49
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	$9.02	$15.32
10	Skilcraft Dry-Erase Markers	$6.06	$10.13
11	G2 Fashion Collection Gel Roller	$6.14	$16.50
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	$9.46	$29.85
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	$8.12	$10.93
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	$6.64	$15.96
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	$7.41	$28.23

Bulk and Quantity Discounts

Table 4 shows a sample of the bulk/quantity discounts offered by each platform. *GSA Advantage* listed quantity discounts by schedule, while individual vendors offered quantity discounts on *Amazon Business*. Of the 60 items, only six *GSA Advantage* schedules offered a quantity discount. On *Amazon Business*, only seven vendors offered quantity discounts. However, while *Amazon Business* offered discounts with lower minimum quantities, *Amazon Business's* prices were still higher than *GSA Advantage's*—even with the discount applied. Both platforms had a quantity discount on item two and item nine, as shown in Table 4.

Table 4. Bulk/Quantity Discounts

Item	Item Description	Quantity Required & Discount			
		GSA		Amazon Business	
1	Smead Mead Heavyweight 2-Pocket Portfolio	-	-	-	-
2	Boise Polaris Premium Multipurpose Paper	3,000 - 4,999 5,000 - 9,999 10,000+	2.00% 3.00% 5.00%	4+	0.98%
3	7510012360059 Document Protector	-	-	-	-
4	Skilcraft Gregg Ruled Steno Book	-	-	-	-
5	7530 Notebook, Steno	-	-	-	-
6	Double Pocket Portfolio, Letter Size, Dk Blue	-	-	-	-
7	Mechanix Wear MP3-F55-010 TAA Compliant	25,000 - 99,999,999	2.00%	-	-
8	Energizer Industrial Alkaline Batteries, AA	-	-	-	-
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	20,001 - 99,999,999	1.00%	4+	0.92%
10	Skilcraft Dry-Erase Markers	-	-	-	-
11	G2 Fashion Collection Gel Roller	-	-	-	-
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	-	-	-	-
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	-	-	-	-
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	-	-	-	-
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	-	-	-	-

Shipping

Table 5 displays *Amazon Business's* shipping time and cost, and Table 6 displays *GSA Advantage's* shipping time and cost. On *Amazon Business*, the average shipping time was 9.25 days, and the average shipping cost was $2.33. For *GSA Advantage*, the average shipping time was 5.45 days, and shipping was free.

Table 5. Amazon Business Shipping

Item	Item Description	Shipping Time (Days)			Shipping Cost
		Min	Max	Avg	Avg
1	Smead Mead Heavyweight 2-Pocket Portfolio	7	14	8.50	$1.19
2	Boise Polaris Premium Multipurpose Paper	3	14	7.60	$0.00
3	7510012360059 Document Protector	2	17	6.20	$1.98
4	Skilcraft Gregg Ruled Steno Book	6	9	7.50	$8.61
5	7530 Notebook, Steno	2	23	9.10	$4.02
6	Double Pocket Portfolio, Letter Size, Dk Blue	6	16	8.38	$4.78
7	Mechanix Wear MP3-F55-010 TAA Compliant	3	14	7.33	$3.64
8	Energizer Industrial Alkaline Batteries, AA	2	14	7.60	$0.00
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	5	14	8.90	$2.09
10	Skilcraft Dry-Erase Markers	5	23	11.50	$1.50
11	G2 Fashion Collection Gel Roller	5	26	11.60	$2.16
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	6	14	9.00	$6.61
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	2	15	6.20	$0.00
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	6	12	8.50	$4.98
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	2	14	7.20	$6.35

Table 6. GSA Advantage Shipping

Item	Item Description	Shipping Time (Days)			Shipping Cost
		Min	Max	Avg	Avg
1	Smead Mead Heavyweight 2-Pocket Portfolio	1	4	2.2	$0.00
2	Boise Polaris Premium Multipurpose Paper	2	5	3.2	$0.00
3	7510012360059 Document Protector	2	7	4.2	$0.00
4	Skilcraft Gregg Ruled Steno Book	1	7	2.8	$0.00
5	7530 Notebook, Steno	1	7	3.6	$0.00
6	Double Pocket Portfolio, Letter Size, Dk Blue	3	7	4.4	$0.00
7	Mechanix Wear MP3-F55-010 TAA Compliant	3	45	14.2	$0.00
8	Energizer Industrial Alkaline Batteries, AA	1	5	3	$0.00
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	1	5	2.6	$0.00
10	Skilcraft Dry-Erase Markers	2	14	5.2	$0.00
11	G2 Fashion Collection Gel Roller	1	5	2.6	$0.00
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	3	7	4.8	$0.00
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	1	4	2	$0.00
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	2	5	3.6	$0.00
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	3	14	6.8	$0.00

Small Business Representation

On *GSA Advantage*, every small business category was represented, and of the 60 items we researched, every item was offered by a socioeconomic business. On *Amazon Business,* most of the small business categories were represented. However, only 35%, (21 of the 60 items) were offered by a small business. Table 7 displays small business category representation of our researched items.

Table 7. Small Business Representation

		GSA	AB
s	Small Business	X	X
o	Other than Small Business	X	
w	Woman Owned Business	X	X
wo	Women Owned Small Business (WOSB)	X	X
ew	Economically Disadvantaged Women Owned Small Business (EDWOSB)	X	
v	Veteran Owned Small Business	X	X
mo	Minority Owned		X
dv	Service Disabled Veteran Owned Small Business	X	
d	SBA Certified Small Disadvantaged Business	X	X
8a	SBA Certified 8(a) Firm	X	X
h	SBA Certified HUBZone Firm	X	X

Vendor Ratings

GSA Advantage does not provide vendor ratings. On *Amazon Business*, the average vendor rating was 93% (out of a possible 100%). Table 8 shows the average vendor ratings for 15 of the 60 researched items.

Table 8. Average Vendor Rating

Item	Item Description	Avg Vendor Rating
1	Smead Mead Heavyweight 2-Pocket Portfolio	88.50%
2	Boise Polaris Premium Multipurpose Paper	90.20%
3	7510012360059 Document Protector	94.20%
4	Skilcraft Gregg Ruled Steno Book	95.00%
5	7530 Notebook, Steno	93.20%
6	Double Pocket Portfolio, Letter Size, Dk Blue	87.00%
7	Mechanix Wear MP3-F55-010 TAA Compliant	99.67%
8	Energizer Industrial Alkaline Batteries, AA	97.00%
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	91.20%
10	Skilcraft Dry-Erase Markers	93.50%
11	G2 Fashion Collection Gel Roller	91.60%
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	88.00%
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	97.00%
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	95.00%
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	94.00%

Product Origin

On *GSA Advantage*, all the products originated from the United States. However, on *Amazon Business*, some of the products originated from another country, or the origin was unidentified. Table 9 shows a sample of the product origins of five of the 60 researched items on *Amazon Business*.

Table 9. Amazon Business Product Origin

Item	Item Description	Supplier	Product Origin
1	Smead Mead Heavyweight 2-Pocket Portfolio	My Office Innovations MYO	USA
		Blue Cow Office Product	USA
		Shoplet	USA
		ReStockIt	USA
		AMAZON.COM	USA
2	Boise Polaris Premium Multipurpose Paper	Office Depot, Inc.	USA
		My Office Innovations	USA
		Shoplet	USA
		Bison Office	USA
		Clean It Supply	USA
3	7510012360059 Document Protector	Queenkim98	USA
		Bargain Bosses LLC	USA
		L Palms LLP	Unknown
		Acedepot	USA
		GTN Office Basics	USA
4	Skilcraft Gregg Ruled Steno Book	Corgi Lamps	China
5	7530 Notebook, Steno	Brian Delrosario	USA
		My Office Innovations	USA
		Rock Shop Central	USA
		Independence Fulfillment Services	USA
		Alliance (SUPPLY)	USA

Minimum Order Requirements

Amazon Business did not have any stated minimum order requirements, but every item we researched on *GSA Advantage* had a minimum order requirement. The minimum dollar amounts are dictated by the schedules. We codified the lowest minimum order requirement, the highest minimum order requirement, and the average minimum order requirements. Table 10 displays a sample of the minimum order requirements for 15 of the 60 items.

Table 10. GSA Advantage Minimum Order Requirements (MORs)

Item	Item Description	Min	Lowest MOR	Highest MOR	Average MOR
1	Smead Mead Heavyweight 2-Pocket Portfolio	$50.00	$50.00	$100.00	$80.00
2	Boise Polaris Premium Multipurpose Paper	$120.00	$100.00	$120.00	$106.00
3	7510012360059 Document Protector	$100.00	$50.00	$100.00	$80.00
4	Skilcraft Gregg Ruled Steno Book	$28.75	$25.00	$100.00	$55.75
5	7530 Notebook, Steno	$28.75	$25.00	$28.75	$25.75
6	Double Pocket Portfolio, Letter Size, Dk Blue	$100.00	$50.00	$100.00	$80.00
7	Mechanix Wear MP3-F55-010 TAA Compliant	N/A	$1.00	$25.00	$19.00
8	Energizer Industrial Alkaline Batteries, AA	$50.00	$25.00	$100.00	$65.00
9	Wilson Jones Basic Round-Ring View Binder Plus Pack, 1" Cap, White	$100.00	$50.00	$100.00	$80.00
10	Skilcraft Dry-Erase Markers	$50.00	$25.00	$100.00	$55.00
11	G2 Fashion Collection Gel Roller	$50.00	$50.00	$100.00	$85.00
12	United Stationers (OP) 8105011958730 Bag Clear 10 Gallon	$100.00	$1.00	$100.00	$80.20
13	Brother P-touch ~3/8" (0.35") Black on White Standard Laminated Tape	$50.00	$30.00	$100.00	$66.00
14	Saalfeld Redistribution Lysol Surface Disinfectant Cleaner	$100.00	$25.00	$100.00	$70.00
15	Accelerator-free Disposable Nitrile Glove, Powder Free, Small	$1.00	$1.00	$100.00	$35.20

Discussion

Our findings highlight the challenges of implementing an online marketplace for federal and defense requirements, where platforms must address the unique requirements of public-sector purchasing (e.g., socioeconomic representations, exclusions for suspended and debarred suppliers, and country-of-origin mandates such as the Buy American Act, Berry Amendment, and Trade Agreements Act). Our qualitative and quantitative results suggest a need to modernize *GSA Advantage* for improved ease of use and to maintain technological footing with private-sector marketplaces.

Do government regulations limit GPC holder's ability to use commercial e-commerce sources?

The Department of Defense GPC Guidebook and FAR do not limit the ability of GPC holders to utilize private-sector online marketplaces. However, before making a purchase, cardholders must screen for and use mandatory sources of supply (FAR Part 8). If the requirement cannot be met by a mandatory source, the cardholder must consider the use of non-mandatory sources of supply (FAR 8.004). If the mandatory sources listed in FAR 8.002 and 8.003 do not meet the need of a cardholder's requirement, users are encouraged to consider the use of non-mandatory sources of FAR 8.004(a)(1) prior to utilizing commercial sources. We were unable to locate a statute or, for the Air Force, a regulation indicating that the cardholders could not go to *Amazon Business* as a first non-mandatory source. In sections A.4.5 and A.1.2.2, the GPC Guidebook specifically references *GSA Advantage* as an available, non-mandatory, but prioritized government source (DoD, 2015). While the GPC Guidebook cites *GSA Advantage* as an available source, the Guidebook does not cite any available commercial sources. To give users additional buying options, we recommend the Defense Procurement and Acquisition Policy explore adding *Amazon Business* (and similar private-sector marketplaces such as Walmart.com) as examples of available, non-mandatory commercial sources within Guidebook Section A.1.2.2. However, additional research should first be conducted to understand if incorporating a private-sector marketplace into the Guidebook would violate the Competition in Contracting Act, unfairly favoring *Amazon Business* over other commercially-available platforms that are not included.

How do government purchasing agents view these marketplaces in relation to GSA Advantage?

Based on the GPC survey data, 78% of users would choose *Amazon Business* or Amazon.com over *GSA Advantage*. Further research should explore why users prefer one online ordering platform over another. We found certain factors, such as age, affect a user's preference, but we did not explore why. Using our survey, we found the online ordering platforms ranked in the following order: Amazon.com, *Amazon Business*, and *GSA Advantage.* A majority (70%) of respondents said that Amazon.com was easier to use, and 68% of respondents said that *Amazon Business's* website was easier to use.

From our results, we infer that users preferred Amazon's platforms for several reasons. First, the platform is used widely in the commercial sector, and many people use Amazon in their personal lives. The platform provides a vast selection of supplies, product ratings, and vendor ratings while also offering two-day delivery for most items. We also found that older respondents were less likely to prefer Amazon.com or *Amazon Business*, compared to their younger counterparts. This is likely because older respondents are more familiar with *GSA Advantage*; they have used or have been exposed to the platform for many years in their work life. It is possible that older respondents may also be less likely to use Amazon.com in their personal lives, compared to younger respondents. Our results also suggest that if users were satisfied with *GSA Advantage*, their odds of preferring Amazon decrease by 47%. This shows that once users become comfortable with a platform, they have a hard time accepting or preferring a new platform. However, our results also show if users are dissatisfied with *GSA Advantage*, their odds of preferring Amazon increased by 177%. This means that it is much easier to change behavior if a user is dissatisfied with their current platform.

Because most cardholders prefer Amazon's platforms, *GSA Advantage* and Amazon could partner by putting federally-negotiated schedules on *Amazon Business's* platform. Government purchasers would benefit from the advantages of Amazon's platform (e.g., product and price search, reviews), while maintaining the continuity, security, and quantity pricing available from *GSA Advantage*. Minimum order requirements will still have to be addressed. However, more research is needed to explore the viability of placing government schedules on a commercial platform.

Are these private-sector online marketplaces positioned to support the unique socioeconomic, environmental, and regulatory requirements of the Department of Defense and other federal agencies?

While using *Amazon Business* for GPC purchases may provide several benefits to the government, *Amazon Business*, in its current state, does not appear to be ready for use on purchases above the micro-purchase threshold. The terms and conditions of the business arrangement must be codified, which should include data collection and distribution to the government, privacy, and security of government transactions. *Amazon Business* also needs to improve catalog characteristics to ensure users can easily identify small business vendors when viewing item details. We found it difficult to identify the socioeconomic characteristics of businesses. FAR 19.502-2 states, "each acquisition of supplies or services exceeding $3,500, but not over $150,000 is automatically reserved exclusively for small business concerns and shall be set aside for small business." Further, vendors on *Amazon Business* must clearly label the country of origin for available items. We found it difficult to identify if a product complied with the Buy American Act (41 U.S.C. §§ 8301–8305), Trade Agreements Act (19 U.S.C. §§ 2501–2581), or other country-of-origin restrictions such as those found in the Berry Amendment (10 U.S.C. § 2533a).

Future Research

Currently, *Amazon Business's* pilot program is underway at a few test bases across the Air Force. While data from the pilot are not yet available for analysis in this research, we recommend future researchers conduct another GPC survey to analyze GPC users' thoughts and preferences of the *Amazon Business* pilot compared to *GSA Advantage*. We also recommend future researchers compare the transactional level data provided by *Amazon Business* to the data provided by *GSA Advantage*. Future research should also compare a breadth of item categories between platforms, including items above the Simplified Acquisition Threshold. Future research should include an analysis of strictly commercially-available items that are not uniquely offered by *GSA Advantage*. Due to the dataset available for our spend analysis, we compared some AbilityOne and Skilcraft items on *Amazon Business* and *GSA Advantage*. We believe these items were more expensive on *Amazon Business's* website because most of these items were sold through third-party vendors. We also recommend researchers should explore supply chain integrity on commercial ordering websites, as counterfeit items have been a problem on Amazon.com.

Conclusion

Our research focused on the benefits and limitations of each platform for government purchase cardholders, comparing prices, shipping costs, shipping time, ease of use, and customer satisfaction, while considering future improvement initiatives. Every attempt was made to objectively assess each online marketplace. Government purchasing agents should utilize the platform that allows them to purchase a reliable product from trusted vendors, at the best price, while maximizing the value of their time. When comparing *Amazon Business* to *GSA Advantage*, we found that each online ordering platform has advantages and disadvantages. *GSA Advantage* offers discounted commodities, strategically sourced contract vehicles, and tailored data for the Air Force; however, the ordering website is not the best source for GPC purchases due to the minimum purchase requirements. While government cardholders preferred Amazon platforms over the GSA, we found *Amazon Business* in its current state does not appear to be ready for use on purchases above the micro-purchase threshold.

References

The Coalition for Government Procurement. (2017). *E-commerce pilot recommendations*. Washington, DC: Author.

DoD. (2015, October). *Government charge card guidebook for establishing and managing purchase, travel, and fuel card programs*. Retrieved January 4, 2017, from https://www.acq.osd.mil/dpap/pdi/pc/policy_documents.html

Floyd, K., Freling, R., Alhoqail., S., Cho, H. Y., & Freling, T. (2014). How product reviews affect retail sales: A meta-analysis. *Journal of Retailing, 90*(2), 217–232.

GAO. (2015). *Federal supply schedules: More attention needed to competition and prices* (GAO-15-590). Retrieved from https://www.gao.gov/assets/680/671309.pdf

GAO. (2016). *Government purchase cards: Opportunities exist to leverage buying power* (GAO-16-526). Retrieved from https://www.gao.gov/products/GAO-16-526

GSA. (2018). *Procurement through commercial e-commerce portals: Implementation plan.* Washington, DC: Author.

Miller, J. (2017, November 6). Industry tries to prune 'Amazon amendment' before NDAA is finalized. Retrieved from https://federalnewsradio.com/reporters-notebook-jason-miller/2017/11/industry-tries-to-prune-amazon-amendment-before-ndaa-is-finalized/

R Core Team. (2017). R: A language and environment for statistical computing. R *Foundation for Statistical Computing*. Vienna, Austria.

Steward, M. D., Narus, J. A., & Roehm, M. L. (2017). An exploratory study of business-to-business online customer reviews: External online professional communities and internal vendor scorecards. *Journal of the Academy of Marketing Science, 11*(2), 131–141.

Wu, C., Che, H., Chan, T. Y., & Lu, X. (2015). The economic value of online reviews. *Marketing Science, 34*(5), 739–754.

Analysis of Contract Prices: Comparing Department of Defense With Local Governments

Latika Hartmann—is an associate professor of economics in the Graduate School of Business and Public Policy at the Naval Postgraduate School. Her research focuses on issues in public finance, regulation, and economic development. She received a PhD in economics from the University of California, Los Angeles in 2006 and a BA in economics from the University of California, Berkeley in 2000. Her dissertation was nominated for the Gerschenkron dissertation prize in 2006. She was an economics fellow at Stanford University from 2007 to 2009, and an assistant professor at Scripps College from 2009 to 2014. [lhartman@nps.edu]

Lt Col Karen Landale, USAF—received a BS in international business from the University of Tampa in 2002, an MBA from the Naval Postgraduate School in 2009, and a PhD in marketing from the University of North Carolina in 2014. Her dissertation was awarded the DeLozier Fellowship. Her research focuses on category management, strategic sourcing, services marketing, and talent management. [karen.landale@us.af.mil]

Rene Rendon—is an Associate Professor in acquisition management in the Graduate School of Business and Public Policy at the Naval Postgraduate School. He received a BBA from Angelo State University in 1981, an MBA from the University of North Dakota in 1985, and a DBA from Argosy University in 2003. A retired Air Force contracting officer, he has received the NCMA Award for Excellence in Contract Management Research and Writing many times. His research focuses on issues in contract management and strategic sourcing. [rgrendon@nps.edu]

Abstract

In this paper, we first compare the costs of waste disposal services across Air Force (AF) bases and then between AF bases and their neighboring cities. Using linear regression analysis, we find the average cost per ton of waste disposal is negatively correlated with the number of containers across bases. But, the average cost is not significantly correlated with local economic factors such as average wages, population density, and the consumer price index of the county where the base is located. We also find no significant difference in costs per ton for waste disposal between AF bases and their neighboring cities. However, when we split the sample, we find smaller bases have significantly higher costs per ton of waste disposal. Costs per ton are almost twice as high for smaller bases with fewer than 75 containers compared to larger bases. We recommend smaller AF bases review their waste disposal contracts, compare and contrast their costs relative to their neighboring cities, and then consider coordinating with their neighboring city to reduce costs.

Introduction

The Department of Defense (DoD) spends billions of tax dollars every year on contracts for base operations support (BOS) services. With an ever-decreasing procurement budget, the federal government must ensure that each tax dollar obligated on contracts provides the best value to the government and its citizens. Best value includes ensuring the prices paid are fair and reasonable. But, how do DoD contracting officers determine whether prices are reasonable? To answer this question, our study compares the prices paid per ton, i.e., average cost, on solid waste disposal services across (1) AF bases and (2) AF bases and their neighboring cities. We focus on waste disposal because it is a uniform BOS category across bases, and prices paid can be readily compared to neighboring cities that also contract for waste disposal. Although we focus on AF bases because the data are readily available, our results have implications for other military bases and DoD agencies.

Background

DoD bases across the United States procure standard BOS services such as waste disposal, custodial, and grounds maintenance, just to name a few. While the nature of the service is essentially the same, there is significant variation in the prices paid (total and per-unit) across bases. Moreover, it is unclear how the prices paid by military bases compare to their neighboring cities. Indeed, the literature to our knowledge is silent on the base-city comparison. In principle, we would expect military bases to pay similar prices for standard services because they are non-profit government entities. But, base prices could exceed those of their neighboring cities if the DoD imposes a disproportionate regulatory burden on private vendors. In that case, we can view the city prices as a floor for comparison.

Recognizing that many agencies within the federal government purchase similar products and services, the Office of Management and Budget (OMB), in particular the Office of Federal Procurement Policy, has promulgated category management and strategic sourcing implementation memorandums and guides (OMB, 2012; OMB, 2014; OMB, 2015). The main goals of category management are to achieve price, process, and demand savings by leveraging volume of spend (buying as one), reducing the number of contracts written (as well as the number of contracting offices writing contracts for similar products and services), and implementing internal controls to shape consumption. Strategic sourcing is one tool category managers can use to implement acquisition solutions within their categories.

Category management and strategic sourcing require extensive market research—a much more comprehensive examination and understanding of the markets for common products and services than the federal government has ever performed in the past. Category management teams compare historical government trends in spend, use, and consumption to historical commercial and near-peer trends. To better support such category management goals within the DoD, we study the prices paid for waste disposal services across AF bases, especially compared to prices paid for similar services by local cities.

Research Approach and Findings

In coordination with the Air Force Installation Contracting Agency (AFICA), we collected the total contract amount paid for solid waste disposal services, total tonnage of waste and total number of containers for each AF base reporting these data. Then, we calculated the distance to landfill for each base using Google Maps. To capture economic factors that may be correlated with local prices, we collected information on the state cost of living adjustment (COLA) for each base, population density, average weekly wages, and the consumer price index of the county where each base is located.

Since bigger bases house more people and generate more waste, we focus on the average cost per ton of waste disposal, that is, the total contract amount paid for waste disposal divided by total tonnage of waste. We refer to the price paid per ton on waste disposal as the cost per ton because prices paid represent the costs of waste disposal for AF bases. Here, we find large differences in cost per ton across AF bases ranging from $44 to $844. Using linear regression analysis, we find these costs are not significantly correlated with local economic factors. Rather, the average cost per ton is negatively correlated with the number of waste containers on the base. Our interpretation is that firms incur high fixed costs of contracting with AF bases, namely many forms and training requirements. Such regulatory costs are the same if the base has 50 containers or 750. This suggests there are economies of scale in waste disposal. And, smaller AF bases could perhaps reduce their costs by coordinating with other government entities under a single waste disposal contract.

In the second part of our analysis, we compare AF bases to local cities. We first matched each AF base to its nearest city. Then, we contacted the city to learn the nature of their waste disposal contract and their costs of waste disposal. While 30 cities responded to our questions, we study only 20 cities where the waste disposal data are comparable to the AF base. Across these 20 cities, we find no significant difference in the price paid per ton across cities and AF bases. But when we split the sample by the number of containers, we find smaller bases with fewer than 75 containers pay almost twice as much per ton for waste disposal compared to their neighboring city. We find no such large base-city difference for bases with more than 75 containers. Smaller cities would do well to contract with their neighboring city for waste disposal.

Our findings benefit the Air Force and DoD by comparing the costs per ton of waste disposal across AF bases and across bases and their neighboring cities. Moreover, the methodology can be applied to study the variation in prices paid between DoD bases and local cities for other BOS type services, such as custodial, grounds maintenance, and perhaps even professional services.

Literature Review

Our research compares the costs of ISWM services across AF bases and then between AF bases and their neighboring cities as an application of category management. Informing our research is a literature review encompassing three different areas: federal government and Air Force category management programs, Air Force base-level procurement research, and ISWM cost reduction research.

Our research adds to the literature on federal government and Air Force category management programs. The federal government's category management programs are focused on reducing costs and increasing efficiency and effectiveness. The OMB initiative on "Buying as One Through Category Management" is focused on "managing commonly purchased goods and services … by implementing strategies to drive performance, like developing common standards in practices and contracts, driving greater transparency in acquisition performance, improving data analysis, and more frequently using private sector (as well as government) best practices" (OMB, 2014, p. 2).

The Air Force's category management program evolved from the DoD's commodity sourcing strategies, which focused on total ownership costs and strategic sourcing strategies (Rendon, 2005). Category management is focused on "leveraging buying power, improving efficiencies, and managing consumption" (Sharkey, 2015, p. 7). The Air Force conducts category management by analyzing major performance levers (demand management, supplier management, strategic sourcing, and total cost management) to identify category improvement initiatives. Within the total cost management lever, the focus is on identification of specific price drivers in the acquisition that can result in increased efficiency and effectiveness and a reduction in costs. Price drivers can be either product/service-related or contract-related and can impact savings associated with rate (getting more for less), process (getting more with less), and demand (getting less) (Sharkey, 2015, pp. 21–24). The product/service-related price drivers impact rate savings, process savings, and demand savings. Contracting-related price drivers impact rate savings. Once these price drivers are identified, the Air Force executes changes to its acquisition strategies for these supplies/services and then conducts performance tracking, benchmarking, and continuous improvement of the management of the specific category of product/services (Sharkey, 2015, pp. 25–33).

Our research also adds to the ISWM literature, specifically analyzing approaches to reducing costs for solid waste management. The recent stream of ISWM research has been centered mainly on exploring the most cost-effective waste collection systems. For example, Boskovic et al. (2016) developed a management tool to determine waste collection costs for different waste collection schemes and input data. The tool can calculate the time and costs of waste collection. Also, Arribas, Blazquez, and Lamas (2010) propose a methodology for designing an urban solid waste collection system which uses combinatorial optimization and integer programming, and geographic information system tools to minimize collection time, and operational and transport costs. Their methodology establishes feasible collection routes, determines an adequate vehicle fleet size, and presents a comparative cost and sensitivity analysis of the results. Their research findings yielded significant cost savings in the total solid waste collection system. Finally, Solano et al. (2002) developed an ISWM model to assist in identifying alternative ISWM strategies that meet cost, energy, and environmental emissions objectives. The model is flexible to allow representation of waste diversion targets, mass flow restrictions and requirements, and targets for the values of cost, energy, and emission.

Specific to Air Force ISWM services, Landale et al. (in press) show how data analytics can be used to identify areas of potential cost savings for ISWM services. Using sequential regression, Wilcoxon Rank-Sum Test, and ordered logistic regression, they investigated the influence of service- and contracting-related variables on price and contractor performance. They found that service-related and contracting-related variables influence price. Specifically, they identified that a service-related variable—number of containers—significantly affects price and that two contracting-related variables—one type of small business set-aside and the number of offers received—also significantly affect price.

With the growth in procurement of base-level services, there is a developing stream of research focused on improving contracting for base-level services. For example, Apte, Rendon, and Salmerón (2011) developed an optimization model for selecting a set of contractor proposals from among multiple offerors for base-level services to be performed at multiple installations. The selection balanced the confidence level in an offeror's past performance with the cost of services to the Air Force, thereby achieving the most favorable objective. Their research findings demonstrate improvements over the traditional sourcing process in both overall performance and cost. Additionally, Boehmke et al. (2017) use a data envelopment analysis (DEA) approach to measure efficiency in installation support services. By focusing specifically on facility sustainment activities, their DEA approach supports decision-making by "quantifying cost savings and performance improvements, and systematically bench-marking to identify best practice peers" (Boehmke et al., 2017, p. 39). Also, Boehmke et al. (2015) apply a growth curve clustering approach to identify cost curve behavior in their research on analyzing cost growth and investigating approaches to reducing cost growth in the Air Force. Their findings indicate that micro-level growth curves vary greatly from the aggregate cost curves. They also found that their clustering approach can help decision-makers direct their focus and policies toward specific growth curves that must be "bent" (Boehmke et al., 2016, p. 126).

Finally, our research includes an analysis of municipal government costs for contracting ISWM services. Although contracting at the municipal and city level is a vast and decentralized effort, it is one of the most "under-studied" aspects of government contracting (Haselmayer, 2018, p. 1). With "557,000 city and municipal governments procuring an estimated 10% of the world GDP in goods and services annually to serve their communities," this area of government contracting is a fertile ground for the application of category management processes (Haselmayer, 2018, p. 1).

Our paper adds to the various research streams related to category management, ISWM cost reduction, and base-level services. We focus on comparing ISWM costs paid by the Air Force and compare to costs paid by their surrounding municipalities. The next section discusses our research data and methodology.

Data and Methodology

Our analysis has two parts. First, we analyze the variation in Air Force prices paid for waste disposal using data specific to AF bases. Second, we compare and contrast the prices of waste disposal between AF bases and their neighboring cities. To this end, we contacted local cities near each base and collected information on their waste disposal contracts and costs. We begin by describing the data on the AF bases below.

Air Force Base Data

We collected data on waste disposal costs for 68 AF bases in the continental United States as of 2017. Our student researchers contacted the Air Force Installation Contracting Agency (AFICA) to collect price data on Integrated Solid Waste Management (ISWM) service, which includes the costs of waste disposal and other categories under ISWM. This data set, organized by base, was jointly compiled by AFICA and the Air Force Civil Engineer Center (AFCEC), the mission owner for facilities-related services. The data set contains (1) annual contract price for solid waste services, which was pulled from the Federal Procurement Data System-Next Generation (FPDS-NG) by AFICA, and (2) number of bins and tons of solid waste, which was collected by AFCEC. Unfortunately, these data were missing for many bases. Our analysis sample focuses on 48 bases with complete data.

AFICA reports total contract costs for solid waste disposal that are annualized based on the past three to five years of contract data. Comparing total prices paid for solid waste disposal across AF bases is not informative because larger bases are likely to generate more waste and hence pay more for solid waste disposal. But, AFCEC reports the annual total tonnage of waste disposal. So we constructed a more informative measure, namely the annual price/cost per ton (total annual contract cost of waste disposal divided by tons of waste disposal). We find significant variation in the cost per ton from a high of $855 per ton in Columbus AFB in Mississippi to $41 per ton in Dyess AFB in Texas.

To understand the variation in waste disposal costs across bases, we collected data on two sets of independent variables. The first were specific to each base, such as the distance to landfill and the number of waste disposal containers. We estimated distance to landfill for each base using Google maps. We expect that bases further away from a landfill pay more for waste disposal, as do larger bases with more people and hence more containers. Though we may expect a non-linear relationship between the number of container and solid waste costs per ton on account of economies of scale. We discuss such economies of scale in the next section.

The second set of variables capture differences in the local environment of the base such as the cost of living index of the state in which the base is located, the average weekly earnings and consumer price index in the county where the base is located, and county population density. We expect that cost of waste disposal is likely higher in places with higher cost of living and wages. With regard to population density, it may be there is more competition among waste disposal companies in denser cities that would translate into lower prices paid for waste disposal in bases located in such centers conditional on cost of living.

We obtained the data on 2017 state cost of living index from Missouri Economic Research and Information Center that constructed the index for each state. The data on average weekly wage by county is from the Bureau of Labor Statistics, while the data on population density per square mile is from the U.S. census.

Table 1 shows the summary statistics. Both the cost per ton of waste disposal and the number of containers vary significantly as do the county population density. Indeed, the average cost per ton has a mean of $270 with a standard deviation of $159. In contrast, we observe less variation in the state cost of living index, consumer price index, and the average county weekly wage.

Table 1. Summary Statistics

	N	Mean	SD
AF Price paid per ton ($)	48	270	159
Num Containers	44	166	136
Distance to Landfill	48	11	7
State COLA	48	104	18
County Population Density	48	474	584
Consumer Price Index	41	246	16
County Weekly Wage	45	902	164

Local Municipality Data

After collecting the necessary Air Force data, we selected local cities near each base based on two rules: the city/municipality should lie within 30 miles of the AF base and must be part of the same county. This ensures we are comparing an AF base to a city that faces similar market conditions. Moreover, the closeness of the city and base suggests they could in principle use the same commercial vendor for waste disposal. Our goal in collecting these data are to compare and contrast the waste disposal costs of AF bases to their neighboring cities. Since both entities are public, we expect the costs to be similar across bases and their neighboring cities. Similar to bases, cities are non-profit government organizations, albeit with fewer regulatory hurdles than bases.

Before our students contacted the cities, they spoke to the City of Monterey to understand their process of solid waste contracting. The idea was that a conversation with local city officials would lead us to design more effective questions for the larger data collection effort. In particular, our students met with the sustainability coordinator for the city. They learned that cities contract for waste disposal in four different ways, namely (1) franchise agreements, (2) city-owned solid waste haulers, (3) three- to five-year term contracts with commercial vendors, and (4) open market with operating permits. Interestingly, AF bases rely only on (3).

Based on this visit, we asked the matched cities the following five questions. The appendix shows the standard email solicitation we sent to each city.

1. What type of contract does your city use for solid waste disposal?
2. What are the negotiated rates for solid waste collection at city-owned buildings, i.e., per bin size and frequency of collection?
3. What are your published prices for commercial business rates for solid waste collection, i.e., per bin size and frequency of collection?
4. Is the local Air Force base's solid waste contract managed by your city? What are the Air Force base's rates? Are the Air Force base's rates the same as the negotiated rates for the city? What is the surcharge rate applied to the Air Force for the city to manage its solid waste contract?
5. Are there standing city regulations requiring the local Air Force base to utilize the same hauler that is already contracted with the city?

Answers to these questions provide information on the contracts used and the prices paid for waste disposal by neighboring cities. Our student researchers used the following protocol in collecting data from these cities. First, they contacted each city's government office by phone and identified the solid waste contract administrator. They introduced themselves as military officers studying at the Naval Postgraduate School and briefly explained the study using a script. After the initial phone introduction, they followed up with a standard e-mail. We were concerned some cities would be reluctant to share their contract information. Hence, we added the following language: "The information you provide will be treated as confidential. Our report will be sanitized of any city, Air Force base, or commercial hauler names. After the report is complete, we would be happy to share our findings with you."

Our student researchers contacted 69 cities matched to adjacent AF bases. Each city was contacted multiple times by phone and e-mail over three months. By the end of the three months, 31 of the 69 cities responded, translating into a 45% response rate. Barring one region, the distribution of responses was uniform in other parts of the country. For example, 60% of cities responded in the south and southwest regions (19/32 bases), and 64% of midwestern cities responded (9/14). Our lowest response rate was in the northeast. Indeed, not one city adjacent to any of the nine AF bases in the northeast responded. Figure 1 displays the cities that responded to our questions. We review findings from these cities in the next section.

Figure 1. City Responses Adjacent to AF Bases

Unlike AF bases that report their annual cost of waste disposal and tons of solid waste, cities without exception report a cost per cubic yard of solid waste along with the frequency of collection (number of times per week) and the size of the waste container (2 cubic yards, 4 cubic yards, and so forth). To ensure an apples-to-apples comparison, we estimated an equivalent cost per ton for each city in the following manner.

City unit costs were measured as costs per cubic yards where cubic yards came in many sizes ranging from 2, 4, 6, and 8 cubic yards. Moreover, the cost of each size container varied by the frequency of scheduled solid waste collection, ranging from one to five times per week. For example, a city in the southwest reported its cost per cubic yard increases from $57 for a collection of once per week to $312 for six times per week. As the cubic yards increased, the cost also increased, though again in a non-linear manner.

Our first challenge was converting the city costs per cubic yard, a measure of volume, to Air Force costs per ton, a measure of weight. We used the Environmental Protection Agency's Office of Resource Conservation and Recovery standard volume-to-weight conversion factors to convert cubic yards to tons of solid waste. These published conversion factors suggest the standard weight per cubic yard of un-compacted, residential, institutional, and commercial solid waste translates into an estimated weight of 250 to 300 pounds, namely 0.125 to 0.15 tons. We created estimates of costs per ton for each matched city using the 300-pound (0.15 tons) estimate of the weight range. Our patterns are similar if we use the lower 250-pound estimate.

We also made a decision to use the city costs for once-a-week collection. We selected once a week to err on the side of constructing a higher estimate for city waste prices. Cities paid more per unit for once a week collection compared to five times a week. Since most of our cities report their data for a two cubic yard container emptied once a week, our choice of once a week collection also ensured a larger matched city sample. We unfortunately had to exclude cities that did not provide data for two cubic yard containers collected once a week and those reporting other units of measure. This left us with 20 cities matched to their nearest AF bases. On account of the small sample, we report t-tests for difference in means of costs per ton between AF bases and their neighboring cities. We turn to these results next.

Findings

We begin by presenting results from linear regressions using the annual AF cost of waste disposal per ton as our dependent variable (see Table 2). In regression (1), we focus on the base specific independent variables and then add more variables in regressions (2)-(4). Across the specifications, the coefficient on number of containers is negative and statistically significant. In terms of magnitude, a one standard deviation increase in the number of containers (136 containers) translates into a decrease of $38 in average cost, an economic effect of roughly 14% given average cost per ton of $270. This coefficient is remarkably robust to the addition of controls. While the coefficient on distance to landfill is positive, it is not robust. The coefficient is statistically significant only in regression (4) when we include other location-specific variables. We also find in regression (4) that the county-specific price index is positively correlated with average costs, which is perhaps unsurprising.

Table 2. Dep. Variable—Average Annual Cost per Ton for Waste Disposal ($)

	(1)	(2)	(3)	(4)
Number—Waste Containers	-0.28*	-0.27*	-0.30*	-0.29*
	[0.15]	[0.15]	[0.16]	[0.16]
Distance to Landfill	5.69	6.04	6.17	7.74*
	[4.72]	[4.70]	[4.71]	[4.09]
State Cost of Living Index		-0.45	-0.85	-0.72
		[1.50]	[1.54]	[1.49]
County Population Density		0.01	-0.01	0.00
		[0.05]	[0.07]	[0.07]
Average County Weekly Wage			0.13	-0.02
			[0.21]	[0.22]
Consumer Price Index				1.32*
				[0.70]
Constant	260.72***	295.10*	236.56	34.86
	[52.97]	[161.59]	[185.71]	[236.06]
Observations	44	44	43	38
Adjusted R-squared	0.061	0.017	0.004	0.070

Robust standard errors in brackets

*** p<0.01, ** p<0.05, * p<0.1

While the results in Table 2 point to a large and negative relationship between average costs and number of containers, it is unlikely that average costs decrease for each extra container in a linear manner. We explore this relationship in more detail in Figure 2, where we plot average cost per ton on the y-axis against the number of waste containers on the x-axis. Indeed, this picture suggests that economies of scale can perhaps account for some of the variation in average costs across AF bases. Economies of scale arise when there are huge fixed costs compared to marginal costs of waste disposal. It is likely that waste disposal companies face high fixed costs of contracting with the DoD and getting on an AF base. Conditional on those costs, it seems the cost of hauling each additional container is low. We use the term average cost per ton interchangeably with annual price per ton, namely the annual price per ton paid by an AF base for waste disposal.

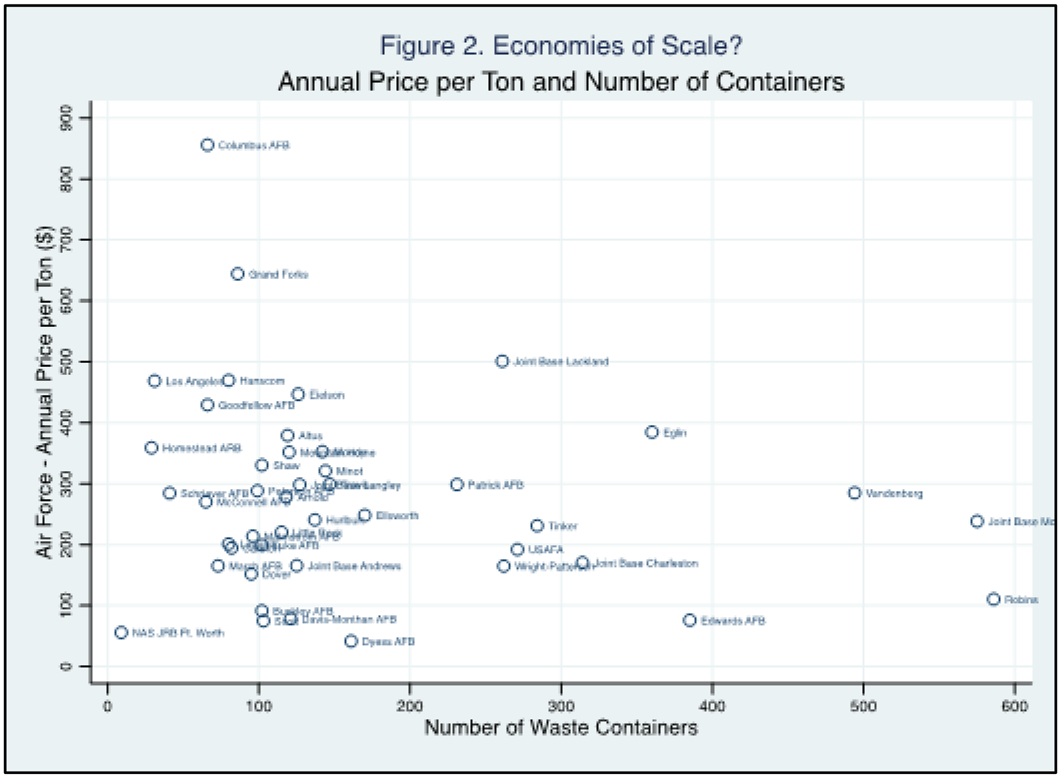


Figure 2. Economies of Scale? Annual Price per Ton and Number of Containers

The presence of economies of scale suggests smaller AF bases could in principle reduce their costs if they joined forces with neighboring cities. This would allow them to leverage their containers with cities and secure a lower price per ton from waste disposal companies. Indeed, smaller bases would have more bargaining power negotiating with commercial waste disposal companies because they would be negotiating over a larger amount of waste disposal (small base plus neighboring city as one entity).

Nonetheless, economies of scale are not the entire story because we observe large differences in costs per ton for bases with the same number of containers. Indeed, there are striking differences in costs for bases with around 100 containers. For example, average costs range from a low of $75 in Scott AFB Illinois to a high of $644 in Grand Forks AFB in North Dakota. To understand this variation, we turn to the matched city comparison next.

As noted earlier, our research team focused on 20 matched comparisons between AF bases and their neighboring cities. Of the 31 cities that responded to our request, three were unable or unwilling to provide cost/price data due to proprietary relationships with their ISWM contractors. Another five cities did not provide detailed cost data and we were unable to normalize their data to costs per ton. The remaining three cities gave us sufficient cost data, but AFICA and AFCEC did not have sufficient cost data for their neighboring AF bases.

Figure 3 shows the comparison of Air Force and matched city price per ton for each AF base. In this sample of 20 matched pairs, cities' average cost per ton is $203, compared to $236 for AF bases. But, this difference is not statistically significant at conventional levels of significance (95% or 90% level of confidence). AF bases have a higher coefficient of variation (81%), suggesting their values are more dispersed compared to cities at 67%.

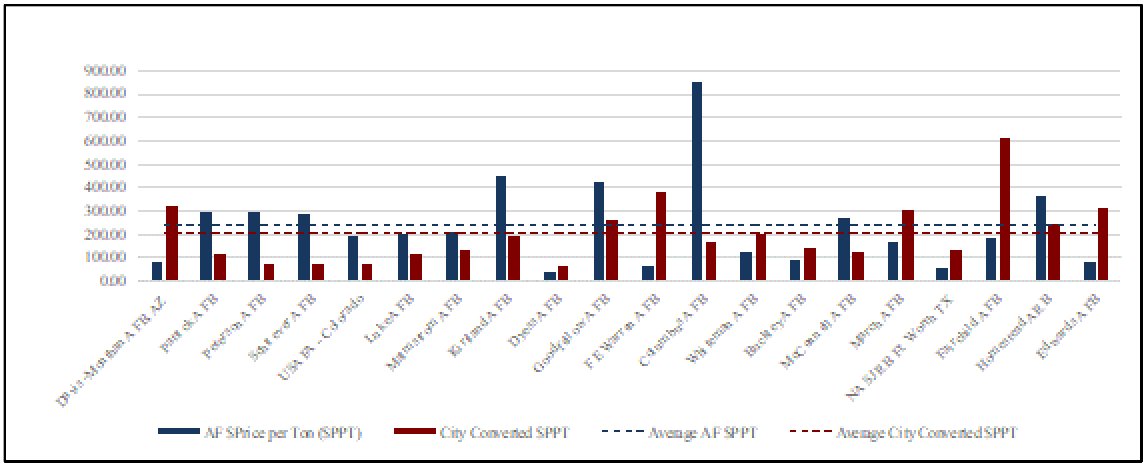


Figure 3. Cost per Ton on Waste Disposal: AF Base and City Comparison

Of the 20 matched cases, the Air Force cost per ton is higher for 11 cities, and in the case of 7 of these 11, the Air Force cost per ton is almost twice as high as the matched city. To assess if there are any systematic patterns in these 20 matched cases, we split the sample by number of containers into big and small AF bases. Bases with fewer than 75 containers were binned as small, while the rest were binned as large. Across the 7 small bases, the Air Force cost per ton averaged $346 compared to $188 in the neighboring city. This is a striking difference, as seen in Figure 4.

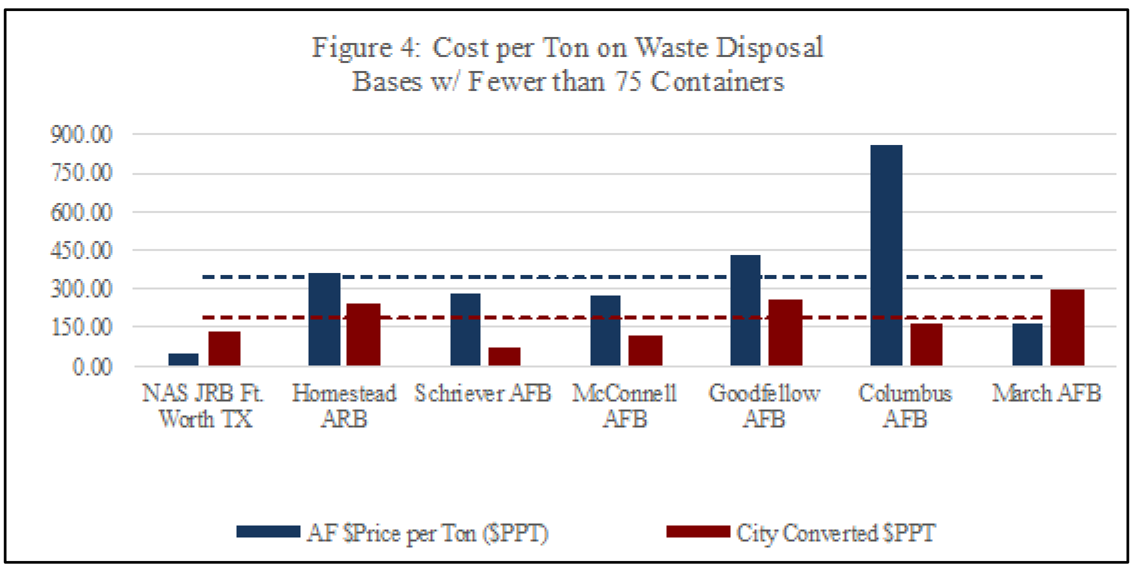


Figure 4. Cost per Ton on Waste Disposal Bases With Fewer Than 75 Containers

Small bases are at a significant disadvantage, most likely on account of their size and the economies of scale associated with waste disposal. Our recommendation is that these bases would be better served if they coordinated with their neighboring city for waste disposal. In contrast, the cost per ton averages $164 for bigger bases compared to $150 for their neighboring cities, as seen in Figure 5. The difference is small and statistically insignificant. Both the matched AF base city and AF base only analysis thus suggest economies of scale may be an important factor in contracting for waste disposal.

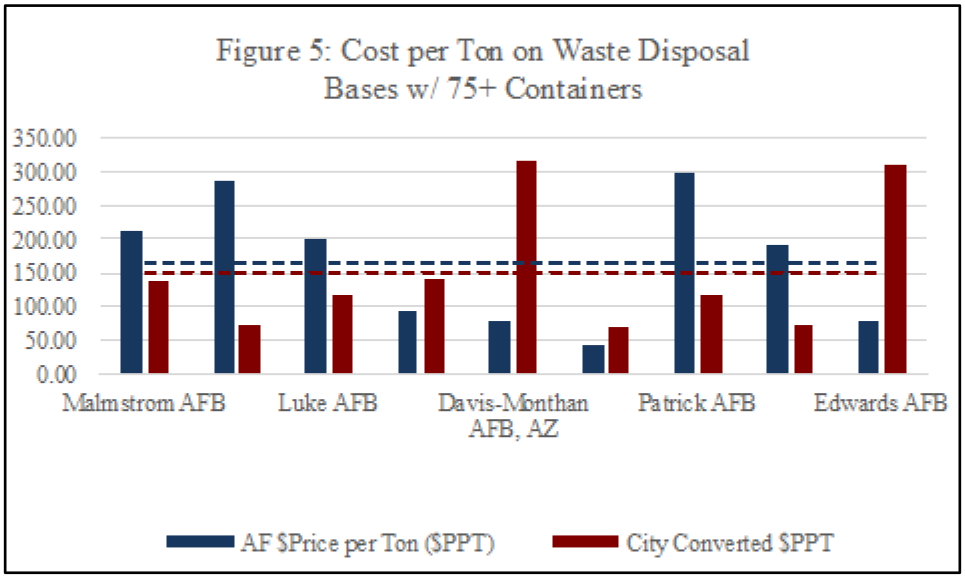


Figure 5. Cost per Ton on Waste Disposal Bases With 75+ Containers

Before concluding, we want to review qualitative findings from the matched AF base city analysis for the 31 cities that responded to our questions. In our sample, cities managed their waste disposal services in five different ways. Of the 31 cities, 10% (3) used a three-year term contract and 16% (5) used a five-year term contract. Such contracts are perhaps most similar to standard Air Force contracts. But, 39% (12) of cities managed ISWM services in house. This could perhaps be an option for larger AF bases that can exploit their economies of scale by moving services in house. Another 22% (7) use a franchise agreement, and finally, 13% (4) use an open market via an annual city-issued permit. In the latter cases, cities revoke a vendor's permit if they receive too many complaints against a vendor and they can validate those complaints.

In response to our question of whether the city managed the local AF base's solid waste contract, only one city answered in the affirmative. We were surprised that AF bases chose not to contract with their neighboring cities even when cities managed their own solid waste program. At least for smaller bases, this would seem like a more cost-effective option. Unlike cities, AF bases have more uniform contracting arrangements that perhaps do not exploit local conditions, leading to franchise contracts by some cities and in-house provision by others. At the very least, we believe Air Force contracting officers would be better served if they had basic information on the type of contract and cost per ton paid by their local city, a non-profit government organization, for waste disposal. We suggest contracting officers contact their local governments to investigate partnering for ISWM services as part of their required market research.

Conclusion and Recommendations

We find no significant difference in costs per ton for waste disposal between AF bases and their neighboring cities. But, when we split the sample by size, we find smaller bases have significantly higher costs (i.e., pay higher prices) of waste disposal. Air Force costs per ton are almost twice as high for smaller bases (fewer than 75 containers) compared to larger bases. Our first recommendation is that smaller AF bases should review their waste disposal contracts, compare and contrast their costs relative to their neighboring city, and then consider coordinating with their neighboring city to reduce their costs of waste disposal.

Our second recommendation is that the Air Force standardize the contract line item number (CLIN) cost data in ISWM contracts across all AF bases. As we collected the cost breakdown from AFICA, we quickly learned that each base formats their CLINs differently. More importantly, none of the bases report the data per industry standards. We had to normalize the data to prices paid per ton/cost per ton to make informed comparisons across bases. Unlike AF bases, U.S. cities record their waste disposal costs in terms of prices paid by bin size, number of bins, and frequency of pick-up. This seems to be the industry standard. We had to make assumptions on conversion from cubic yards (volume measure) to tons (weight). In an ideal world, AFCEC would be collecting the data by industry standard.

Finally, we had to find appropriate conversion factors to compare the data to local cities. It is hard to imagine Air Force contracting officers undertaking such research before they award contracts. We believe Air Force officers need comparable information on prices paid by neighboring non-profit and even for-profit entities to make informed decisions on what is a fair and reasonable price. We hope the Air Force makes such information available to their contracting officers.

References

Achillas, C., Moussiopoulos, N., Karagiannidis, A., Banias, G., & Perkoulidis, G. (2013). The use of multi-criteria decision analysis to tackle waste management problems: A literature review. *Waste Management & Research, 31*(2), 115–129.

Apte, A., Rendon, R. G., & Salmerón, J. (2011). An optimization approach to strategic sourcing: A case study of the United States Air Force. *Journal of Purchasing & Supply Management, 17*(4), 222–230.

Arribas, C. A., Blazquez, C. A., & Lamas, A. (2010). Urban solid waste collection system using mathematical modelling and tools of geographic information systems. *Waste Management & Research, 28*(4), 355–363.

Boehmke, B. C., Jackson, R. A., Johnson, A. L, White, E. D., Weir, J. D., & Gallagher, M. A. (2017). Measuring U.S. Air Force installation support activities via data envelopment analysis. *Military Operations Research, 22*(1), 39–58.

Boehmke, B. C., Johnson, A. L, White, E. D., Weir, J. D., & Gallagher, M. A. (2015). Bending the cost curve: Moving the focus from macro-level to micro-level cost trends with cluster analysis. *Journal of Cost Analysis and Parametrics, 8*(2), 126–148.

Boskovic, G., Jovicic, N., Jovanovic, S., & Simovic, V. (2016). Calculating the costs of waste collection: A methodological proposal. *Waste Management & Research, 34*(8), 775–783.

GAO. (2016, October). *Federal procurement: Smarter buying initiatives can achieve additional savings, but improved oversight and accountability needed* (GAO-17-164). Washington, DC: Author.

Haselmayer, S. (2018). The de-globalized city. *New Global Studies*. doi:10.1515/ngs-2018-0013

Henderson, B. D. (1975). The coming revolution in purchasing. *Journal of Purchasing and Materials Management, 11*(2), 44–46.

Kraljic, P. (1983). Purchasing must become supply management. *Harvard Business Review, 61*(5), 109–117.

Landale, K. F., Apte, A., Rendon, R. G., & Salmerón, J. (in press). Using analytics to inform category management & strategic sourcing. *The Journal of Defense Analytics and Logistics.*

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2015*). Purchasing and supply chain management*. Independence, KY: Cengage Learning.

Office of Management and Budget (OMB). (2012, December 5). *Improving acquisition through strategic sourcing* [Memorandum]. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-02_0.pdf

Office of Management and Budget (OMB). (2014, December 4). *Transforming the marketplace: Simplifying federal procurement to improve performance, drive innovation, and increase savings*. Retrieved from https://www.whitehouse.gov/blog/2014/12/04/transforming-marketplace-simplifying-federal-procurement-improve-performance-drive-i

Office of Management and Budget (OMB). (2015). *Government-wide category management document—Version 1.0.* Retrieved from https://hallways.cap.gsa.gov/information/Gov-wide_CM_Guidance_V1.pdf

Rendon, R. G. (2005). Commodity sourcing strategies: Processes, best practices, and defense initiatives. *Journal of Contract Management, 3*(1), 7–20.

Rendon, R. G., & Templin, C. R. (1992, July). Corporate procurement strategy: An analysis of supply line management. *Contract Management, 32*(7), 18–25.

Sharkey, J. (2015, November 4). Buying as one through category management. Wright-Patterson AFB, OH: Air Force Base Contracting Agency, Enterprise Sourcing Directorate [SJP1].

Solano, E., Ranjithan, S. R., Barlaz, M. A., & Brill, E. D. (2002). Life-cycle-based solid waste management. I: Model development. *Journal of Environmental Engineering, 128*(10), 981–992.

Tabachnick, B. G., & Fidell, L. S. (2007). *Experimental designs using ANOVA*. Boston, MA: Pearson.

USA Spending. (2016). Overview of awards—FY 2016. Retrieved from https://www.usaspending.gov/Pages/Default.aspx

Appendix: E-Mail Sent to City Governments

Dear XXXX,

My name is Lieutenant Commander XXXX. I am a student at the Naval Postgraduate School conducting research on solid waste contract costs. Our research team is comparing and contrasting the costs of solid waste disposal between Air Force bases and their neighboring cities. We hope this will help military leadership in deciding the types of service contracts they should pursue across bases.

To that end, my team and I are inquiring as to how municipalities manage their solid waste contracts in an effort to use them as a model to improve Air Force contract processes.

I am hoping that you can answer a few quick questions in support of our research for the military and the federal government.

- What type of contract does your city use for solid waste disposal?
- What are the negotiated rates for solid waste collection at city owned buildings, i.e., per bin size and frequency of collection?
- What are your published prices for commercial business rates for solid waste collection, i.e., per bin size and frequency of collection?
- Is the local Air Force installation's solid waste contract managed by your city? What are the Air Force base's rates? Are the Air Force base's rates the same as the negotiated rates for the city? What is the surcharge rate applied to the Air Force for the city to manage its solid waste contract?
- Are there standing city regulations requiring the local Air Force installation to utilize the same hauler that is already contracted with the city?

The information you provide will be treated as confidential. Our report will be sanitized of any city, Air Force installation, or commercial hauler names. After the report is complete, we would be happy to share our findings with you.

Thank you very much for your time and support; your vital contribution to our research will help improve Department of Defense contract processes.

Sincerely,

XXXX

Panel 13. Applying Model-Based Systems Engineering to Defense Acquisition

Wednesday, May 9, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: Captain Jonathan Garcia, USN,** Major Program Manager, Program Executive Officer for Integrated Warfare Systems ***Testing Whether the Adoption of Model-Based Systems Engineering Influences How Stakeholders Think About Systems*** Ronald Giachetti, Naval Postgraduate School Mollie McGuire, Naval Postgraduate School Karen Holness, Naval Postgraduate School ***An MBSE Methodology to Support Australian Naval Vessel Acquisition Projects*** Brett Morris, Australian Defence Science and Technology Group Stephen Cook, University of Adelaide Stuart Cannon, Australian Defence Science and Technology Group Dylan Dwyer, Australian Defence Science and Technology Group ***Enabling Operationally Adaptive Forces*** Dan Boger, Naval Postgraduate School Charles Deleot, Patriot Foundation Norman Eaglestone, Eaglestone Consulting Scot Miller, Naval Postgraduate School Scott Rosa, Naval Postgraduate School

Testing Whether the Adoption of Model-Based Systems Engineering Influences How Stakeholders Think About Systems

Ronald E. Giachetti—is the Chair and Professor of the Systems Engineering Department at the Naval Postgraduate School (NPS) in Monterey, CA. He teaches and conducts research in the design of enterprise systems, systems modeling, and system architecture. He has published over 50 technical articles on these topics, including a textbook, Design of Enterprise Systems: Theory, Methods, and Architecture. Prior to joining NPS, he was at Florida International University in Miami, FL. He earned his BS in mechanical engineering from Rensselaer Polytechnic Institute, MS in manufacturing engineering from Polytechnic University (at NYU), and PhD in industrial engineering from North Carolina State University. [regiache@nps.edu]

Karen Holness—is an Assistant Professor in the Department of Systems Engineering at the Naval Postgraduate School (NPS) in Monterey, CA. She holds a BS, MS, and PhD in industrial engineering from the University at Buffalo. Prior to NPS, she worked as a Navy Civilian in the acquisition workforce for eight and a half years in various industrial engineering, systems engineering, and human systems integration roles. She also previously worked for three years as an Industrial Engineer at Corning, Incorporated in Corning, NY. [kholness@nps.edu]

Mollie McGuire—is a Research Associate in the Department of Information Systems at the Naval Postgraduate School, Monterey, CA.

Abstract

The Department of Defense is adopting model-based systems engineering in which models will replace the extensive amounts of documentation generated in developing a new system. This research examines how this shift from textual description of the system and its requirements to a model-based description will affect the acquisition process. Specifically, we ask whether engineers and other stakeholders will be able to extract the same understanding of the system requirements from the models as they can from the traditional textual requirements specifications. We propose a theory called Model Relativity Theory, saying that the language used to represent and communicate system design and requirements influences how people think about the system. In this presentation, we describe the theory, present our exploratory research studies, discuss our research protocol, describe the research plan, and present the current status of our study.

An MBSE Methodology to Support Australian Naval Vessel Acquisition Projects

Brett Morris—is a Naval Architect/Systems Engineer who joined the Defence Science and Technology Group in 2007. He has previously worked for the RAN in the Directorate of Navy Platform Systems and has conducted research in the fields of naval ship concept design, modelling and simulation of ship performance, along with MBSE. Morris has a Grad. Dip. in Systems Engineering, a BE (Nav. Arch.) and is currently undertaking part-time research towards a PhD on Model-Based Systems Engineering. [brett.morris@dst.defence.gov.au]

Stephen Cook—is the Professor of Defence Systems at the University of Adelaide where he works in the Entrepreneurship, Commercialisation, and Innovation Centre undertaking research and teaching in system of systems engineering and complex project management. He is also a Systems Engineering Advisor with Shoal Engineering Pty Ltd where he applies his knowledge to a range of systems engineering management and research challenges. Until June 2014, he was the Professor of Systems Engineering at the University of South Australia where he led a number of research concentrations for more than 15 years. Preceding this he accumulated 20 years of industrial R&D and SE experience spanning aerospace and defence communications systems. Cook has a PhD, is an INCOSE Fellow, a Fellow of Engineers Australia, and a Member of the Omega Alpha Association. [stephen.cook@adelaide.edu.au]

Stuart Cannon—is a Naval Architect and is currently the Program Leader for Surface Ship Science and Technology at the Defence Science and Technology Group. In this role, he is the adviser to many of the surface ship projects for the RAN. He has a PhD in ship structures and is a Fellow of the Australian Academy of Technology and Engineering. [stuart.cannon@dst.defence.gov.au]

Dylan Dwyer—graduated from the Australian Maritime College with a Bachelor of Engineering (Naval Architecture) with Honours. He joined DST Group in 2016 as a graduate Naval Architect working under the discipline of Platform Systems Analysis. In his time at DST Group, Dylan has developed, and is continuing to expand his knowledge of Systems Engineering practices. Dylan is currently undertaking studies towards a masters in systems engineering. [dylan.dwyer@dst.defence.gov.au]

Abstract

This paper covers research to construct a Model-Based Systems Engineering (MBSE) methodology to support above-the-line, or left-of-contract stakeholders during the early stages of Australian naval vessel acquisition projects. These projects now adopt off-the-shelf (OTS) acquisition strategies as the default approach. OTS acquisition strategies change the nature of defence acquisition projects from the traditional top-down, requirements-driven approach to a middle-out approach. In the middle-out approach, the required functions are decomposed from the capability needs, whilst existing OTS offerings are scrutinised to find those that best satisfy the capability needs with minimal design changes. This scrutiny of the OTS solution space is generally undertaken without extensive design data being available to the acquirer.

The MBSE methodology that has been constructed comprises two main parts. The first part of the MBSE methodology is a concept and requirements exploration approach, which is the focus of this paper. Of significance, this stage of the methodology incorporates set-based design principles, model-based conceptual design, and design patterns. MBSE is used as the backbone of the methodology to manage and guide the early stage acquisition and analysis activities, whilst maintaining traceability to strategic needs. The paper includes an example implementation of the methodology for an indicative Hydrographic and Oceanographic Survey vessel capability.

Introduction

In the latest of a long line of reviews of the Australian Department of Defence (ADOD) undertaken on behalf of the government of the day, the ADOD was described as having a capability acquisition and sustainment system where there is a "persistence of fundamental problems … from capability planning to acquisition, delivery and finally sustainment" (Peever, 2015, p. 14). This review also noted that in the next 10 to 20 years, the ADOD acquisition system

> must deliver a significant capability modernisation program against a backdrop of strategic uncertainty including, but not limited to: rapid technological change; budget uncertainty; substantial economic growth in our region; and increasing demand for military responses. (Peever, 2015, p. 13)

Following this latest review of the ADOD, a new acquisition manual, the *Interim Capability Life Cycle Manual* (ICLCM; Defence, 2017a), was released. Compared to both the previous ADOD acquisition manual and current U.S. DoD acquisition manuals, DoDI 5000.02 (DoD, 2015b) and JCIDS (DoD, 2015a), the ICLCM provides a far less structured approach to acquisition. The ICLCM (Defence, 2017a) also provides far less guidance than the U.S. acquisition manuals on satisfying the newly established ADOD oversight function, called "contestability," that seeks to ensure that the acquisition project will acquire a capability that addresses the strategic needs of Australia. This means that ADOD acquisition professionals have been given an additional layer of oversight, whilst at the same time they have been provided with less guidance on how to produce defensible decisions based on solid, traceable evidence.

An important constraint on Australian naval vessel acquisitions is the adoption of the off-the-shelf (OTS) acquisition strategy as the default approach. This strategy is perceived as a means of reducing the acquisition cost and schedule risk (Saunders, 2013). The trade-off of in reducing these risks is that the capability option selected may not fully meet all of the user's operational needs, may not fully integrate with other in-service capabilities, and may not fully suit the local geographic and strategic circumstances (SFAD&TC, 2012). In 2017, the ADOD released its Naval Shipbuilding Plan (Defence, 2017b) that effectively mandated the acquisition of OTS naval vessels. The guiding principles of implementing the plan included the following:

- Selecting a mature design at the start of the build and limiting the amount of changes once production starts;
- Limiting the amount of unique Australian design changes. (Defence, 2017b, p. 105)

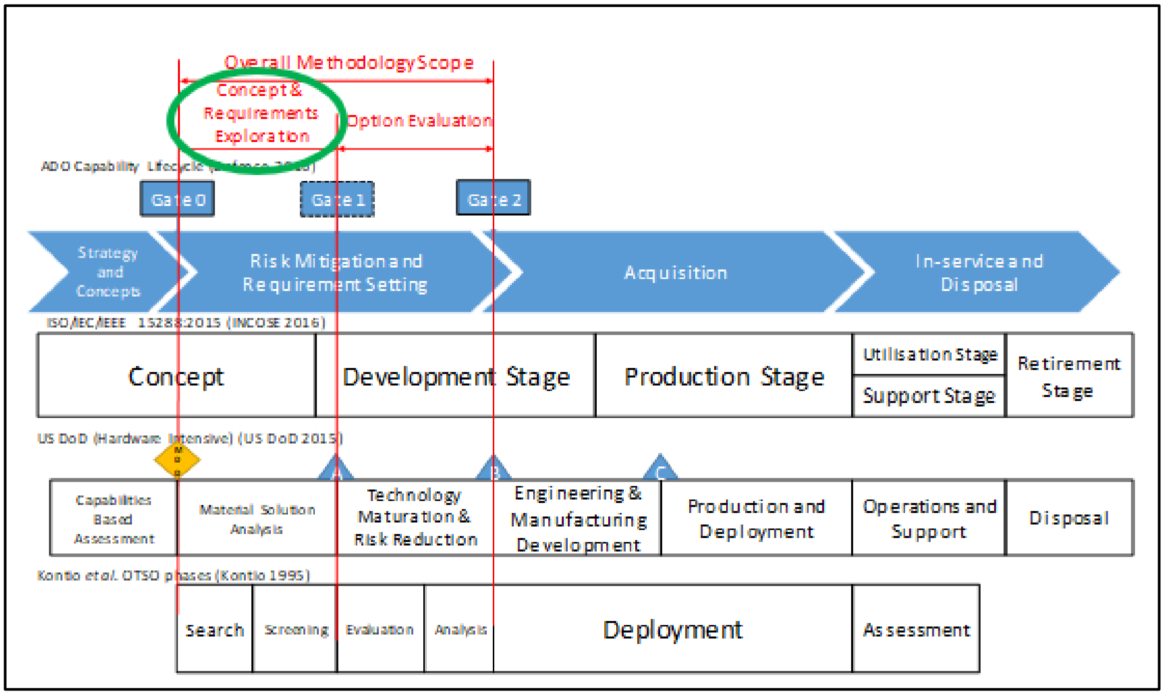
The OTS strategy appears to be analogous to the "modified-repeat" ship design strategy, where a parent design is modified, due to the perception that both the OTS strategy and modified-repeat design approach reduce acquisition cost and schedule risk (Morris, Cook, & Cannon, 2018, p. A-22). The modified-repeat design approach has, however, only been found to realise the benefits of lower acquisition costs and schedule risks, when the operational and legislative requirements are nearly identical to those that shaped the original design (Covich & Hammes, 1983). Hence, to achieve the benefits of lower acquisition cost and schedule risks in OTS naval vessel acquisitions, the project will need to identify existing OTS designs with very similar operational and legislative requirements to those for the vessel being acquired, and then specify tender requirements accordingly. Unlike a navy undertaking a modified-repeat design, the OTS acquirer will not have knowledge of the parent design's requirements, or access to detailed design data. This means the traditional "top-down" acquisition approach needs to be adjusted for OTS vessel acquisitions due to the constraint placed on the solution system by the available OTS solutions (Saunders, 2013). A “middle-out” systems engineering (SE) approach that combines top-down decomposition from strategy to functions and key performance parameters (KPPs), with bottom-up mapping from OTS naval vessel designs through the KPPs to the functions, could provide a means of enhancing rigour in contestability of OTS Defence acquisitions. A “middle-out” SE approach could also help provide an early understanding of any capability risks due to the OTS constraint.

The situation outlined above gives rise to the research issue investigated in this paper. The research issue is as follows:

> In the early stages of Australian Defence Organisation off-the-shelf naval vessel capability acquisition projects, support for traceable, defensible requirement development activities is often lacking. Concurrently, these projects are facing shortages of skilled staff and constrained financial resources. The OTS constraint also changes the nature of the acquisition’s SE approach in acquisitions that adopt this strategy.

The focus of the research covered in this paper is the activities within the early stages of Australian OTS naval vessel acquisition projects, since performing these stages well is vital for the success of any system development or acquisition project. Naval vessels, like all man-made systems have a lifecycle (Walden et al., 2015), several examples of which are shown in Figure 1. The lifecycle used in the ADOD is described in the ICLCM (Defence, 2017a). The early stage of interest for this research in the ADOD lifecycle is termed the Risk Mitigation and Requirement Setting Phase (Defence, 2017a). This phase “involves the development and progression of capability options through the investment approval process leading to a government decision to proceed to acquisition” (Defence, 2017a, p. 28). The early stages of Defence acquisitions can also be seen as a design activity (Hodge & Cook, 2014; Coffield, 2016; Cook & Unewisse, 2017), where the initial activities correspond to the concept design stage as shown in Figure 1. There is a growing understanding within the SE discipline that the process of requirements definition should include design activities. This understanding is evidenced by the statement by Crowder, Carbone, and Demijohn (2016, p. 105), “In the end, the activities which we would call design are nothing different from the activities required to create the ‘to-be’ requirements.”

The research is targeted at supporting “above-the-line” (acquirer) naval vessel acquisition stakeholders to perform the key activities of **requirements definition, requirements setting, and options refinement** in a traceable, defendable manner, during the ADOD Risk Mitigation and Requirements Setting phase.



Note. The Concept and Requirements Exploration part of the MBSE methodology in the green oval is the focus of this paper

Figure 1. Various System Lifecycles and the Stages of Interest for the Research

.

This paper covers the latest iteration of research undertaken to construct a Model-Based Systems Engineering (MBSE) methodology that supports acquisition stakeholders during the early stages of Australian OTS naval vessel acquisitions. The MBSE methodology is built around two main parts. The first part is a concept and requirements exploration approach tailored for OTS acquisitions and is the focus of this paper as shown inside the green oval in Figure 1. The second part of the MBSE methodology is a model-based approach to option evaluation that leverages the MBSE model built during the concept and requirements exploration part. The model-based option evaluation method has been covered elsewhere (see Morris & Cook, 2017; Morris et al., 2018). In this paper, a high-level overview of the research approach and the concept and requirements exploration part of the MBSE methodology is provided. The paper then steps through an example implementation of the concept and requirements exploration approach for an indicative Hydrographic and Oceanographic Survey Vessel capability acquisition. The paper concludes with some observations from the example implementation and recommendations for future work.

Research Approach

The research covered in this paper can be classed as being in the field of SE. The primary purpose of SE research has been identified as being to improve SE methods, tools, and techniques (Ferris, Cook, & Honour, 2005). This means the interventionist research paradigm, which includes action research, design science, and constructive research approaches, is well suited. Interventionist research has also been described as *development* research, since common characteristics of these methods include "design, constructed artefacts, and/or interventions" (Viliers, 2012, p. 240). The research methodology selected for the research covered in this paper is the constructive research approach (CRA). The CRA "implies building of an artefact (practical, theoretical, or both) that solves a domain specific problem in order to create knowledge about how the problem can be solved (or understood, explained, or modelled) in principle" (Crnkovic, 2010, p. 363). The problem in the case of the research described in this paper is the research issue given in the introduction. The CRA comprises the following features as espoused by Piirainen and Gonzalez (2013):

1. The focus is on real-life problems.
2. An innovative artefact, intended to solve the problem, is produced.
3. The artefact is tested through application.
4. There is teamwork between the researcher and practitioners.
5. It is linked to existing theoretical knowledge.
6. It creates a theoretical contribution.

The creation of a theoretical contribution that can improve SE methods, tools and techniques, makes the CRA well suited to SE research. The artefact produced in this research is the MBSE methodology.

Proposed MBSE Methodology

MBSE is used as the foundation of the methodology constructed for this research because it inherently supports traceability and provides numerous other benefits. Specifically, it enhances communications among the development team, improves specification and design quality, and promotes reuse of system specification and design artefacts (Friedenthal, Moore, & Steiner, 2009, p. 15). Morris et al. (2016) also report that applying MBSE during the early stages of the system lifecycle has yielded benefits associated with a clearer understanding of the problem space and facilitation of requirements development. In 2012, The U.S. Government Accountability Office (GAO) made a strong case for the use of MBSE in Defence acquisition projects: "Positive acquisition outcomes require the use of a knowledge-based approach to product development that demonstrates high levels of knowledge before significant commitments are made. In essence, knowledge supplants risk over time."

The MBSE methodology constructed for this research incorporates several features. The features were incorporated after assessing each for adherence to three guiding principles. These guiding principles are related to recurring issues in ADOD acquisitions identified by Peever (2015). The guiding principles are as follows:

1. Maintain traceability to the original, strategic intent of the vessel being acquired in order to ensure a defensible outcome.
2. Assist the stakeholders to make defensible decisions that account for competing goals and objectives.

3. Maximise the capacity to reuse elements, thereby reducing subsequent acquisition efforts to implement the methodology and the resources required to manage these projects.

Six key approaches were included in the MBSE methodology after assessing each against the guiding principles: model-based conceptual design (MBCD), modelling and simulation (M&S), design space exploration (DSE), resilient systems, pattern-based methods, and multi-criteria decision making (MCDM). MBCD is implemented through integrating MBSE with M&S and DSE within the concept and requirements exploration part of the methodology. Resilience is incorporated into the MBSE methodology through the use of set-based design (SBD) principles. This means ranges of design parameters are used during the concept and requirements exploration in order to ensure all feasible regions of the design space are explored prior to setting requirements. Pattern-based methods are implemented through the use of patterns of naval operations, such as that given in the Universal Naval Task List (CNO, 2007) and a functional architecture based on the "float, move, and fight" top-level functions. A MCDM approach (multi-attribute value analysis) is included in the option evaluation part of the MBSE methodology.

When implementing MBSE, a methodology comprising a collection of processes, methods and tools is used (Morris & Sterling, 2012). A metamodel, or schema, that defines the MBSE model element's concepts, terminology, characteristics and interrelationships is also used when implementing MBSE. It has been noted that "the metamodel is the method by which the underlying structure is embedded into the methodology" (Morris, 2014, p. 3). Furthermore, Logan et al. (2013) state, "The principal reason for using metamodels in MBSE is to create structure and consistency in the model and associated products" (p. 3).

During the research described in this paper, the metamodel underpinning the MBSE was refined over several iterations. The metamodel is based on the Whole-of-System Analytical Framework (WSAF) metamodel because it has gained increasing acceptance within the ADOD from repeated usage (Logan et al., 2013, p. 3). The WSAF metamodel is one of three components of the WSAF framework that has been used to support requirements definition in ADOD acquisition projects. The WSAF metamodel is also consistent with the CORE DODAF 2.02 schema (Cook et al., 2014). Several extensions to the WSAF metamodel were made during the research. A key extension was the introduction of the "analysis domain." The analysis domain allows executable analyses to be conducted, managed and the results stored within the MBSE model. A high-level overview of the key parts of the MBSE metamodel developed for the research is shown in Figure 2. The operational domain shown in green in Figure 2 allows strategic guidance from the capability needs statement to be traced to system functions and requirements. The analysis domain shown in red in Figure 2 allows executable analyses to be conducted, managed and stored within the MBSE model. The "vessel properties" element within the blue oval in Figure 2 is discussed further in later sections and detailed in Figure 8.

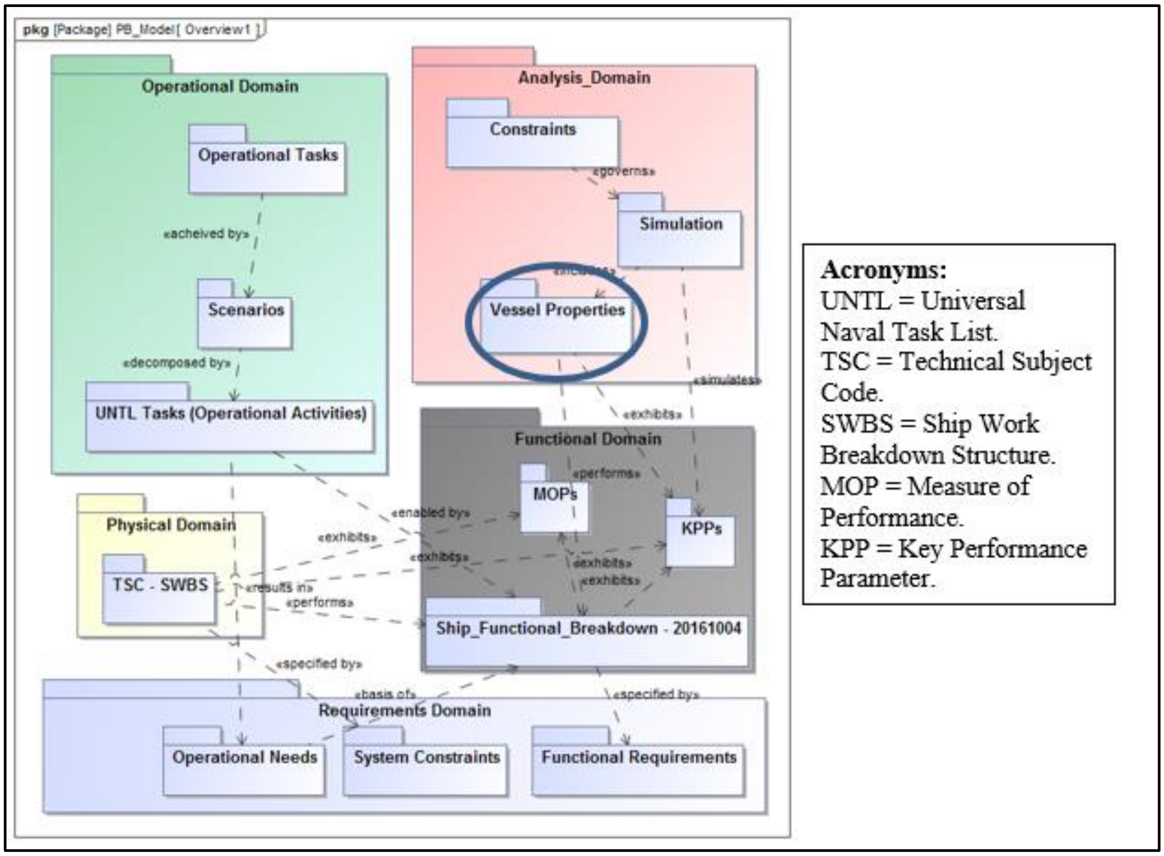


Figure 2. Overview of the MBSE Metamodel Developed as Part of the Research to Construct the MBSE Methodology

Concept and Requirements Exploration

Concept and requirements exploration (C&RE), or requirements elucidation, is an approach to early stage naval vessel design that “responds to a stated mission need with an early high-level assessment of a broad range of ship design options and technologies” (Brown, 2013). A review of the open literature found that several C&RE approaches to support the early stages of naval vessel acquisition projects have been developed in recent years. A summary of the naval vessel C&RE methodologies identified within the open literature and reviewed for this research, along with the features, or approaches they comprise is given in Table 1. The C&RE approaches in Table 1 are typically focused on identifying optimal concept designs for the operational missions the vessel will perform. This knowledge can then be used to ensure the emergent requirements are “elucidated” (McDonald, Andrews, & Pawling, 2012) in an iterative manner, through engagement between the acquirers and designers.

Table 1. Summary of Naval Vessel C&RE Methodologies Reviewed and the Approaches They Include: Model-Based Systems Engineering (MBSE), Modelling and Simulation (M&S), Design Space Exploration (DSE), and Multi-Disciplinary Analysis and Optimisation (MDAO)

C&RE Approach and Key References	MBSE	M&S	DSE	MDAO	Other	Comments
Virginia Tech. Concept & Requirements Exploration (C&RE) Brown & Thomas, 1998; Kerns, Brown, & Woodward, 2011a; Kerns, Brown, & Woodward, 2011b; and Brown, 2013	X	X	X	X	Value model (Analytical Hierarchy Process [AHP]) used for overall measure of effectiveness (MOE).	Uses MBSE to manage ship and mission architecture, separate ship synthesis, operational effectiveness models (OEMs), and MDAO models to analyse effectiveness and optimise.
Response Surface Methods (RSM) Approach Hootman, 2003 and Fox, 2011		X	X		Includes AHP for "rolling up" lower level MOPs	Approaches use separate ship synthesis and OEMs to build concept design space. No explicit link to requirements.
SubOA/IPSM Nordin, 2015; Harrison et al., 2012		X	X			Both approaches use OEMs for submarine option/configuration evaluation during conceptual design. No integration with MBSE models.
Design Building Block (DBB) Andrews, 2006 and McDonald et al., 2012		X	X		Hullform performance (e.g., seakeeping, resistance, and stability) can be simulated using a synthesised CAD model.	Approach facilitates rapid synthesis of a computer aided design (CAD) hullform based on ship functions.

The OTS constraint on the solution space, which is limited to the range of existing designs in the market, arguably not only changes the nature of the required SE approach to middle-out, but it also changes the nature of the C&RE. The need to optimise concept designs is negated and the discussion between stakeholders (especially the navy users) and acquirers changes from eliciting needs and requirements to identifying KPPs and discussing the degree to which existing designs may satisfy them. To inform this discussion, a market survey activity needs to be incorporated into the concept and requirements exploration approach in order to identify whether suitable designs for the operational needs already exist. If they do not, the needs will need to be revisited and adjusted until they reflect the marketplace, or a case needs to be made that the capability risk is unacceptable and a developmental acquisition strategy, rather than OTS, is required. An overview of the C&RE part of the MBSE methodology to support Australian OTS acquisitions, which includes its latest refinements, is shown in Figure 3.

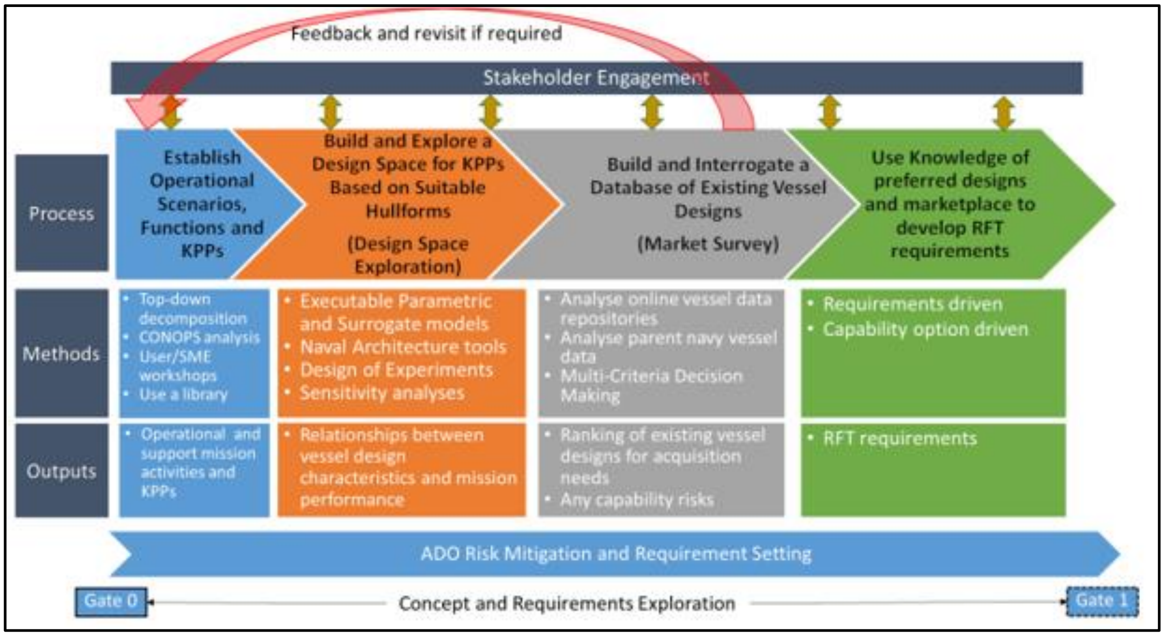


Figure 3. Overview of the Off-the-Shelf Concept and Requirements Exploration Methodology

From Figure 3 it can be seen that three of the features from the existing C&RE approaches in Table 1, MBSE, design space exploration, and modelling and simulation, can be used in the OTS C&RE approach. It is also noteworthy the OTS C&RE approach can be used to support activities and tasks within the ISO/IEC/15288:2015 (ISO/IEC/IEEE, 2015) technical processes: Business or Mission Analysis (e.g., defining the problem space), Stakeholder Needs and Requirements Definition (e.g., analyse stakeholder requirements), System Requirements Definition (e.g., maintain traceability of requirements) and Architecture Definition (e.g., relate the architecture to design). Rather than discuss each stage of the C&RE approach in detail here, in the following section an example implementation of the C&RE part of the MBSE methodology to support Australian OTS naval vessel acquisitions is covered. This provides an overview of each step and the methods that can be used to generate the necessary outputs in the context of an indicative acquisition of a hydrographic and oceanographic survey capability.

Hydrographic and Oceanographic Survey Capability Example Implementation

The example implementation covered in this section was undertaken as part of the constructive research approach, where the artefact (in this case the MBSE methodology) is tested through application. The case study is based on an exemplar strategic need for a military hydrographic and oceanographic survey capability. The assumed solution system concept employs a ship in combination with an array of uninhabited systems that perform the survey functions. This concept could use a range of vessel types, so part of the study involved investigating the suitability of three hullform types currently in service with the Royal Australian Navy. To bound the design space, several assumptions were made: firstly, the vessel hullform was assumed to be monohull; secondly, the vessel length was constrained to be a maximum of 95 metres; and finally, the area of operations was assumed to have sea-state four conditions as the most commonly occurring conditions. Constraints such as these would typically be imposed on a naval acquisition due to considerations such as the planned area of operations and the need to utilise existing port infrastructure.

Step 1: Establish the Mission Scenario and Key Performance Parameters

The first step in the C&RE part of the MBSE methodology is to identify the missions, scenarios, and key performance parameters (KPPs) for the capability being acquired. This step is performed in a top-down manner, where the top-level needs are decomposed into mission scenarios comprising the required operational activities. The operational activities can then be traced through the system functions to the KPPs for the capability. The KPPs are considered to be "a critical subset of the performance parameters representing those capabilities and characteristics so significant that failure to meet the threshold value of performance can be cause for the concept or system selected to be re-evaluated or the project reassessed or terminated" (Roedler & Jones, 2005).

As shown in the suitable methods for step one in Figure 3, the top-down decomposition of the top-level capability needs to establish the mission scenarios and KPPs can be undertaken using information developed and captured in a concept of operations, or by consulting subject matter experts (SMEs). The use of MBSE enables this top-down decomposition to be captured in a model, which can then be linked to the potential design space via the KPPs as discussed in the next step. Using MBSE also enables the model to be reused for subsequent naval vessel acquisitions. In line with guiding principle number three above, MBSE models can be collected over several acquisitions to form a repository, or library, containing SME knowledge of the mission scenarios and KPPs for naval missions.

Figure 4 is a partial view from the MBSE model developed during the example implementation that shows the top-down decomposition from the strategic needs to the KPPs for the "move" and "launch and recover objects to/from the sea" system functions (only some of the operational needs, system functions, and performance characteristics are shown for clarity). In the example implementation the representative mission scenario (the "operational activity" stereotype elements within the blue rectangle in Figure 4) and KPPs (the "MOP [Performance Characteristic]" stereotype elements in the red rectangle in Figure 4) were elicited from SMEs in a workshop setting. In this manner, the design space exploration process undertaken in the next step of the methodology allows capability acquisition stakeholders to trace design decisions through to the capability need. Hence, stakeholders will gain a better understanding of the relationship between design decisions and the requirements, assisting the requirements definition process.

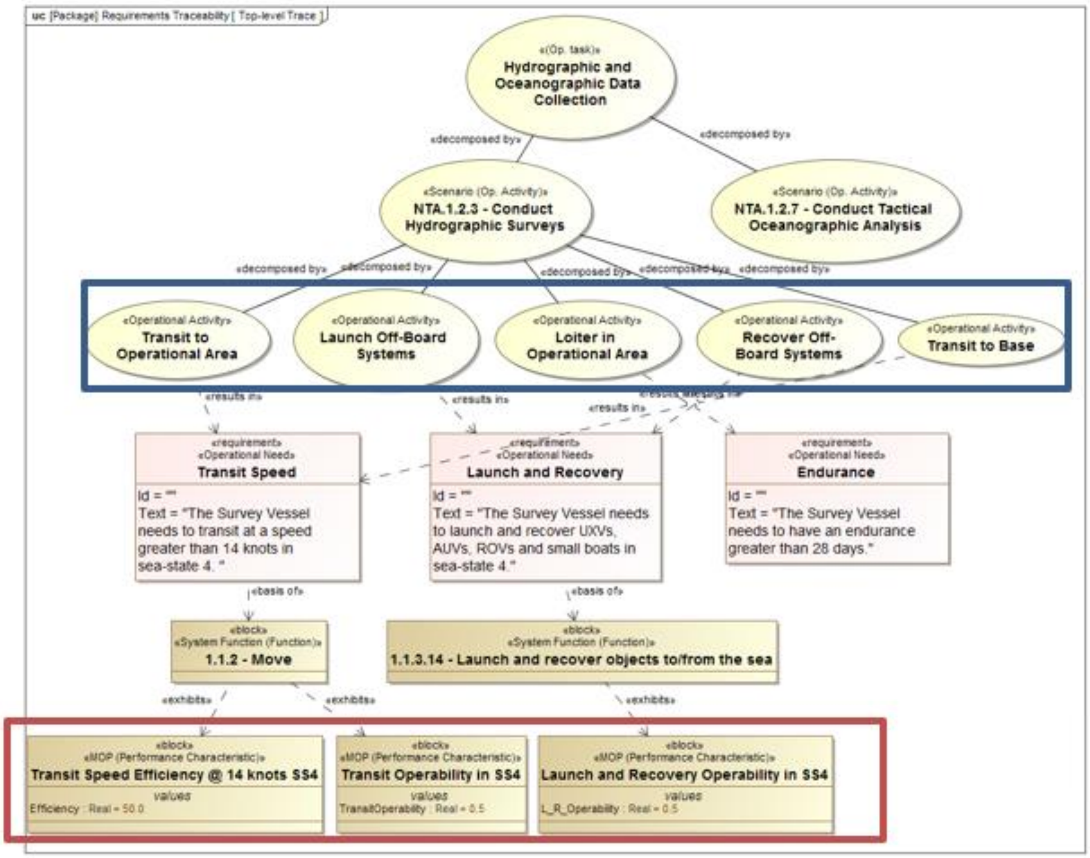


Figure 4. Decomposition From High-Level Guidance Through to the KPPs Related to the Transit Speed and Launch and Recovery Operational Needs

Step 2: Generate and Explore a Design Space Based on Existing Hullforms

In this step, models to calculate KPPs for vessel designs are developed and used to generate a design space that provides stakeholders with insights into relationships between vessel design characteristics and mission performance. These models can range from low-fidelity parametric and surrogate models of relationships between MOPs and ship design parameters, to higher fidelity simulation models that use three-dimensional ship geometries and linear or non-linear solvers. A multi-fidelity approach that uses a combination of high and low-fidelity models can be adopted for this step as the computational and human effort required to implement only high-fidelity simulations at this early stage of the lifecycle is not practical. Basing the models on existing hullforms ensures realistic, feasible design spaces are generated with the OTS constraint in mind. Again, libraries of models can be built over time and reused in subsequent acquisitions.

After tracing in a top-down manner from high-level guidance to the KPPs in the MBSE model during the previous step of the MBSE methodology, in this step, a representation of an existing vessel is captured as value properties in an instantiation of a “vessel properties” stereotype element in the MBSE model. The vessel properties element can then be traced through simulation model element, and KPPs calculated for the instantiation. This is shown in Figure 2 in the red analysis domain elements, where the vessel properties package containing a representation of a vessel “exhibits” the KPPs. The simulation element in Figure 2 (within the red analysis domain package) is linked to executable models through parametric diagrams containing the “constraints” that are built within the MBSE model. Used in conjunction with model integration software or parametric diagram solving software, this approach enables analyses to be conducted, managed and stored from within an MBSE model.

In the example implementation for the hydrographic survey capability, a multi-fidelity approach was used. This approach included the use of the low-fidelity empirical model given by Mennen (1982) to predict the calm-water resistance of the ship representation, as well as the use of a higher fidelity frequency domain seakeeping program (McTaggart, 1997) to predict the motions, as well as the added resistance of the ship representation in waves. The ship representation was a set of roughly 20 design parameters that were extracted from a three-dimensional CAD model. To build views of the design space for the KPPs identified in the previous step, three parent hullforms were systematically varied between the upper length constraint of 95 metres and a lower limit of 65 metres in length. The three hullforms investigated were a hydrographic survey vessel hullform, a frigate hullform, and an offshore patrol vessel hullform. These hullforms were selected as the concept of using a range of uninhabited systems to undertake the data collection activities could conceivably use any available navy ship as a transport platform provided the uninhabited systems are modular in nature. To help ensure the generated design spaces were realistic, the hydrographic vessel and frigate hullforms currently in service with the Royal Australian Navy were used as the parent hullforms that were systematically varied.

A Design of Experiments (DOE) approach (1000 run orthogonal array) was adopted to create a matrix of vessel designs across the design space that were run through the seakeeping and resistance simulation models to calculate their KPP values. This investigation, which was covered in Dwyer and Morris (2017), identified the hydrographic survey hullform as having superior performance with respect to the launch and recovery and transit operability KPPs, as well as being a more efficient hullform when transiting in 14 knots in sea state 4. This means the hydrographic survey hullform is the most suitable for the operational needs in this example implementation. A scatterplot of the results for the hydrographic survey vessel hullform’s seakeeping operabilities during transit and launch and recovery operations, as well as the transit speed efficiency (a measure of the total vessel resistance relative to its displacement) at a transit speed of 14 knots, are shown in Figure 5. The data from the DOE shown in the scatterplot can be used to ascertain the vessel particulars of the best performing generated designs on the pareto front (designs inside the red triangle in Figure 5). These designs exhibit the combinations of highest operabilities and lowest total resistance per tonne of displacement. Some of the vessel particulars for the best performing designs that were generated in the DOE from the pareto front within the red triangle on Figure 5 are shown in Table 2. The block coefficient of these designs is provided to give an indication of the hullform fullness.

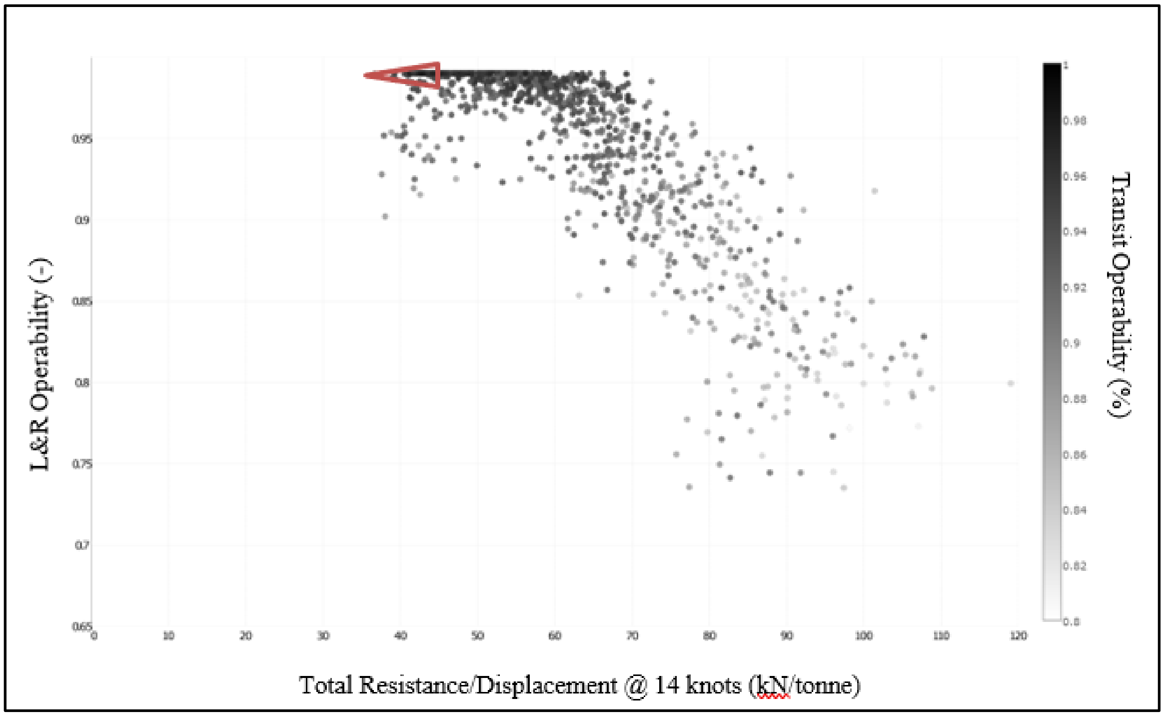


Figure 5. Scatterplot of the 1000 Run DOE for the Hydrographic Survey Vessel Hullform in Sea State 4

Table 2. Vessel Particulars of the Best Performing Designs From the DOE in Sea State 4

Generated Design	Length Overall (m)	Length/Beam	Beam/Draft	Displacement (tonnes)	Block Coefficient
775	95	4.12	3.34	9135	0.6089
337	94.2	4.26	3.47	7871	0.5926
786	95	4.09	3.97	7850	0.6085
796	95	4.48	3.50	7301	0.5971
334	94.2	4.36	3.59	7018	0.5840
785	95	4.26	3.86	7055	0.5785
135	93.4	4.12	3.95	7252	0.6024
482	87.7	4.05	3.56	6443	0.5628
317	90.2	4.16	3.70	7155	0.6322

Furthermore, by analysing the vessel data from the design space using standard correlation techniques, the sensitivity of the vessel performance relative to its design parameters can be established. This sensitivity can be used to identify favourable combinations of design parameters that maximise mission performance. Figure 6 shows the design parameter sensitivities for the transit operability in sea state 4 KPP. This shows that vessel length has a large positive influence on transit operability as it increases and that the length-to-beam ratio has a negative influence as it increases. This shows that as both the vessel length and length-to-beam ratio increase there is a positive influence and negative influence on transit operability respectively.

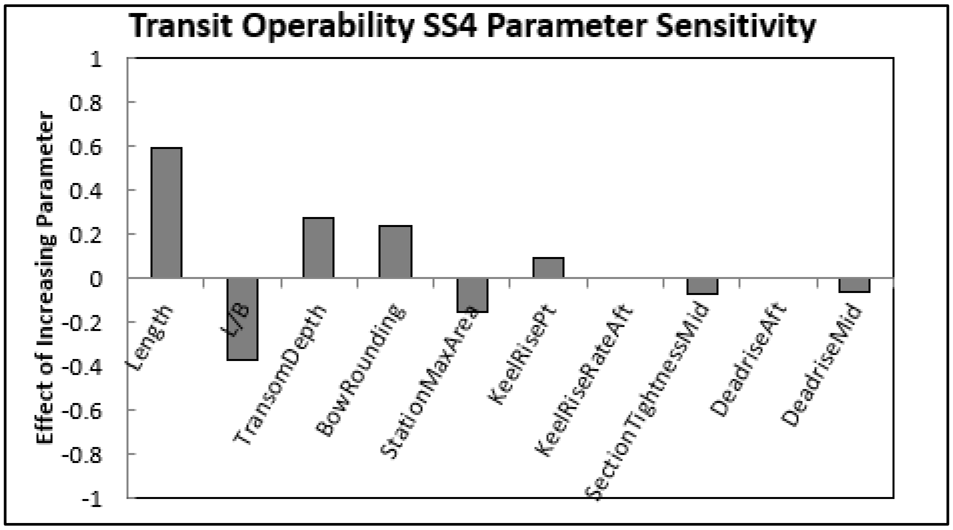


Figure 6. Vessel Design Parameter Sensitivity for the Transit Operability in Sea State 4 KPP

Figure 7 shows the vessel design parameter sensitivities for the launch and recovery operability in sea state 4 KPP. Figure 7 also shows that like the transit operability, increasing both the vessel length and length-to-beam ratio has a positive influence and negative influence on the launch and recovery operability respectively, even though the limits are different for launch and recovery. These aspects are likely to be intuitive to the naval architect, however, this exploration of the design space allows other stakeholders to quantify the effects and make decisions on requirements definition based on robust analysis.

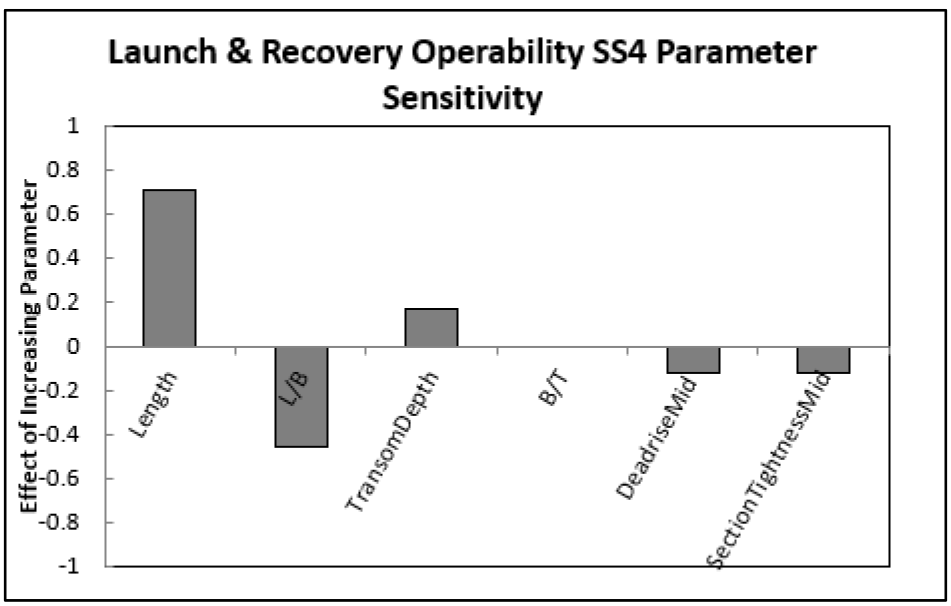


Figure 7. Vessel Design Parameter Sensitivity for the Launch and Recovery Operability in Sea State 4 KPP

Step 3: Build and Interrogate Database of Existing Designs

This step within the concept and requirements exploration part of the MBSE methodology is a preliminary market survey activity. This activity supports the definition of requirements that reflect the OTS naval vessel design marketplace in a bottom-up manner by constraining the solution space to existing designs. Furthermore, this step in the methodology can assist in identifying any capability risks associated with the OTS constraint, as the mission performance of OTS can be estimated using the data from the previous step.

This step uses the knowledge gained from the previous step to build, then rank a database of existing vessel designs based on the preferred combinations of design parameters. For the hydrographic survey vessel example implementation, a database of existing designs was built from relevant existing vessel design data contained in the Janes IHS database (IHS, 2017). Then, using the knowledge gained about the vessel design parameter sensitivities in the previous step of the MBSE methodology, the vessels in the database were ranked. Two key design parameters were used to rank the designs. The first ranking criterion was vessel length, since increasing vessel length had the highest sensitivity metric and therefore the greatest influence on both operabilities, as well as the transit efficiency. The second ranking criterion is the length-to-beam ratio, since the length-to-beam ratio had the second greatest sensitivity metric considered in the example implementation. Other vessel design parameters could have been used to rank the designs, however, a shortcoming of the database used in this example implementation was the limited number of vessel design parameters it contained. This will be a shortcoming present in most OTS acquisitions as the acquirer is unlikely to have access to extensive OTS vessel design data.

In the hydrographic survey vessel example implementation, the vessel ranking was performed using the multi-attribute value analysis method, where the overall weighted value of each vessel in the database was calculated based on a summation of the swing weights of its length and length-to-beam ratio. The weights were calculated from the ranks of the sensitivities of the vessel design parameters (vessel length first and length-to-beam-ratio second) using the Rank Order Centroid technique from Buede (2000). Value curves for length (greater value as it increases) and the length-to-beam ratio (greater value as it decreases) were assumed to be linear with a positive and negative gradient respectively. Design data for the top 10 vessels in the database with lengths between 65 and 95 metres is shown in Table 3.

Table 3. Top 10 Entries in the Existing Vessel Database Based on the Vessel's Length and Length-to-Beam Ratio

Rank	Displacement (tonnes)	Length (m)	Beam (m)	Length/Beam	Speed (knots)	Range (nm)	Crew
1	6421	89.9	19.1	4.71	15	12000	33
2	2889	87	14.6	5.96	15	12000	31
3	3477	85.7	15	5.71	14	11000	58
4	3455	83.5	16	5.21	15	11300	22
5	2991	85	14.1	6.03	14	10060	23
6	3024	72.5	15.24	4.76	12	10500	20
7	2164	76.8	12.8	6.00	14.5	10000	24
8	2205	71.2	15.2	4.68	14	18000	61
9	2382	67.5	15.3	4.41	16.5	22000	22
10	2298	68.3	13.1	5.21	11	19000	49

The database and interrogation tool were set up in a spreadsheet application, which was then wrapped into the MBSE model as an external analysis via model integration software. The key vessel design parameter's ranks and the gradients of the values curves are held as SysML value properties in a Block type element, "key design parameters" within the "vessel properties" package in the MBSE model as shown in Figure 8. The "vessel properties" package is an element within the analysis domain in the metamodel as shown in the blue oval in Figure 2.

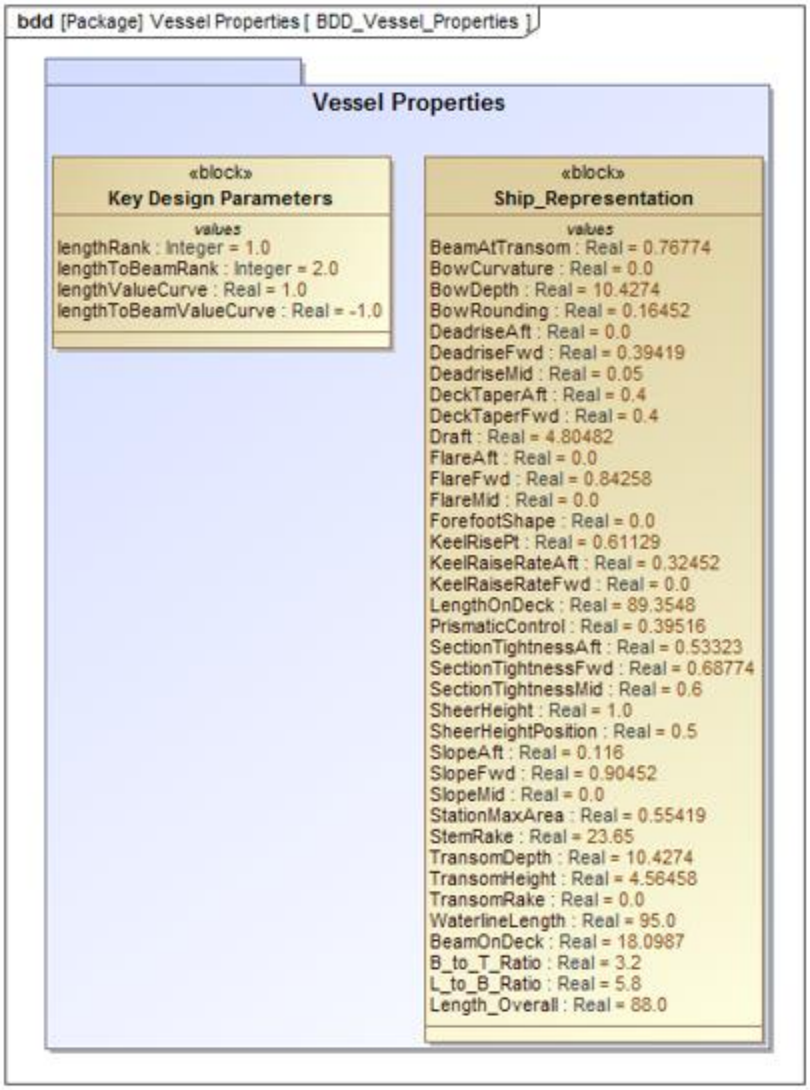


Figure 8. Vessel Properties Package Within the MBSE Model Built During the Example Implementation

The top-ranked designs from the database can be investigated further to establish their suitability for the capability needs. In this stage of the investigation, aspects such as the operating navy, year of design, and country of origin of the designer can be established, as well as refinement of the top-ranking vessels based on any key criteria, such as the range and crew size. The year of design should be an important consideration, since, as the aforementioned analogy between the OTS strategy and "modified repeat" ship design approach highlighted, the approaches work best when the follow-on ships have nearly identical legislative and operational requirements.

In considering whether there are any capability risks for the operational needs due to the OTS constraint for the hydrographic and oceanographic survey vessel example implementation, the data from the top-ranking existing vessels can be cross-checked against the data from the design space generated in the previous step. By comparing the top-ranked existing designs in Table 3 with the top performing generated designs in Table 2, some inferences can be drawn. Firstly, there does not appear to be many existing designs with vessel particulars similar to the optimal designs in Table 2. This could suggest some of the top performing generated designs may be unrealistic, or conversely, there is a gap in the marketplace. To investigate further, relationships between vessel length and the KPPs were generated from the 1000 run hydrographic survey vessel hullform DOE as shown in Figure 9. From Figure 9, it can be seen that the slope of both the launch and recovery (L&R) and transit operabilities decreases as the vessel length grows from approximately 85 metres to 95 metres. This means there is likely to be only marginal improvements in the operability of hullforms to be gained in acquiring a design longer than 90 metres up to the 95 metre limit used in this implementation. This provides a degree of confidence, that the existing vessels larger than roughly 85 metres in length, provided they have a typical hydrographic survey vessel hullform, will have high L&R operability and be capable of meeting the operational needs for the example implementation. This implies there is only low capability risk and that there is no need to revisit the missions and KPPs established in the first step of the MBSE methodology as shown in Figure 3. However, it is a concern that only the top-ranked existing design in Table 3 appears to be close to the optimal region of the design space for the KPPs considered in this example implementation. In a full implementation there would be other KPPs such as acquisition and through-life costs that would impact the decision on whether to revisit the missions and KPP and step through the methodology again.

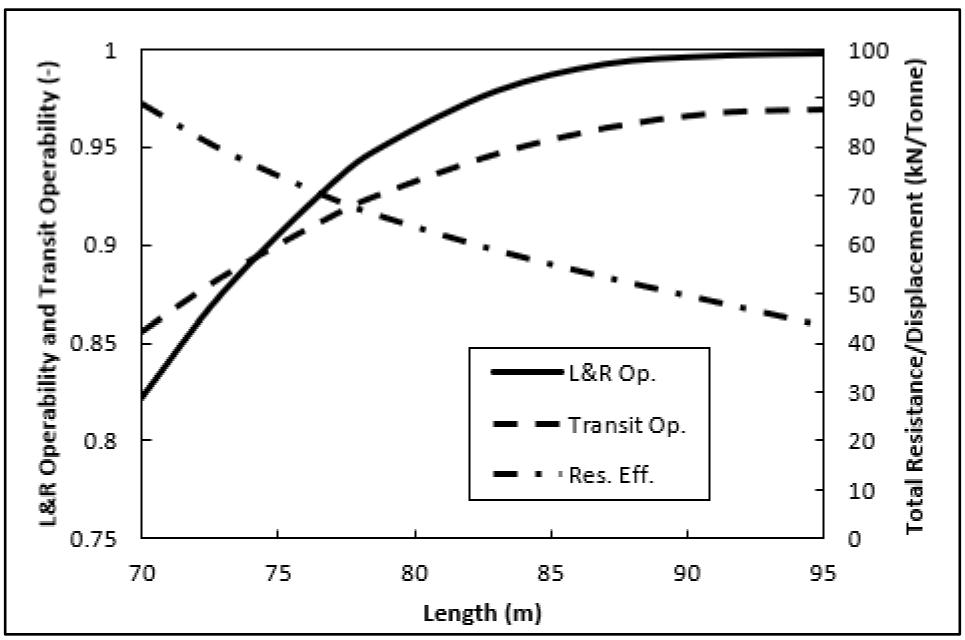
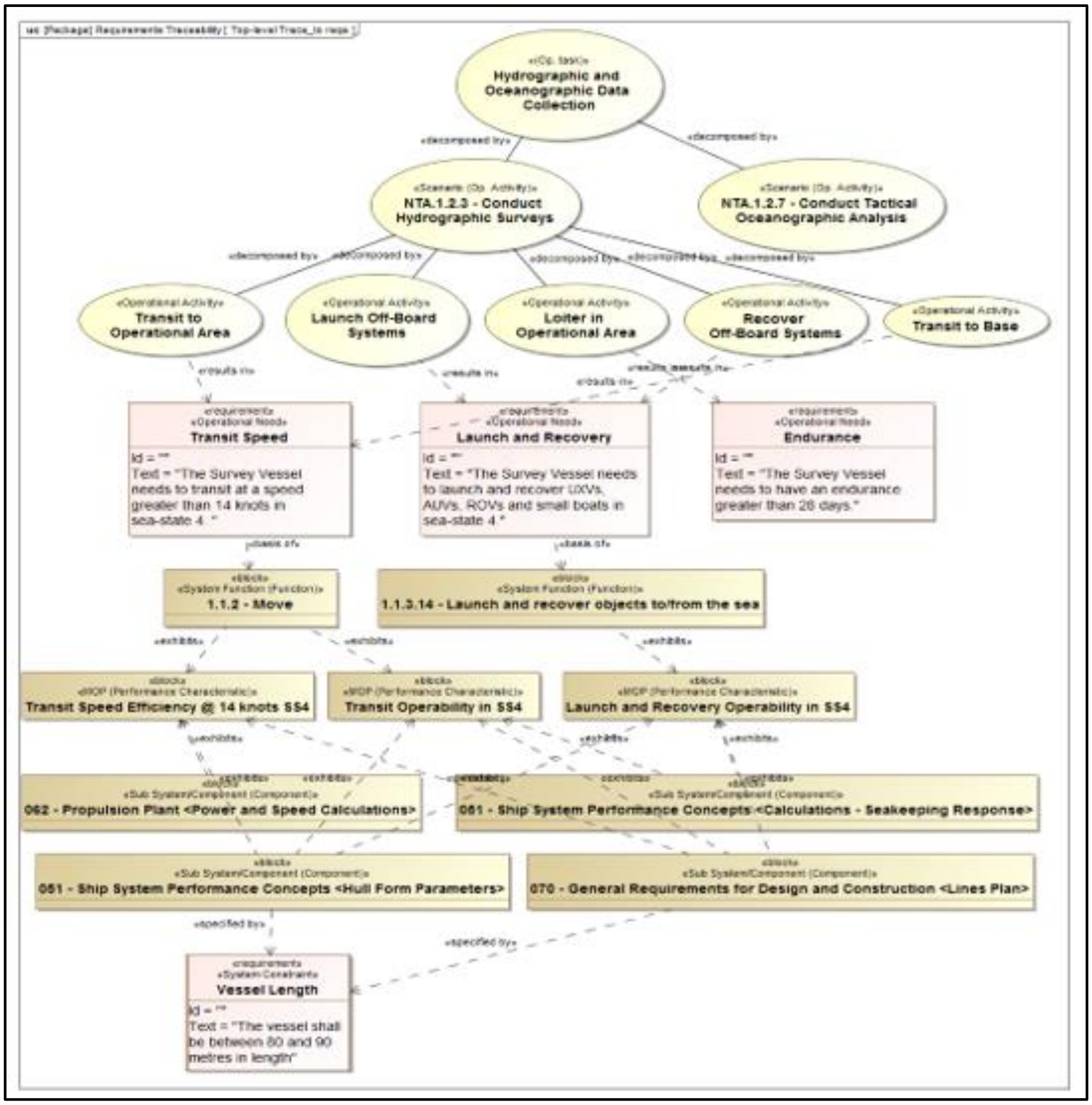


Figure 9. Relationships Between Vessel Length and the Operabilities (L&R Op. and Transit Op.) in Sea State 4 KPPs and Transit Speed Efficiency (Res. Eff.) in Sea State 4 KPP for the Hydrographic Survey Hullform

A final point worth noting in this step is that differences between the optimal combinations of vessel design parameters identified in the design space exploration and the suitable existing vessel designs identified in this step could provide opportunities for design changes. Although this technically violates the OTS constraint, some design changes from the existing design are typically made due to legislative and other requirements differences. If the design changes are affordable, it seems to make sense to pursue changes that could increase performance for the KPPs of the naval vessel being acquired. These design changes could be driven by the requirements released to industry as discussed in the next step.

Step 4: Set Request-for-Tender Requirements

For the hydrographic and oceanographic survey vessel example implementation, the design space exploration (Step 2) and interrogation of existing designs (Step 3) have shown that we can be reasonably confident there are vessels in the marketplace that have been designed to meet similar needs. We can narrow the field of potential respondents to the request for tender by including a constraint on the vessel size to be between 80 and 90 metres in length. We can do this with a degree of confidence that there are existing designs in the marketplace within this range and it will also limit responses to those that are most likely to meet the operational needs. Including the constraint in the request for tender (RFT) requirements can be done in a traceable manner within the MBSE model by continuing the traceability to the KPPs shown in Figure 4, through the ship systems that exhibit the KPPs to the system constraint or requirement. As an example, the vessel length constraint can be included in the MBSE model as shown in Figure 10. Other constraints and requirements can be set and included in the RFT in a similar manner.



Note. Only a partial mapping is shown for clarity.

Figure 10. MBSE Model View Showing the Traceability of the "Vessel Length" Constraint to Be Included in the RFT Requirements to the High-Level Guidance That Triggered the Acquisition

By imposing constraints in the request for tender requirements using the knowledge gained of optimal designs during the design space exploration step, it could encourage designers to propose variants of existing designs that are already close to the optimum. This should not pose a significant risk to the acquisition provided the designer is an established and reputable designer.

Conclusions

This paper covered the latest iteration of research to construct an MBSE methodology to support Australian OTS naval vessel acquisitions. The focus was on the concept and requirements exploration part of the methodology, which was refined to include an explicit market survey activity during this latest iteration. Previously, the C&RE approach relied on parametric and surrogate models based on existing vessel design data to generate a design space representative of the OTS vessel marketplace.

Two main recommendations for further work arose during the research covered in this paper. Firstly, it is recommended to test the MBSE methodology for an actual acquisition in order to satisfy the "holistic market test" part of CRA. This would gain valuable insights into the utility MBSE methodology and provide data for further refinements. Secondly, further research is required to investigate techniques that could be used to estimate the value of KPPs for existing designs based on a low-level of design data being available. This is the situation the above-the-line acquirer is faced with during the early stages of naval vessel acquisitions. Generally, the acquirer will only have access to publicly available design data, which is often insufficient (as shown during the market survey step in the example implementation above) to make a robust estimate of the design's performance.

In response to the research problem identified in the introduction to this paper, an easily implementable MBSE methodology has been developed that supports knowledge generation, capture and reuse during Australian off-the-shelf naval vessel acquisitions. The methodology supports defensible decision making through evidence-based analysis and traceability to the strategic capability needs.

References

Andrews, D. J. (2006). Simulation and the design building block approach in the design of ships and other complex systems. *Proceedings of the Royal Society A, 462,* 3407–3433.

Brown, A., & Thomas, M. (1998). *Reengineering the naval ship concept design process*. Paper presented at the From Research to Reality in Ship Systems Engineering Symposium.

Brown, A. J. (2013). *Application of operational effectiveness models in naval ship concept exploration and design*. Paper presented at the Third International Ship Design and Naval Engineering Congress, Cartagena, Columbia.

Buede, D. M. (2000). *The engineering design of systems: Models and methods*. New York, NY: John Wiley & Sons.

CNO. (2007). *Universal naval task list (UNTL)* (OPNAVINST 3500.38B). Washington, DC. Retrieved from https://www.uscg.mil/directives/cim/3000-3999/CIM_3500_1B.pdf

Coffield, B. (2016). The system of systems approach (SOSA). *INSIGHT, 19,* 39–42.

Cook, S., Do, Q., Robinson, K., Lay, M., & Niedbala, M. (2014). *Progress on using MBSE models as key information artefacts in project tendering.* Paper presented at the SETE2014, Adelaide, Australia.

Cook, S. C., & Unewisse, M. H. (2017). *A SoS approach for engineering capability programs.* Paper presented at the 27th Annual INCOSE International Symposium, Adelaide, Australia.

Covich, P., & Hammes, M. (1983). Repeat ship designs facts and myths. *Naval Engineers Journal, 95*(3), 101–108. doi:10.1111/j.1559-3584.1983.tb01629.x

Crnkovic, G. D. (2010). Constructive research and info-computational knowledge generation. In *Model-based reasoning in science and technology* (pp. 359–380). Springer.

Crowder, J. A., Carbone, J. N., & Demijohn, R. (2016). The overall systems engineering design. In *Multidisciplinary systems engineering: Architecting the design process* (pp. 105–127). Springer International.

Defence. (2017a). *Interim capability life cycle manual*. Canberra, Australia: Department of Defence.

Defence. (2017b). *Naval shipbuilding plan*. Canberra, Australia: Commonwealth of Australia. Retrieved from http://www.defence.gov.au/navalshipbuildingplan/Docs/NavalShipbuildingPlan.pdf

DoD. (2015a). *Manual for the operation of the Joint Capabilities Integration and Development System* (CJCSM 3170.01H). Retrieved from www.dtic.mil/cjcs_directives/cdata/unlimit/3170_01.pdf

DoD. (2015b). *Operation of the defense acquisition system—Change 2 2017*. Retrieved from http://www.dtic.mil/whs/directives/corres/pdf/500002p.pdf

Dwyer, D., & Morris, B. A. (2017). *A ship performance modelling and simulation framework to support requirements setting*. Paper presented at the Pacific 2017 International Maritime Conference, Sydney, Australia.

Ferris, T. L. J., Cook, S. C., & Honour, E. C. (2005). *Towards a structure for systems engineering research*. Paper presented at the 15th Annual International Symposium of the International Council on Systems Engineering, Rochester, NY.

Fox, J. (2011). *A capability-based, meta-model approach to combatant ship design* (Master's thesis). Monterey, CA: Naval Postgraduate School.

Friedenthal, S., Moore, A., & Steiner, R. (2009). *A practical guide to SysML: The systems modeling language.* Burlington, MA: Morgan Kaufmann OMG Press.

Harrison, S., Rodgers, J., Wharington, J., & Demediuk, S. (2012*). Analysis of platform configurations using the integrated platform system model*. Paper presented at the Pacific 2012, Sydney, Australia.

Hodge, R. J., & Cook, S. C. (2014). Australian national security and the Australian Department of Defence. In A. Gorod, B. E. White, V. Ireland, S. J. Gandhi, & B. Sauser (Eds.), *Case studies in system of systems, enterprise systems, and complex systems engineering*. CRC Press.

Hootman, J. C. (2003). *A military effectiveness and decision making framework for naval ship design and acquisition* (Master's thesis). Cambridge, MA: Massachusetts Institute of Technology.

IHS. (2017). Jane's. Retrieved from http://janes.ihs.com/janes/home

ISO/IEC/IEEE. (2015). ISO/IEC/IEEE 15288:2015 *Systems and software engineering–Systems life cycle processes* (Vol. ISO/IEC/IEEE 15288:2015). Switzerland: ISO/IEC/IEEE.

Kerns, C., Brown, A., & Woodward, D. (2011a). *Application of a DoDAF total-ship system architecture in building a design reference mission for assessing naval ship operational effectiveness.* Paper presented at the ASNE Global Deterrence and Defense Symposium, Bloomington, IL.

Kerns, C., Brown, A., & Woodward, D. (2011b). *Application of a DoDAF total-ship system architecture in building naval ship operational effectiveness models*. Paper presented at the MAST Americas, Washington, DC.

Logan, P. W., Morris, B., Harvey, D., & Gordon, L. (2013). *Model-based systems engineering metamodel: Roadmap for systems engineering process*. Paper presented at the SETE 2013, Canberra, Australia.

McDonald, T. P., Andrews, D. J., & Pawling, R. D. (2012). A demonstration of an advanced library based approach to the initial design exploration of different hullform configurations. *Computer-Aided Design, 44*(2012), 209–223.

McTaggart, K. A. (1997). *SHIPMO7: An updated strip theory program for predicting ship motions and sea loads in waves*.

Mennen, H. (1982). An approximate power prediction method. *International Shipbuilding Progress.*

Morris, B., & Sterling, G. (2012). *Linking the defence white paper to system architecture using an aligned process model in capability definition*. Paper presented at the SETE APCOSE 2012, Brisbane, Australia.

Morris, B. A. (2014). *Blending operations analysis and system development during early conceptual design of naval ships*. Paper presented at the SETE2014, Adelaide, Australia.

Morris, B. A., & Cook, S. C. (2017). *A model-based method for design option evaluation of off-the-shelf naval platforms*. Paper presented at the INCOSE IS2017, Adelaide, Australia.

Morris, B. A., Cook, S. C., & Cannon, S. M. (2018, January–February). A methodology to support early stage off-the-shelf naval platform acquisitions. *International Journal of Maritime Engineering, 160.*

Morris, B. A., Harvey, D., Robinson, K. P., & Cook, S. C. (2016). *Issues in conceptual design and MBSE successes: Insights from the model-based conceptual design surveys.* Paper presented at the 26th Annual INCOSE International Symposium (IS 2016), Edinburgh, Scotland, UK.

Nordin, M. (2015). Operational analysis in system design of submarines during the early phases. *Transactions of the Royal Institution of Naval Architects Part A: International Journal of Maritime Engineering, 157*, A65–A84. doi:10.3940/rina.ijme.2015.al.312

Peever, D. (2015). First principles review: Creating one defence. Retrieved from http://www.defence.gov.au/publications/reviews/firstprinciples/

Piirainen, K. A., & Gonzalez, R. A. (2013). *Seeking constructive synergy: Design science and the constructive research approach.* Paper presented at the DESRIST 2013: 8th International Conference on Design Science at the Intersection of Physical and Virtual Design.

Roedler, G. J., & Jones, C. (2005). *Technical measurement*.

Saunders, S. J. (2013). *A framework for harmonising systems engineering and off-the-shelf procurement processes*. Paper presented at the INCOSE 2013 International Symposium, Philadelphia, PA.

SFAD&TC. (2012). Procurement procedures for defence capital projects. Retrieved from http://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Foreign_Affairs_Defence_and_Trade/Completed_inquiries/2010-13/procurement/report/index

Viliers, M. R. D. (2012). Models for interpretive information systems research, Part 2. In M. Mora, O. Gelman, A. Steenkamp, & M. S. Raisinghani (Eds.), *Research methodologies, innovations and philosophies in software systems engineering and information systems*. Hershey, PA: IGI Global.

Walden, D. D., Roedler, G. J., Forsberg, K. J., Hamelin, R. D., & Shortell, T. M. (Eds.). (2015). *Systems engineering handbook: A guide for system life cycle processes and activities* (4th ed.). San Diego, CA: INCOSE.

Enabling Operationally Adaptive Forces

Dan Boger—served in the U.S. Navy from 1968 to 1975. He earned his PhD at the University of California, Berkeley in 1979, when he joined the faculty at the Naval Postgraduate School. He currently serves as the Chairman of the Information Sciences Department. [dboger@nps.edu]

Charles Deleot—served in the U.S. Navy from 1967 to 1990. He holds three master's degrees in management and IT related fields. He served as a senior civilian advisor on the staff of the Commander, Pacific Fleet from 1970 to 2002. He serves as the President of the Patriot Foundation in Southern Pines, NC. This nonprofit has raised over $5 million for dependents of killed or wounded Army Special Force soldiers. [cdeleot@nc.rr.com]

Norman Eaglestone—is a senior software engineer who has supported Open Management Group standards development since the 1990s. He served at various IT related companies, including Sun Microsystems, Northup Grumman, SAIC, and Sierra Nevada. [eaglestoneconsultingllc@gmail.com]

Scot Miller—is a Faculty Associate–Research at the Naval Postgraduate School. He served in the U.S. Navy from 1978 to 2007 and at the Naval Postgraduate School since. He holds master's degrees in three separate fields. [samille1@nps.edu]

Scott Rosa—is currently a Major serving in the U.S. Marine Corps. He is a graduate student at the Naval Postgraduate School. [sjrosa@nps.edu]

Abstract

Emerging warfare trends demand an operationally adaptive force, ready to adjust material solutions, such as systems and systems integrations, in near real time. Software is the most important element of those changes. The DoD has a poor track record with software development, as well as in requirement development, semantic interoperability, and cyber awareness and defense, to name just a few. Mechanical and aeronautical engineering migrated to machine-based designing and testing two decades ago, with transformative results. Software engineering has lagged in this transformation, but our research shows that it has reached the tipping point. What does this mean for the DoD? It means that formal models will enable very rapid capability development, integration, test and evaluation, semantic interoperability, and cyber assessment and remediation, changing the way the entire DoD acquisition enterprise performs. We envision a virtuous improvement cycle where costs spiral down, speed to capability accelerates, and performance increases, all due to formal models.

Introduction

General McChrystal authored a book in 2016 about fighting the global war on terror. He concluded his best successes occurred when he and his forces rapidly adapted, since his lesser experienced, lesser resourced foes always changed their tactical approach (Collins et al., 2015).

These foes enjoyed a faster adaptiveness loop at first than McChrystal's forces, but as his experience grew, his forces, too, learned to be adaptive—adaptive not only in adjusting tactics, but in using equipment, systems, and applications differently (Collins et al., 2015). This often was a struggle, as the cumbersome acquisition and support processes struggled to keep up. Engaging emerging near-peer threats in the future will demand an even more resilient and adaptive U.S. military force.

Our research team postulates that if reforming this lumbering set of processes (requirements, procurement, test, cyber, etc.) was possible, it would enable all operators to be adaptive in near real time. Serving as acquisition professionals, IT engineers, and operators at the tactical edge, this research team has over 200 years of related experience. Our first realization was that nearly every technical requirement needing adaptiveness relied in whole or in part on software, so that is where we focused our efforts.

We also learned, and proved, that yes, it is quite possible to grow an adaptive acquisition enterprise. What follows is why, how, and what remains to be done.

Operational Adaptability

We start with *what is operational adaptability*? The research team established four broad requirements.

First, achieve rapid requirements collection. Operators often know soonest when something needs help, change, or adjustment. We still need to capture that emerging requirement and insert it into the acquisition-related processes, including validation and funding.

Second, given a validated and funded requirement, accelerate problem fixing or new capability delivery. Today that means months or years; we need to be like Google, incorporating a nearly continuous cycle of improvement! That implies an equally responsive test and evaluation approach as well.

Third, leverage current legacy applications and data sources. This means rapid integration as well as platform provisioning. The DoD invests billions in systems, so reusing them makes sense as they do have value.

Finally, account for cyber impacts. Introducing new capabilities and novel integrated systems-of-systems mash-ups means delivering potentially vulnerable systems, where the operational risk is not understood. That would be unsatisfactory. Any reform to our acquisition processes means enabling rapid risk management and corrections.

These four components need to work together. When an actual requirement is identified, the DoD needs rapid validation and funding to support the agents responsible for an attentive response. Careful consideration of legacy apps and data needs to be included, while none of this should proceed without including cyber defense and awareness as part of the overall process. Our DoD processes need to be highly integrated and supportive of one another.

Acquisition and Related Processes as an Adaptiveness Enabler Today

So how are our monolithic acquisition processes doing now, compared to these four components? In a word, terrible. We won’t repeat the disaster stories of many programs, but there are many. How do our current approaches match up to the vision outlined above?

Most agree that requirements are tricky. Operator input is a must, yet often operators just want to slightly improve their current capabilities, and are completely unaware of technical opportunities. That makes sense, of course, since they do have real work to do. Incorporating emerging capabilities is a must as well, though. Too often it seems that the DoD wants to adopt the newest IT technologies without enough thought. For instance, in 2003 FORCEnet was the big C4I theme of the day in the Navy. Gray beard technologists recommended a service-oriented architecture. The first page of every SOA book says, “Naturally, don’t try SOA if you are not working in a business with a well-connected network and well understood business processes” (Brauel et al., 2009). The Navy discovered that intermittent satellite links do not equate to a well-connected network. Our doctrine was well understood, but seldom followed! This was not a recipe for SOA implementation. The research team believes the same misunderstandings exist today in the DoD for jumping on the Cloud and AI “bandwagons.” We are on the precipice of making grave mistakes in our IT investments. We certainly do not have the rapid requirements capture and funding process needed.

For rapid capability delivery, there are actually many examples of success. However, just about all of these required high-level phone calls, tons of money, and the transport of expensive technical tiger teams to faraway places. That approach is unsustainable. For most programs, the answer is no, it takes much longer than it should, by anyone’s measure. Consider friends of the researchers in PEO C4I’s PMW-150. In 2008 they rapidly built the Command and Control Rapid Prototyping Continuum, with software engineers embedded with operators. The operational level customer was thrilled with the rapid delivery and impressive results (Fein, 2011). Yet, has that translated to an afloat capability? Almost. But it is a decade later despite the fact that this organization is filled with consummate professionals, forward leaning technicians, and outstanding leadership.

Integrating capabilities today is quite the challenge. Two factors play here. First, if one just integrated one additional system, it is straightforward. Integrating to two is a bit harder. Integrating to five though, proves an N-squared relationship between the number of systems/data sources to be integrated and the number of connections needed. Add maintaining configuration management of all this, and the challenge grows geometrically in yet another dimension.

For cyber, can we agree that our older approaches leave much to be desired? The Risk Management Framework (RMF) process today adds emphasis on early cyber engineering and requires continuous monitoring, steps in the right direction (DoD Instruction, 2014). Yet the process still is time consuming.

Today, each of these components is an independent silo. Yes, RMF is designed to be included in the capability delivery silo, and this is slowly occurring. The remaining silos barely touch, yet are completely dependent on each other. Each silo has a different boss. For instance in the Army, TRADOC owns requirements. Capability delivery is the province of service acquisition and program executive officers. Integration is also their province. Cyber approval for RMF is led by service cyber commands, completely independent of the service acquisition executive and PEOs. Making this responsive is difficult even in the best of times.

Foundational Approach

We researched if our imagined operationally adaptive acquisition process was technically feasible. Our findings were successful. Figure 1 shows the traditional approach at the top and the revised approach at the bottom. This is an example drawn from the mechanical engineering community. Traditionally (at the top of the figure), engineers designed parts and drafted formal drawings for the machinist who converted the drawings into actual prototypes. Next, the part was iteratively tested and improved, until a set of standards were met. Often mechanical engineering students were required to intern on the machine shop floor so they could appreciate the difficulties of translating their drawing into actual parts.

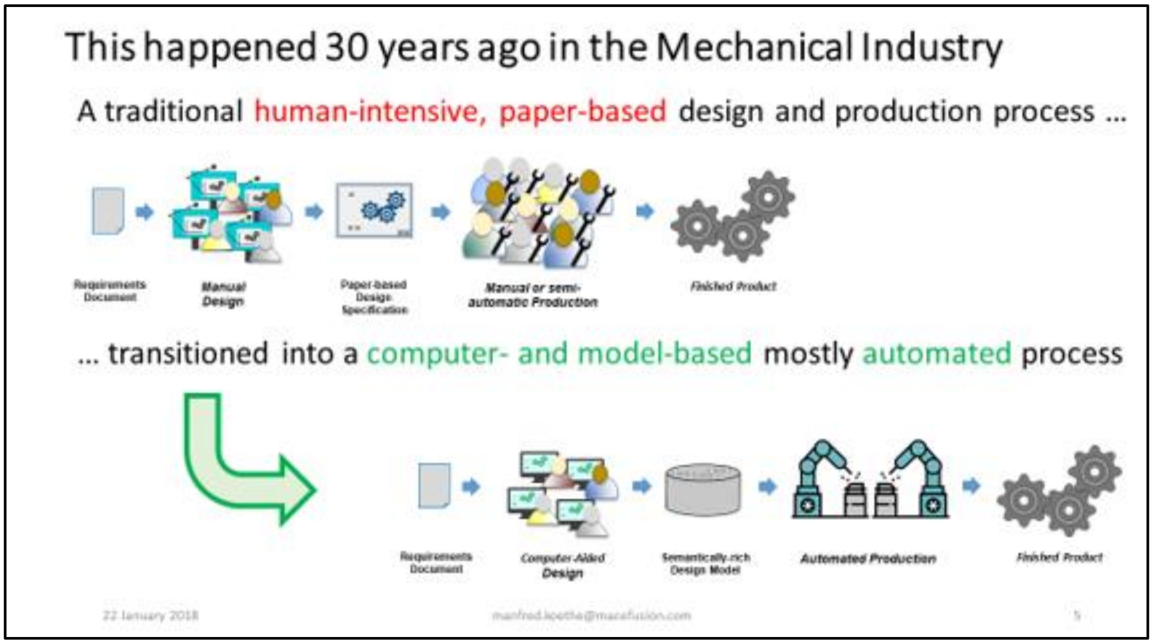


Figure 1. Machine-Based Engineering Transformation
(Koethe, 2017)

The bottom of Figure 1 shows how parts are produced today. Yes, the mechanical engineer produces drawings, but these leverage computer aided design/computer aided manufacturing (CAD/CAM) practices. What comes to the machine shop floor is a digital product. This product undergoes extensive testing in a virtual environment, so there is high confidence the part will work. The machinist programs a robot machinist that produces the part to exacting standards.

We learned that as late as the 1990s, the transmission in most cars was unique, because of the variability in the precision of the manufacture of that transmission. Fixing a transmission meant identifying that a part in the transmission was bad, then replacing the whole transmission. Since tolerances were so tight, parts were not interchangeable (M. Koethe, personal communication, 2017). That is not the case today, no matter what AAMCO says.

Read *Aviation Week* and you will realize that the aeronautical engineering community does the same (Bozdoc, 2006). Even the prototype for our most modern fighter, the F-35, flew within three years of contract award. It took an additional 14 years to create the operational software ("F-35 Initial," 2013)!

We asked ourselves, where is the software engineering equivalent to this mechanical/aeronautical engineering approach? We found it under our noses. The answer is formal software modeling, which is the software development equivalent to engineering CAD/CAM development.

In this approach, software engineers, coders, etc., create models of the functionality they want to develop. Once this formal representation is achieved, it is transformed, depending on the hardware selected, into a true formal model. Automatic code generators produce code, then assess code quality. Once satisfied, this code is provisioned onto the designated platform.

This formal model, just like the digital representation of a mechanical part, can be tested in a virtual environment, including a cyber-environment. "What-if" and engineering trade-off analyses can be easily achieved using parametric modeling. A library of endpoints, which are agreements between this kind of formal model and existing application protocols and data sources, makes integration much faster (N. Eaglestone, personal communication, 2014).

Automatic code analysis tools ensure the code produced is optimized for quality, which is directly related to security. New tools in the development environment enable each integration of systems of systems to achieve semantic interoperability, which is central to achieving success in machine learning, deep learning, and other artificial intelligence techniques.

This approach allows humans to do what they do best: consider all options and employment considerations, understand the operating environment, and address constraints. It allows machines to do what they do best, which is to keep this information for future use and reuse, and to produce code at least 100,000 times faster than humans (Eaglestone, 2012).

What is most promising about these tools and their power is that it will easily enable collaboration between each of the tools' users. Integrating these tools means an enterprise approach to solving the operational adaptation challenges.

While not complete, this approach also enables an emerging capability that is more icing on the cake. Additions to systems engineering processes include designing for man-machine interdependence. Achieving such interdependence, through careful consideration of how the observability, predictability, and directability between men and machines can be achieved, is difficult. This new systems engineering addition allows for establishing the requirements to achieve this interdependence. Such an approach can be easily incorporated into the tools described above (Johnson, 2014). An ability to achieve interdependence (another word is *collaboration*) between man and machine might even support a fourth offset strategy.

Additional Details

Our research uncovered the technical tools to produce a revolution in military affairs. Imagine actually being able to respond in hours or at most, days, to pressing operational needs? This would be a game changer.

These tools are based on open standards developed by the Object Management Group (OMG). Many of your contractor companies send representatives to their meetings (R. Soley, personal communication, 2015). This is not magic, but rather a set of tools that have evolved over the past 20 years and have reached maturity. They are ready to be employed today!

These tools now enable semantic interoperability between integrated systems. This is a huge accomplishment, yet few programs are leveraging this capability. Previously, semantic interoperability could be achieved through very expensive and time consuming one off programming and was brittle to configuration changes. OMG adopted a new Archetype Modeling Language, born from efforts to integrate various health care systems, to achieve semantic interoperability. Our research shows that creating meaning between medical systems is at least as hard as doing so for DoD systems (N. Eaglestone, personal communication, 2016).

We reviewed six separate efforts that used formal modeling approaches. Their project requirements varied from building a simple set of models evaluating counter-battery fire, developing a web portal that assesses software for quality and cyber resilience, and translating Chinese notice to mariner's messages for U.S. nautical chart changes. Table 1 summarizes the type of capabilities and integration required, how long it took to produce, and estimated costs.

As this table shows, these are remarkable project achievements executed in very short amounts of time. Keep in mind that these projects required expert formal modelers. Much of their time was spent building the model; they draft very little actual code. What code they do write is often associated with transforms in the modeling process, not actual functionality of the systems.

Four of the six were purely proofs of concepts, where in every case the sponsor was very satisfied with the results. The cartography project has continued at the National Geospatial Intelligence Agency (NGA), since it was the only viable solution to a growing cartography correction crisis (R. Wicks, personal communication, June 12, 2015). This code assessment portal is under refinement, including a component that would enable continuous code monitoring, an RMF requirement.

One final exciting piece of formal modeling is the ease in producing transforms that convert the formal model into any required program documentation (N. Eaglestone, personal communication, October 12, 2017). For instance, most programs are required to deliver various DoDAF views. That is a simple, minute-long process using formal models.

Table 1. Formal Model Proofs of Concept, 2012–2018
(N. Eaglestone, personal conversation, 2017)

Formal Model Proofs of Concept, 2012–2018				
Project Title/Date	General capability	Systems Integrated	Time	Cost
Counter battery 2012	Parametric modeling of sensors and networks	Four models	Three man months	$100k
Social network analysis 2014	Sensors to computer vision to tweet based alert network	Two sensors, facial recognition software, data bases, basic semantic interoperability; network integration; alert development	Three man months	$150k
Unmanned robot collaboration 2015	Enable air and ground robot to collaborate on finding target of interest; reduce Marine cognitive load	Four sensors, robot operating system, developed robotic command and control, networks integration, user interface on to iPad	Four man months	$350k
Nautical chart correction process prototype 2015	Character and feature recognition, translation, and work flow support	Several databases, character recognition software, semantic interoperability, user interface	Five man months	$100k
Digital Fires 2016	Facial recognition generates call for fire to afloat platform radar and combat system	Ship combat system, missile launcher, radar, facial recognition, and ground robot operations	Three man months	$200k
Code Assessment 2018	Enable code devleopers to upload and assess code	17 different code assessment tools; semnatic interoperability between six different cyber vulnerability data sources, semantic interoperability between all of the above	Seven man months	$200k

Challenges

New processes are not without challenges. We uncovered five significant issues to start with; no doubt there may be others.

First, as one might imagine, creating a formal model is not easy. It takes many iterations between operators and modelers to get the model right. Many program managers grow very impatient with the rate of progress. Therefore, many are unwilling to risk trying an approach that promises such great deliverables but has "nothing" (since there is only a model) to show for months. Our research shows that patience does pay off. Proper prior preparation prevents poor performance. That has been an axiom for software development since the 1950s. Formal modeling is just that to an extreme. But how can that be proven to program managers?

Observers of model-based systems engineering will point to large DoD efforts that focused on using formal models for their objectives, but without any success. For instance, in the early 2000s, the Navy led the Single Integrated Air Picture initiative, which was model-based. No doubt quality engineers and modelers combined with operational experts to create the models. However, the modeling expertise, the standards, and the tools were not quite mature enough to guarantee success (Dinkins, 2006). We believe that the standards and the tools are more rigorous now, and creating the formal models needed to achieve rapid capability delivery are now present.

This leads directly to the third problem. Even today, creating formal models requires the work of master modelers, not journeymen engineers fresh from engineering school. OMG is working on methods to improve the tools that would help solve this conundrum (M. Koethe, personal communication, December 14, 2017). Right now, expert formal modelers demand high hourly rates and receive them. The DoD's own contracting guidelines often prevent us from being able to hire these masters. It's a chicken and egg challenge. How can we train apprentice engineers to be masters if we cannot hire the masters to train them?

The fourth issue relates back to empowering the operators as requirements creators. In a perfect world, operators, working directly with modeling masters, would give input. The modelers would then tease out exactly the meaning, then iterate again (and again) with the operators to ensure correctness. What tools exist to support this process? So far, the answer is very few. In the cartography proof of concept a PowerPoint-based tool, generated by the model itself, was used to provide the operators an idea of what was going on, enabling them to provide feedback (Wicks, 2015). More intuitive tools are needed. Of course, direct interaction between modelers and operators cuts out the entire requirements validation and funding process.

This points toward the last big challenge: Who is in charge of all of this? The service acquisition executive? The individual program managers? The Program Executive Officers? The service's Pentagon staffs? Our research uncovered several possible options, but the common thread was that someone must lead the effort for a long period of time, at least 10 years. This implies someone of passion who is unusually adept in DC political wrangling. The Army is discussing a new command that might actually try to do this (L. Brown, personal communication, January 17, 2018). It bears watching over this summer to see what they decide.

Conclusions and Next Steps

It is not just our research team that explores formal models. An Army team investigated helicopter flight control software by releasing a formal model in their request for proposals, as a sort of acquisition experiment. They used a cost-estimating team to predict what the bids would be, both in time and money. Four proposals said they would do the work for one-fourth the anticipated amount, in half the time. The Army funded three and they all beat their predictions with working software (A. Farrar, personal communication, May 17, 2017).

Formal modeling is real. It works, and it will reduce costs, accelerate delivery, and improve operational performance. Achieving formal models enables parametric modeling, rapid test and evaluation, semantic interoperability, and improved code, all while enabling operators more intimate requirements inputs. This creates a virtuous cycle of continuous improvement in all phases of the requirements, procurement, test, sustainment, and cyber defense processes, with automated document generation.

We invite the readers to join us in our quest to make acquisition the chief enabler of operationally adaptive forces. Please send us any other ideas on how this can occur.

References

Bozdoc, M. (2006). History of computer aided design. Retrieved from http://www.thocp.net/software/software_reference/cad_bozdoc.htm

Brauel, B., Matsumura, M., & Shah, J. (2009). *SOA adoption for dummies*. Hoboken, NJ: Wiley.

Collins, T., Fussel, C., & McChrystal, S. (2015). *Team of teams: New rules of engagement for a complex world*. London, UK: Portfolio.

Dinkins, R. (2006). *Single Integrated Air Picture (SIAP)* [PowerPoint slides]. Retrieved from https://www.ll.mit.edu/HPEC/agendas/proc06/Day2/03_Fairbairn_Pres.pdf

DoD. (2014). *Risk Management Framework (RMF) for DoD information technology (IT)* (DoD Instruction 8510.01) Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/851001_2014.pdf

Eaglestone, N. (2012). *Universal command and control* [PowerPoint slides].

F-35 initial operational capability [Report to Congressional Defense Committees]. (2013). Retrieved from https://breakingdefense.com/documents/services-set-ioc-dates-for-f-35s-confidence-is-the-watchword/

Fein, G. (2011). *ONR develops new acquisition model for delivering information to the fleet.* Retrieved from https://www.onr.navy.mil/en/Media-Center/Press-Releases/2011/C2RPC-Information-Fleet

Johnson, M. (2016). *Co Active Design* [PowerPoint slides]. Retrieved from http://www.ihmc.us/users/mjohnson/publications.html

Koethe, M. (2017). *Model oriented development environment* [Power Point slides].

Wicks, R. (2015). *Notice to Mariners (NtM) Production Automation Limited Objective Experiment (LOE)* [PowerPoint slides].

Panel 14. Ethics, Auditability, and Retention in the Acquisition Workforce

Wednesday, May 9, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: David Slade, Col, USAF (Ret.),** Director, Air Force Acquisition Career Management ***Analysis of Procurement Ethics in the Workplace*** Rene G. Rendon, Naval Postgraduate School ***Auditability in Procurement: An Analysis of DoD Contracting Professionals' Procurement Fraud Knowledge*** Juanita M. Rendon, Naval Postgraduate School ***Contract Compliance and Audit*** Charlie Williams Jr., Section 809 Panel

David Slade, Col, USAF (Ret.)—is the Director of Acquisition Career Management, Assistant Secretary of the Air Force for Acquisition (SAF/AQH). Slade is responsible for the integrated management of the acquisition workforce across all functional areas. He provides acquisition human resources policy and strategic planning while managing the training and development of civilian and military acquisition personnel Air Force–wide. Additionally, Slade ensures Air Force compliance and implementation of the Defense Acquisition Workforce Improvement Act (DAWIA) through management of the Acquisition Professional Development Program (APDP) and the Defense Acquisition Workforce Development Fund (DAWDF). Slade is also designated as the Career Field Manager for both military and civilian Scientists, Engineers, and Acquisition Program Managers. His team also provides personnel management services for the SAF/AQ Headquarters staff.

Slade received an aerospace engineering degree from the University of Colorado and was commissioned through the Reserve Officer Training Corps in 1983. Following pilot training, he served as a forward air controller, flying the O2-A and an F-15C and AT-38 instructor pilot. He served as a Commander at the Squadron and Group levels. As a command pilot with over 3,600 flying hours, he flew 32 missions over Iraq during Operation Desert Storm and has participated in Operations Noble Eagle, Northern Watch, and Southern Watch.

Prior to his current assignment, Slade served as Director of Assignments, Headquarters Air Force Personnel Center, Randolph Air Force Base, TX. He was responsible for the assignment of more than 65,000 officers below the grade of colonel and 285,000 enlisted personnel below the grade of chief master sergeant.

Slade retired from active duty after 29 years in November 2012 in the rank of colonel and entered civil service in January 2013.

Analysis of Procurement Ethics in the Workplace[1]

Rene G. Rendon—is an associate professor at the Naval Postgraduate School in Monterey, CA, where he teaches in the MBA Contract Management program. Prior to joining NPS, he served for more than 20 years as a contracting officer in the United States Air Force. His Air Force career included assignments as a contracting officer for the Peacekeeper ICBM, Maverick Missile, F-22 Raptor, Space Based Infrared Satellite program, and the Evolved Expendable Launch Vehicle rocket program. Dr. Rendon's publications include *Management of Defense Acquisition Projects*, *U.S. Military Program Management: Lessons Learned & Best Practices*, and *Contract Management Organizational Assessment Tools*. Rendon has been published in the *Journal of Contract Management, Journal of Public Procurement, Journal of Purchasing and Supply Management, Journal of Public Affairs Education, Project Management Journal*, and the *Defense Acquisition Research Journal.* [rgrendon@nps.edu]

Introduction

In 2017, the DoD obligated more than $330 billion in contracts for mission-critical supplies and services. This includes the planning, awarding and administering of more than three million contract actions (USA Spending, 2018). DoD contracting officers play a critical role in the contracting process. Contracting officers are the only individuals authorized to award and administer contracts and make related determinations and findings (FAR, 2018). Additionally, contracting officers serve as the primary focal point for contractual issues, managing horizontal interfaces with external organizations, as well as vertical interfaces with internal organization (Rendon & Wilkinson, 2016). This role places contracting officers, in comparison to other members of the DoD workforce, in a challenging position from the perspective of ensuring contracts comply with laws, codes, and regulations. The DoD has established ethical codes of conduct to be observed by every member of the defense workforce. Additionally, the National Contract Management Association (NCMA) has also established a code of ethics for the members of the contract management profession. However, not everyone in the DoD, including senior government officials, or members of the acquisition workforce, may be aware, knowledgeable, or even in compliance with established ethical standards of conduct (Rendon & Rendon, 2015, 2016; Whitely et al., 2017). Thus, contracting officers face additional ethical challenges in ensuring contract management processes are performed in an ethical manner, compared to other members of the DoD workforce. The purpose of this research was to explore ethics and compliance strengths and challenges in the contract management workforce (Rendon & Wilkinson, 2016, pp. 49–50).

Research Approach

The research was supported by the Ethics Research Center (ERC) and used their National Business Ethics survey. The ERC collaborated with the National Contract Management Association (NCMA) to survey the NCMA membership on their current ethics environment and to identify possible ethics risks and challenges. The NCMA membership, which includes buyers and sellers from government and industry, represents the contract

[1] Author's Note: This proceedings paper contains excerpts from Rendon, R. G., & Wilkinson, J. W. (2016, July). Ethics in the workplace: A comparison between the contract management and general business workforces. *Contract Management, 56*(7), 49–58.

management workforce in this comparison. The survey results were then compared to ERC's National Business Ethics Survey (NBES) database of past survey results from the general business workforce. The survey is a voluntary, anonymous, online survey that was deployed to approximately 18,000 NCMA members representing buyers and sellers. Of the eligible survey participants invited to take the survey, 897 responded, resulting in a response rate of 4.9%. The sampling error of the findings is +/- 3.2% at the 95% confidence level. The purpose of this paper is to present the results of the survey and the implications of these results on the DoD contracting workforce (Rendon & Wilkinson, 2016, pp. 49–50).

Research Findings

The survey items focused on four measurable outcomes related to ethics and compliance:

1. Pressure to violate the law
2. Observe misconduct
3. Reported the observed misconduct
4. Experienced retaliation for reporting the misconduct

The survey also included items related to the ethical culture of their organizations (from the perspective of top management, supervisors, and coworkers), strengths of the ethical culture of their organizations, and organizational independence.

As reflected in Figure 1, the survey findings reveal that the contract management workforce, as represented by the responding NCMA membership, felt pressure to violate the law, observed ethical misconduct, and reported the observed misconduct at a higher rate than the general business population (Rendon & Wilkinson, 2016, pp. 50–54). Specifically, the findings indicate the following:

- 23% of the CM workforce respondents experienced pressure to compromise ethical standards or violate the law compared to 9% in the NBES. The greatest sources of pressure included meeting deadlines, satisfying performance goals, interpreting requirements loosely, and allowing vaguely worded contracts. The sources of pressure were reported by more than 40% of the CM workforce respondents.
- 45% of CM workforce respondents observed misconduct compared to 37% in the NBES. The five most-observed procurement-related types of misconduct included improper contract awards (28%), improper use of single source awards (13%), misuse of contract change orders (12%), contract violations (11%), and improper provision of personal services (11%).
- 77% of CM workforce respondents reported the misconduct that they observed compared to 63% in the NBES, and 84% of the survey respondents felt prepared to handle potential misconduct.
- 14% of the CM workforce respondents experienced retaliation compared to 21% in the NBES, and 75% of those respondents reported that retaliatory behavior to their organization's attention.

As reflected in Figure 2, the survey findings reveal that the buyers and sellers differed in their responses to the survey questions (Rendon & Wilkinson, 2016, pp. 50–54). Specifically, the findings indicate the following:

- 30% of buyers felt pressure to violate the law compared to 19% of sellers.
- 55% of buyers observed misconduct compared to 40% of sellers.

- 74% of buyers reported the observed misconduct compared to 82% of sellers.
- 18% of buyers experienced retaliation for reporting the observed misconduct compared to 11% of sellers.

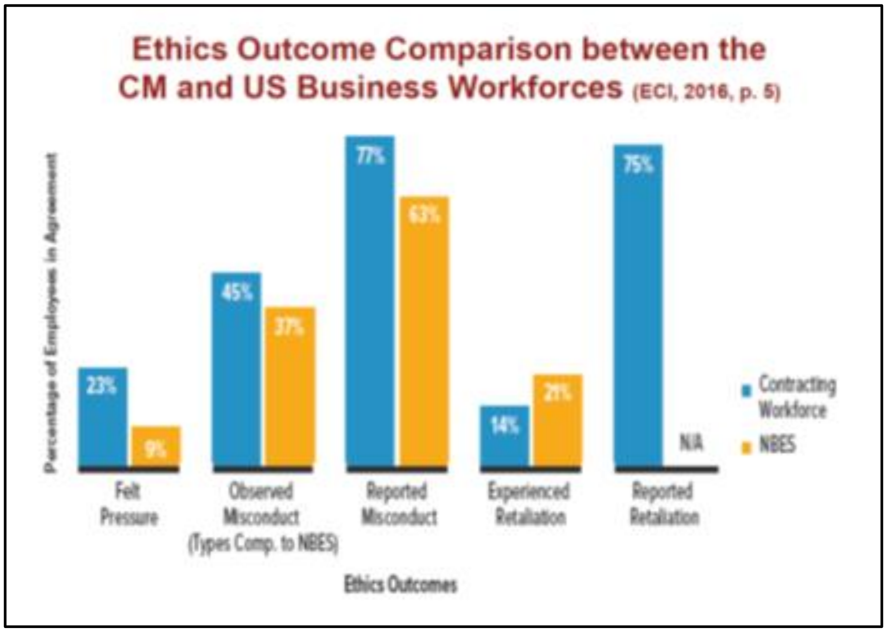


Figure 1. Ethics Outcome Comparison Between the CM and U.S. Business Workforces
(ECI, 2016, p. 5)

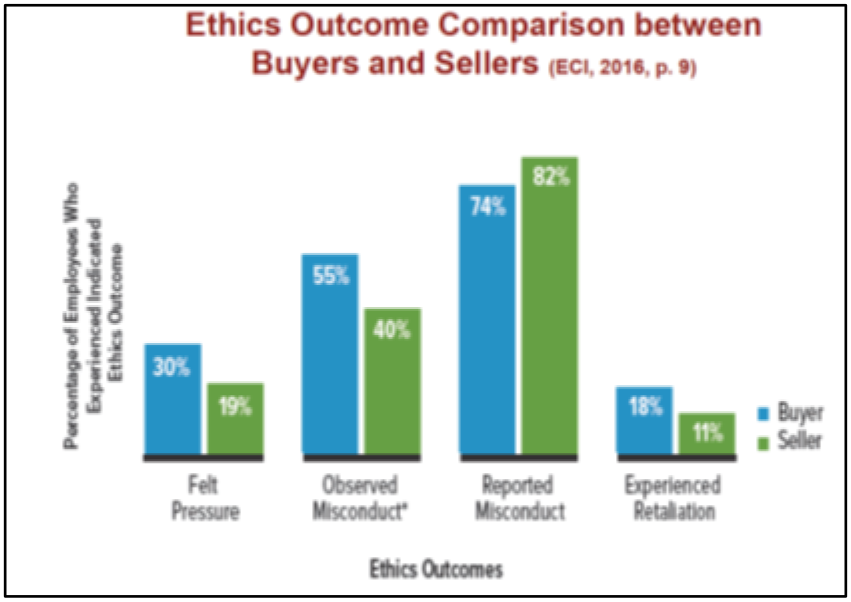


Figure 2. Ethics Outcome Comparison Between Buyers and Sellers
(ECI, 2016, p. 9)

As reflected in Figure 3, the survey findings reveal that the contract management workforce, as represented by the responding NCMA membership, had differing views of their top management, supervisors, and co-workers ethics culture (Rendon & Wilkinson, 2016, pp. 50–54). Specifically, the findings indicate the following:

- 53% of CM workforce respondents feel that the ethics culture among their top managers is strong/strong-leaning compared to 69% in the NBES.
- 63% of CM workforce survey respondents feel that the ethics culture among their supervisors is strong/strong-leaning compared to 68% in the NBES.
- 72% of the CM workforce respondents feel that the ethics culture among their co-workers is strong/strong-leaning compared to 65% in the NBES.

As reflected in Figure 4, the survey findings reveal that the buyers and sellers differed in their responses to the survey questions related to ethics culture (Rendon & Wilkinson, 2016, pp. 50–54). Specifically, the findings indicate the following:

- 58% of the sellers feel that top management ethics culture is strong/strong-leaning compared to 46% of the buyers.
- 66% of sellers feel that the supervisor ethics culture is strong/strong-leaning compared to 60% of buyers.
- 75% of sellers feel that the co-worker ethics culture is strong/strong-leaning compared to 72% of the buyers.

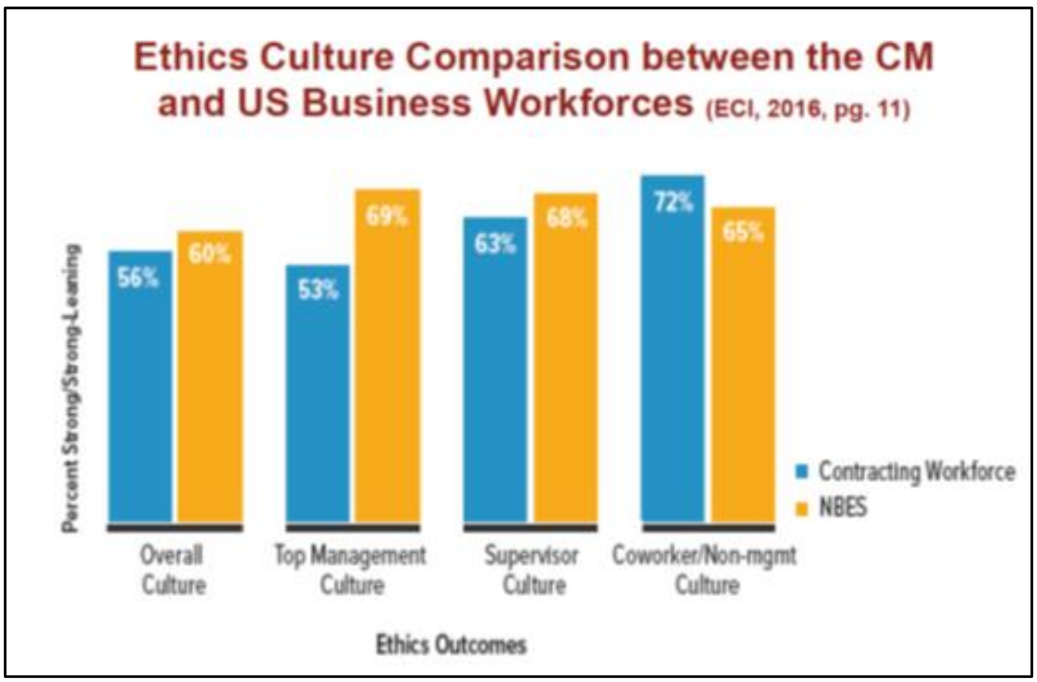


Figure 3. Ethics Culture Comparison Between the CM and U.S. Business Workforces
(ECI, 2016, p. 11)

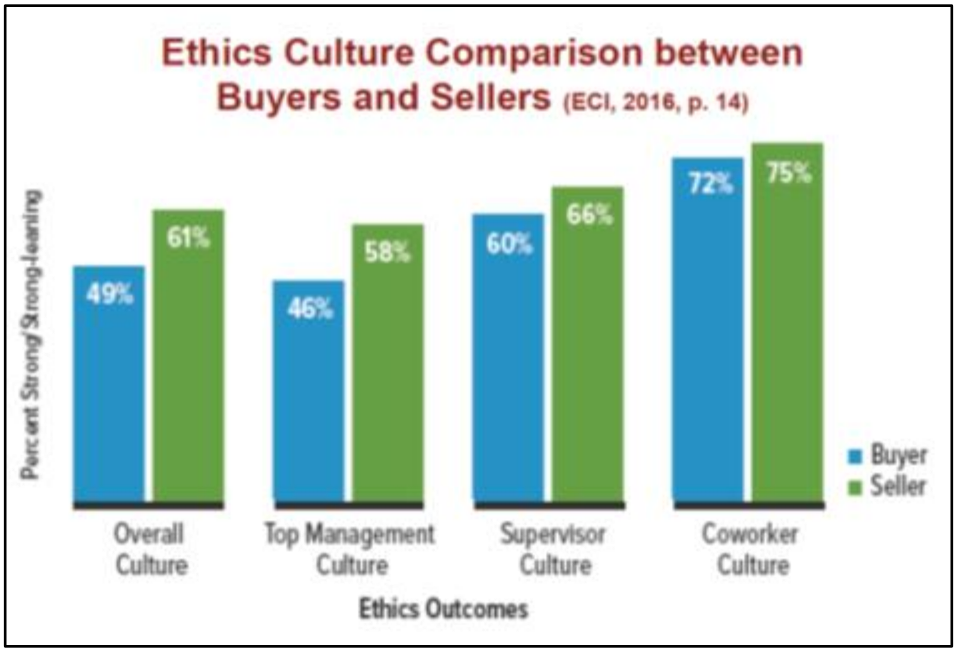


Figure 4. Ethics Culture Comparison Between Buyers and Sellers
(ECI, 2016, p. 14)

Discussion

The survey findings reveal that the contract management workforce felt pressure to violate the law, observed ethical misconduct, and reported the observed misconduct at a higher rate than the general business population. These findings are supported by the previous discussion of the contracting officers' position in the organization. Contract managers are positioned at a pivotal point, interfacing with both internal and external organizations on all contractual matters, giving them a unique vantage point for identifying any ethical violations or procurement fraud "red flags" compared to the general business population. Additionally, since the contract management workforce receives extensive training on procurement integrity and ethical rules, they have a heightened awareness of these ethical requirements and an increased sensitivity to violations in the workplace. The lower percentage of retaliation experienced by the contracting workforce, when reporting misconduct, compared to the general business workforce is also interesting. This finding may indicate that members of the contract management workforce have a stronger commitment to procurement integrity and have a higher level of credibility in the eyes of the senior government officials (Rendon & Wilkinson, 2016, pp. 54–56).

The comparison of these survey results between buyers and sellers is most interesting. The percentage of buyers experiencing the four main ethics outcomes targeted by this survey was higher compared to the sellers. Only the percentage of buyers reporting the observed unethical conduct was lower compared to sellers. One area of further research would be to explore what percentage of buyers in the survey work for government agencies compared to buyers working for industry. A difference in the ethics culture between the contract management workforce in the government and the non-contract-management workforce in the government may be resulting in increased instances of pressure, observed ethical misconduct, and experienced retaliation. Another area of further research would be to explore the extent that the government contract management workforce receives extensive training on procurement integrity and ethical rules as opposed to the non-contract-management workforce within the government. This may lead to determining if only the government contract management workforce has the heightened awareness of these ethical requirements and the increased sensitivity to violations in the workplace. After all, it is the government contracting officer who signs the contract and who is responsible for ensuring the contract management process is conducted in accordance with procurement statutes and regulations. While other members of the acquisition workforce have the same goal to achieve the expected contract performance, they are typically measured by different metrics and may not necessarily share the concerns of the contracting officer (Rendon & Wilkinson, 2016, pp. 54–56).

It is also interesting to note that the survey findings reveal that a lower percentage of the contract management workforce, compared to the general business population, perceives that the ethical culture of top management and supervisors are “strong/strong-leaning.” Yet, a greater percentage of the contract management workforce perceives its co-workers as having a “strong/strong-leaning” ethical culture. Once again, this may be because the members of the contract management workforce perceive themselves as having a stronger commitment to procurement integrity and having a higher level of credibility among their contract management peers (Rendon & Wilkinson, 2016, pp. 54–56).

It is also noteworthy that the strength of the top management, supervisor, and coworker ethical cultures are consistently higher among sellers than among buyers. However, the survey results do not indicate whether the top management and supervisors are within the contracting authority chain of command or within the organizational chain of command. Within the government, many members of the contract management workforce report to more than one supervisor and top manager, some within the contracting chain of authority (e.g., the procuring contracting officer, the level above the procuring contracting officer, the chief of the contracting office, the director of contracting, etc.) as well as within the organizational chain of command (e.g., the project manager, program manager, program executive officer, etc.). Once again, these questions (i.e., what is the percentage of buyers that work for the government and what is the percentage of top managers and supervisors that are part of the contracting chain or the organizational chain) deserve additional investigation to further explore this area (Rendon & Wilkinson, 2016, pp. 54–56).

Conclusion

Contract managers, because of their position in the organization, face additional ethical challenges than many other employees in any given organization. The ERC and NCMA survey of the NCMA membership on their current ethics environment identified possible ethics risks and challenges. The survey findings revealed that the majority of the surveyed contract management workforce did feel pressure to violate the law, did observe ethical misconduct, and reported the observed misconduct at a higher rate than the general business population. The survey also showed that the contract management workforce perceives that the ethical culture of top management and supervisors as “strong/strong-leaning” at a lower rate than the general business population. The implications of the survey findings point to the importance of the DoD workforce (both contracting and non-contracting) being trained in ethics rules and compliance requirements; the importance of contract management processes that are mature, aligned, and supportive of ethics rules and compliance requirements; and the importance of internal controls that are effective in ensuring that the personnel comply with the required contract management processes (Rendon, 2015; Rendon & Wilkinson, 2016, p. 58).

References

Ethics & Compliance Initiative. (2016). *National business ethics survey of the contracting workforce. Arlington, VA: Ethics Research Center*. (To view the complete report, visit the NCMA website—https://www.ncmahq.org.)

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018).

Rendon, J. M., & Rendon, R. G. (2016). Procurement fraud in the U.S. Department of Defense: Implications for contracting processes and internal controls. *Managerial Auditing Journal, 31*(6/7), 748–767.

Rendon, R. G. (2015). Benchmarking contract management process maturity: A case study of the U.S. Navy. *Benchmarking: An International Journal, 22*(7), 1481–1508.

Rendon, R. G., & Rendon, J. M. (2015). Auditability in public procurement: An analysis of internal controls and fraud vulnerability. *International Journal of Procurement Management, 8*(6), 710–730.

Rendon, R. G., & Wilkinson, J. W. (2016, July). Ethics in the workplace: A comparison between the contract management and general business workforces. *Contract Management, 56*(7), 49–58.

USA Spending. (2013). Retrieved from http://www.usaspending.gov/

Whiteley, J. T., Foster, J. A., & Johnson, K. A. (2017). *Contracting for Navy husbanding services: An analysis of the Fat Leonard case*. Monterey, CA: Naval Postgraduate School.

Auditability in Procurement: An Analysis of DoD Contracting Professionals' Procurement Fraud Knowledge

Juanita M. Rendon—is a Certified Public Accountant (CPA), a Certified Fraud Examiner (CFE), and a Lecturer at the Naval Postgraduate School (NPS). Dr. Rendon teaches auditing, finance, and accounting courses at NPS. She is a former IRS Revenue Agent. She is a member of the American Institute of Certified Public Accountants and the Association of Certified Fraud Examiners. She has published articles in the J*ournal of Contract Management, the International Journal of Procurement Management,* and the *Managerial Auditing Journal*, as well as book chapters in *Cost Estimating and Contract Pricing* (2008) and in *Contract Administration* (2009) by Gregory A. Garrett. She has presented at the National Contract Management Association (NCMA) World Congress conferences as well as at NCMA's National Education Seminars and at the Institute of Supply Management (ISM) conference. She is also the recipient of the 2008 NCMA Award for Excellence in Contract Management Research and Writing. [jmrendon@nps.edu]

Abstract

The purpose of this research was to assess Navy contracting professionals' procurement fraud knowledge, as well as contract management processes and related internal controls, and to analyze their perceptions regarding their organization's procurement fraud susceptibility. This research study utilized a previously developed web-based survey designed to assess the DoD procurement workforce's knowledge of procurement fraud schemes, internal controls, and contract management processes as well as their perceptions of fraud susceptibility in each of these areas. Based on the research findings, the Navy may be lacking auditability in their organizations due to a lack of procurement fraud knowledge. Recommendations are provided to the Navy and DoD regarding increasing the procurement fraud knowledge of their contracting professionals in order to help decrease procurement fraud vulnerabilities within their organizations. As DoD agencies continue to strive for accountability, integrity, and transparency in their procurement of goods and services, procurement fraud knowledge and auditability will continue to increase in importance.

Background

The Department of Defense (DoD) procurement workforce manages millions of contract actions, and billions of public dollars are spent on supplies and services in order to achieve the mission of the DoD ("Federal Procurement," 2017). The DoD must ensure that each tax dollar, hard-earned by the American people, is being spent appropriately with the highest degree of public trust.

The Department of Defense Inspector General (DoDIG) and the Government Accountability Office (GAO) have identified issues such as lack of adequately trained contracting personnel, lack of capable contract management processes, and lack of effective contract management internal controls within the federal government (DoDIG, 2009, 2014; GAO, 2013). These contract management issues may make the DoD vulnerable to procurement fraud (Rendon, R. G., & Rendon, 2015).

With procurement fraud cases on the rise, in order to achieve its mission, it is important that the DoD procurement workforce have the necessary procurement fraud knowledge to properly manage the procurement function with integrity, accountability, and transparency (Cohen & Eimicke, 2008; Thai, 2014). Analyzing the procurement fraud knowledge level of Navy contracting professionals and making recommendations for improvement of procurement fraud education within the Navy, as well as within the DoD, can help ensure that taxpayer funds are used effectively and help ensure the public interest is protected. Integrity, accountability, and transparency in federal government procurement are crucial.

The purpose of this research was to assess Navy contracting professionals' procurement fraud knowledge, as well as contract management processes and related internal controls, and to analyze their perceptions regarding their organization's procurement fraud susceptibility. The research questions for this study include the following:

1. What is the Navy contracting professionals' procurement fraud knowledge level of procurement fraud schemes as related to contract management processes, internal control components, and procurement fraud scheme categories?
2. What is the Navy contracting professionals' perception of procurement fraud as related to the contract management processes, internal control components, and procurement fraud scheme categories?

As the DoD works toward being audit ready for a financial statement audit in FY2018, auditability is of utmost importance. For an organization to be auditable, it should ensure that its people are competent, its processes are capable, and its internal controls are effective (Rendon, J. M., & Rendon, 2016). Competent people, which is the focus of this research, are one of the components of the auditability triangle (Rendon, J. M., & Rendon, 2016).

Literature Review

This section provides a brief literature review that sets the groundwork for this research study. Scholarly journal articles, professional journal articles, government reports, and previous research studies in the areas of auditability, contract management phases, internal controls, and procurement fraud scheme categories are discussed. The following sections address auditability, contract management phases, internal control components, procurement fraud scheme categories.

Auditability

Auditability occurs at different levels of an organization and flows from the lowest level of an organization upwards. The process of "making things auditable" requires organizations to establish and actively manage an institutionally acceptable knowledge management system supporting its governance of processes and practices (Power, 1996, p. 289). Rollins and Lanza (2005) support the need for an increased emphasis on effective internal controls due to an increase in procurement fraud cases. In addition, Crawford and Helm (2009) contend that public sector governance is important to ensure a commitment to compliance, accountability, and transparency. Prior research supports the importance of competent personnel and competent organizations related to capable processes in order to ensure successful procurement projects (Frame, 1999).

In response to internal control weaknesses and resulting procurement process deficiencies, the DoD is trying to increase its emphasis on procurement training and the development of procurement workforce competencies (GAO, 2002) as well as auditability in its procurement organizations. Auditability within federal government organizations is necessary in order to ensure the integrity, accountability, and transparency of its procurement programs, fight the battle against procurement fraud, and ensure value for money (Rendon, R. G., & Rendon, 2015).

As reflected in Figure 1, R. G. Rendon and Rendon (2015) contend that auditability encompasses competent personnel, capable processes, and effective internal controls. Having competent people includes personnel having appropriate education, adequate training, and relevant experience. The focus of this research is on competent personnel in terms of procurement fraud knowledge.

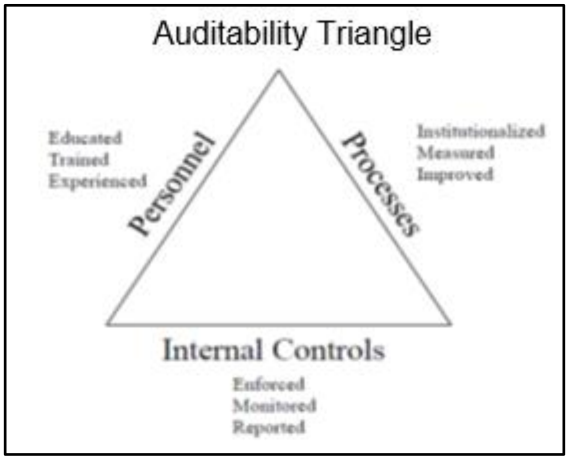


Figure 1. Auditability Triangle
(Rendon, R. G., & Rendon, 2015)

Contract Management Phases

Rendon and Snider (2008) state that the contract management phases include pre-award, award, and post-award. The pre-award phase consists of the procurement planning, solicitation planning, and solicitation processes, which are discussed in the following sections.

Pre-Award: Procurement Planning. Procurement planning is a vital aspect of contract management as it encompasses key activities such as defining the requirement, conducting market research, developing budgets and cost estimates, and conducting risk analysis (Rendon, J. M., & Rendon, 2015).

Pre-Award: Solicitation Planning. Solicitation planning includes key activities such as determining the procurement method and contract type, developing the solicitation document, determining the contract-award strategy, and finalizing the solicitation (Rendon, J. M., & Rendon, 2015).

Pre-Award: Solicitation. The solicitation process involves obtaining information (proposals) from the sellers regarding how project needs can be met (Rendon, R. G., 2008).

Award: Source Selection. The source selection process includes key activities such as applying evaluation criteria to the management, cost, and technical proposals, negotiating with suppliers, and executing the contract award strategy (Garrett, 2013; Rendon, R. G., 2008).

Post-Award: Contract Administration. The contract administration process involves key activities such as conducting a pre-performance conference, monitoring the contractor's work results, measuring the contractor's performance, and managing the contract change control process (Rendon, R. G., 2008).

Post-Award: Contract Closeout. The contract closeout process includes key activities such as processing government property dispositions, finalizing acceptance of products or services, making final contractor payments, and documenting the contractor's final past-performance report (Rendon, R. G., 2008). In addition to capable contracting processes, effective internal controls are also important for federal agencies to become more auditable (Rendon, R. G., & Rendon, 2015).

Internal Control Components

Effective internal controls ensure the organization is "[complying] with laws and regulations, monitoring procedures to assess enforcement, and reporting material weaknesses" (Rendon, R. G., & Rendon, 2015, p. 716). In May 2013, the Committee of Sponsoring Organizations of the Treadway Commission (COSO) updated its internal control integrated framework, which now includes 17 principles within the five components of internal control (COSO, 2013). In September 2014, the GAO updated its *Standards for Internal Control for the Federal Government* (Green Book; GAO, 2014). Figure 2 illustrates the five components of internal control (COSO, 2013, p. 6).

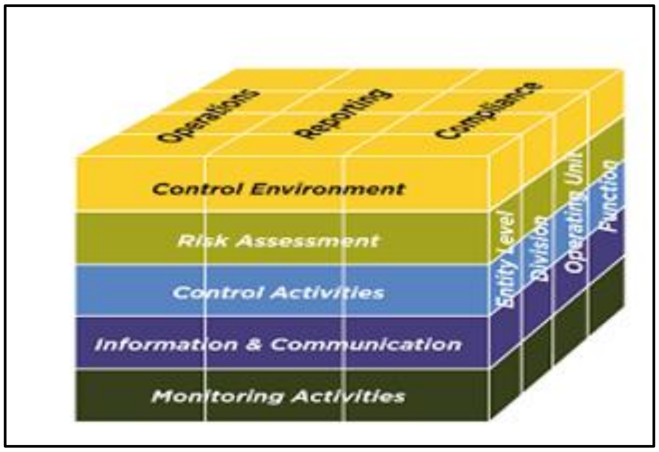


Figure 2. Relationship of Internal Control Objectives and Components
(COSO, 2013, p. 6)

Figure 3 illustrates all of the 17 principles associated with each of the five internal control components. The five components of the integrated internal control framework are discussed in the following sections (COSO, 2013).

Control Environment. The control environment component of the integrated internal control framework sets the tone at the top and is related to the integrity and ethical behavior of the organization's management (COSO, 2013).

Risk Assessment. The risk assessment component of the integrated internal control framework involves assessing what could go wrong within the organization and what management can do to mitigate any potential risks, including fraud risks (COSO, 2013; GAO, 2014).

Control Activities. The control activities component of the integrated internal control framework incorporates all of the control procedures that the organization needs to implement in order to reach its goals and objectives (COSO, 2013).

Information and Communication. The information and communication component of the integrated internal control framework includes internal and external communications as well as the accounting system (COSO, 2013).

Monitoring Activities. The monitoring activities component of the integrated internal control framework entails the close observation of all of the other internal control components to ensure that the controls are being practiced appropriately (COSO, 2013).

Control Environment	Risk Assessment	Control Activities	Information and Communication	Monitoring Activities
• Demonstrates commitment to integrity and ethical values • Exercises oversight responsibility • Establishes structure, authority, and responsibility • Demonstrates commitment to competence • Enforces Accountability	• Sets specific suitable objectives • Identifies and analyzes risk • Assesses fraud risk • Identifies and analyzes significant changes	• Selects and develops control activities • Selects and develops general control over technology • Deploys through policies and procedures	• Uses relevant information • Communicates internally • Communicates externally	• Conducts ongoing and/or separate evaluations • Evaluates and communicates deficiencies

Figure 3. COSO's 17 Fundamental Principles
(COSO, 2013, p. 6)

Procurement Fraud Scheme Categories

Internal controls that are not appropriately mandated and implemented may leave the federal government vulnerable to procurement fraud. Tan (2013) found that incidents of procurement fraud in the DoD and the federal government could be traced to ineffective internal controls, which left government organizations vulnerable to fraud, waste, and abuse. The Association of Certified Fraud Examiners (ACFE), a fraud-fighting organization, defines fraud as "a knowing misrepresentation of the truth or concealment of a material fact to induce another to act to his or her detriment" (ACFE, 2016, para. 2). In the 1940s, after interviewing embezzlers in jail, Cressey (1972), a criminologist, found that that the embezzlers had a perceived pressure (motivation), a perceived opportunity, and a justification (rationalization) in common, now known as the fraud triangle (Wells, 2001). The fraud triangle is illustrated in Figure 4 (Albrecht, 2014).

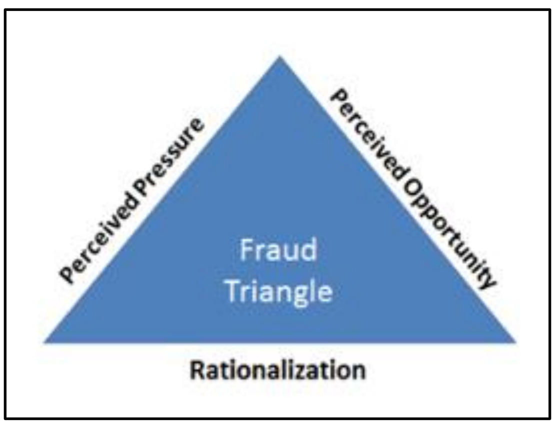


Figure 4. Fraud Triangle
(Albrecht, 2014, para. 1)

While there are numerous fraud schemes, they can be categorized into six major procurement fraud scheme categories, which are illustrated in Table 1 (Rendon, J. M., & Rendon, 2015).

Collusion. Collusion is "a situation where two or more employees work together to commit fraud by overcoming a well-designed internal control system" (Wells, 2005, p. 122). The collusion fraud scheme category includes procurement fraud schemes such as kickbacks, bribery, and deliberate split purchases (Rendon, J. M., & Rendon, 2015).

Bid Rigging. Bid rigging "is a process by which an employee assists a vendor to fraudulently win a contract through the competitive bidding process" (Wells, 2005, p. 283). Bid-rigging schemes include collusion bidding by contractors, excluding qualified bidders, leaking bid data, manipulation of bids, rigged specifications, and unbalanced bidding (Rendon, J. M., & Rendon, 2015).

Conflict of Interest. Conflict of interest is "when an employee, manager, or executive has an undisclosed economic or personal interest in a transaction that adversely affects the company" (Wells, 2005, p. 273). Conflict of interest fraud schemes include conflicts of interest, unjustified sole source awards, and phantom vendors (Rendon, J. M., & Rendon, 2015).

Billing, Cost, and Pricing Schemes. Billing, cost, and pricing schemes involve "fraudulent payment by submitting invoices for fictitious goods or services, inflated invoices, or invoices for personal purchases" (Wells, 2005, p. 98). Billing, cost, and pricing schemes include such things as cost mischarging; defective pricing; change order abuse; co-mingling of contracts; false, inflated, or duplicate invoices; and false statement claims (Rendon, J. M., & Rendon, 2015).

Fraudulent Purchases. Fraudulent purchases involve purchasing "personal items with company money" (Wells, 2005, p. 114). Fraudulent purchases include purchases for personal use or resale, unnecessary purchases, and imprest fund abuse (Rendon, J. M., & Rendon, 2015). GAO (2002) found that fraudulent purchases occur in the government purchase card programs within the federal government.

Fraudulent Representation. Fraudulent representation includes failure to meet contract specifications and product substitution (Rendon, J. M., & Rendon, 2015). Product substitution is also known as "bait and switch."

Table 1. Categories of Procurement Fraud Schemes
(Rendon, J. M., & Rendon, 2015)

Categories of Procurement Fraud Schemes					
Collusion	**Conflict of Interest**	**Bid Rigging**	**Billing/Cost/ Pricing Schemes**	**Fraudulent Purchases**	**Fraudulent Representation**
Bribes & Kickbacks	Conflict of Interest	Collusive Bidding by Contractors	Cost Mischarging	Purchases for Personal Use or Resale	Failure to Meet Contract Specifications
Split Purchases	Unjustified Sole Source Awards	Excluding Qualified Bidders	Defective Pricing	Unnecessary Purchases	Product Substitution
	Phantom Vendor	Leaking Bid Data	Change Order Abuse	Imprest Fund Abuse	
		Manipulations of Bids	Co-mingling of Contracts		
		Rigged Specifications	False, Inflated, or Duplicate Invoices		
		Unbalanced Bidding	False Statement and Claims		

Conceptual Framework

The conceptual framework for this research entails the auditability triangle. R. G. Rendon and Rendon (2015) contend that "the theory of auditability incorporates aspects of governance which emphasizes effective internal controls, capable processes, and competent personnel" (p. 715). These major elements of the auditability triangle, which are illustrated in Figure 1, present the conceptual framework for this research and focus on the competent personnel element.

In order for federal procurement organizations to be auditable, they need to have competent people, capable processes, and effective internal controls. Since contracting professionals play an essential role in the procurement process, they have unique opportunities for detecting and deterring procurement fraud. However, without proper and adequate knowledge of procurement fraud schemes, as well as effective internal controls and capable contracting processes, these contracting professionals may not be able to deter or detect significant procurement fraud activities within the federal government organizations.

Research Methodology

The research methodology for this research study includes a literature review covering contract management phases, internal control components, and procurement fraud schemes. The literature review consists of the GAO reports as well as nongovernmental literature and scholarly articles. Furthermore, this research methodology involved the use of a previously developed knowledge assessment tool that was used to assess Navy contracting professionals.

The web-based assessment tool includes 27 knowledge-based questions regarding contracting processes, internal controls, and procurement fraud schemes. In addition, the assessment tool also includes 12 organization-based questions related to the contracting officers' perceptions of internal controls within their organizations. These survey questions were designed to assess the contracting officers' perceptions of their organizations regarding susceptibility to fraudulent activity. The organization-based items were adopted and modified from the Internal Control Survey developed by the New York State Internal Control Association (NYSICA, 2006).

After following the appropriate Institutional Review Board (IRB) procedures and obtaining the protocol approval from the Naval Postgraduate School IRB office, the web-based assessment tool was deployed using the Naval Postgraduate School online survey-hosting service LimeSurvey. The survey link was e-mailed to a Navy-designated person who was not in the chain of command, who forwarded the e-mail message with the web link to the Navy contracting professionals at a Navy contracting command. The web-based assessment tool was available for a four-week period. Based on the research findings, recommendations are made to the Navy and the DoD for improving its contracting professionals' procurement fraud knowledge as well as its contract management processes and internal controls.

Research Findings

The web-based assessment tool was deployed on January 26, 2016, to a total eligible population of 82 Navy contracting professionals located at a Navy contracting command. The assessment tool was initiated by 44 respondents, and was completed by 32 respondents, resulting in a response rate of 39% (Grennan & McCrory, 2016).

All of the 32 respondents were Navy civilian contracting professionals. Figures 5–7 reflect demographics of the respondents. The figures show the number of respondents as well as the percentage. For example, 1, 3% for the 11 to 20 years category in Figure 5 indicates one respondent, which was 3% of the total respondents had 11 to 20 years of experience. Regarding the experience level, the majority of the respondents (10 respondents, 32%) had 0–2 years of experience. Regarding their DAWIA levels shown in Figure 6, the majority of the respondents (15 respondents, 47%) had DAWIA Certification Level II, and 22% (7 respondents) had no DAWIA certification levels. Regarding their warrant status shown in Figure 7, the majority of the respondents (78.13%) did not have a warrant.

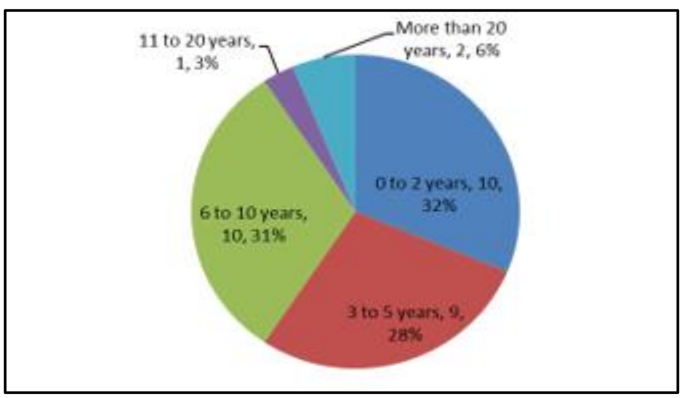


Figure 5. Number of Participants by Years of Experience
(Grennan & McCrory, 2016)

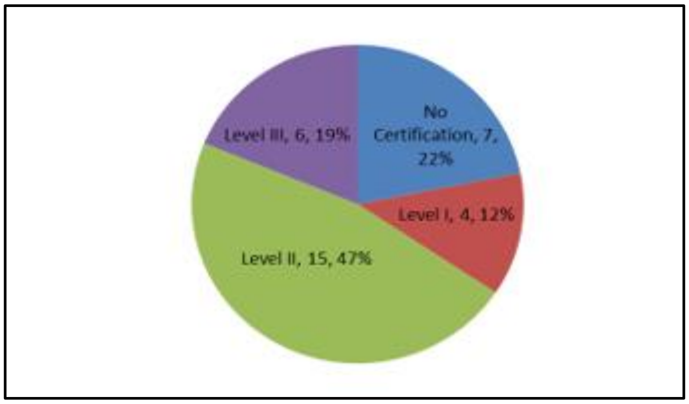


Figure 6. Number of Participants by DAWIA Certification Level
(Grennan & McCrory, 2016)

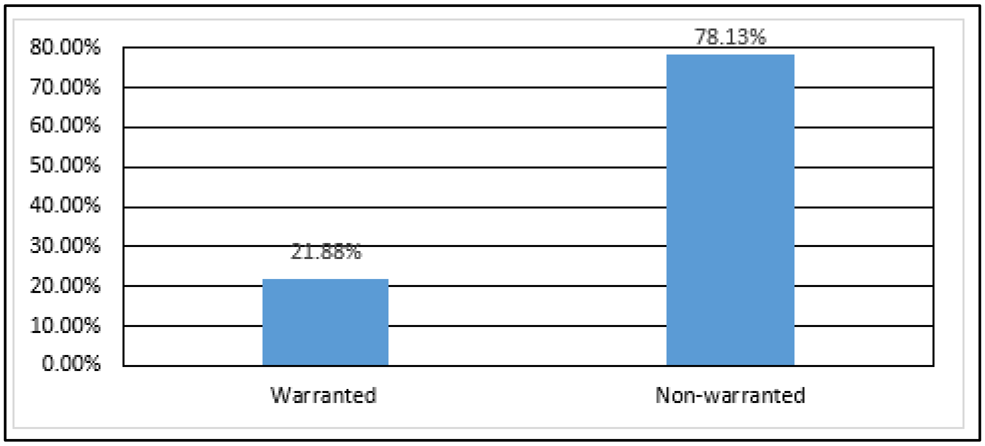


Figure 7. Percentage of Participants by Warrant Status
(Grennan & McCrory, 2016)

The average score on the knowledge portion of the web-based assessment tool was 58% correct of the 27 knowledge-based questions. Figures 8–10 reflect the average score based on years of experience level, DAWIA certification level, and warranted contracting officer status. As contracting experience and DAWIA level increases, so does the average score on the knowledge assessment.

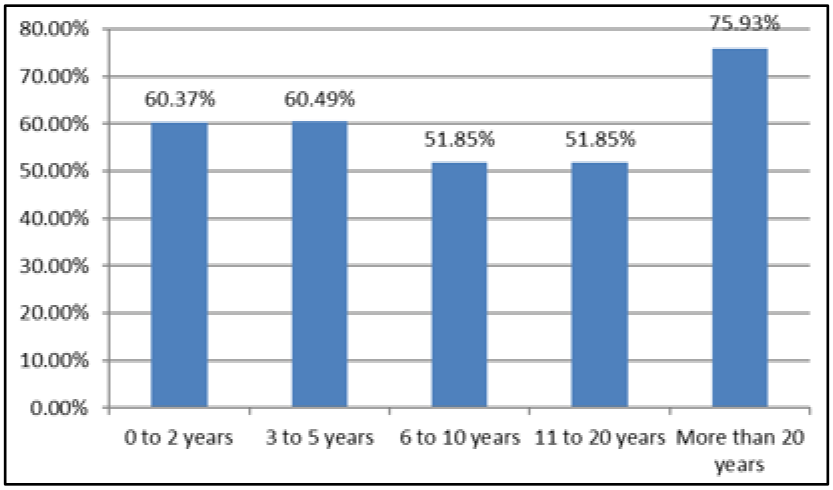


Figure 8. Average Score by Years of Experience
(Grennan & McCrory, 2016)

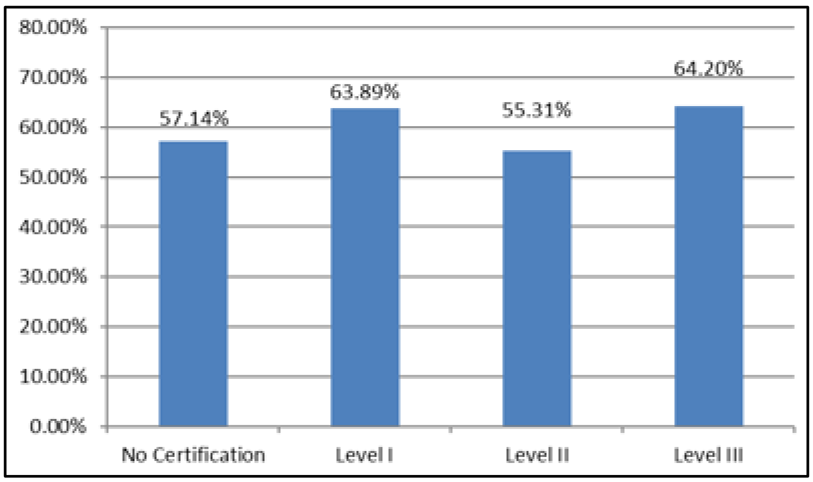


Figure 9. Average Score by DAWIA Level
(Grennan & McCrory, 2016)

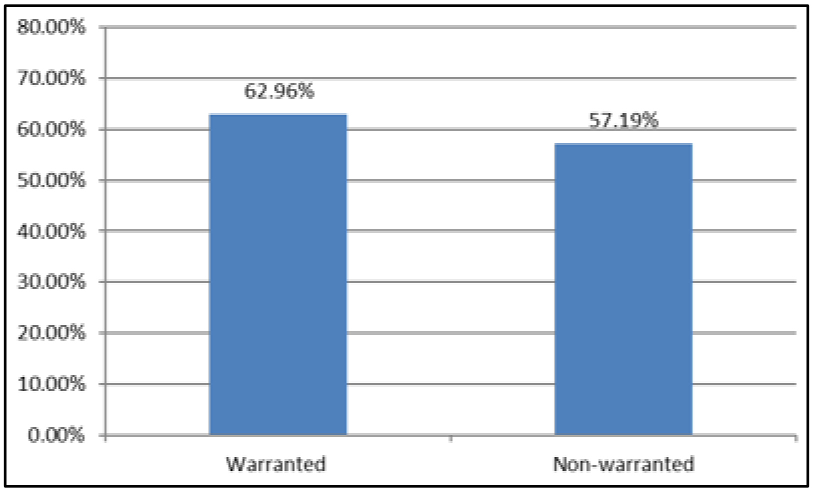


Figure 10. Average Score by Warrant Status
(Grennan & McCrory, 2016)

As previously stated, each knowledge assessment question was related to contract management processes, internal control components, and procurement fraud schemes. Figures 11–13 reflect the average score based on each of these areas. From the perspective of the contract management process as shown in Figure 11, assessment knowledge questions related to the procurement planning process had the highest average score, compared to questions related to contract closeout, which had the lowest score. From the perspective of the internal control components as shown in Figure 12, assessment knowledge questions related to the control environment component had the highest average score, compared to questions related to information and communication, which had the lowest score. From the perspective of procurement fraud schemes as shown in Figure 13, assessment knowledge questions related to bid rigging scheme had the highest average score, compared to questions related to conflict of interest schemes, which had the lowest score.

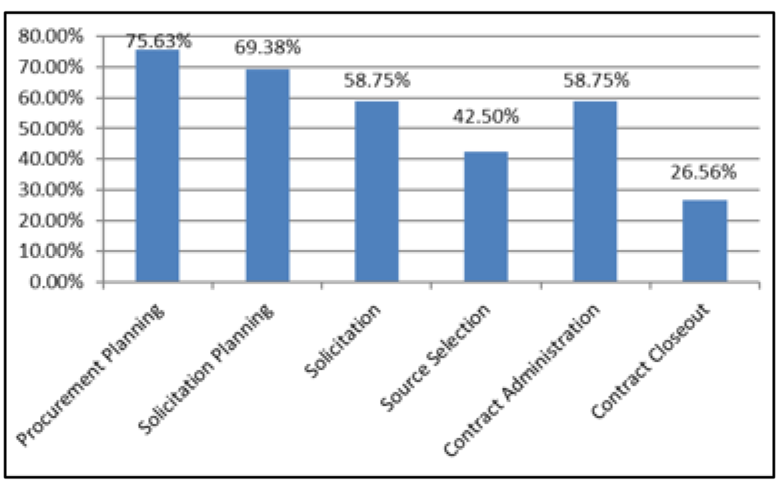


Figure 11. Average Score by Contract Management Process
(Grennan & McCrory, 2016)

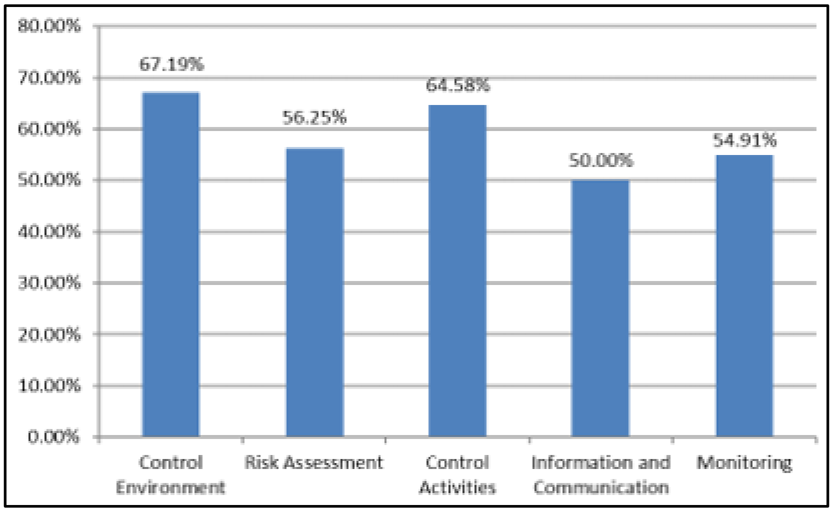


Figure 12. Average Score by Internal Control Component
(Grennan & McCrory, 2016)

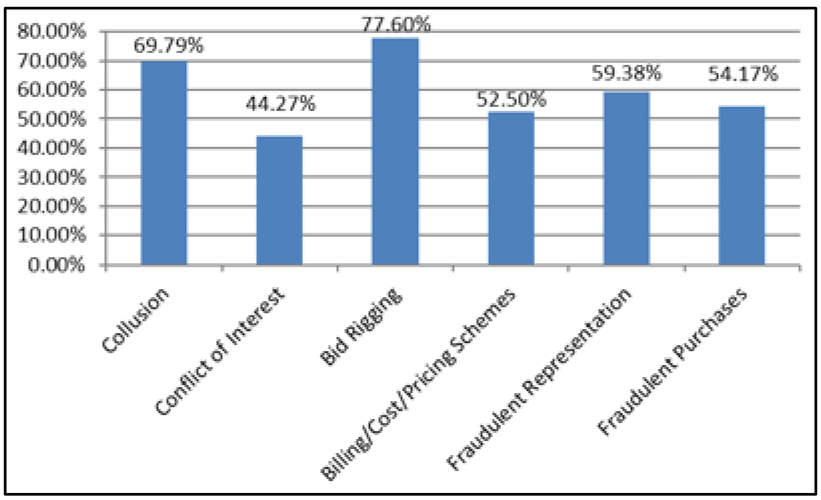


Figure 13. Average Score by Procurement Fraud Scheme
(Grennan & McCrory, 2016)

Analysis of Organizational Perception Findings

The web-based assessment tool also included survey questions related to the participants' perceptions of their organization's susceptibility to vulnerabilities to procurement fraud within the contract management phases, internal control components, and procurement fraud schemes. Figures 14–16 reflect the responses to these assessment questions.

As shown in Figure 14, when asked which contract management phase is most vulnerable to fraud in their organization, the contract administration phase was selected the most often (21.88%) and procurement planning, solicitation planning, and source selection were all selected the least often (0% for each one). Approximately 19% responded that they did not know, approximately 44% of the respondents stated they did not suspect fraud, and approximately 3% responded that they preferred not to answer.

As shown in Figure 15, when asked which internal control component is most vulnerable to fraud in their organization, the monitoring activities component was selected the most often (13%) and control environment was selected the least often (0%). Approximately 22% responded that they did not know, approximately 47% of the respondents stated they did not suspect fraud, and approximately 6% responded that they preferred not to answer.

As shown in Figure 16, when asked to which procurement fraud scheme they perceived their organization was most susceptible, collusion and conflict of interest were selected the most often (6.25% each) and bid rigging was selected the least often (0%). Approximately 19% responded that they did not know, approximately 53% of the respondents stated they did not suspect fraud, and approximately 6% responded that they preferred not to answer.

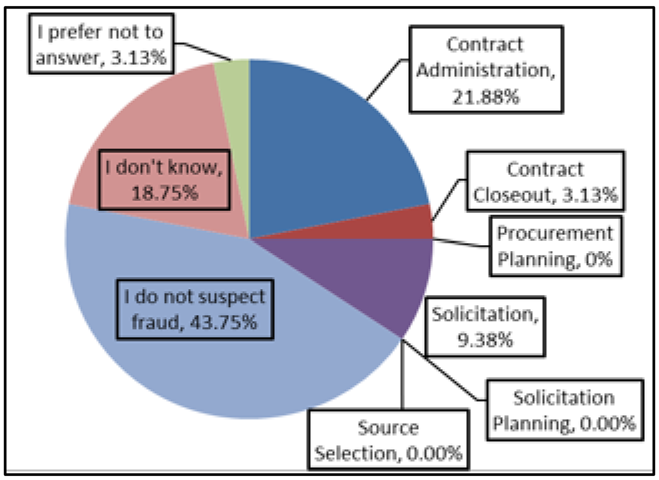


Figure 14. Percentage of Responses to Contract Management Phase Perception Question
(Grennan & McCrory, 2016)

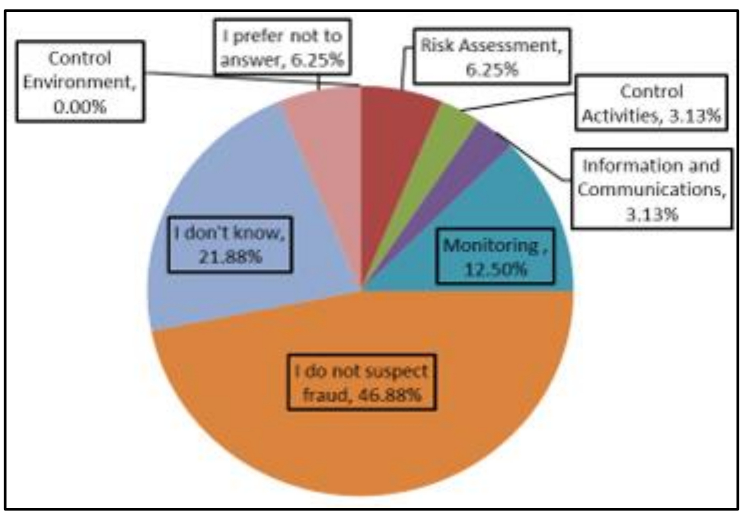


Figure 15. Percentage of Responses to Internal Control Component Perception Question
(Grennan & McCrory, 2016)

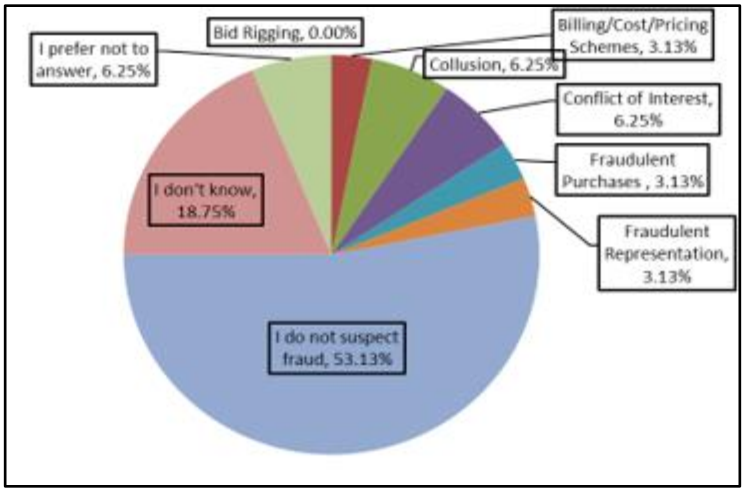


Figure 16. Percentage of Responses to Procurement Fraud Scheme Perception Question

(Grennan & McCrory, 2016)

Nine of the organizational questions were related to the contracting professionals' perceptions of their organization's internal controls and were designed to determine if any aspects of the organizations' internal control structure, processes, or culture made the organization more susceptible to fraudulent activity. The Likert Scale responses ranged from 1 (Strongly Disagree) to 5 (Strongly Agree). The average response mean to all of the nine questions was 4.24, and the range of responses was from 3.66 to 4.72. The lowest response mean (3.66) was for the item "I have adequate knowledge of contracting fraud schemes to perform my duties." Zero respondents answered "I Don't Know." The highest response mean (4.72) was for the item "I would report fraudulent or suspicious activity if I saw or suspected it." Zero respondents answered "I Don't Know."

Implications of Findings

The results of both the knowledge assessment and the organization perception assessment have interesting implications. The contracting professionals' average score on the overall knowledge assessment (58%) indicates a possible knowledge deficiency in procurement phases, internal controls, and procurement fraud schemes. This finding, along with the average response mean to the organization perception item "I have adequate knowledge of contracting fraud schemes to perform my duties" of 3.66, suggests that perhaps the contracting professionals are overly optimistic in self-assessing their knowledge of procurement fraud schemes.

Furthermore, a significant percentage of the respondents indicated "I do not suspect fraud" in relation to the organization's contracting phases (43.75%), internal control components (46.88%), and procurement fraud scheme susceptibility (53.13%). These findings, along with the low scoring knowledge assessment may indicate that although the majority of contracting professionals do not suspect fraud in their organizations, they also do not have a sufficient working knowledge of procurement fraud. The contracting professionals' limited knowledge of procurement fraud and their perception that their organization is not susceptible to fraud may reveal that the organization could in fact be vulnerable to some form of procurement fraud. An example of this type of vulnerability to procurement fraud can be found in the Fat Leonard case, which is still currently under investigation.

Recommendations

The results of the knowledge-based assessment indicated that, although the average score was 58%, the contracting professionals' knowledge of contracting processes, internal controls, and procurement fraud schemes increases as years of experience and DAWIA certification level increase. Recent research shows that the DAWIA required courses for contracting certification do not include a mandatory fraud training or awareness course (Castillo & Flannigan, 2014). The first recommendation is for the Defense Acquisition University (DAU) to incorporate coverage of internal controls and procurement fraud schemes in the mandatory contracting curriculum.

Another recommendation is to further explore the organization's information and communication internal control component and improve monitoring activities. Yet another recommendation is for the Navy and DoD as a whole to place serious emphasis on educating its contracting professionals regarding procurement fraud schemes and fraud awareness as well as areas vulnerable to procurement fraud.

Conclusion

In an environment of increased spending in government contracting for goods and services in the DoD, there is also an increased risk of public dollars being vulnerable to fraud, waste, and abuse (GAO, 2006). In addition, there is an increased risk of contracting organizations not getting the best value and not having contracting requirements met.

Contract management deficiencies and related internal control weaknesses have resulted in procurement fraud within the DoD (GAO, 2006; DoDIG, 2009). The results of this research indicate that contracting professionals in the Navy scored low in their knowledge of procurement fraud (Grennan & McCrory, 2016). At the same time, the contracting professionals self-assessed that they had sufficient procurement fraud knowledge to deter and detect procurement fraud. The implications of the results of the analysis indicate that there is a need for making procurement fraud education available to contracting personnel in order to make them more aware of vulnerabilities to fraud in federal government procurement.

This research investigated the Navy contracting professionals' perception of their organization's vulnerability to procurement fraud. This research indicates that the Navy contracting professionals' limited knowledge of procurement fraud and their perception that their organization is not susceptible to fraud may reveal that the organization could in fact be vulnerable to procurement fraud as in the case of the Fat Leonard incidents, which are still under investigation.

Overall, competent personnel, capable processes, and effective internal controls, which are the three components of the auditability triangle, may help federal agencies in their efforts to reduce, detect, and deter procurement fraud in their organizations throughout the Navy and DoD. In light of the potential fraud vulnerabilities within federal government contracting organizations, it is crucial that the Navy and DoD acquisition workforce have the necessary knowledge of procurement fraud schemes and procurement fraud indicators in order to help deter and detect procurement fraud and attain the best value for the government. As the federal government continues to increase procurement of goods and services, the pressure to reduce costs warrants federal agencies to strive to decrease its vulnerability to procurement fraud.

References

Albrecht, W. S. (2014). *The compromise triangle*. Retrieved from http://wheatley.byu.edu/fellow_notes/individual.cfm?id=3

Association of Certified Fraud Examiners (ACFE). (2012). *Report to the nations on occupational fraud and abuse: 2012 global fraud study.* Retrieved from http://www.acfe.com/uploadedFiles/ACFE_Website/Content/rttn/2012-report-to-nations.pdf

Association of Certified Fraud Examiners (ACFE). (2016). What is fraud? Retrieved from http://www.acfe.com/fraud-101.aspx

Castillo, J. C., & Flanigan, E. M. (2014*). Procurement fraud: A knowledge-level analysis of contracting personnel* (Master's thesis, Naval Postgraduate School). Retrieved from http://hdl.handle.net/10945/44533

Chang, P. (2013, June). *Analysis of contracting processes, internal controls, and procurement fraud schemes* (Master's thesis). Monterey, CA: Naval Postgraduate School.

Cohen, S., & Eimicke, W. (2008). *The responsible contract manager: Protecting the public interest in an outsourced world*. Washington, DC: Georgetown University Press.

Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2013). *Internal control—Integrated framework. Executive summary*. Retrieved from http://www.coso.org/documents/Internal%20Control-Integrated%20Framework.pdf

Cressey, D. R. (1972). *Other people's money: The psychology of embezzlement*. Montclair, NJ: Wadsworth Publishing Company.

DoD Inspector General (DoDIG). (2009). *Summary of DoD Office of Inspector General audits of acquisition and contract administration* (Report No. DoDIG-2009-071). Washington, DC: Author.

DoD Inspector General (DoDIG). (2014). *DoD needs to improve processes for issuing and managing cost-reimbursement contracts* (Report No. DoDIG-2015-029). Washington, DC: Author.

DoD Inspector General (DoDIG). (2015). *Semiannual report to the Congress: April 1, 2015 to September 30, 2015*. Retrieved from https://oig.state.gov/system/files/oig_fall_2015_sar.pdf

Department of Justice (DoJ). (2015). Former supervisory contracting officer arrested in Navy bribery scandal. *Justice News.* Retrieved from http://www.justice.gov/opa/pr/former-supervisory-contracting-officer-arrested-navy-bribery-scandal

Doss, M., & Jonas, G. (2004). *Section 404 reports on internal control: Impact on ratings will depend on nature of material weaknesses reported*. Boston, MA: Moody's Investors Service, Global Credit Research.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2017). Retrieved from http://farsite.hill.af.mil/

Federal Procurement Data System–Next Generation. (2017). Retrieved from https://www.fpds.gov/fpdsng_cms/index.php/en/

Frame, D. L. (1999). *Project management competence: Building key skills for individuals, teams, and organizations.* San Francisco, CA: Jossey-Bass.

GAO. (2002, July). *Acquisition workforce: Agencies need to better define and track the training of their employees* (GAO-02-737). Washington, DC: Author.

GAO. (2006, July 7). *Contract management: DoD vulnerabilities to contracting fraud, waste, and abuse* (GAO-06-838R). Washington, DC: Author.

GAO. (2009, March). *Department of Defense: Additional actions and data are needed to effectively manage and oversee DoD's acquisition workforce* (GAO-09-342). Washington, DC: Author.

GAO. (2013, February). *High-risk series: An update* (GAO-13-283). Washington, DC: Author.

GAO. (2014, September). *Standards for internal control in the federal government* (GAO 14-704G). Retrieved from http://www.gao.gov/products/GAO-14-704G

Garrett, G. A. (2013). How to evaluate a purchasing system: Tools, techniques and best practices. *Contract Management, 53*(1), 41–51.

Grennan J. A., & McCrory, M. A. (2016, December). *Auditability in the U.S. Navy: A knowledge assessment of the contracting workforce* (Master's thesis). Monterey, CA: Naval Postgraduate School.

New York State Internal Control Association (NYSICA). (2006). Internal control survey. Presented at the 2006 Conference NYS OSC. Retrieved from http://www.nysica.com/library.php

Power, M. (1996). Making things auditable. *Accounting, Organizations and Society, 21*(2), 289–315.

Rendon, J. M., & Rendon, R. G. (2015, March 22). *Defense procurement: An analysis of contract management internal controls* (NPS-CM-15-003). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Rendon, J. M., & Rendon, R. G. (2016). Procurement fraud in the U.S. Department of Defense: Implications for contracting procedures and internal controls. *Managerial Audit Journal, 31*(6/7), 741–767.

Rendon, R. G. (2008). Procurement process maturity: Key to performance measurement. *Journal of Public Procurement, 8*(2), 200–214.

Rendon, R. G. (2009, August). *Contract management process maturity: Empirical analysis of organizational assessments* (NPS-CM-09-124). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Rendon, R. G. (2010, June). *Assessment of Army Contracting Command's contract management processes* (NPS-CM-10-154). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Rendon, R. G. (2011, April). *Assessment of Army Contracting Command's Contract Management Processes (TACOME and RDECOM)* (NPS-CM-11-019). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Rendon, R. G., & Rendon, J. M. (2015). Auditability in public procurement: An analysis of internal controls and fraud vulnerability. *International Journal of Procurement Management, 8*(6), 710–730.

Rendon, R. G., & Snider, K. F. (Eds.). (2008). *Management of defense acquisition projects*. Reston, VA: American Institute of Aeronautics and Astronautics.

Tan, L. H. J. (2013, December). *An analysis of internal controls and procurement fraud deterrence* (Master's thesis). Monterey, CA: Naval Postgraduate School.

Thai, K. V. (2001). Public procurement re-examined. *Journal of Public Procurement, 1*(1), 9–50.

Thai, K. (2004). *Introduction to public procurement*. Herndon, VA: National Institute of Governmental Purchasing.

Thai, K. V., & Grimm, R. (2000). Government procurement: Past and current developments. *Journal of Public Budgeting, Accounting & Financial Management, 12*(2), 231.

Wells, J. T. (2001). Why employees commit fraud. *Journal of Accountancy, 191*(2), 89–91. Retrieved from http://search.proquest.com/docview/206773009

Wells, J. T. (2005). *Principles of fraud examinatio*n (1st ed.). Hoboken, NJ: John Wiley and Sons.

Wells, J. T. (2008). *Principals of fraud examination* (2nd ed.). Hoboken, NJ: John Wiley and Sons.

Panel 15. Capital Investment Strategies in DoD

Wednesday, May 9, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: Michael McGrath,** Consultant and Senior Technical Advisor, McGrath Analytics LLC ***Fixed vs. Flexible Approaches to Improving Capital Investment in Military Depots*** William Lucyshyn, University of Maryland John Rigilano, University of Maryland ***Preliminary Findings: Is the Ratio of Investment Between R&D to Production Experiencing Fundamental Change?*** Rhys McCormick, Center for Strategic and International Studies Andrew Hunter, Center for Strategic and International Studies Gregory Sanders, Center for Strategic and International Studies ***Extending an Econophysics Value Model for Early Developmental Program Performance Prediction and Assessment*** Raymond D. Jones, COL, USA (Ret.), Naval Postgraduate School Thomas Housel, Naval Postgraduate School

Michael McGrath—is an independent consultant. As a former Vice President at Analytic Services Inc. (ANSER), he led business operations in systems and operations analysis. He previously served as the Deputy Assistant Secretary of the Navy for Research, Development, Test and Evaluation, where he was a strong Navy proponent for improvements in technology transition, modeling and simulation, and test and evaluation. In prior positions, he served as Vice President for Government Business at the Sarnoff Corporation (former RCA corporate lab); ADUSD for Dual Use and Commercial Programs in the Office of the Secretary of Defense (OSD) with responsibility for industrial base and commercial technology investment programs; Program Manager at the Defense Systems Research Projects Agency (DARPA), where he managed manufacturing technology programs; and Director of the DoD Computer-Aided Acquisition and Logistics Support program, automating the interface between DoD and industry for technical data interchange and access. His early government career included positions in Logistics Management at Naval Air Systems Command and in Acquisition Management in OSD. He has served on Defense Science Board and National Academies studies, and is an active member of the National Defense Industrial Association (NDIA), the National Materials and Manufacturing Board, the Board on Army Science and Technology, and several university and not-for-profit advisory boards.

Dr. McGrath holds a BS in Space Science and Applied Physics and an MS in Aerospace Engineering from Catholic University and a doctorate in Operations Research from George Washington University (where he also served as adjunct faculty).

Fixed vs. Flexible Approaches to Improving Capital Investment in Military Depots[1]

William Lucyshyn—is the Director of Research and a Research Professor at the Center for Public Policy and Private Enterprise in the School of Public Policy at the University of Maryland. In this position, he directs research on critical policy issues related to the increasingly complex problems associated with improving public-sector management and operations and with how government works with private enterprise.

His current projects include modernizing government supply-chain management, identifying government sourcing and acquisition best practices, and analyzing Department of Defense business modernization and transformation. Previously, Lucyshyn served as a program manager and the principal technical advisor to the Director of the Defense Advanced Research Projects Agency (DARPA) on the identification, selection, research, development, and prototype production of advanced technology projects.

Prior to joining DARPA, Lucyshyn completed a 25-year career in the U.S. Air Force. Lucyshyn received his bachelor's degree in engineering science from the City University of New York and earned his master's degree in nuclear engineering from the Air Force Institute of Technology. He has authored numerous reports, book chapters, and journal articles. [lucyshyn@umd.edu]

John Rigilano—is a Researcher at the Center for Public Policy and Private Enterprise. He earned his Master of Public Policy degree from the University of Maryland, College Park, in 2011, and he holds a Bachelor of Arts degree in anthropology from Pennsylvania State University. He is pursuing a career in policy and program analysis. [jprig@umd.edu]

Introduction

The military depots form a vital component of America's defense capability, providing for the repair, rebuilding, and major overhaul of weapon systems (e.g., ships, armored vehicles, missile systems, and aircraft), their parts, assemblies, and subassemblies[2]. In FY 2014, the DoD spent $31.4 billion on depot-level maintenance and repair work (Office of the Assistant Secretary of Defense, Logistics & Materiel Readiness [OASD(L&MR)], 2015).

As of 2007, each of the three military departments is required by law to make an annual capital investment in its depots at a rate of at least 6% of their combined average revenue.[3] The required investments have been made to support military construction, facilities maintenance and repair, and equipment procurement and process installation.

This requirement was enacted in response to the deteriorating capabilities of depots during the 1990s, which lawmakers and military leaders attributed to insufficient investments

---

[1] This is an abridged, preliminary version of the final report, which will be released in June 2018.
[2] The DoD maintains a wide range of weapon systems, including 237 ships, 14,444 aircraft/helicopters, 884 strategic missiles, and 391,520 ground combat and tactical vehicles (OASD[L&MR], 2015). In FY 2014, approximately 53% of the depot-level workload was accomplished in government-owned facilities; the remainder was accomplished by the private sector in commercial facilities (OASD[L&MR], 2015).
[3] Through the use of revolving fund structures (i.e., working capital funds), the depots earn revenue via the "sale" of their services to military customers (i.e., military operating units); see Part IV "Funding Capital Investment."

in facilities, equipment, and human capital. According to the Government Accountability Office (GAO; 2001a), this lack of investment could be traced to the "DoD's downsizing of its depot infrastructure and workforce since the end of the Cold War, [which] was done without sound strategic planning" (p. 3).

Indeed, by the end of the millennium, the DoD had outsourced a number of logistical support functions to the private sector, including weapon system maintenance and repair activities, with some arguing that inadequate consideration was being given to the definition and protection of so-called "core" capabilities.[4]

In light of increasing budgetary pressure at all levels of government, improving strategic investment decision-making—the process of correctly identifying, evaluating, and selecting among projects that will have the greatest impact on the organization's ability to perform its mission—is of critical importance. In the case of military depots, there is debate over whether the mandated minimum investment requirement facilitates or inhibits strategic investment decision-making.

For instance, although the law does not place upper limits on annual capital investment, there is an implicit assumption that 6% is and will continue to represent an adequate (minimum) level of investment *and* that previous years' revenues represent the appropriate sum upon which to base the 6%. This is unlikely to be the case. At the same time, it might be argued that in the absence of dedicated funding, routine investment in depots will be overlooked—that it, in fact, *was* overlooked—to fund more visible, higher-profile programs and projects. This report explores the impact of funding mechanisms on decision-making, investment levels, and capabilities.

Barrett and Greene (2013) assert that "when funds are dedicated, often from a special revenue stream," the advantage of consistent funding "buffers a program from the powerful wind of changing political climate" (p. 1). They contrast dedicated funding (or "earmarking") with "one-fund-fits-all" (i.e., general fund financing), which gives legislators and managers more financial flexibility to move funds as needs change.

They conclude, "Unfortunately, there's no overridingly best practice here—no black or white… but understanding the pros and cons of both routes to funding holds out the hope of coming to the right answer for a particular project" (p. 1). This report evaluates these pros and cons, as well as any barriers to change, within the context of military depot funding. Ultimately, it seeks to determine if and how current capital investment policy should be modified in order to optimize depot capabilities.

---

[4] Since 1984, law has required that the DoD maintain a government owned and operated logistics capability (including personnel, equipment, and facilities) to ensure "a ready and controlled source of technical competence and resources necessary to ensure effective and timely response to a mobilization, national defense contingency situations, and other emergency requirements" (10 U.S.C. § 2464).

Capital Investment Requirement

The 2007 National Defense Authorization Act (NDAA; 10 U.S.C. § 2476), *Minimum Capital Investment for Certain Depots*—like the statues (Core[5], 50/50[6]) that preceded it—was enacted to safeguard and strengthen the DoD's organic capabilities in the face of downsizing, base closures, and the preference for increased contracting.

Prior to its enactment, the Air Force, in its 2002 *Depot Maintenance Master Plan,* committed to allocate $150 million each fiscal year for six years, beginning in 2004, in order to correct for years of underinvestment (DoD, 2006). The Air Force noted that past capital investment, which had averaged 3% of total depot revenue, led to a significant equipment purchase backlog of approximately $200 million. *The Air Force Depot Maintenance Strategy,* published for the first time in 2002, envisioned an annual capital investment level of 6% of revenue (DoD, 2006). According to the Air Force, this level of investment was in line with levels seen in the private sector (DoD, 2006).[7]

In 2005, Congress commended the Air Force for its proactive capital investment strategy. In Section 324 of the 2006 NDAA, entitled *Sense of Congress Regarding Depot Maintenance,* Congress stated that "the *Depot Maintenance Strategy* and *Master Plan of the Air Force* reflect the essential requirements for the Air Force to maintain a ready and controlled source of organic technical competence, thereby ensuring an effective and timely response to national defense contingencies and emergency requirements." It is perhaps unsurprising, then, that a version of the Air Force plan made its way into law a couple years later.

Meanwhile, in 2006, the DoD issued an overarching *Depot Maintenance Strategic Plan,* which articulated its plans for "ensuring its organic depot maintenance infrastructure is postured and resourced to meet the national security and materiel readiness challenges of the 21st century." The *Strategic Plan* formalized the 6% investment figure cited by the Air Force across the DoD:

> Each DoD Component that operates organic depot-level maintenance activities will establish a programming goal for depot maintenance capital investment. The minimum annual funding target for each DoD Component

---

[5] 10 U.S.C. § 2464—core logistics capability statute—reads, in part, as follows: "It is essential for the national defense that the Department of Defense maintain a core logistics capability that is Government-owned and Government-operated (including Government personnel and Government-owned and Government-operated equipment and facilities) to ensure ready and controlled source of technical competence and resources necessary to ensure effective and timely response to a mobilization, national defense contingency situations, and other emergency requirements."

[6] 10 U.S.C. § 2466 states that "not more than 50 percent of the funds made available in a fiscal year to a military department or a Defense Agency for depot-level maintenance and repair workload may be used to contract for the performance by non-Federal Government personnel of such workload for the military department or the Defense Agency."

[7] The 6% figure was based on an Air Force study that examined capital investment levels in commercial firms engaged in maintenance, repair, and overhaul (MRO). The study concluded that MRO firms' capital investments averaged out to about 6% of revenue. Commercial firms make capital investments to further business objectives; previous years' revenues may be a consideration, but do not form the explicit basis upon which investments are made.

will be an amount equal to six percent of its combined funded core-sustaining workload. Expected implementation is not later than the FY 2009–14.

The *Strategic Plan,* like the 1996 policy, used core-sustaining workload as the basis for investment. A year later, the *Strategic Plan* was superseded by the minimum investment requirement, and, like the Air Force strategy, used total revenue as the basis.

The 2007 statute, 10 U.S.C. § 2476[8], reads as follows:

> Each fiscal year, the Secretary of a military department shall invest in the capital budgets of the covered depots of that military department a total amount equal to not less than six percent of the average total combined maintenance, repair, and overhaul workload funded at all the depots of that military department for the preceding three fiscal years.

The statute warrants a second read in order to appreciate the details and their implications. One should note the following:

- The 6% requirement is a "floor," as opposed to a "ceiling." Some of the military departments have invested well over 6% in a given year.
- The statute does not require uniform investment across a military department's covered depots.
- The basis for the calculation is the "workload funded at *all* the depots of that military department [emphasis added]," but only investments made in the "covered" depots count toward meeting the 6% requirement.[9]

[8] The Secretary of Defense may waive the requirement for reasons of national security.
[9] For example, the Army's organic industrial base comprises 13 depots and arsenals, but only investments made in the eight "covered" depots count toward meeting the 6% requirement.

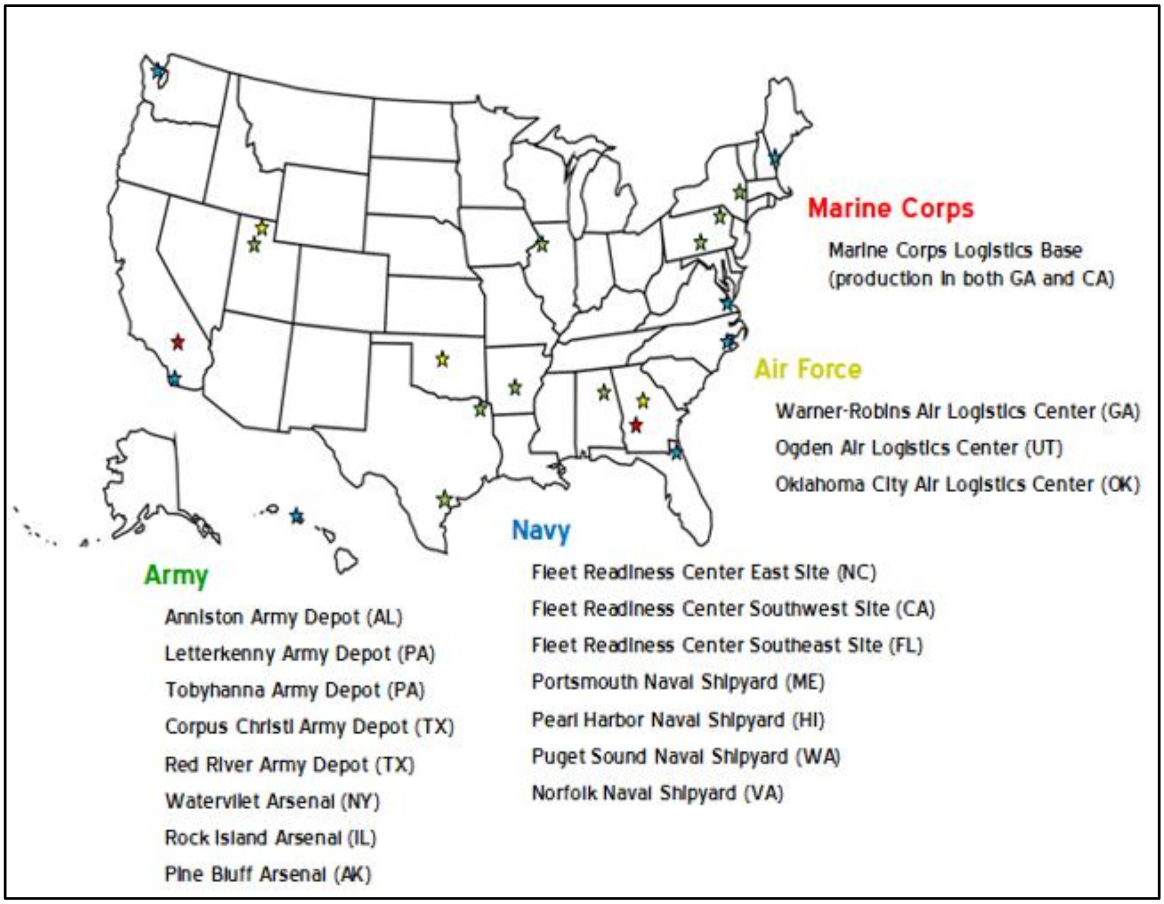


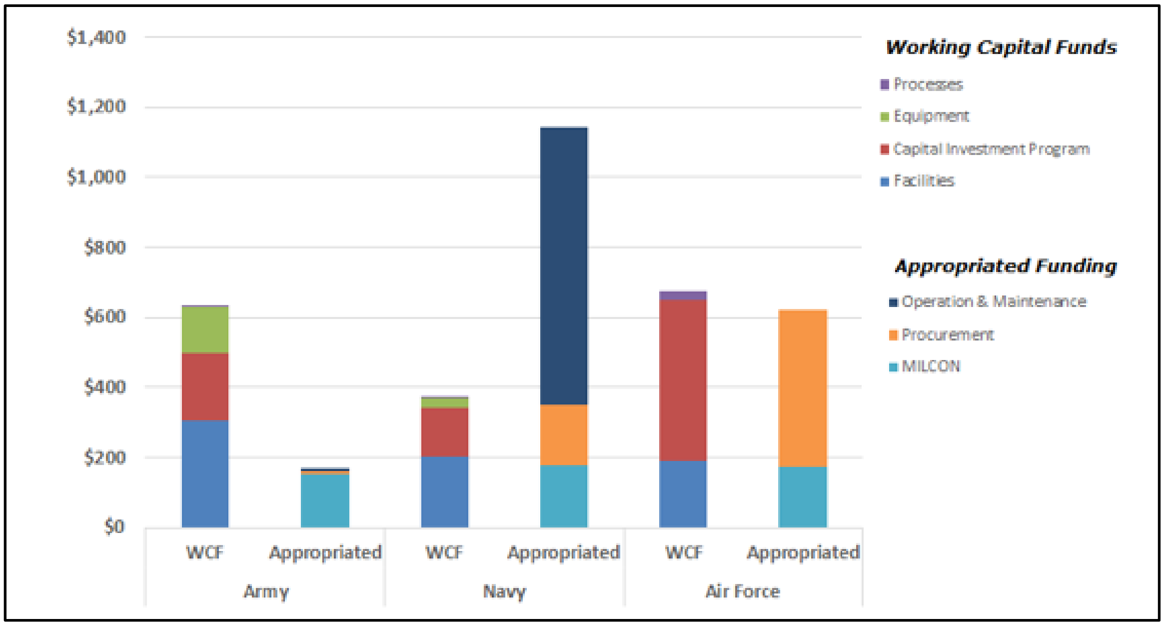
Figure 1. 20 Covered Depots

Funding Capital Investment

Capital investment in the military depots is financed through direct congressional appropriation and through the military departments' working capital funds.[10] Working capital funds have been in use by the DoD, and other government organizations, for several decades. Support organizations (e.g., the depots) set their own rates and manage cash flow. In effect, each depot relies on revenue from the "sale" of its services (to DoD customers and, less often, to other U.S. government and foreign customers) in order to finance their operations.

Figure 2 compares the sources of capital investment funding within the Air Force, Navy, and Army over a three-year period between FY 2015 and FY 2017. Note the relative lack of uniformity among the three departments in terms of the composition of investment sources. In particular, the Air Force and Navy have relied significantly more on appropriated funding in recent years. In fact, representatives from the Air Force have noted that if not for the high levels of appropriated funding, the department would have found it very challenging to meet the 6% requirement (DoD, 2014).

[10] The Navy's shipyards are not funded through working capital funds, but through appropriations.



Note. Navy data was obtained from Navy Working Capital Fund Budget Justifications, FY 2017 and FY 2018. Air Force data was obtained from Air Force Working Capital Fund Budget Estimates, FY 2017 and FY 2018. Army data was obtained from Army Working Capital Fund Budget Estimates, FY 2017 and FY 2018.

Figure 2. Sources of Capital Investment ($ Millions), FY 2015–2017

The capital investment program (CIP) is an important component of any department's working capital funds. The program allows the depot to depreciate a capital asset by reallocating its cost over its useful life. In effect, the depot is able to acquire needed assets without having to significantly increase customer rates. Purchases funded through the other components of the depot's working capital fund (i.e., processes, equipment, and facilities accounts; see Figure 8) are expensed (i.e., the whole cost amount is placed on the depot's income statement).

The CIP provides "the framework for planning, coordinating, and controlling NWCF resources and expenditures to obtain capital assets" (DoN, 2016). The four approved capital budget investment categories within the CIP are Automated Data Processing (ADP) and Telecommunications Equipment; Non-ADP Equipment; Software Development; and Minor Construction (DFMR, 2016).[11] Equipment purchased through the CIP have a unit cost greater than $250,000 and a useful life of two or more years (FY 2017 AWCF Budget; DoN, 2016).

The military depots must use a cost comparison or a pre-investment economic analysis to justify proposed capital investments under the program. For investments with an estimated cost of under $1,000,000, a cost comparison must be included in the depot's capital budget submission. The comparison must present a differential cost display (i.e., the

[11] Minor construction is generally limited to projects that cost $750,000 or less; for projects impacting health, safety or environment, the figure is $1,500,000 or less. Larger construction projects are funded through the military construction appropriation (see Figure 2).

total costs attributed to each alternative) using the payback period capital budgeting procedure.[12] For capital investment projects with a cost of $1,000,000 or more, the depot must submit a pre-economic analysis that presents a differential cost display using the net present value approach.[13]

The development of pre-economic analyses can pose a challenge to the depots. At some depots, there are a limited number of personnel capable of developing the analyses. A greater challenge is that cost tracking at the depot level is limited and often does not provide the data necessary to develop a timely, comprehensive analysis. As a result, the military command with jurisdiction over the depot may reject the economic analysis. This can be a major setback given that the CIP approval process is seen by the depots as slow and inefficient to begin with; in fact, the turnaround time for approval often extends to three years. During this period, the depot is required to update its analysis to reflect changing costs and assumptions. Contractor quotes, when updated, often exceed the 10% tolerance permitted by law. Depot personnel must then develop a new analysis or "down scope" the project.

This lengthy and, at times, bureaucratic process can lead depot personnel to try to reduce the purchase cost of a capital asset to below the CIP minimum threshold of $250,000, especially when the estimate only narrowly exceeds this minimum in the first place. Indeed, one can find examples of facilities and pieces of equipment that cost just under $250,000 at some depots. In some cases, these may be suboptimal solutions.

Investment Limitations

Prior to 2012, 10 U.S.C. § 2476 stated, "The capital budget of a depot includes investment funds spent on depot infrastructure, equipment, and process improvement in direct support of depot operations." Concerned that some depot operating expenses were being funded under the guise of capital investment, Congress sought to clarify the law. The 2012 NDAA was amended to read: "The capital budget of a depot includes investment funds spent to *modernize or improve* the efficiency of depot facilities, equipment, work environment, or processes in direct support of depot operations, *but does not include funds spent for sustainment of existing facilities, infrastructure, or equipment* [emphasis added]."

This constraint has generated some confusion over what, exactly, constitutes a capital investment. The Defense Executive Steering Committee (DoD, 2014) provides some examples of projects that, under the current definition, *cannot* be justified as capital expenditures: the replacement of the roof and fire suppression system of an aircraft hangar; renovation of an avionics repair shop; or a new corrosion control building. Many would argue that these types of expenditures *necessarily* "modernize or improve efficiency." Indeed, under criteria used by the Internal Revenue Service—which include "rebuilding property after the end of its economic useful life" and restoring property or equipment to "like new" condition—examples such as these would be considered capital expenditures. In effect, the

[12] Payback period shows the number of years it takes to break even from undertaking the initial expenditure, by discounting future cash flows and recognizing the time value of money.
[13] Net present value analysis evaluates the cash flows forecasted to be delivered by a project by discounting them back to the present using the time span of the project and the firm's weighted average cost of capital.

law limits the depots' ability to finance certain projects through the CIP, projects that in the commercial sector would almost certainly fall under the category of capital investment and whose costs would be depreciated over time. Thus, new investment in the depots may come at the expense of needed maintenance—a problem that 10 U.S.C. § 2476 was originally enacted to address.

Depot Capabilities

Despite some recent improvements and with some notable exceptions, depot capabilities remain at suboptimal levels. In July 2017, *National Defense* reported that "certain service chiefs, the administration, and some in the media have stated that U.S. military mission capability and readiness could increasingly be considered a national security problem" (Captain, 2017, p. 17). *U.S. Naval Institute Proceedings* reports from 2016 and 2017 have stated that "virtually all of the naval services' helicopters, the F/A-18, and Harriers are at or below 50% readiness levels" (Captain, 2017, p. 17). Of course, not all of the services' readiness challenges can be traced to inadequate capital investment in depots.

However, many of them can be. At the department level, the absence of strategic investment planning, in particular, has contributed to declining readiness levels. A cursory examination of the *Army's Depot Maintenance Enterprise Strategic Plan, 2008–2025* reveals it to be less of a strategy than a to-do list. Cited objectives include "update infrastructure planning" and "establish an integrated human capital plan." The plan does not state how or when these are to be accomplished. In 2009, the GAO stated that the lack of a "meaningful department wide assessment" of the shortcomings of organic depots has left the DoD with no way to accurately determine whether they have the resources and capabilities to meet sudden threats and warfighter needs (GAO, 2009a).

In 2010, the GAO published another report entitled *Improved Strategic Planning Needed to Ensure That Air Force Depots can Meet Future Requirements*. The report found that the Air Force's failure to use benchmarks to evaluate the adequacy of investment funding called into question "its assertion that its depots are postured and resourced to meet future maintenance challenges." A year earlier, the GAO released a similar report that questioned the capabilities of Army and Marine Corps depots (GAO, 2009b). All three of the services' strategic plans were criticized for not using a results-oriented management framework to help ensure that they were positioned to meet future needs.

As recently as September 2017, the GAO found that despite the Navy's development of an improved investment plan in 2013, its shipyards and equipment remain in poor condition, with backlogged maintenance projects having grown by 41% over five years to a Navy-estimated $4.86 billion that will require 19 years to complete. The poor condition of the shipyards has contributed to the Navy's inability to meet operational needs. According to the GAO, "In fiscal years 2000 through 2016, inadequate facilities and equipment led to maintenance delays that contributed in part to more than 1,300 lost operational days—days when ships were unavailable for operations—for aircraft carriers and 12,500 lost operational days for submarines" (GAO, 2017, p. 1). The GAO concluded that unless the Navy adopts a "comprehensive, results-oriented approach to addressing its capital investment needs, [it] risks continued deterioration of its shipyards, hindering its ability to efficiently and effectively support Navy readiness over the long term" (GAO, 2017).

The lack of adequate strategic planning by the military departments has led to "at least seven instances of recommendations to create a single depot maintenance command or manager as the preferred direction in the evolution of the organic depot maintenance capability and as a way to achieve the desired performance" (Avdellas et al., 2011, p. 1-2). Avdellas et al. (2011) notes that those recommendations have been advanced by various bodies including Congress, DoD review panels, the GAO, and the Joint Staff. Yet none has been implemented. Instead, according to Avdellas et al. (2011), “We observe a continuation of the multitude of customer-provider encounters playing out in weapon system acquisition and sustainment, without a consistent or integrated strategic vision” (p. 1-2). Some capabilities are far from “world class” or “best of breed.” Often, the distribution of capabilities is uneven if not bimodal: At one depot, state-of-the-art equipment and some new facilities stand out against a landscape of aging buildings, near-obsolete testing equipment, and shelves of metal parts left exposed to the elements.

Whether, and to what extent, the present lack of strategic planning can be attributed to inadequate military leadership, cultural artifacts within the DoD, segmented lines of authority, congressional interventions, or other causes can be debated. It should also be recognized that solutions that work in the commercial sector, especially those aimed at improving economic efficiency, may not work well within large public organizations. According to Nutt (2005),

> The external environment of a public organization is littered with political considerations. The views of opinion leaders, outright manipulation by legislators and interest groups, and opposition to an agency’s prerogatives are more important than economic issues, which are crucial for private organizations (Levine et al., 1975). Disagreements, reciprocity, and quid pro quos can occur at any time and, within limits, are permissible ingredients in public decisions. Bargaining is required to find the permissible arenas of action. How things are viewed and understood by stakeholders holds more salience than the accuracy of claims. The meaning of a claim is derived from opinions as well as facts. *If economic reasoning, such as efficiency, is applied, it must be preceded by a decision to deal with efficiency questions, which often has political undertones* [emphasis added]. (p. 293)

Nutt (2005) goes on to say that public sector decision-makers generally “have weaker power bases” and that they “lack the funds to make investments that reshape systems they manage” (p. 297). He concludes that decision-makers in public organizations are “more apt to use consultative or networking practices to make decisions” and, critically, “less apt to make decisions using analytical and speculative practices, seeing them as more risky” (Nutt, 2005, p. 298).

To improve depot capabilities, approaches to funding capital investment must be considered carefully. As Nutt (2005) suggests, “oft-repeated call[s] for public-sector organizations to adopt private sector practices” (p. 292), though well-intentioned, may be misguided. The optimal approach must balance private sector practices with public sector realities.

Fixed Funding

In principle, the government should allocate funds, irrespective of their source, in such a way as to maximize benefits to the citizenry. Critics of earmarking—the legislative provision mandating that approved funds be spent on specific projects—support their position by arguing that an earmarking provision is an “unnecessary constraint in the utility-maximization problem of allocating the last dollar to yield equal marginal utility in every direction” (Teja, 1988, p. 523).

The advantages of earmarking include a guarantee of funding, predictability and budget planning, and the potential to depoliticize future funding decisions. The primary disadvantages revolve around budgetary inflexibility: “Earmarked revenues, not program needs or benefits relative to the competing priorities, may determine overall funding levels for the programs" (Michael, 2015, p. 5). Public spending, it is argued, should be determined by deliberate policymaking.

However, it is unclear whether and to what extent such deliberation occurs, even in the absence of earmarking. As Teja (1988) has observed, "it is implicitly assumed that expenditures under general fund financing are indeed periodically reviewed and adjusted to ensure that no program is under- or overfunded," an assumption he describes as "highly questionable."

In any case, earmarking some percentage of revenue for capital investment is not unique to the military depots. Because earmarks that are derived from recurring sources of revenue (e.g., annual taxes) "implicitly promise funding of at least the level of the earmark" (Michael, 2015, p. 1), they provide some measure of predictability, which can improve budgeting, planning, and decision-making. The state of Missouri amended its constitution in 1996 to create a separate Facilities Maintenance Reserve Fund to dedicate general fund dollars toward maintenance. The fund was gradually phased in from 1998 to 2007, dedicating 0.1% of the state's general revenue to the fund in its first year, and increasing by 0.1% over the next 10 years. Since 2007, 1% of the general revenue is transferred into the fund each year.

Of the various earmarking schemes that have been tried, there is a general consensus that earmarking "benefit taxes" or user fees for related expenditures is preferable (Wilkinson, 1994; Transport Research Center, 2008). A 2008 study by the Transport Research Center notes that earmarking can have "an element of the benefit approach[14] to equity in taxation, i.e., the idea that people should be paying according to the benefits they receive from consuming a commodity" (p. 150). Taxes levied by the state on gasoline, which are then used to fund transportation infrastructure, are often cited as examples. Not only does this type of earmark link supply and demand, but it informs the taxpayers of the cost of the services that they are consuming.

On this basis, the depot investment requirement might be viewed quite favorably given that mandatory spending is a function of funded workload (i.e., supply and demand are linked). Moreover, because much of the capital investment requirement is funded through the working capital funds, the military activities that rely on the depots have some visibility into the cost of their operations (through the rates that they pay), which, in principle, serves to further ensure that the earmarked funding is used effectively and efficiently.

Some have argued that general fund financing inhibits sound capital investment decision-making. Bratland (2010), for example, has asserted that the public sector simply does not have the ability to invest effectively in public infrastructure. He points to the sustained lack of investment in transportation infrastructure throughout the country, which, though often politicized, is a real and growing problem. According to the American Society of Civil Engineers, cumulative infrastructure investment needs will total $2.7 trillion by 2020, rising to $10 trillion by 2040 (Cullen, 2013). Anticipated funding will cover only 60% of these needs through 2020, dropping to 53% by 2040. The corresponding investment gaps are

[14] The benefit theory of taxation states that each citizen should be called upon to pay taxes in proportion to the benefits derived by him from services provided by the government.

estimated to total $1.1 trillion by 2020, growing to $4.7 trillion by 2040 (Cullen, 2013). Bratland (2010) asks the question, "Is the neglect of public infrastructure endemic to its governmental provision and management and thus inevitable?" The answer, according to Bratland, is "Yes." He writes,

> The maintenance problem arises from the absence of ownership of public infrastructure and the fact that the infrastructure's benefits yield no appropriable sales revenue that can serve as a guide to maintenance. Hence, neglect appears to be inherent in the fact of government provision. Labeling components of infrastructure as public capital is simply a metaphor that misleads the electorate into thinking public infrastructure can be successfully maintained. (p. 38)

Bratland (2010) concludes, "Legitimate capital concepts suggest that ownership and maintenance of infrastructure facilities should never be placed within the government's scope of responsibility" (p. 41). Again, there is reason to be more optimistic with regard to depot investment; as discussed, 10 U.S.C. § 2476 links investment to revenue through the working capital fund structure in a way that imitates, albeit imperfectly, the private sector. In other words, the depots *do* "yield an appropriable sales revenue that can serve as a guide" to capital investment.

The precise role for government earmarks may turn on whether and to what extent infrastructure—and the tendency to neglect it—is representative of public sector capital assets generally. If we are destined to neglect the maintenance, recapitalization, and capital improvement of public sector assets, including military facilities and equipment, then earmarking funds for these purposes may be the only acceptable recourse outside of privatization.

Flexible Funding

Earmarking lies on the far end of a continuum spanning fixed and flexible approaches to capital investment. On the other end lies real options analysis, which applies option valuation techniques to capital budgeting decisions. Traditionally, managers in the public and private sectors have relied on discounted cash flow techniques[15] in order to determine whether a proposed capital investment should be made. Future net cash flows are estimated over the anticipated life of a given project; if the value that is obtained is higher than the current cost of the investment, then (in theory) the investment should be made. In practice, however, this coarse-grain approach to investment decision-making fails to take into account a number of variables that may influence a project's profitability vis-à-vis the status quo or other investment possibilities.

Specifically, traditional cash flow techniques fail to capture the benefits associated with *flexibility* as it pertains to project size, timing, and process (i.e., the so-called "real options" available to management; Schubert & Barenbaum, 2007). By assigning value to flexibility, private and public sector organizations can make more informed capital budgeting decisions.

---

[15] The most commonly used techniques include net present value, internal rate of return, profitability index, breakeven time, and payback period (Chan, 2004).

However, real options are often given little consideration because the value of said benefits is far more difficult to assess relative to the costs. Thus, public sector managers, in particular, tend to view capital investment decision-making as an exercise in "straightforward cost minimization" (Schubert & Barenbaum, 2007). In the private sector, on the other hand, the relevant benefits are quantified in terms of a discrete figure—profit—the motivation for which lends itself more readily to the real options approach. Today, firms rely on several different option-pricing models.

Many factors reinforce the public sector's tendency to resist real options analysis. For instance, the pressure "to use it or lose it" strongly discourages the value of waiting. Rather, there is a tendency to spend as investment funding becomes available, which invariably leads to under or overinvestment. Schubert and Barenbaum (2007) describe the tendency for public sector managers to "overbuild":

> A public sector manager is likely to design a budget that overbuilds assets such as schools and water treatment facilities in order to serve future growth potential rather than to wait and see if such potential growth becomes more likely. In the scenario where the manager waits, and the potential growth occurs, the manager will need to go back and argue for more resources, when in the overbuilding scenario they need only argue for the financial resources once. (p. 144)

In other instances, where there is pressure to obligate limited funds quickly, investments are likely to be narrowly conceived and, hence, less effective in terms of contributing to strategic objectives.

Historically, real options, even in their crudest form (e.g., wait vs. invest), have seldom been considered within the context of depot capital investment. The portrayal provided by Glass and Schwartz of the Logistics Management Institute in 1988 paints an unflattering picture:

> Capital investments [in the military departments' depots], by and large, are made piecemeal, primarily to enhance peacetime operating efficiency or capability. They are biased toward projects that provide quick payback. Pressure to obligate funds quickly exacerbates the tendency to undertake small, easily justified, short-term projects. By using this piecemeal approach, the military services are missing the benefits of an integrated series of investments following a planned, technological direction. Most importantly, they are risking their depots' abilities to accomplish essential wartime missions. (p. iii)

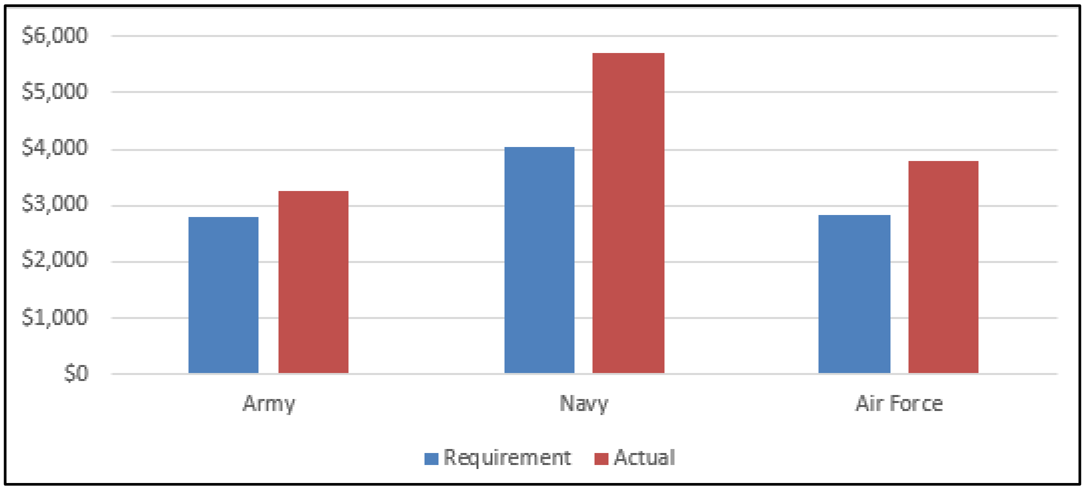
The military services have argued that the minimum investment requirement, by its very nature, discourages and undervalues investment flexibility. According to the DoD's Maintenance Executive Steering Committee (DoD, 2014), the military departments view the minimum investment requirement as a needless burden that "forces" investment in lower priority projects while discouraging or delaying investment in more costly, higher priority programs. In other words, the requirement undermines the ability to engage in strategic investment decision-making. However, the portrayal by Glass and Schwartz suggests that better investment decisions would not necessarily have been made in the absence of the requirement. Better strategic investment planning is needed at the department level in order to benefit from more flexible approaches to capital investment.

Trends by Military Department

Note. USMC data was included in Navy figure. Navy data from 2008–2013 was obtained from DoD (2014). Navy data from 2013–2017 obtained from Navy Working Capital Fund Budget Justifications. Air Force data was obtained from Air Force Working Capital Fund Budget Estimates, 2008–2017. Army data was obtained from Army Working Capital Fund Budget Estimates, 2008–2017.

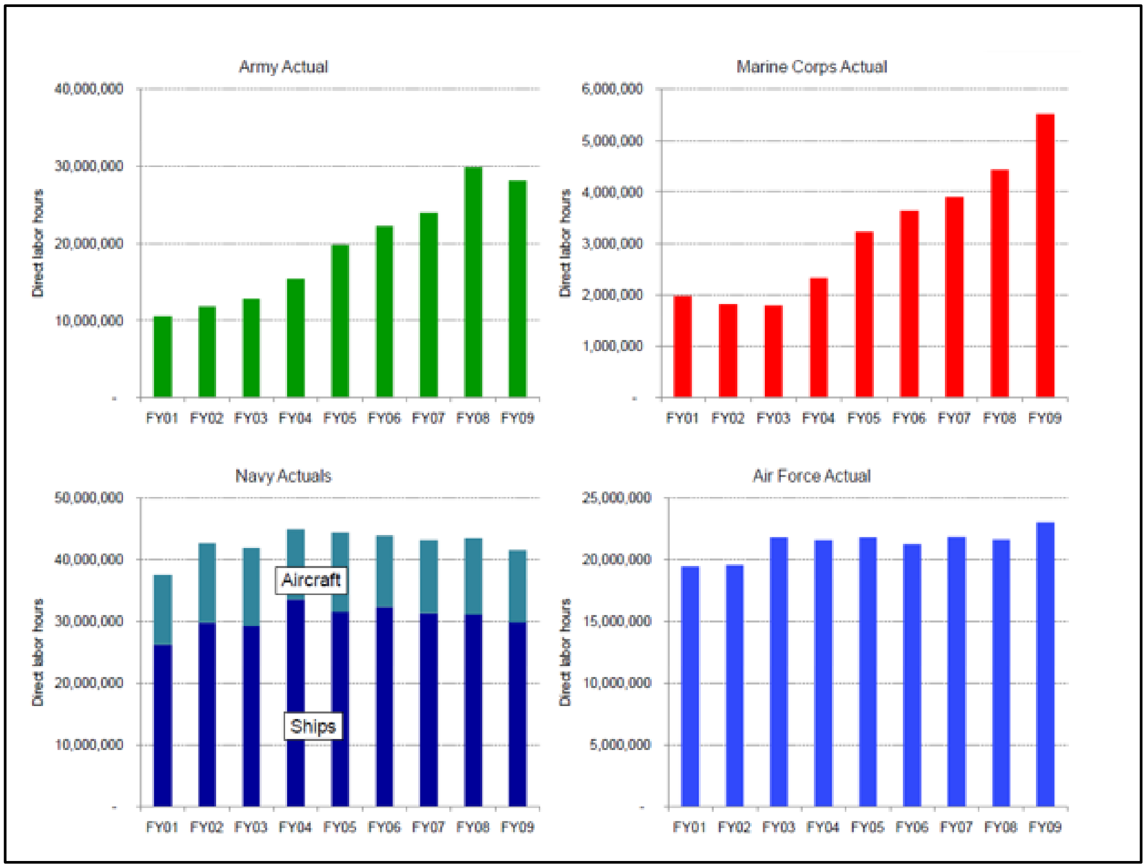
Figure 3. Capital Investment in Depots by Military Department ($ Millions), Actual and Requirement, 2008–2017

Figure 3 shows the actual annual capital investments made by each of the military departments and the corresponding annual investment requirements since the law came into effect. As discussed previously, representatives from all of the military departments have stated that it has been a challenge to meet the minimum investment requirement. The Navy and the Air Force, however, have not only met, but have exceeded, the minimum requirements. As for the Army, the investment landscape exhibits a significant peak in 2009, but also valleys corresponding to years in which the investment requirement was not met. Although it should be noted that the Army's cumulative investment has exceeded 6% of total revenue since the minimum requirement was put into effect; in total, the Army has invested $3.2 billion, or 6.9% of total revenue between 2008 and 2017 (see Figure 4).



Note. USMC data was included in Navy figure. Navy data from 2008–2013 was obtained from DoD (2014). Navy data from 2013–2017 was obtained from Navy Working Capital Fund Budget Justifications. Air Force data was obtained from Air Force Working Capital Fund Budget Estimates, 2008–2017. Army data was obtained from Army Working Capital Fund Budget Estimates, 2008–2017.

Figure 4. Cumulative Capital Investment in Depots ($ Millions) Between 2008 and 2017, Requirement and Actual



Note. The Army DLHs reflect work at the five major organic depots only.

Figure 5. Depot Repair and Maintenance Workloads by Military Service, Expressed in DLHs, 2001–2009
(Avdellas et al., 2011)

Interestingly, the relationship between weapon system use and required maintenance is not straightforward; moreover, this relationship varies considerably among the military services. Figure 5 compares the depot repair and maintenance workloads by military service, expressed in DLHs, between 2001 and 2009, a period marked by high levels of overseas military engagement. Whereas the Army and Marine Corps exhibited significant sustained increases in depot workloads, the Navy and Air Force workloads remained relatively stable, following modest post-2001 increases. Avdellas et al. (2011) explain that "this level of demand from the Air Force and Navy reflects the operation of an essentially constant inventory of aircraft and ships" (p. 1-4). In a RAND report, Cook, Ausink, and Roll (2005), writing about the Air Force, provided some additional insight:

> Surge has become part of regular ongoing depot activity instead of an unusual event. Furthermore, recent contingencies in which there have been increases in flying hours have not led to overwhelming increases in depot repair. Depot work is not necessarily linked to actual demand at a fixed point in time; appropriate planning can help the depots proactively prepare for expected conflicts. (xiii)

In fact, an earlier RAND report, Keating and Camm (2002) could not find "any category of organic [Depot Maintenance Activity Group; DMAG] expenditures that is consistently positively correlated with flying hours across multiple weapon systems" (p. xv).

Figure 6, which compares C-135 flying hours and organic repair expenditures, illustrates this lack of correlation.

In contrast, increasing workloads within Army and Marine Corps depots were attributed directly to "the added intensity of equipment operation in combat" (Avdellas et al., 2011, p. 1-4). These differences in workload (steady and predictable vs. unsteady and unpredictable) have obvious implications with regard to the 6% investment requirement given in that it is based on depot revenue, which, in turn, is a reflection of workload (specifically, direct labor hours). Needless to say, maintaining adherence to the investment requirement is likely less challenging when demand is steady and predictable in that investments can be made in conjunction with long-term strategy, rather than in response to a changing workload.

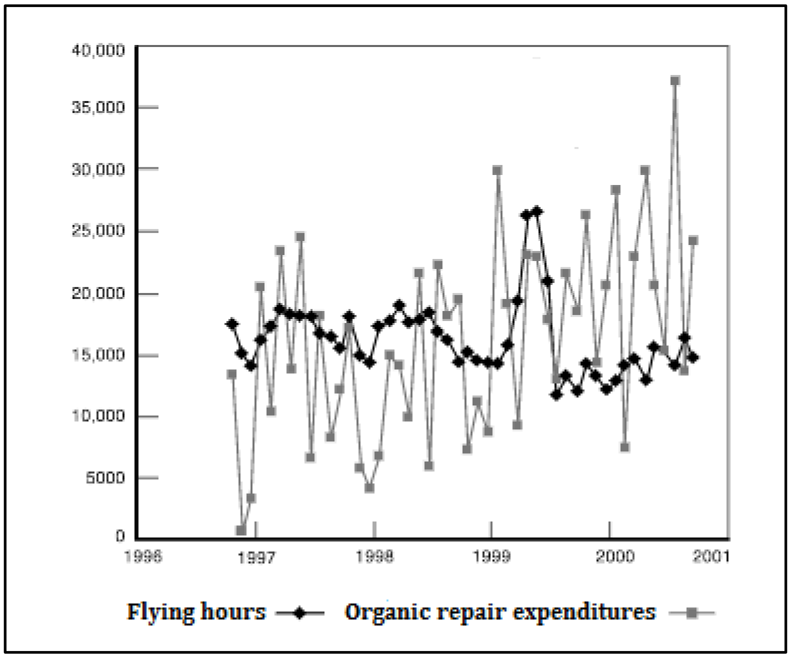


Figure 6. C-135 Flying Hours and DMAG Organic Repair Expenditures
(Keating & Camm, 2002)

The Army, in particular, may find it challenging to meet the investment requirement (if based on higher wartime revenues) following a drawdown from combat (when workloads are declining). Recall that the minimum investment requirement is based on total average revenue from the *preceding three years*. In effect, the law can force overinvestment during a period of declining resources, which is not an enviable position for an organization to find itself in. The challenge is even greater given that capital investment within the Army has been financed primarily through the working capital fund (rather than appropriations) in recent years. In response to this challenge (i.e., overinvestment during periods of declining workloads), OSD has proposed a forward-looking calculation method that bases the 6% target on total average revenue from the previous year, the execution year, and the following three years (i.e., the budgeted, planned, and programmed revenue).

It is not immediately clear whether this proposal represents a durable solution to the problem of overinvestment. One can envision a situation in which revenues are projected to increase rapidly following a prolonged period of operational stability. The military department may not need to make the required capital investments based on increased revenue projections; rather, greater investment may be needed following a drawdown from combat in order to recapitalize worn assets—in which case the current calculation method may prove preferable. In any event, it is unlikely that Congress would support a calculation method that relies so heavily on projected revenue.[16]

The Army, for its part, has proposed a reasonable compromise that bases the 6% target on average revenue from the *previous year, the execution year, and the future budget year.* During periods of steadily declining revenues, the “straddle” method generates minimum investment requirements that are lower than those generated by the current method, but higher than what would be generated by OSD’s forward-looking method. Conversely, during periods of increasing revenues, “straddle” would generate minimum requirements higher than the current method but lower than the forward-looking approach.

Figure 7 compares the effect of the current, straddle, and forward-looking methods on the Army’s minimum investment requirement in light of actual revenues generated between 2008 and 2016 and projections between 2017 and 2019. Revenues during this period declined significantly (from $5.9 billion in 2008 to $3.7 billion in 2016). As the graph indicates, using the straddle method would have resulted in a reduction to the minimum investment requirement of about $20 million annually.

Because the forward-looking method’s basis for investment spans five years, the line that is generated is comparably smoother, which translates to an investment requirement that is more consistent over time. By altering the Army’s proposal to include the preceding *two* years of revenue, the execution year, and the following *two* years, the peaks and valleys generated by the straddle method could be made similarly less prominent.

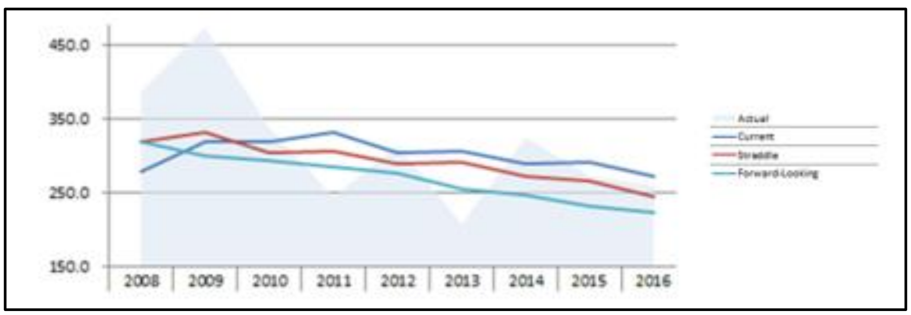


Figure 7. A Comparison of the Three Calculation Methods During a Period of Declining Revenues ($ Millions)

Note, however, that the shortfalls in actual investment (in 2011 and 2013) occur even when the forward-looking method is employed. This is not to say that the minimum investment requirement should necessarily be altered to accommodate such shortfalls in the future; at the same time, they draw attention to the reality of competing priorities and budgetary unpredictability. The military departments should have some added flexibility to

[16] *Eliminate OCO funding from the requirement?* The Army has also proposed that funding provided through the overseas contingency operations (OCO) fund be eliminated from the calculation method. This proposal undermines the linkage between revenue and investment that, as described previously, serves to justify fixed investment strategies in the first place; in other words, the proposal ignores the reality that investment and recapitalization needs are driven, in large measure, by the use, and subsequent wear and tear, of existing capital assets.

adjust to an unpredictable environment, especially considering that such flexibility could also serve to strengthen investment options analysis and facilitate strategic decision-making.

Recommendations and Conclusion

Based on the history of capital investment in military depots, our examination of the positives and negatives associated with fixed and flexible funding, the discussion on real options analysis, and trends in depot investment, we offer the following recommendations.

Recommendations

Develop and implement detailed strategic plans to properly guide capital investment.

- In some cases, the military departments lack detailed and comprehensive strategic investment plans. Without these plans, it will be challenging for depot leadership to pursue, develop, and execute integrated series of investments that ensure that the depots are able to provide the needed capabilities to meet future organic maintenance and repair requirements. Without these, it is difficult to convince Congress that the military departments have a plan to make the needed investments in the depots.

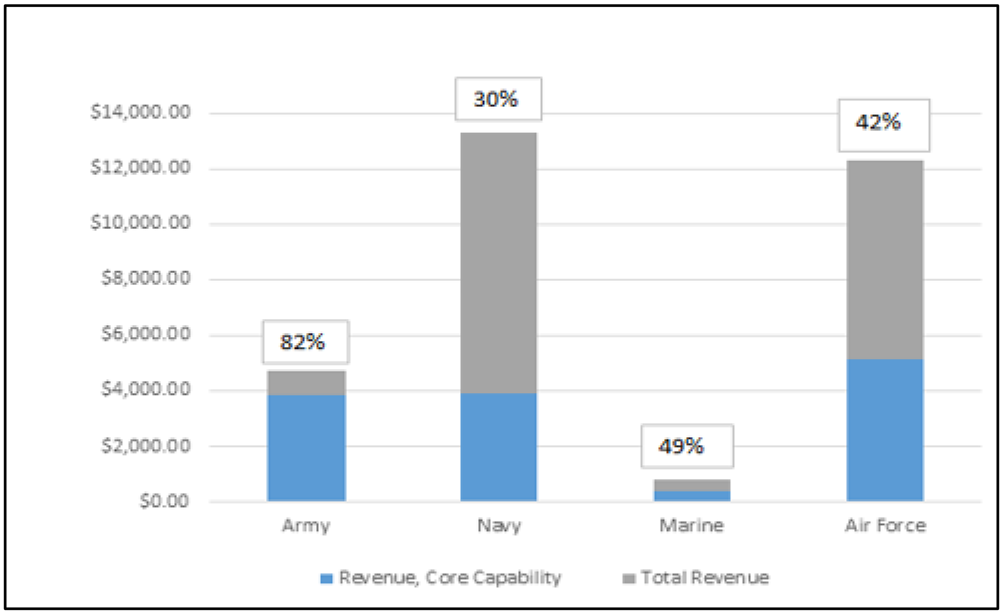
Maintain—but modify—the minimum investment requirement to encourage strategic investment decision-making.

- Given the lack of detailed and comprehensive investment plans and the historical challenges in making adequate investments that ensure the sufficiency of organic capabilities, a minmum investment requirement is warranted. However, in its current form, the requirement can lead to overinvestment during periods of declining revenues and potential underinvestment during periods of increasing revenues. A minimum requirement that bases the 6% target on revenue from the preceeding year, the year of execution, and the following year, or—to further reduce year-to-year fluctuations in the requirement—the preceeding two years, the year of execution, and the future two years, will improve investment effectiveness.

  This change alone may not provide the flexibility necessary to faciliate strategic decision-making and ensure that the planned intergrated series of investments are made. As discussed, investments have historically taken the form of small, short-term projects—a tendency that may be exacerbated by the annual investment requirement. Lawmakers may wish to consider modifying the requirement to enhance flexibility by, for example, allowing the military departments to credit any annual investment in excess of 6% to the future minimum requirement, thereby providing department leadership some additional leeway in formulating their investment strategy.

Continue to base the minimum investment requirement on total revenue.

- Proposals to base the investment requirement solely on revenue generated by “core” workload, or those that seek to eliminate from consideration OCONUS funding, represent misguided attempts to reduce the *required level of investment* by narrowing the *basis for investment*. If the required funding level (6% of revenues) is believed to be too high, then the 6% figure should be reconsidered at some point in the future. Narrowing the basis for investment has the potential to mask investment needs, if, for example, non-core or OCONUS-generated workload increases relative to core workload (a problem that is exacerbated by the fact that defining core requirements is a largely subjective enterprise that relies on methodolgies that are not consistently applied).



Note. The information in this figure came from data submitted in the DOD 2014 Biennial Core Report (GAO, 2016) and OUSD(AT&L; 2016).

Figure 8. Revenue Generated by Core Capabilities as a Percentage of Total FY 2016 Revenue

Finally, eliminating core from the investment basis would have a highly disparate impact on the military services. Figure 8 shows revenue generated by core capabilities as a percentage of total FY 2016 revenue for each of the military services. Were the requirement to be based solely on core-sustaining workload, the Army, which has struggled the most to meet the investment requirement would see minimal relief, whereas the Air Force and Navy requirements would fall considerably.

Modify the minimum investment requirement so that qualifying investments are not limited to the covered depots.

- The revenue generated by *all* of a military department's depots forms the basis for the 6% investment requirement; hence, it stands to reason that all of a department's depots should be made eligible for investment under the requirement. At present, a significant amount of the basis for the investment requirement is generated by software maintenance, yet many of the facilities that perform this maintenance are not "covered" under the current requirement.

Widen and clarify the definition of capital investment.

- The depots should rely on a standard definition of capital investment to ensure that the investment requirement does not inadvertently lead to increases in deferred maintenance. As discussed, rebuilding infrastructure after the end of its economic useful life or restoring it to “like new” condition constitute capital investment under standard defintions. Consequently, the depots must “expense” equipment and facilities that, under a standard defintion of capital investment, would be allocated over time. Relying on a standard defintion also helps reduce any grey area that might lead to needless bureaucratic meddling, added expense, or schedule delay.

Streamline the CIP approval process.

- The approval process for depot-level capital investments should be made flatter and faster. In some cases, CIP expenditures must be approved by a 4-star command. The structure of the working capital fund system may already provide sufficient constraints on capital investment decision-making at the depot level. In other words, customer sensitivity to increasing rates may serve to adequately promote sound capital investment at the depot level. Could not the subordinate commands, to which the depots already report, provide the necessary approval? The higher-level commands should devote more time and resources to developing long-term strategic investment plans that guide depot-level decision-making.

St**udy the potential for funding larger construction projects through the CIP.**

- Recall that, at present, construction projects valued at more than $750,000 can only be funded through congressional appropriation (which is often difficult to obtain). Consequently, there has been a longstanding tendency—which persists to this day—to “build groups of very small facilities” (Glass & Schwartz, 1988), when larger facilities would have been better economic investments. Funding larger construction projects through the working capital funds would provide military customers, the DoD, and Congress with a better understanding of the true cost of depot maintenance and repair, while improving the cost efficiency and effectiveness of capital investments.

Continue to pursue public-private partnerships.

- Public-private partnerships have allowed the DoD to harness the best mix of capabilities from the government and commercial sectors in many areas, including depot maintenance. The DoD should continue to pursue appropriate partnerships to the extent possible.

  Depot labor rates do not fully reflect the associated indirect costs; as a result, the rates are often lower than those seen in the commercial sector (Captain, 2017), which can provide an incentive for firms already performing depot-level maintenance to partner with the DoD (through a direct sales agreement[17]) in order to gain access to depots' personnel, equipment, and facilities. PPPs provide a "win-win" for both parties, improving depot capabilities, reducing costs, and enabling compliance with 50/50 and core requirements.

Conclusion

In the absence of dedicated funding, needed investment in capital assets can be overlooked. When funding is dedicated, unnecessary or shortsighted investments are sometimes made. Fortunately, fixed and flexible funding strategies are the two end-points on a continuum that spans a significant middle ground. Within the context of depot investment, the optimal balance has not yet been achieved. We believe that the above recommendations will generate the necessary shift along the continuum toward increased flexibility, thereby strengthening the military depots' capabilities and ensuring that their vital role in safeguarding America's security is maintained.

---

[17] Under a direct sales agreement, the contractor is held accountable for accomplishing the depot's funded workload via an outcome-based support contract. The contractor, in turn, "subcontracts" with the depot to acquire organic repair and maintenance services at the depot's hourly labor rate.

Reference List

10 U.S.C. § 2464. Core depot-level maintenance and repair capabilities.

10 U.S.C. § 2466. Limits on the performance of depot-level maintenance of materiel.

10 U.S.C. § 2474. Centers of Industrial and Technical Excellence: Designation; Public-private partnerships.

10 U.S.C. § 2476. Minimal capital investment for certain depots.

Avdellas, N., Berry, J., Disano, M., Oaks, D., & Wingrove, E., III. (2011*). Future capability of DoD maintenance depots* (Report No. LG901M2). Retrieved from http://www.acq.osd.mil/log/mpp/.plans.html/1_LG901M2_REPORT_FINAL_02-14-11.pdf

Avdellas, N., & Erickson, S. (2012). Limits of competition for depot maintenance contracting. Retrieved from Defense Acquisition University website: http://dau.dodlive.mil/2015/06/25/limits-of-competition-for-depot-maintenance-contracting/

Barrett, K., & Greene, R. (2013). Is earmarking the best way to fund projects? Retrieved from http://www.governing.com/columns/smart-mgm

Bratland, J. (2010). Capital concepts as insights into the maintenance and neglect of infrastructure. *Independent Review, 15*(1), 35–51. Retrieved from http://www.independent.org/pdf/tir/tir_15_01_3_bratland.pdf

Captain, T. (2017). Performance-based logistics: An answer to the readiness conundrum. Retrieved from http://www.nationaldefensemagazine.org/articles/2017/6/20/performancebased-logistics-an-answer-to-the-readiness-conundrum

Commission on Roles and Missions of the Armed Forces. (1995). *Directions for defense*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a402681.pdf

Cook, C. R., Ausink, J. A., & Roll, C. R., Jr. (2005). *Rethinking how the Air Force views sustainment surge*. Retrieved from https://www.rand.org/content/dam/rand/pubs/monographs/2005/RAND_MG372.pdf

Cullen, D. (2013). ASCE: Infrastructure-funding gap must be bridged. Retrieved from http://fleetowner.com/fleet-management/asce-infrastructure-funding-gap-must-be-bridged

Defense Acquisition University. (2015). Life cycle sustainment. In *Defense Acquisition Guidebook.* Retrieved from https://www.dau.mil/guidebooks/Shared%20Documents/Chapter%204%20Life%20Cycle%20Sustainment.pdf

DoD. (1996, March). *Policy regarding performance of depot-level maintenance and repair*. Retrieved from http://handle.dtic.mil/100.2/ADA314841

DoD. (2007). *Depot maintenance strategic plan*. Retrieved from http://www.acq.osd.mil/log/mpp/.plans.html/3_PartI_DMSP_SEMI_02_23_07.pdf

DoD. (2007, January 5). *Depot maintenance core capabilities determination process* (DoDI 4151.20). Retrieved from http://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/415120p.pdf

DoD. (2014, October 9). *10 USC 2476: Minimum capital investments at certain depots*. Presentation by Maintenance Executive Steering Committee and the Joint Group on Depot Maintenance.

DoD. (2016a). *Maintenance fact book*. Retrieved from http://www.acq.osd.mil/log/mpp/.factbook.html/Fact_Book_2016_Ecopy.pdf

DoD. (2016b). *Operation and maintenance overview: Fiscal year 2017 budget estimates*. Retrieved from http://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2017/fy2017_OM_Overview.pdf

DoD. (2017). *DoD financial management regulation,* Volume 2B, Chapter 9. Retrieved from http://comptroller.defense.gov/Portals/45/documents/fmr/archive/02barch/chapter09.pdf

DoN. (2016, February). *Department of the Navy fiscal year (FY) 2017 budget estimates*. Retrieved from http://www.secnav.navy.mil/fmc/fmb/Documents/17pres/NWCF_Book.pdf

GAO. (1997). *Defense depot maintenance: Uncertainties and challenges DoD faces in restructuring its depot maintenance program* (GAO/T-NSIAD-97-112). Retrieved from https://www.gpo.gov/fdsys/pkg/GAOREPORTS-T-NSIAD-97-112/html/GAOREPORTS-T-NSIAD-97-112.htm

GAO. (2001a). *Defense logistics: Actions needed to overcome capability gaps in the public depot system*. Retrieved from http://www.gao.gov/assets/240/232874.pdf

GAO. (2001b) *Sustaining readiness support capabilities requires a comprehensive plan*. Retrieved from http://www.gao.gov/assets/110/108755.pdf

GAO. (2009a). *Depot maintenance: Actions needed to identify and establish core capability at military depots* (GAO-09-83). Retrieved from http://www.gao.gov/products/GAO-09-83

GAO. (2009b). *Improved strategic planning needed to ensure that Army and Marine Corps depots can meet future maintenance requirements* (GAO-09-865). Retrieved from http://www.gao.gov/products/GAO-09-865

GAO. (2010). *Improved strategic planning needed to ensure that Air Force depots can meet future maintenance requirements* (GAO-10-526). Retrieved from http://www.gao.gov/products/GAO-10-526

GAO. (2014). *Depot maintenance: Accurate and complete data needed to meet DoD's core capability reporting requirements*. Retrieved from http://www.gao.gov/assets/670/665915.pdf

GAO. (2016). *Defense budget: DOD needs to improve reporting of operation and maintenance base obligations*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/1014201.pdf

GAO. (2017, September). *Naval shipyards: Actions needed to improve poor conditions that affect operations* (GAO-17-548). Retrieved from https://www.gao.gov/assets/690/687105.pdf

Glass, D., & Schwartz, L. (1988). *Depot maintenance modernization* (DTIC-ADA197948). Bethesda, MD: Logistics Management Institute. Retrieved from http://www.dtic.mil/get-tr-doc/pdf?AD=ADA197948

Jones, G., White, E., Ryan, E. T., & Ritschel, J. D. (2014). Investigation into the ratio of operating and support costs to life-cycle costs for DoD weapon systems. *Defense Acquisition Research Journal, 21*(1), 442–464. Retrieved from http://dau.dodlive.mil/files/2014/11/ARJ68_Jones.pdf

Jones, L., Candrva, P., & Devore, M. (2012). *Financing national defense: Policy & process*. Charlotte, NC: Information Age Publishing.

Keating, E., & Camm, F. (2002). *How should the U.S. Air Force Depot Maintenance Activity Group be funded?* Retrieved from https://www.rand.org/pubs/monograph_reports/MR1487.html

Martin and Martin, Inc. (2016). *Letterkenny Army Depot: Joint land use study.* Retrieved from http://www.fcadc.com/wp-content/uploads/2016/09/1378.1-JLUS-Final-Report-2016.pdf

Michael, J. (2015). *Earmarking state tax revenues* [Policy brief]. St. Paul, MN: Research Department, Minnesota House of Representatives. Retrieved from http://www.house.leg.state.mn.us/hrd/pubs/earmarking.pdf

Miller, T. D. (2010, June 30). *The Defense sustainment industrial base—A primer.* Brookings Institute. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/0630_defense_industrial_base_miller.pdf

National Defense Authorization Act for Fiscal Year 2006. Retrieved from https://www.gpo.gov/fdsys/pkg/PLAW-109publ163/pdf/PLAW-109publ163.pdf

National Research Council. (2011). *Examination of the U.S. Air Force's sustainment needs in the future and its strategy to meet those needs*. Washington, DC: The National Academies Press. Retrieved from https://www.nap.edu/catalog/13177/examination-of-the-us-air-forces-aircraft-sustainment-needs-in-the-future-and-its-strategy-to-meet-those-needs

Nutt, P. C. (2005, March 30). Comparing public and private sector decision-making practices. *Journal of Public Administration Research and Theory, 16*, 289–318. doi:10.1093/jopart/mui041

Office of the Assistant Secretary of Defense, Logistics & Materiel Readiness (OASD[L&MR]). (2015). Maintenance overview. Retrieved from http://www.acq.osd.mil/log/mpp/maintenance_overview.html

Office of the Assistant Secretary of Defense, Logistics & Materiel Readiness (OASD[L&MR]). (2016). *DoD maintenance 2016 fact book*. Retrieved from http://www.acq.osd.mil/log/mpp/.factbook.html/Fact_Book_2016_Ecopy.pdf

OUSD(AT&L). (2016). *Report to Congress on distribution of Department of Defense depot maintenance workloads for fiscal years 2015 through 2017*. Retrieved from http://www.acq.osd.mil/log/MPP/.plans.html/50-50_Reports/FY15-17_50-50_DOD_Workloads.pdf

Schubert, W., & Barenbaum, L. (2007). Real options and public sector capital decision-making. *Journal of Public Budgeting, Accounting, & Financial Management, 19*(2), 139–152. Retrieved from http://pracademics.com/attachments/article/707/Capital%20Budgeting.pdf

Teja, R. S. (1988). The case for earmarked taxes. *IMF Staff Papers, 35*(3), 523.

Transport Research Center. (2008). *Transport infrastructure investment: Options for efficiency.* OECD/International Transport Forum.

Wilkinson, M. (1994). Paying for public spending: Is there a role for earmarked taxes? *Fiscal Studies, 15*(4), 119–135. Retrieved from https://econpapers.repec.org/article/ifsfistud/v_3a15_3ay_3a1994_3ai_3a4_3ap_3a119-35.htm

Preliminary Findings: Is the Ratio of Investment Between R&D to Production Experiencing Fundamental Change?

Rhys McCormick—is an Associate Fellow with the Defense-Industrial Initiatives Group (DIIG) at CSIS. His work focuses on unmanned systems, global defense industrial base issues, and U.S. federal and defense contracting trends. Prior to working at DIIG, he interned at the Abshire-Inamori Leadership Academy at CSIS and the Peacekeeping and Stability Operations Institute at the U.S. Army War College. He holds a bachelor's degree in security and risk analysis from Pennsylvania State University and a master's degree in security studies from Georgetown University.

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. From 2011 to 2014, he served as a Senior Executive in the Department of Defense, serving first as Chief of Staff to Under Secretaries of Defense (AT&L) Ashton B. Carter and Frank Kendall, before directing the Joint Rapid Acquisition Cell. From 2005 to 2011, Hunter served as a Professional Staff Member of the House Armed Services Committee. Hunter holds a master's degree in applied economics from Johns Hopkins University and a bachelor's degree in social studies from Harvard University.

Greg Sanders—is a Fellow in the International Security Program and Deputy Director of the Defense-Industrial Initiatives Group at CSIS, where he manages a research team that analyzes data on U.S. government contract spending and other budget and acquisition issues. In support of these goals, he employs SQL Server, as well as the statistical programming language R. Sanders holds a master's degree in international studies from the University of Denver, and he holds a bachelor's degree in government and politics and a bachelor's degree in computer science from the University of Maryland.

Abstract

With the advent of the information age, both commercial industry and the Department of Defense are moving towards complex R&D-intensive systems over the simpler, mass-produced systems of the industrial age. This paper uses budgetary and program data to better understand the historical trends in the relationship of production costs to development costs in complex acquisition programs.

Introduction

This paper presents preliminary analysis of the historical trends in the relationship of production costs to development costs in complex acquisition programs. To understand this phenomenon, the study team examines it at two different levels. The first is the macro investment level where portfolio management trade-offs are made between aggregate development and procurement and between programs. The second level are individual programs where the ambitions of the program and the underlying technology shape the resources required for a program to complete development.

Starting with the macro level, for all militaries, finding the proper investment balance between the needs of the current force structure and the potential future force structures is a recurring challenge. Militaries must find a balance between the procurement of existing systems with the development of new platforms and technologies. In the United States, this dynamic has followed a cyclical historical pattern in the ratio of procurement to research, development, test, and evaluation (RDT&E) in the Department of Defense (DoD) budgets. When the overall DoD budget increases, procurement rises disproportionately and thus the ratio of procurement to RDT&E also increases. Inversely, when the defense budget falls, procurement spending falls faster than the overall budget and the ratio of procurement to RD&TE also falls. Overall, since fiscal year (FY) 1955, the DoD has spent an average of 2.2 dollars on procurement for every dollar it spent on RDT&E. However, as shown in Figure 1, the center of the range of this cycle abruptly shifted downwards following the peak of the Regan buildup when the DoD was spending 3.25 dollars on procurement to one dollar on RDT&E.

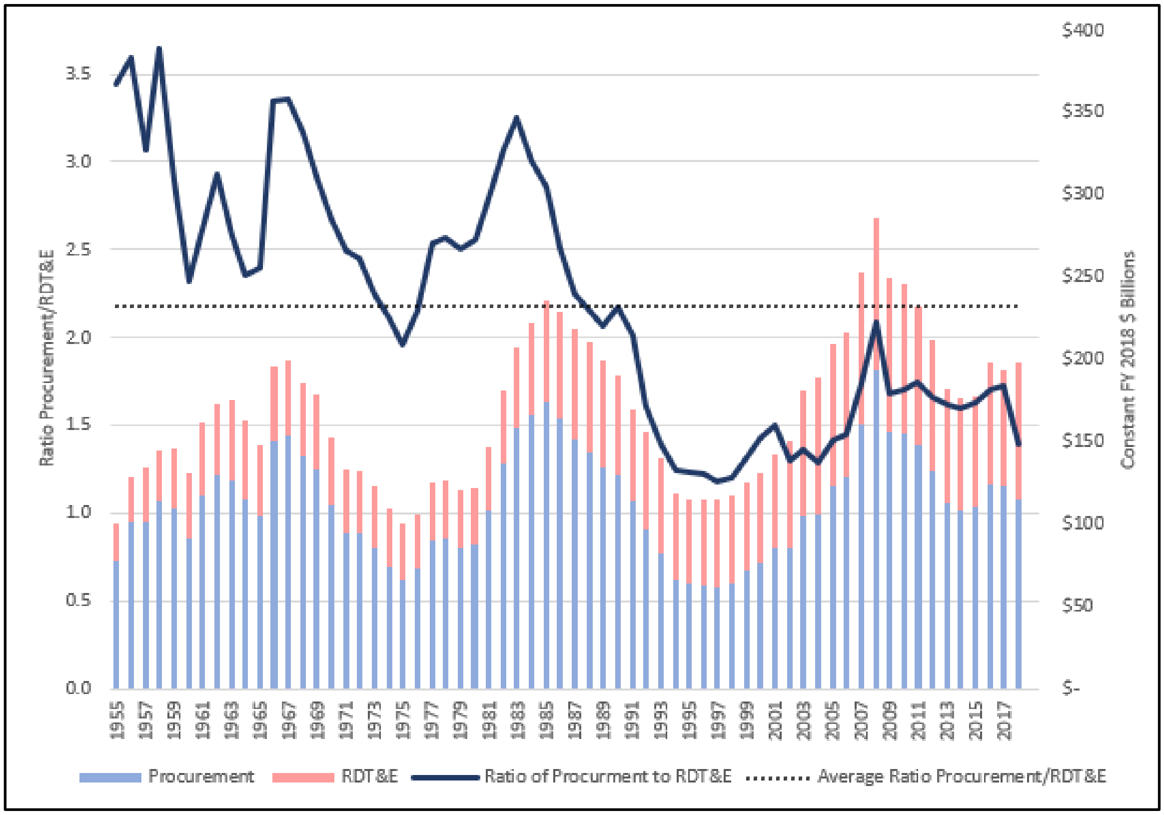


Figure 1. DoD Ratio of Procurement to RDT&E, 1955–2018
(OUSD[C], 2017; CSIS analysis)

Following the peak of the Reagan buildup, the ratio of procurement to RDT&E fell as the overall defense budget declined, but this time it fell more sharply than previous drawdowns and failed to rebound to expected levels during subsequent budget increases. The ratio of procurement to RDT&E fell to a historic low of 1.22 in FY 1998 compared to the previous historic low: 1.95 in FY 1975, as a result of the 1990s "procurement holiday" that led to sharp cuts in procurement spending and only relatively modest declines to RDT&E. Furthermore, when defense contracting rebounded in the 2000s, the 2.17 ratio of procurement to RDT&E in FY 2008, remained well below the historical average and significantly below the 3.24 to 3.64 ratios of previous peak buildup periods. The historically low ratio of procurement to RDT&E seen during the 1990s can be explained by the decisions made following the end of the Cold War, the success of the Gulf War in prioritizing the development of next-generation weapon systems, and the "procurement holiday" in the 1990s that slashed procurement budgets. But why did the ratio of procurement to RDT&E remain below historical averages during the mid-to-late 2000s, despite historic modernization budgets? Is there something different about this generation of major defense acquisition programs (MDAP) or are there other factors at play?

Weapon systems, and other complex acquisition programs, have always grown in complexity from generation to generation, but has the information age brought about a fundamental change in the relationship between R&D and production? Compared to the simpler mass-produced systems of earlier generations, today's systems are exponentially more complex and heavily leveraged on software. In 1960, software performed only 8% of an F-4's functions. By 1982, software performed 45% of the functions in the F-16, and by 2000, software performed 80% of the functions in the F-22 (DSB, 2000). The DoD's software development and maintenance requirements have been estimated to be growing somewhere between 15% to 20% annually (Tate, 2017). These trends are not unique to defense. Boeing's 787 Dreamliner required 14 million source lines of code (SLOC). Some of today's premium-class cars utilize up to 100 million SLOC. When compared to the 400,000 SLOC in the original Space Shuttle, the importance of software is evident.

The growth in software requirements is staggering, but is a fundamental change in the relationship between production and development underway? Has the information age changed the importance of R&D, or has it just shifted the focus of R&D efforts to software? For companies, this question could have business-altering dynamics. For firms like Boeing and others in the defense marketplace, the business model has been to conduct development and early production at a financial loss before turning a profit as production ramps up. However, if a fundamental change in the relationship between development and production is underway, these business models may no longer be sustainable.

Declining Procurement–RDT&E Ratios: The Result of a Broken Acquisition System?

The problems with the current MDAP portfolio are well-known and have led many to state that the defense acquisition system is broken. However, are there truly significant differences between this generation of MDAPs and previous generations? In 2008, the Government Accountability Office (GAO) made headlines when it reported that 70% of the DoD's MDAPs were over budget and behind schedule. Cumulative MDAP cost-growth totaled $295 billion and the average MDAP was 21 months behind schedule (GAO, 2008). Are the delays and cost growth associated with today's MDAPs higher than historic rates?

Policymakers and analysis have been concerned about the development and procurement of major weapon system platforms since the advent of the modern defense industrial base at the end of World War II. Despite the ever-increasing complexity of weapon systems generation to generation, studies of the changes of the procurement system show that the management problems have been remarkably consistent. Multiple studies have shown that weapon system cost growth during development and procurement and schedule growth has remained largely consistent over time (Bolten et al., 2008; Drezner et al., 1993; IDA, 2010; Jarvaise, Drezner, & Norton, 1996; Marshall & Meckling, 1959; Younossi et al., 2007). Additionally, recent analysis both in and out of government shows that cost growth today is similar to historical rates (OUSD[AT&L], 2016; Watts & Harrison, 2011). These studies suggest that most cost-growth in MDAPs occurs during the development phase. Average overall cost growth ratio during development totals approximately 1.6, but there are differences between types of platforms (Younossi et al., 2007).

Concerning cycle times, Tate found that "highly-visible programs," those with the greatest total acquisition costs, are driving a false perception that cycle times have been increasing over the past 25-plus years (Tate, 2016). Instead, Tate (2016) found that for all commodity types, including "highly-visible programs," cycle growth over the past 25 years has been statistically insignificant. Scholars have found a difference in the cost growths of programs resulting from the differing conditions of the funding climate for when programs achieve Milestone B status. McNicol and Wu (2014) found that if a program attained Milestone B status when the budget climate was relatively constrained, the program could be burdened by overly optimistic costing assumptions. When these optimistic assumptions fail to pan out, the result is significant cost growth. Thus, McNicol and Wu (2014) conclude that programs that attain Milestone B status in "bust" periods are more likely to experience cost growth than programs that attain Milestone B status in "boom" periods.

The literature suggests that the cost and schedule growth in MDAPs since the 1990s is not beyond historical norms and does not explain the top-line trends in the declining ratio of procurement to RDT&E. Although these previous studies extensively studied cost and schedule growth in MDAPs, there has been little analysis of the ratio of procurement to RDT&E. Previous analysis has largely focused on the topline budget trends previously highlighted (Harrison, 2013; Harrison, 2016) or topline MDAP data. The *Performance of the Defense Acquisition System: 2015 Annual Report* stated that for the 76 active MDAPs, "at the median, the procurement share is more than six times larger than the RDT&E share" (Office of the Under Secretary of Defense for Acquisition Technology and Logistics [OUSD(AT&L)], 2015). However, these reports contained no further breakdowns of the data service or platform.

Research Approach

Given the literature suggesting that these trends are not necessarily the result of a broken acquisition system, but other factors, this paper seeks to further investigate the potential sources of the declining ratio of procurement to RDT&E across the DoD. Additionally, given the software growth trends occurring in defense and non-defense complex acquisition programs, this paper looks to see if there are similar trends occurring in the broader economy. Specifically, this paper seeks to answer the following questions:

- Has the relative importance of R&D changed across the broader economy over the past four decades?
- What are the historical trends in the ratio of procurement to R&D funding in the military services? Are there significant differences between the military services?
- What are the top-line historical trends in the ratio of procurement to RDT&E funding for the MDAP portfolio?

Data Methodology

To compare the DoD's budget trends to broader economic trends, this paper looks at the historical R&D intensity trends in select industries. The study team selected industries that are similar in nature to the defense industry. R&D intensity is measured by total expenditure of all firms on R&D over total net sales in an industry. Although R&D intensity is not perfectly analogous to the DoD's budgetary trends, it provides a rough approximation given limited visibility into more specific budgetary trends within private companies. Additionally, while R&D intensity is not without issues (Hughes, 1988), it is a commonly used method of measuring "the relative importance of R&D across industries and among firms in the same industry" (National Science Board, 2008).

To measure the top-line historical ratio trends for the MDAP portfolio, this paper uses the data from Selected Acquisition Reports (SAR). The issues associated with SARs have been well-noted[1], but they still provide the most reliable source of data (Hough, 1992). This paper uses the annual SAR data accessed through Defense Acquisition Management Information Retrieval (DAMIR) for MDAPs from FY 1997 to FY 2017. For analytical purposes, the CSIS team excluded any program in each year that reported incomplete procurement or RDT&E data. Additionally, the study team focused only on the current estimated ratio in a given SAR. Future analysis will expand this analysis to compare the current estimates against projected ratios at different acquisition milestones.

To enable preliminary historical comparisons, the DAMIR SAR data is supplemented with historical data from RAND's Defense System Cost Performance Database (DSCPD) made available in *The Defense System Cost Performance Database: Cost Growth Analysis Using Selected Acquisition Reports* (Jarvaise et al., 1996). The DSCPD provides a summary of SAR data from the 1960s to FY 1994. The DSCPD provides funding breakdowns, both the stated planned estimates at different acquisition milestones and "current" for historical MDAPs.

From the historical dataset, this paper looks at the historical trends in the ratio of procurement to RDT&E funding for MDAPs. This paper begins by looking at the historical ratio trends estimated at Milestone B/II, the "official start of a program." Next, this paper looks at the actual spending for programs that are completed, or have largely been completed and are no longer submitting annual SARs.

Due to the gap in data from FY 1995 to FY 1996, this paper only presents the study team's preliminary findings which are subject to change. The study team has identified additional data sources to be used to supplement and validate existing data, as well as addressing the gaps in the data. CSIS will incorporated the additional data sources, where available, into its final technical report to be released in late 2018.

Analysis

The following sections present analysis of the data related to the three research questions. This section begins with analysis of the historical R&D intensity across the broader economy, followed by analysis of ratio trends within the services and concludes by analyzing the MDAP portfolio topline trends.

Historical R&D Intensity Trends by Industry

How do these trends compare to the broader marketplace? Although software and high-tech intensive devices get much of the media attention, are we seeing shifts in the importance of R&D to industry generally? These questions are challenging to address. The data on civilian firms is not as thorough as the defense budget which makes perfectly analogous comparisons difficult for outside researchers. Instead, the study team looks to R&D intensity which is an economic metric that aids in understanding the general importance of R&D to firms in a certain sector.

[1] Some of the problems with utilizing SARs include, but are not limited to, inconsistent baseline cost estimates, exclusion of some significant cost elements, exclusion of special access programs, constantly changing preparation guidelines of SARs, inconsistent interpretations of preparation guidelines across programs, cost sharing in joint programs, and the reporting of the effects of costs changes rather than their root causes.

Analysis of historical analysis of R&D intensity trends within industries is complicated by the creation of the North American Industry Classification System (NAICS) in 1997 and subsequent move away from, and eventually elimination of the Standard Industrial Classification (SIC) system. This report uses the R&D intensity data from the National Science Foundation who used SIC codes up until 1998, before switching to NAICS codes in 1999. Given the shift from SIC codes to NAICS codes, this paper focuses on the general trends from 1970 to 1998 and then from 1998 to 2014.

R&D Intensity: 1970–1998

The data show that R&D intensity in the manufacturing sector did not following a singular long-term trend, but a series of intermediate trends as shown in Figure 2 below. Throughout the 1970s, manufacturing R&D intensity gradually fell from approximately 3.7 in 1970 to 2.6 in 1979. Then from 1980 to 1986, manufacturing R&D intensity grew at 7.68% Compound Annual Growth Rate (CAGR). Finally, from 1986 to 1996 manufacturing R&D declined at -2.84% CAGR.

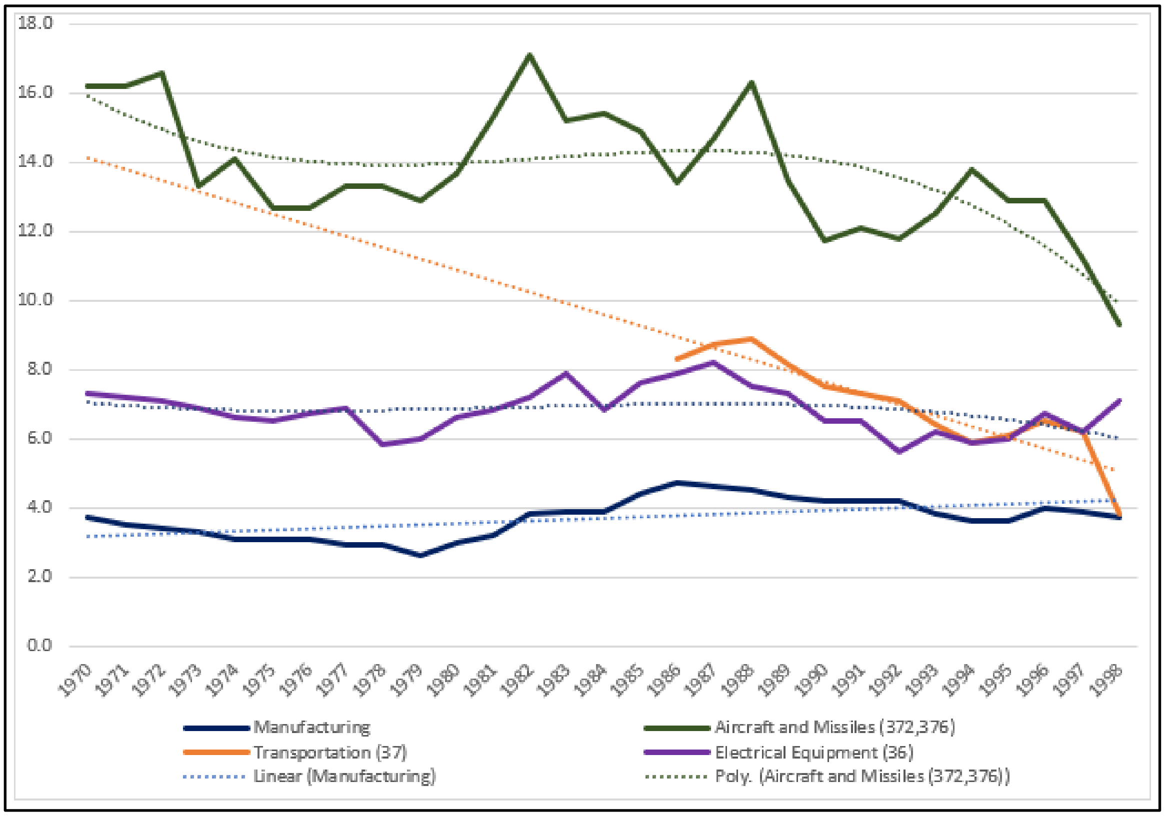


Figure 2. R&D Intensity in Manufacturing and Select Industries, 1970–1998
(National Science Foundation Industrial Research and Development Information System; CSIS analysis)

Beyond the top-line manufacturing R&D intensity trends, the data show that the trends could vary between industries. Across the broader Transportation industry, the data show that R&D intensity was on a downward trend since the late 1980s. R&D intensity in the Transportation industry declined at -6.3% CAGR from 1987 to 1998. Comparatively, the Electrical Equipment industry followed a more cyclical pattern, but remained relatively steady.

One subset of the Transportation industry is of particular interest to this paper, Aircraft and Missiles. Of note, the data show that R&D intensity trends in the Aircraft and Missile industry followed a cyclical pattern, relatively similar to the DoD's trend in the ratio of procurement to RDT&E. In the Aircraft and Missile industry, R&D had cyclical periods of growth followed by periods of decline and vice versa, but has been broadly trending downward since the mid-1980s.

R&D Intensity: 1999–2014

The data show that there is not an overall trend in the importance of R&D, as measured by R&D intensity, but that there is more uncertainty at lower levels.

As shown in Figure 3, R&D intensity in the manufacturing industry was on an upward trajectory until the onset of the fiscal crisis, but fell sharply in the following years. Manufacturing R&D intensity has started trending back upwards in the last two years of available data, but sustained growth is necessary before any definitive conclusions can be drawn. Comparatively, R&D intensity in the non-manufacturing sector had been gradually declining even prior to the fiscal crisis and the one-year sharp decline. However, R&D intensity in the non-manufacturing industry rebounded quicker than the manufacturing sector and has been on a steady growth pass since, but still remains below historical averages since 1999.

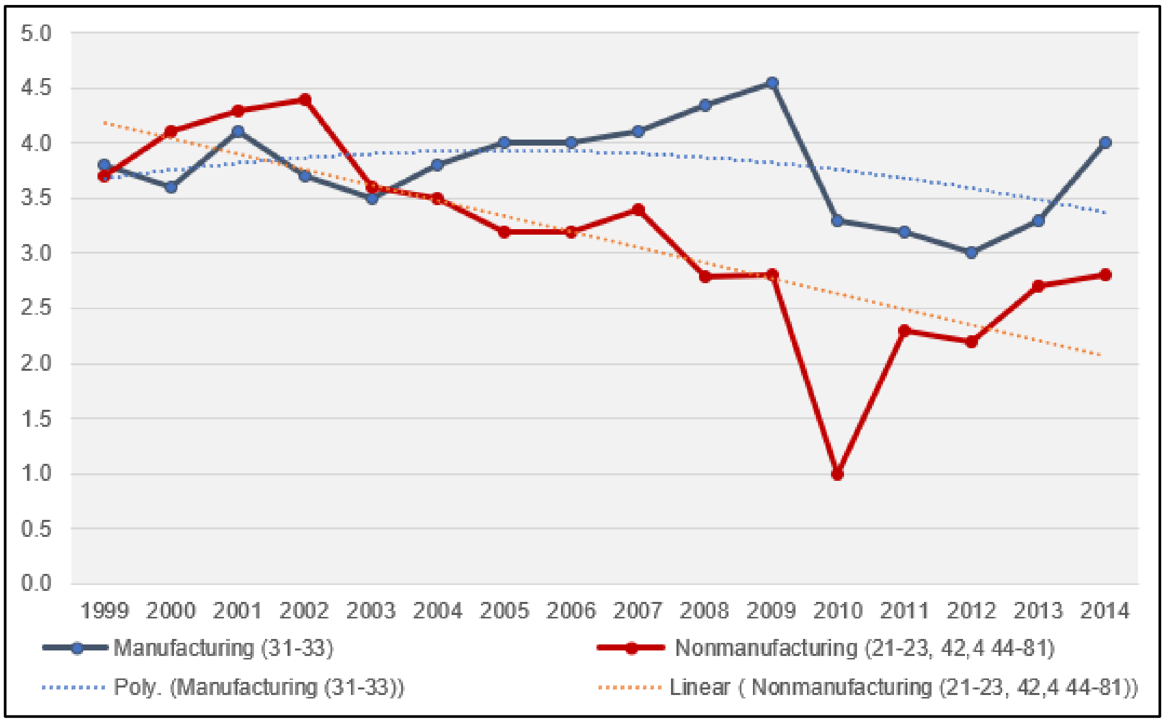
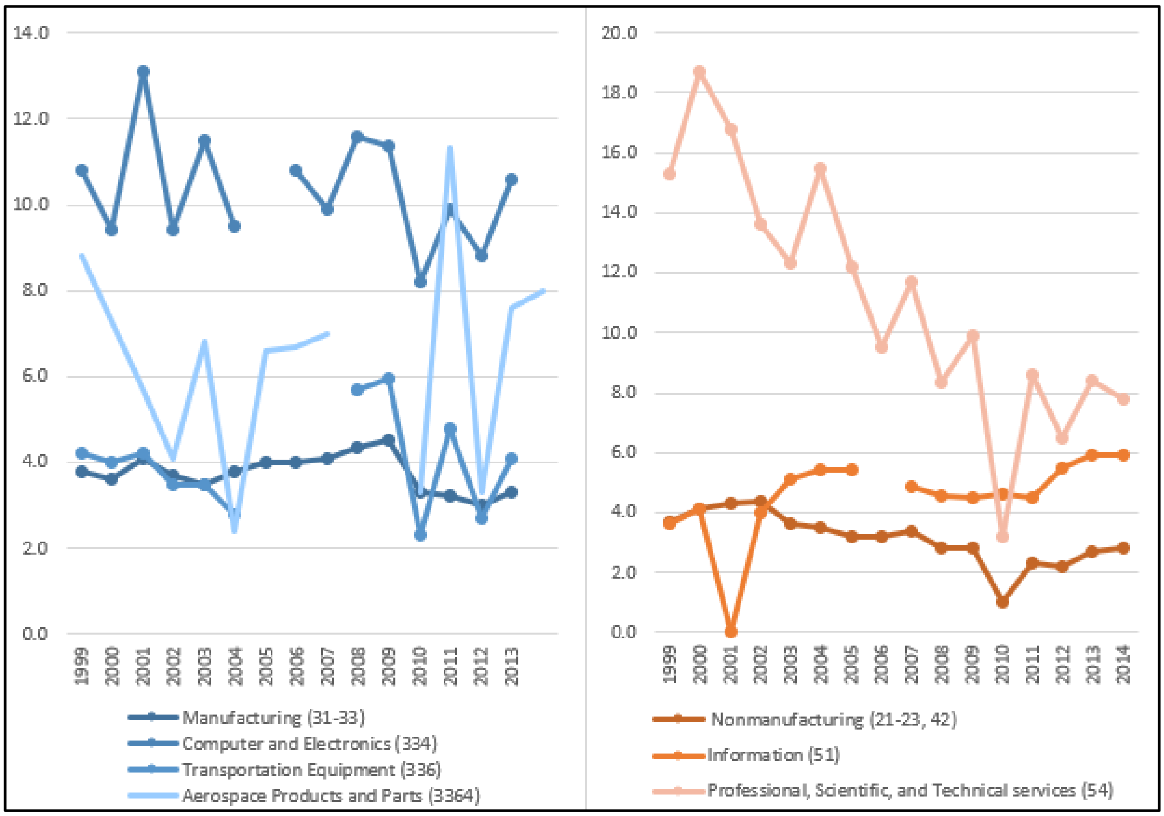


Figure 3. Manufacturing and Non-Manufacturing R&D Intensity, 1999–2014
(National Science Foundation Business Research and Development Innovation Survey; CSIS analysis)

The data show that below the Manufacturing, there are no obvious trends in the data. Amongst the selected manufacturing industries and sub-industries (Computer and Electronics, Transportation Equipment, Aerospace Products and Parts), the data is often noisy, with significant variance from year to year.

The data for selected industries in the non-manufacturing sector is less noisy overall and does suggest more definitive certain trends. Figure 4 shows the selected industries within the Manufacturing and Non-Manufacturing sectors from 1999 to 2014. The data show that, in general, R&D intensity in the Professional, Scientific, Technical Services industry (NAICS code 54), is on a downward trend since 2000. From a peak of 18.7 intensity in 2001, R&D intensity for that sector fell as low as 3.2 in 2010, and is currently 7.8 in the last available data. In the other selected non-manufacturing industry, information, after holding steady throughout the years reporting data, R&D intensity has been slightly trending upward in the past few years. However, just as the trend for the broader manufacturing sector, it is too early to draw definitive conclusions.



Note. Gaps in the data in certain years are due to the NSF masking that year's data.

Figure 4. R&D Intensity in Select Industries, 1999–2014
(National Science Foundation Business Research and Development Innovation Survey; CSIS analysis)

In general, the data do not show an overall shift in the importance of R&D across the broader market over the past 40 years as measured using R&D intensity. There are certain trends in the broader Manufacturing sector and the Information industry that suggests a shift could be occurring, it remains too early to draw definitive conclusions.

Ratio of Procurement/RDT&E by Service

The topline data shows that the ratio of procurement to RDT&E is down across the board in the DoD, but are similar trends occurring in the parts of the DoD making the actual investment decisions—the military services?

Navy

Amongst the military services, the Navy has the highest historical average ratio of procurement to RDT&E: 3.38. Figure 5 shows the ratio of procurement to RDT&E in the Navy's budget from FY 1955 to FY 2018.

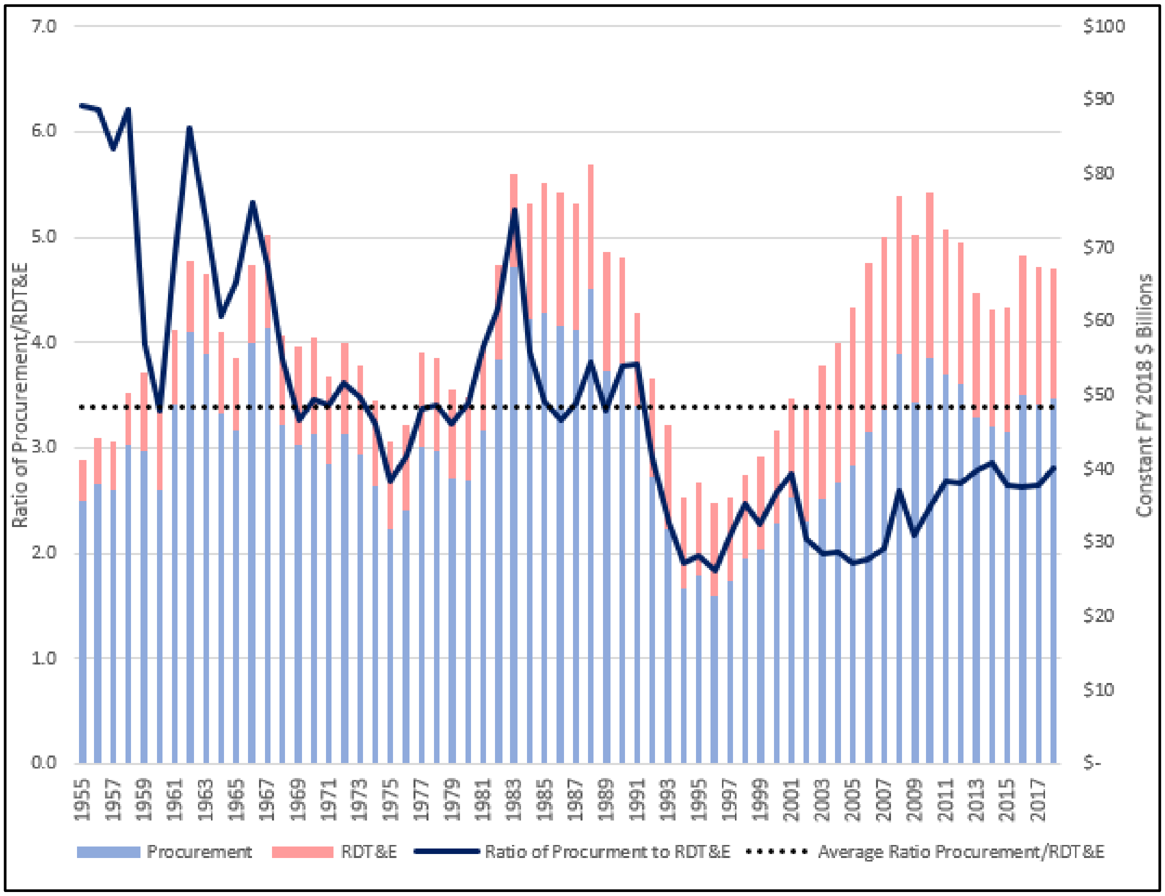


Figure 5. Navy Ratio of Procurement to RDT&E, 1955–2018
(OUSD[C], 2017; CSIS analysis)

The Navy ratio trends generally followed overall DoD trends, with a few notable differences. After the ratio of procurement to RDT&E in the Navy peaked in 1983, that ratio fell precipitously in the following years despite near-historic procurement budgets. Whereas overall DoD budget trends in those years were largely driven by procurement funding declining quicker than RDT&E funding, that was not the case in the Navy. Instead, the declining ratio in the Navy was driven by a $5 billion increase in RDT&E over a five-year period while the procurement budgets stayed relatively flat. In recent years, the ratio of procurement to RDT&E in the Navy fell from FY 2000 to FY 2005, during which time Navy RRDT&E funding grew at nearly three times the rate of procurement funding. Since FY 2005, the Navy's ratio of procurement to RDT&E has been on a stable, but gradual growth path.

Air Force

Of the three military services, the Air Force has the lowest historical ratio of procurement to RDT&E, spending 1 dollar on development for every 2.05 dollars spent on production. This is not surprising given the Air Force's cultural preference for new, high-tech solutions even before it became its own military service in 1947. Figure 6 shows the ratio of procurement to RDT&E in the Air Force's budget from FY 1955 to FY 2018.

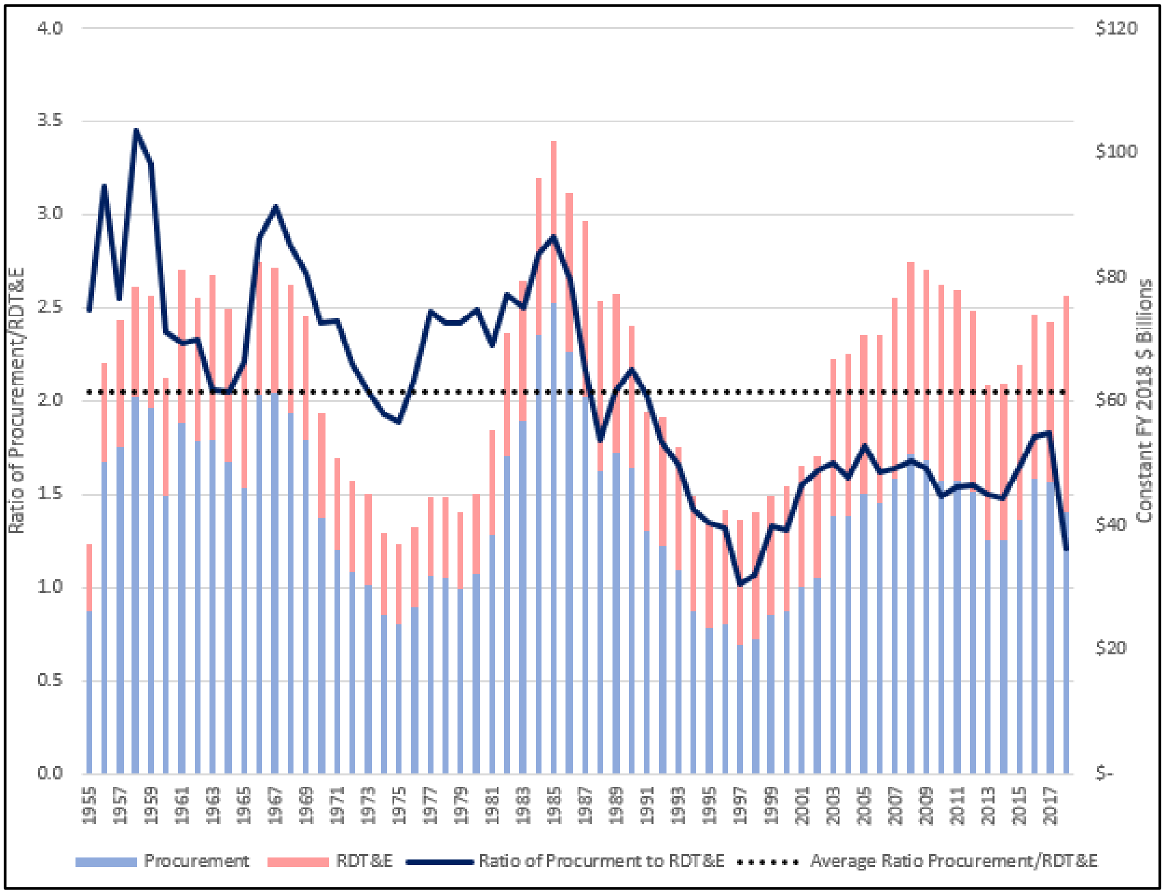


Figure 6. Figure 6: Air Force Ratio of Procurement to RDT&E, 1955–2018
(OUSD[C], 2017; CSIS analysis)

The Air Force differs from the other services in a few notable ways. First, whereas the other services saw a rapid spike followed by a quick decline during the Reagan buildup, the Air Force experienced a smaller spike earlier, followed by a more gradual rise to its crescendo. Second, the Air Force ratio fell lower than any of the services during the 1990s, falling to as low as 1.07 in 1997. That year, the Air Force spent nearly one dollar on RDT&E for each dollar it spent on procurement.

After hitting a historic low in FY 1997, the Air Force's ratio of procurement to RDT&E gradually grew from FY 1998 to FY 2008. The ratio then began to fall again from FY 2008 to FY 2014, before increasing from FY 2014 to FY 2017.

Army

The data show that the Army, unlike the Navy and Air Force, has continued to follow cyclical historical patterns. Since FY 1955, the Army has spent on average, 2.36 dollars on procurement for every dollar spent on RDT&E. Figure 7 shows the ratio of procurement to RDT&E in the Army's budget from FY 1955 to FY 2018.

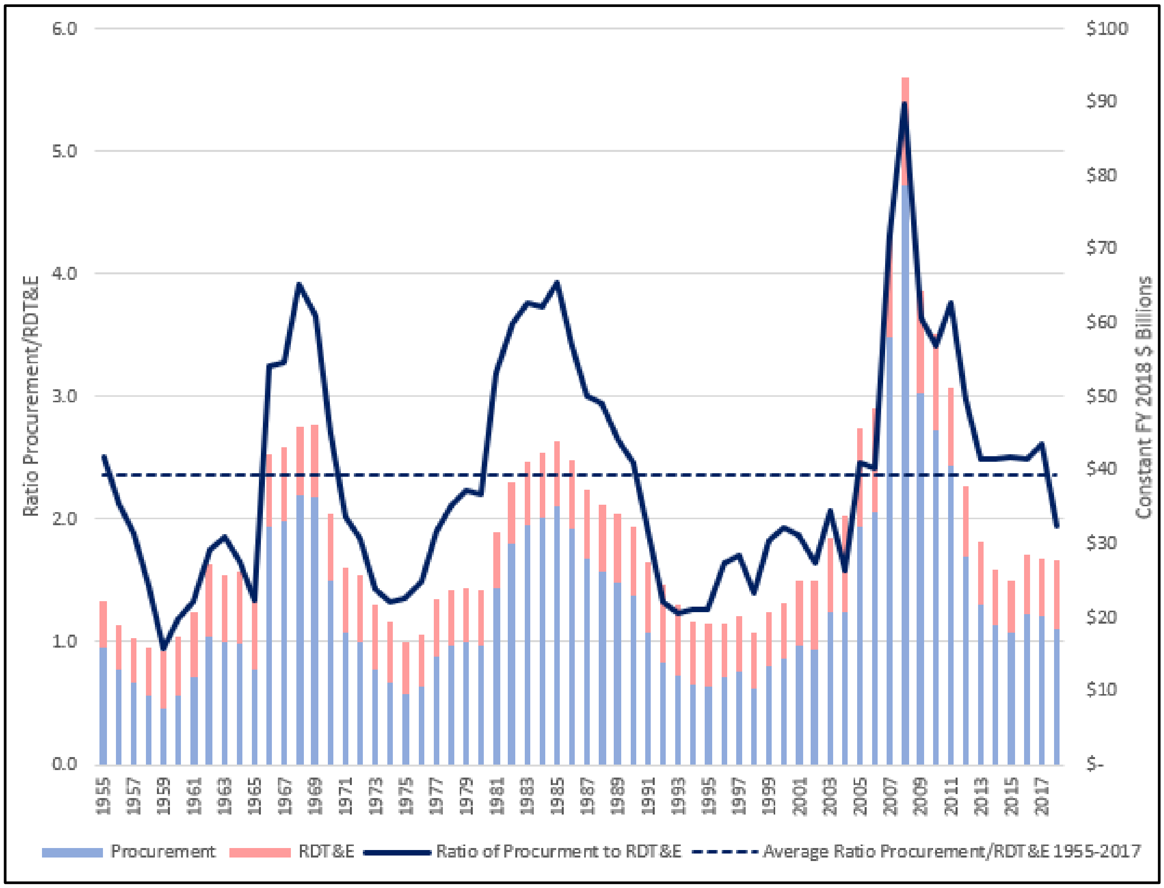


Figure 7. Army Ratio of Procurement to RDT&E, 1955–2018
(OUSD[C], 2017; CSIS analysis)

The Army was the only service that returned to the historical cyclical pattern during the mid-2000s. This trend was driven, in a large part due to the operations in Iraq and Afghanistan, but these trends are interesting given the failures of nearly all the Army's marquee acquisition programs over that period. Additionally, as the wars in Iraq and Afghanistan began to wind down and end, the ratio of procurement to RDT&E remained above historical averages contrary to previous cycles. During this most-recent drawdown, Army RDT&E fell much more sharply than in previous cycles.

Annual Current Estimates MDAP Portfolio Ratio of Procurement/RDT&E

Figure 8 displays the statistical distribution of the current reported ratio of procurement to RDT&E for MDAPs in each year's December SAR from FY 1997 to FY 2017.[2]

[2] The 2008 data is excluded because of a temporary policy guidance change that led the DoD to only submit SAR for programs with a Nunn-McCurdy breach, which was only the H-1 Upgrade that year.

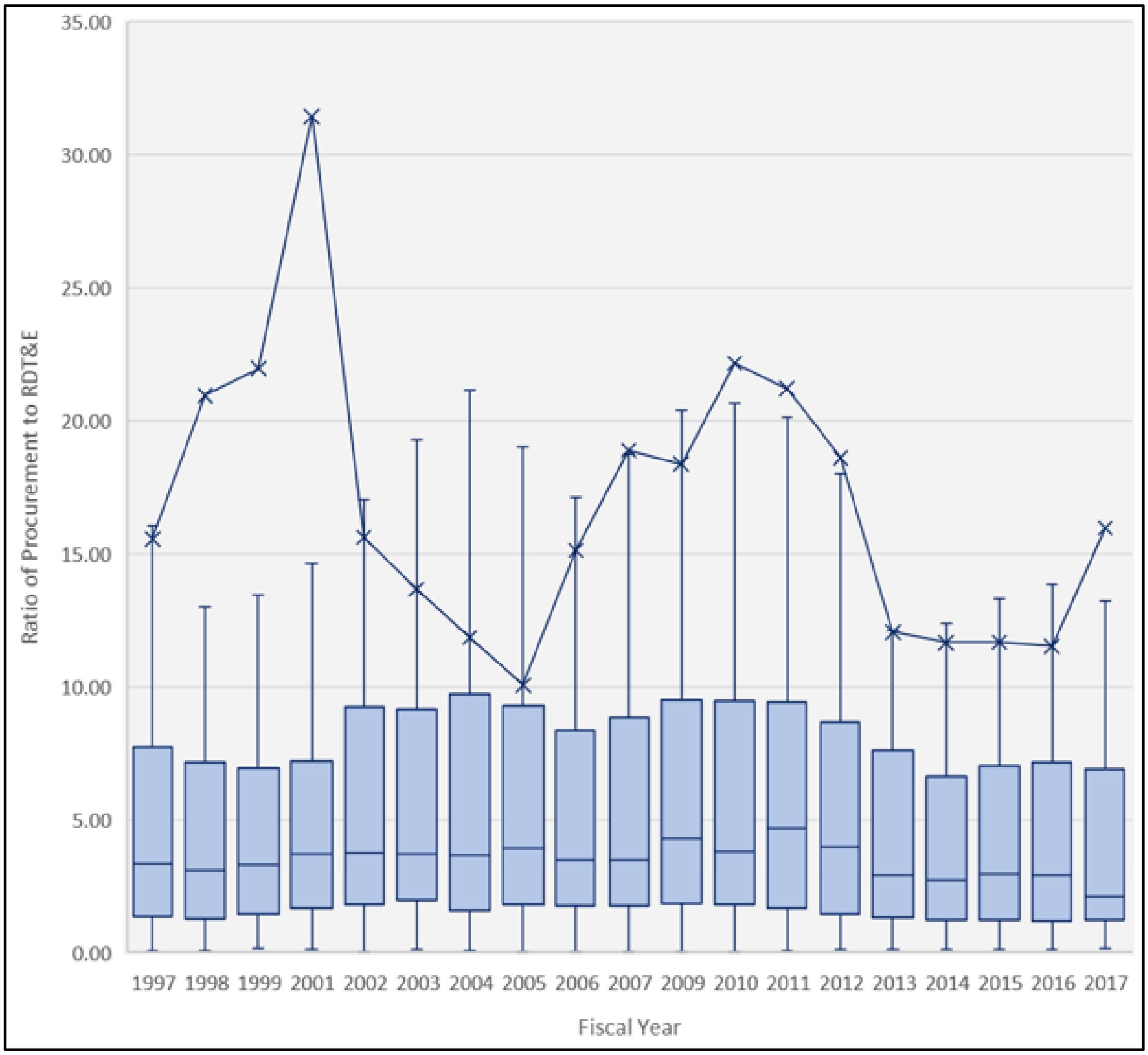


Figure 8. Ratio of Procurement to RDT&E Across the MDAP Portfolio, 1997–2017
(December Selected Acquisition Reports accessed through DAMIR; CSIS analysis)

The topline data is inconclusive and varies based on the measure used. The data show that from FY 1997 to FY 2012, excluding FY 2009, the topline median remained relatively steady.[3] It has only really been since FY 2013 that the median has started to fall. The mean of the ratio is far more volatile, as it is influenced by extremes at both ends of the scale, including outliers not shown in the box-and-whiskers plot. More pertinent are the upper and lower quartiles, shown as the top and bottom of each blue box. The lower quartile has a slow rise, starting in 1998, is stable from 2002 to 2010, before beginning a slow decline that flattens out in 2014. The pattern for the upper quartile is similar but more volatile. Starting in FY 2002, the upper quartile of the distribution of MDAP ratio of

[3] FY 2009 was an outlier due to the rapid procurement of large quantities of Mine-Resistant Ambush Protected vehicles.

procurement to RDT&E rose sharply, and with two exceptions, remained at that higher level for the next 10 years before gradually declining and leveling off over the last five or so years. There is possibly a cycle at work here, but not one that aligns with the account-level defense trends seen in Figure 1. The elevated period between 2002 and 2011 does align with rising wartime procurement spending, but shows no sign of the rapid rise from 2004 to 2008.

It is possible the decline in recent years across multiple measures might indicate a longer-term shift in the ratio, but if so, the magnitude is minimal and the timing does not clearly align with a greater role for software. Moreover, DAMIR coverage begins with programs still active in 1997, which is after the larger shift in the ratio as that coincided with the end of the Reagan build up and post-Cold War draw down had already taken place. To give a greater historical perspective on the change in individual programs, the study team turned to a dataset produced by RAND in cooperation with the office of Program Analysis and Evaluation, the predecessor to today's office of Cost Assessment and Program Evaluation.

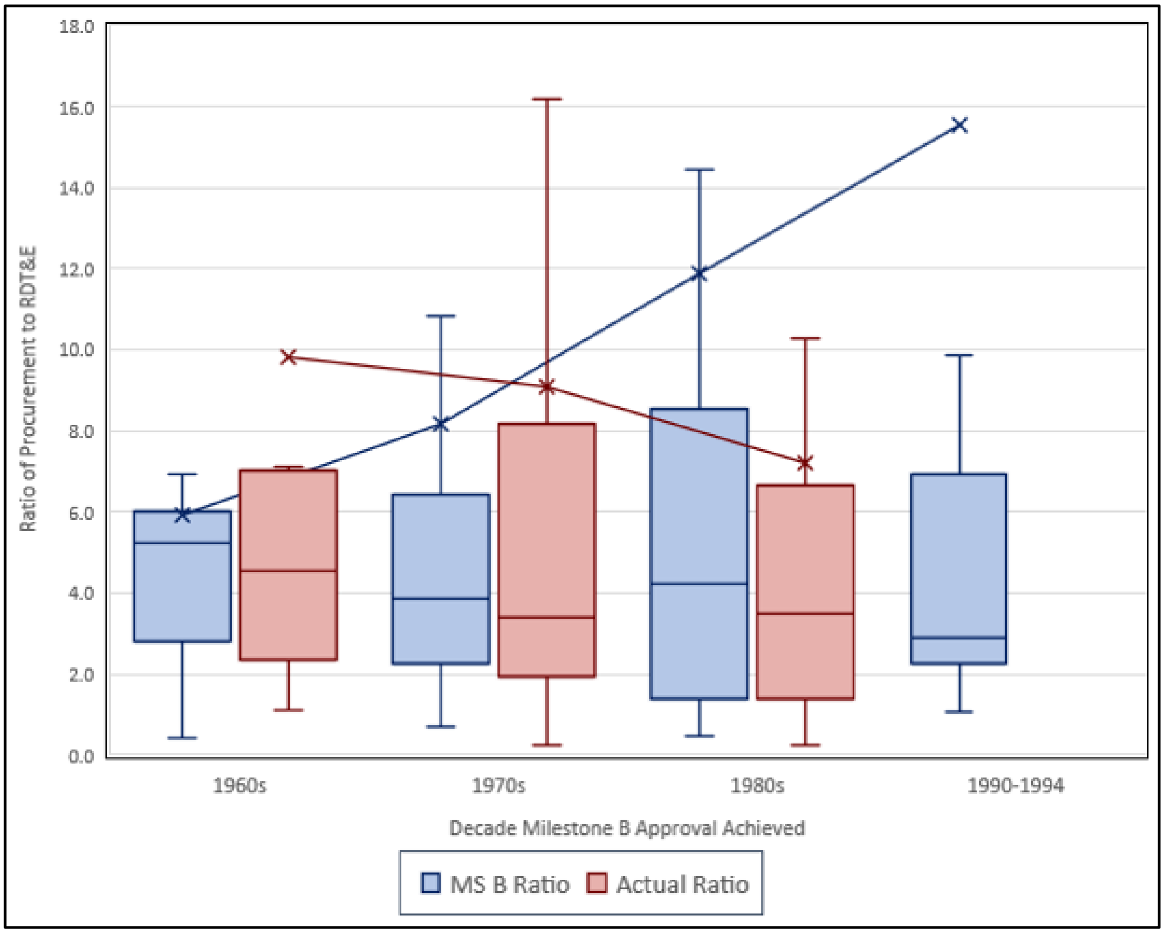


Figure 9. Comparison of Estimated Ratio of Procurement to RDT&E at Milestone B v. Production

(Jarvaise et al., 1996; CSIS analysis)

Figure 9 shows the statistical distribution of the ratio of procurement to RDT&E for MDAPs estimated at Milestone B versus the actual rates. MDAPs were categorized in both groupings based on the decade the program achieved Milestone B approval.

The historical SARs data show two critical trends. First, the data show significant variance between the ratio of procurement to RD&TE estimated at Milestone B and the final report ratio of procurement to RDT&E. Changes in the ratio of procurement to RDT&E are not necessarily the result of negative factors (cost growth, schedule slippage, etc.), but could also be the result of other factors such as significant increases in procurement quantities. For example, the F-16 Milestone B approval only planned on procuring 650 planes, but the Air Force ended up purchasing more than 2,000 by the end of the program. Second, the data show that while the average planned estimated ratio of procurement to RDT&E at Milestone B has been trending upwards, the actual average ratio of procurement to RDT&E has been on a gradual downwards trend. This finding is consistent with the existing literature showing that acquisition program estimates are often based on optimistic assumptions that aren't reflected in the final programs.

Building on the declining ratios for actual expenditures, the decade to decade trends also align with the account level trends seen in Figure 1. The sixties and eighties had higher median ratios while the seventies and nineties had lower ones. For the final technical report, the study team seeks to combine this dataset with the DAMIR dataset to gain a better understanding of trends across the entire period.

Conclusion

Has the relative importance of R&D, as measured by R&D intensity, changed across the broader economy over the past four decades?

The R&D intensity trends shows no overall trend in the change of importance of R&D, across the broader economy over the past four decades. The data show that from 1970 to 1998, R&D intensity in the manufacturing sector followed a series of cyclical up and down intermediate trends, but remained relatively steady in the aggregate. During that period, R&D intensity in the Aircraft and Missiles industry followed a cyclical pattern not dissimilar from the DoD's ratio of procurement spending to RDT&E, and had been on a downward trend since the mid-1980s.

Since 1999, there has been no change in the overall R&D intensity trends, but there is more uncertainty at lower levels. Prior to the fiscal crisis, R&D intensity within the manufacturing sector was increasing, before plummeting during the fiscal crisis. The R&D intensity in this sector has since recovered in recent years and begun to rise again, but additional data is still necessary to confirm that these current trends are not an anomaly. The information industry in the non-manufacturing sector has also shown a similar positive trend in recent years, but additional data is also still required.

There has been a more definitive downward trend in the Professional, Scientific, and Technical Services industry. At the start of the century, R&D intensity in the Professional, Scientific, and Technical Services industry was comparable to the Computer and Electronics industry, but has since fallen sharply. From 2000 to 2014, R&D intensity in the Professional, Scientific, and Technical Services industry declined at -5.7% CAGR.

What are the historical trends in the ratio of procurement to R&D funding in the military services? Are there significant differences between the military services?

The data show that there are significant differences between the different military services. Across the DoD historically, the Navy has the highest ratio of procurement to RDT&E amongst the military. The Navy's 3.38 ratio of procurement to RDT&E is 43% higher than the Army's (2.36) and 65% higher than the Air Force's (2.05).

The Navy and the Air Force generally followed the same cyclical historical trends as the overall trends, but with a few points of interest worth noting. In the Navy, the ratio of procurement to RDT&E began to decline despite continued, near-historic procurement budgets as a result of increases to the RDT&E budget. In the other services, the declining ratio was largely the result of procurement funding falling more sharply than RDT&E funding, but that wasn't the case with the Navy. Of note in the Air Force, the ratio of procurement to RDT&E fell as low as 1.07 in 1997.

Meanwhile, the Army did not see a shift away from the historical cyclical pattern, unlike the Navy or the Air Force, and returned to levels above historical averages during the mid-2000s. These trends are heavily influenced by operations in Iraq and Afghanistan, but are still interesting given the failure of the Army modernization's programs since the end of the Cold War. The Army, more so than any other service, has been maligned for the failures and problems of its acquisition system.

What are the historical trends in the ratio of procurement to RDT&E funding for the MDAP portfolio?

The SARs data from the 1970s to FY 1994, show that, historically, there is notable variance between the ratio of procurement to RDT&E estimated at Milestone B and the program's final ratio during production. This is neither a good nor bad trend as there are many influencing factors, but instead highlights that the initial estimates are often inaccurate. Second, these data show that the ratio for actual expenditures gradually declined from the 1960s to the 1980s in alignment with the account level trends.

Looking at the DAMIR data since 1997, the topline trends are inconclusive. Only in recent years, the ratio of procurement to R&D has declined across multiple measures and even the magnitude of those declines was relatively minimal and not below past low water marks.

The study team is not yet prepared to make definitive conclusions until it has had more time to explore the details closely to isolate potential findings hidden in the topline data noise.

Next Steps

Moving forward, the CSIS study team will focus its efforts on expanding its analysis of the MDAPs data to further explore the trends in the historical relationship between R&D and production. The study team will begin by incorporating the previously mentioned additional data sources to bridge the three-year gap in SARs data between RAND's DSCPD and DAMIR. These datasets will then be combined into a singular, standardized dataset.

From this dataset, the study team will expand its analysis of the MDAPs to include additional program characteristics that enable the team to provide more granular analysis of the data. These additional characteristics include, but are not limited to, the military service responsible for the program, the platform type, whether the program had a prototyping phase, and the program's length. With the addition of these program characteristics, the study team can better isolate potential changing dynamics that don't appear in the topline data. For example, the study plans to closely examine the historical trends in the Aircraft sector after the addition of these variables.

References

Bolten, J. G., Leonard, R. S., Arena, M. V., Younossi, O., & Sollinger, J. (2008). *Sources of weapon system cost growth.* Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG670.pdf

Drezner, J. A., Jarvaise, J. M., Hess, R., Norton, D. M., & Hough, P. G. (1993). *An analysis of weapon system cost growth.* Retrieved from https://www.rand.org/pubs/monograph_reports/MR291.html

Defense Science Board (DSB). (2000, November). Defense Science Board Task Force on Defense Software. Retrieved from https://doi.org/10.1080/09700160108455338

GAO. (2008). *Defense acquisitions: Assessments of selected weapon programs.* Washington, DC: Author.

Harrison, T. (2013). *Chaos and uncertainty: The FY2014 defense budget and beyond.* Washington, DC: Center for Strategic and Budgetary Assessment. Retrieved from http://csbaonline.org/uploads/documents/Analysis-of-the-FY-2014-Defense-Budget.pdf

Harrison, T. (2016). *Analysis of the FY2017 defense budget*. Center for Strategic and International Studies.

Hough, P. G. (1992). Pitfal*ls in calculating cost growth from selected acquisition reports.* Retrieved from http://www.rand.org/pubs/notes/N3136.html

Hughes, K. (1988). The interpretation and measurement of R&D intensity—A note. *Research Policy, 17*(5), 301–307. https://doi.org/10.1016/0048-7333(88)90010-8

IDA. (2010). *The major causes of cost growth in defense, II* (IDA Paper P-4531). Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA519883%5Cnhttp://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA519883

Jarvaise, J. M., Drezner, J. A., & Norton, D. (1996). *The defense system cost performance database: Cost growth analysis using selected acquisition reports*. Retrieved from http://books.google.com/books?id=d5XE-St9v44C

Marshall, A. W., & Meckling, W. H. (1959). Predictability of the costs, time, and success of development. In *The rate and direction of inventive activity: Economic and social factors.* Retrieved from http://www.nber.org/chapters/c2138.pdf

National Science Board. (2008). *Science and engineering indicators, 1.* Retrieved from https://wayback.archive-it.org/5902/20150818072529/http://www.nsf.gov/statistics/seind08/pdf/volume1.pdf

McNicol, D. L., & Wu, L. (2014). *Evidence on the effect of DoD acquisition policy and process on cost growth of major defense acquisition programs*. Institute for Defense Analyses. Retrieved from https://www.acq.osd.mil/parca/docs/ida-p5126.pdf

Office of the Under Secretary of Defense for Acquisition Technology and Logistics (OUSD[AT&L]). (2014). *Performance of the defense acquisition system: 2014 annual report*. Retrieved from http://www.defense.gov/Portals/1/Documents/pubs/Performance-of-Defense-Acquisition-System-2014.pdf

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2015). *Performance of the defense acquisition system: 2015 annual report*. Retrieved from https://www.defense.gov/Portals/1/Documents/pubs/Performance-of-Defense-Acquisition-System-2015.pdf

Office of the Under Secretary of Defense for Acquisition Technology and Logistics (OUSD[AT&L]). (2016). *Performance of the defense acquisition system: 2016 annual report.* Retrieved from http://bbp.dau.mil/docs/performance-of-defense-acquisition-system-2016.pdf

Office of the Under Secretary of Defense (Comptroller; OUSD[C]). (2017). *National defense budget estimates for FY 2018*. Retrieved from http://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2018/FY18_Green_Book.pdf

Tate, D. M. (2016). *Acquisition cycle time: Defining the problem*. Retrieved from https://www.ida.org/idamedia/Corporate/Files/Publications/IDA_Documents/CARD/2016/D-5762.ashx

Tate, D. M. (2017, March). *Software productivity trends and issues* [Conference paper]. Retrieved from https://www.ida.org/idamedia/Corporate/Files/Publications/IDA_Documents/CARD/2017/D-8367.pdf

Watts, B. D., & Harrison, T. (2011). *Sustaining critical sectors of the U.S. defense industrial base.* Retrieved from http://csbaonline.org/uploads/documents/2011.09.20-Defense-Industrial-Base.pdf

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. S., Jr., Jain, A., & Sollinger, J. M. (2007). *Is weapon system cost growth increasing? A quantitative assessment of completed and ongoing programs*. Retrieved from http://www.rand.org/pubs/monographs/MG588.html

Disclaimer

The Center for Strategic and International Studies (CSIS) does not take specific policy positions; accordingly, all views expressed in this presentation should be understood to be solely those of the author(s).

Extending an Econophysics Value Model for Early Developmental Program Performance Prediction and Assessment

Raymond D. Jones, COL, USA (Ret.)—served in the Army for nearly 30 years and is currently a lecturer with the Graduate School for Business and Public Policy at the Naval Postgraduate School in Monterey CA. His final assignment in the Army was as the Deputy Program Executive Officer for the Joint Tactical Radio System (JTRS) in San Diego CA. He has twice served as an ACAT ID Program Manager and has had multiple operational and acquisition related tours. He is a 1995 graduate of the U.S. Naval Test Pilot School with multiple flight test assignments. He has a BS in Aerospace Engineering from the United States Military Academy, an MS in Aeronautical Engineering from the Naval Postgraduate School, an MBA from Regis University, and an MS in National Resource Strategy from the Industrial College of the Armed Forces.

Thomas J. Housel—specializes in valuing intellectual capital, knowledge management, telecommunications, information technology, value-based business process reengineering, and knowledge value measurement in profit and non-profit organizations. He is a tenured full professor for the Information Sciences (Systems) Department at NPS. He has conducted over 80 knowledge value added (KVA) projects within the non-profit, Department of Defense (DoD) sector for the Army, Navy, and Marines. Dr. Housel also completed over 100 KVA projects in the private sector. The results of these projects provided substantial performance improvement strategies and tactics for core processes throughout DoD organizations and private sector companies. [tjhousel@nps.edu]

Abstract

This study is focused on presenting the viability of an econophysics theory of value as a means for creating a quantitative value metric to estimate the future value of Department of Defense (DoD) technology acquisition programs. We will describe a simple value model and further definitize this model into a DoD acquisition framework to illustrate the utility for developmental programs within the DoD and defense industrial base. This paper will describe a method by which a metric for surrogate financial value can be allocated across a program, allowing program managers to assess the surrogate return on investment (s-ROI) of their programs and providing greater flexibility in managing program risk. Additionally, we introduce a new program performance index that reflects s-ROI which incorporates a risk-based measure that modifies and extends the traditional earned value management (EVM) cost and schedule indices and provides an earlier indication of program challenges. We refer to this index as the s-ROI Performance Index (RPI), which has the potential of being a leading program indicator on overall program value and performance. Recommendations for the use of this model in DoD acquisitions, in general, are provided at the conclusion of this study.

Introduction

The research problem is that the Department of Defense (DoD) is not able to predict the **value** to risk relationship of technology acquisitions under development. Current metrics used in DoD acquisition programs are not sufficient to adequately predict program performance early enough for decision makers to objectively influence program outcomes. DoD programs tend to be managed using cost as an independent variable (CAIV), limiting the program managers' (PMs') flexibility with regard to managing risk. Exacerbating this problem is the lack of quantitative economically based value metrics for use in estimating the future value of DoD acquisition programs[1]. This leaves the PM to focus on cost growth, often at the expense of system performance capabilities. Additionally, the primary index by which PMs gain insight into cost variance is through EVM cost and schedule indices, which tend to be lagging indicators due to the latency in the data and the lack of predictive power on future performance. Hence, the PM is driven to making performance trades to reduce cost growth at the expense of capabilities.

When there is no unique quantitative value metric with which to take advantage of commonly used financial ratios, such as ROI, the PM is forced to use metrics that do not have the predictive power because they lack insight into the value per unit cost being realized in the program. As a performance measure, ROI is useful in evaluating the efficiency of an investment or to compare the efficiencies of several different investments (”Return on Investment,” n.d.).

When these ratio estimates are properly constituted, the PM can make more accurate predictions of the future value of product/service acquisitions, leading to more informed investment trades between cost and the value of operational capabilities. These summary performance ratios are useful in making defensible investment decisions because they are broadly accepted and can be used to feed a more sophisticated analysis for investment/acquisition decision making, such as portfolio optimization and real options analysis (Mun, Housel, & Wessman, 2010). Additionally, predicting the value performance of DoD technology acquisitions is necessary in optimizing acquisition investment portfolios before further investments in the more codified, restrictive acquisition stages.

The purpose of this study is to extend the econophysics model to the DoD acquisition program life cycle in order to create a practical quantitative value metric that can be used to better understand and predict the s-ROI estimates of an acquisition program prior to contract award and to provide an early indicator of program performance during program execution. This is important because predicting the quantitative value of future DoD technology programs, prior to contract award, will allow for a more productive use of DoD investments. Additionally, gaining early insight into program performance will mitigate cost and schedule variance throughout the program development phase of the acquisition life

---

[1] Any form of cost will not provide a unique value metric. Measures of cost savings, while useful in evaluating investments in technology, do not provide unique value metrics. For example, if the numerator of a return on investment or cost/benefits ratio is cost savings, then the astute investor would fire everyone and sell all the tangible assets, producing an infinite return with cost savings in the numerator and zero in the denominator. Value estimates must be made independent of cost estimates to ensure a legitimate performance ratio.

cycle. This insight will allow DoD decision-makers to more clearly understand the risk and reward of future systems.

This study applies the econophysics model (Housel, Baer, & Mun, 2015; Baer & Housel, 2016; Baer, Bounfour, & Housel, in press) to generate estimates of the financial value of a given DoD technology investment. We use a basic example to explain the relationships between theoretical physics principles and economic measures frameworks. Our example presents the basic concepts using an applications (app) program for the DoD and relates the variables to a more traditional program acquisitions strategy.

The literature is replete with cost studies describing the cost analysis and program measurement milestones process. Much of this research is retrospective in nature and attempts to use historical data to predict future performance using models that focus exclusively on cost and schedule. Over the past 13 years of the Acquisition Research Program annual symposium, there have been numerous studies of how to estimate acquisition program costs, from activity-based costing to earned value measurement (EVM). In spite of these substantial research efforts, no widely accepted method for estimating costs has won out over all the others. None of these methods has proven to be exceptionally insightful with regard to predicting actual program life-cycle costs or performance with any degree of certainty. The lack of viable cost data prior to program start and no quantitative predictable value data by which to compare with program cost estimates has left decision-makers even more challenged in making reasonable forecasts based on economic program performance. By applying a surrogate measure of revenue to the same cost centers measured by EVM, a value metric can be used to assess overall program performance. This paper will address this gap in literature with regard to value estimation within a system's developmental program life cycle. Ultimately, this approach can be used across many industries and program contexts.

EVM is the program performance model most often used for major defense programs. The name of the model suggests that actual value is being measured, but from an economic perspective, this view would be incorrect. In effect, EVM is a cost model based upon prior cost and schedule predictions. Ultimately, this approach does not make predictions or assess whether the investment in a program, during the development life cycle, yields reasonable returns that are worth the investments in systems. Essentially, the DoD has no idea how much quantitative value it is getting from the investment of a dollar into a program of record under development.

The premise of the current research is that not having an accepted quantitative revenue estimate precludes program managers and program milestone decision authorities (MDAs) from making decisions that are based on a program's projections of overall value within threshold and objective cost boundaries. In order to accurately assess value and the resulting s-ROI, a quantitative surrogate revenue estimate needs to be allocated across a program in addition to the allocation of cost for the program. During the development of the performance measurement baseline, financial value needs to be allocated at the same level of detail as program cost allocation estimates. This would allow PMs to more effectively manage risk and make program decisions within the value and cost trade space.

Previous research on value-based management (VBM) suggested that having an unambiguous quantitative value metric would allow decision-makers to measure the performance of their company from a value maximization perspective, which is the ultimate economic objective for an organization. Since traditional financial performance measures, such as earnings or earnings growth, are not always good proxies for value creation, VBM focused more on the value creation process. Organizations tend to set goals in terms of discounted cash flow (DCF) value, the most direct measure of value creation. VBM takes this a step further by requiring targets to be translated into shorter-term, more objective financial performance targets (Koller, 1994). While this approach begins to address the issue of assessing program performance relative to value creation, it does not go far enough in identifying a commonly unitized measure for value. It simply requires that **qualitative** metrics be established by which an organization can measure “goodness” of performance. These value metrics are not normalized with a common unit of value measure that can be quantitatively compared to cost and subsequent ROI estimates.

Additionally, the lack of a common quantitative surrogate revenue parameter (i.e., quantitative value parameter) that is not directly derived from the cost estimate means that costs cannot be compared across a portfolio of project investments. In turn, the ROI of a portfolio of projects cannot be determined since there is no unitized value metric by which cost can be compared. ROI is a ratio of revenue to cost as expressed in Equation 1.

$$\mathrm{ROI} = [(\mathrm{Revenue} - \mathrm{Cost})/\mathrm{Cost}] * 100 \quad (1)$$

Absent a definitive measure for revenue, a portfolio is simply a conglomeration of costs that provide little insight into whether the portfolio is actually worth the overall investment relative to the portfolio forecasts. From a DoD acquisition perspective, this means that investments in enterprise program organizations are measured against cost and the relative qualitative estimates of the utility these programs provide for the customer. While some may argue that the economic value of a system lies in the operational utility of that system, without a common unit measure of surrogate revenue and therefore ROI, the customer might be overpaying for the expected utility and subsequently impacting the overall operational environment in which the system will operate. By having a higher ROI per system, the DoD will be in a better position to allocate scarce resources across a much larger portfolio of warfighting capability.

The search for a practical value metric has been going on for some time in the field of economics. Interfield theory provides an interesting opportunity for investigating the viability of other scientific theories and principles that might be applied to the field of economics. In the history of economics and physics, economists borrowed the energy concept from physics to develop value theories (Beinhocker, 2006; Mirowski, 1989). The econophysics model used in this study will take advantage of this mapping of energy theory from physics to develop a quantitative value estimate for the pre-contract award of DoD acquisition programs. This interfield approach to developing a methodology for quantitatively measuring value is consistent with many fields that use analogic extensions of physics models. This analogic reasoning is useful in developing more analytical and testable theory propositions (Kuhn, 1970). The mapping of physics-based terms to economic concepts, and subsequently to defense acquisition programmatic concepts, requires a proof of concept modeling demonstration case to test the viability and practicality of the derived value metric. Such a metric must be defensible as well as useful to acquisition professionals when generating investment productivity ratios such as ROI, which is an elegant, intuitively appealing productivity ratio and is applicable across acquisition portfolios.

The value theory demonstrated in this research will bear directly on public procurement policy and management as well as contracting and program/project management. Additionally, the application of value theory within program management introduces information sciences concepts, in that we are dealing with the collection and analysis of critical information within management, physics, and social sciences paradigms. From a policy perspective, this theory will provide a new measure by which to assess the relative value of warfighting systems compared to other system investment options. By understanding the ROI of acquisition programs and comparing them on a portfolio basis, more informed economic trades can be made relative to their overall perceived operationally valued utility. Additionally, at the program level, contracting and program managers' decision-making will be aided by having a robust estimate of the economic value of a given acquisition/procurement over time to compare to the investment costs of the program. Given the extreme riskiness of investments in programs such as information technology, acquisition executives would benefit from a clear understanding of the investment to performance productivity ROI, risk-reward ratios that a system will have over time, and whether that investment return is acceptable.

Research Questions and Objectives

This research addressed the following research questions:

1. Can an econophysics value theory model be used to predict the value of a proof-of-concept pre-contract award technology acquisition in the DoD?
2. How might an econophysics value theory model be used in a DoD acquisition context to aid in investment decisions?

The objective of this study is to test the use of an econophysics value theory model to create a defensible value metric that can be used to predict the performance of future DoD acquisitions in order to optimize acquisition investment portfolios.

Methodology

In what follows, we will provide a rationale and method for identifying and measuring non-monetized quantitative surrogate financial value. We label this value "proto-value" or prototype value (PV) metric. Our econophysics framework identifies the production of proto-value using analogies to a comprehensive physics conceptual model. This model is operationalized using PV calculations for which the case examples provide estimates for the model parameters. By establishing proto-value as a surrogate for allocated revenue, we are able to definitize the required parameters for a surrogate ROI (s-ROI) term in an acquisition program. Plotted over the life of the program, s-ROI reflects the baseline of investment return expected for the program.

The s-ROI performance measurement baseline (PMB) is analogous to the EVM PMB in that it provides a measure of work accomplished over time. However, while the EVM PMB measures the cost of work over time, the s-ROI PMB measures the expected value of the investment relative to the level of effort over time informed by a risk metric. For each increment in time, the s-ROI PMB will provide the decision-maker a unit of value relative to investment cost and risk, providing a more informed measure by which the program can be evaluated for relative worth and practicality. With a surrogate value for revenue, the s-ROI PMB can be operationalized at the work breakdown structure cost center level. Similar to the EVM PMB, the s-ROI PMB will provide indices of performance such as cost performance index (CPI), s-ROI performance index (RPI), and schedule performance index (SPI). Current EVM indices only provide CPI and SPI and provide no analytical index for the quantitative value of the program, whereas RPI provides an additional metric based upon value rather than just cost and risk. CPI and SPI are calculated using Equations 2 and 3, which are based on standard EVM calculation methodologies.

$$\mathrm{CPI} = \mathrm{BCWP}/\mathrm{ACWP} \tag{2}$$

$$\mathrm{SPI} = \mathrm{BCWP}/\mathrm{BCWS} \tag{3}$$

Where

BCWP – Budgeted cost of work performed

ACWP – Actual cost of work performed

BCWS – Budgeted cost of work scheduled

Since RPI is a function of surrogate revenue expressed in terms of proto-value, it may be expressed in the following terms:

$$\mathrm{RPI} = [(\mathrm{PV})(\mathrm{BCWS}) - \mathrm{ACWP}]/\mathrm{ACWP} \qquad (4)$$

where

PV – Proto-value is a non-dimensional value representing allocated surrogate revenue allocated throughout the program work breakdown structure.

The existing econophysics model uses terms from physics to define relationships between individuals and processes in an economic supply and demand framework. Terms such as mass and distance are used to explain product performance and quality as well as the level of consumer attraction toward the product in the context of distance. The consumer attraction toward a product is defined in DoD requirements documents such as the Capability Development Document (CDD), which specifies the systems requirements and critical attributes. These attributes represent the level of demand or attraction the DoD user has toward the specific requirement. Critical attributes with the highest demand are delineated as Key Performance Parameters (KPP) which specify both threshold and objective values that must be met by the program manager of the system under development. If a system under development is close to the objective for the KPP, then the distance between the operational user (consumer) is very small. However, if the system under development is closer to the threshold, then the distance between the user and the product is larger. If the system is below the threshold, then the distance between the user or customer and the product under development approaches infinity.

A fitness matrix can be subsequently generated to map customer need vectors to program value vectors within the context of the relative distance (e.g., cost, ease of use, riskiness) between the two. Additionally, a series of non-linear matrices with associated first order derivatives can be developed that reflect the changing nature of the variables that affect the need and value vectors between the customer and the product. For a DoD program, these derivatives are representative of the vast number of variables that might affect the relationship between the requirement and the intended capability to be provided by a contractor. For DoD acquisition programs, these vectors and derivatives are extracted from requirements documents such as the CDD, program acquisition strategy (AS), technical proposals, proposal evaluations, and cost documentation. Additionally, intervening processes that might affect the AS could be considered in the establishment of derivatives that might impact the relative attraction between a user need and the prospective capability that satisfies that need. During the pre-contract award phase, the relative value of multiple offers from various vendors in industry can be used to compare the value of satisfying the specified requirements in the government request for proposal (RFP), thereby quantitatively establishing priorities in order to forecast the financial value impact of cost, schedule, and requirements changes during contract management of the program life cycle.

In the context of a non-monetized quantitative value theory, there was a need to create new categories for common units of value. One promising common unit candidate for proto-value is a unit of complexity (Housel & Kanevsky, 1995; Housel & Bell, 2001). Complexity theory has been touted as foundational for a new theory of economics (Beinhocker, 2006) even though this prior work did not posit a unit of complexity as central to this argument. Our analysis offers a physics-based framework where we rely on the concepts of mass, potential field, force, momentum, velocity, total energy, and work extracted from total energy. In the example that follows, we have aggregated a number of the physics concepts into a simplified form to show how it is possible to use the resulting framework for a rough-cut analysis of the velocity of adoption rate of the information technology (IT). Table 1 shows the relationship between physics variables and organizational variables.

Table 1. Concept Definitions

mass {**m**} = relative richness of services, measured in common units of complexity. Richness refers to the desirability and perceived availability of a product, capability, or service.
Position of m = name of node in a network of the entity that is offering the service established by the force of the pull of the business and the pull of the customer for the service
Force = the pull of the mass of the company {**Mb**} – the pull of the client mass {**Mc**} (**Force = m*Mc/**r^2 **[distance squared] * K (constant) or client desire/**r^2**)**
Business Mass {**Mb**} has a given strength of pull on the service {m}
Client Mass {**Mc**} pull on the service offered by the business
Number of services {**N**} = total number of {**m**} services at a given point in time (e.g., email, search)
Total Potential Field (TPF) ≈ (**m*Mc*N/distance between the Mc and Mb**) the number of services of a given complexity (m) offered by an organization to a field of potential clients (TPF = total proto-value field)
Velocity (V) = Change in rate of position of **m**
Momentum {Mo} = rate at which service {m} moves from Mb to Mc
TPF*Mo = total Energy (E) ≈ proto-value
Work = total actual value extracted from proto-value

Conceptual Example

In order to better understand the aforementioned concepts, we present a simplified example with a subsequent alignment to the broader DoD acquisition environment. The simplified model uses an example of a pre-award IT contract for a defense intelligence community service program. A quantitative proto-value estimate was derived for this example program by applying the concepts defined in the econophysics model. Throughout this example, we will further definitize the terms in order to explain their association with other DoD developmental programs. Additionally, we will show how this approach is a viable strategy for developing a predictive model to assess the s-ROI PMB and subsequent value targets that are informed by not only cost, but also surrogate revenue and risk.

In this example, the acquisition leadership wished to take advantage of the potential social media apps (i.e., defined as Facebook + Twitter + Snapchat). Table 2 is a summary of the key econophysics terms that we use to demonstrate how our interfield theory approach to economics and program management can be used to better understand program performance through other disciplines such as physics.

Table 2. Framework for Simple Model to Estimate the Proto-Value of a Pre-Award Contract

mass {m} = The term mass is typically used when describing a property of a physical body. It also refers to the strength of the gravitational forces between bodies. We similarly use mass to reflect the relative complexity between services, which is a direct analogy to the strength of ties between the various services or products and their intended user.
Number of apps (**N**) = total number of apps (or products) at a given point in time (i.e., email, texting, image sharing). This is essentially, the number of capabilities being provided by the developer at a specific point in the schedule of the program.
Potential Field (PF) = (**m*N**) the number of apps of a given complexity (m) offered by the organization to a field of customers/clients/users
Velocity = Rate of PF over time period of three years

Mass

In this notional example of the simplified econophysics model, operationally defining mass was done using an interval complexity scale. There are several options for operationalizing mass as delineated in Table 1. The definition of mass depends on the context of the model. Several possibilities, when considering options for defining mass for the simplified model, include

- a 1–10 complexity interval scale
- a more detailed ratio level scale (e.g., lines of code [bits], embedded algorithms)
- a knowledge-based estimate (e.g., amount of knowledge embedded in the IT [learning time] created from intellectual-social capital

Additionally, mass within the context of a more traditional DoD program can be operationalized through interval scales of complexity such as Technology Readiness Levels (TRLs) or other similar complexity scales that rate the level of complexity or richness of specific requirements within the CDD. These scales provide the relevant level of readiness of the desired requirement being asked for by the user and translates into mass within the tenets of the econophysics theory. A useful tool for assessing technology complexity is the TRL assessment rating scale described in Figure 1.

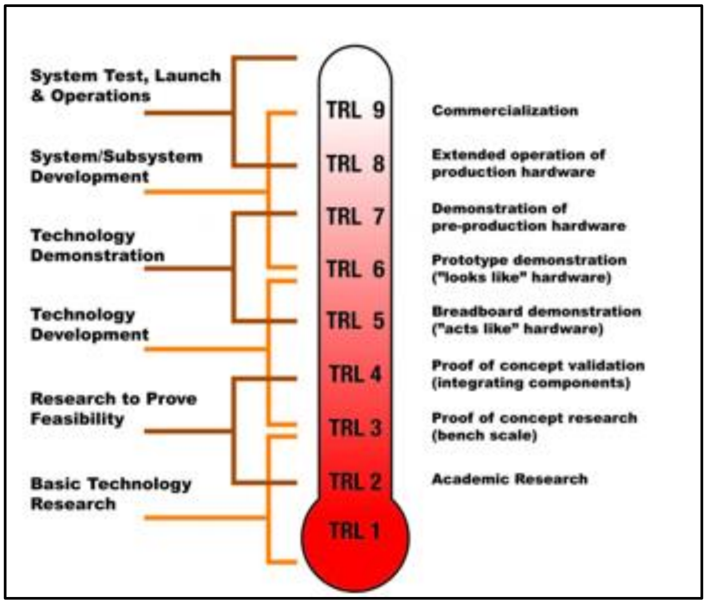


Figure 1. Technology Readiness Level Descriptions
(DoD Technology Readiness Assessment Guide)

The TRL level describes a standard by which the technology should be measured in terms of its readiness to be accepted by the user. The higher the TRL, the more ready the technology is for operational use and the more mass the requirement has from the users' perspective.

Potential Field

Potential field (PF), in the current example, is represented by the number of potential modified social media apps that would be acquired and offered to a given field of user groups. For example, in the case of Facebook, it would be represented as all the modified apps that would be produced and offered to its user groups. Potential Field would be quantified as the total number of modified apps (that had a given mass measurement) offered to the potential user groups at a given point in time.

Within the broader DoD perspective, the PF represents the total number of capabilities (N) the contractor offers in response to a government RFP times the relative mass of these capabilities. The RFP specifies the requirements being asked for by the user and the relative performance, or mass, required to meet these requirements. If the contractor offers all of the capabilities being asked for by the government, the PF would be 100%. times the mass of the capability, as defined previously.

Work

Estimating the amount of work that can be extracted from the total potential proto-value (total potential energy) in the simplified model can be represented as the actual usage of the modified apps by the user groups. This part of the simplified model becomes useful once the apps are offered to the user groups. It then becomes possible to determine the yield rate from potential to actual usage (i.e., amount of realized proto-value, kinetic energy).

Simplified Framework for Estimating Proto-Value and Work

Continuing with our simplified example, DoD acquisition leadership would like a quick, rough-cut estimate of the yield of proto-value from actual usage (i.e., amount of realized proto-value or work) from the apps over a three-year actual adoption rate time period. This provides a means to compare the expected actual adoption rate to the potential adoption rate (calibrated in terms of proto-value) to determine the accuracy of the forecasts for the program. Estimates of potential proto-value over the three-year period provide an estimate of the modified apps adoption rate calibrated in terms of potential value to the user groups. The realized proto-value of the apps provides a measure of the actual value to the user groups calibrated in terms of their usage of the apps over the three-year period. The simple model estimates are summarized in Table 3: Customer Usage of Modified Social Media App Offerings. This kind of adoption rate information provides a means of measuring the value yield of these apps that allows an assessment of the accuracy of the adoption rate forecasts.

Table 3. Customer Usage of Modified Social Media App Offerings

Number of potential uses of apps proto-value (PP) (based on potential uses of the modified apps within the three-year projected adoption rate time-frame)
Number of realized proto-value of the apps (RP) (e.g., based on number of actual downloads, clicks, page views within the time frame of the availability of the modified apps)
PP * PF = Proto-value = total potential energy
RP * PF = Realized Proto-value ≈ Work

The **RP** metric includes how many times users actually used the apps that have a given mass. The equation **RP * PF** can be used to derive the measure of the yield extracted from **PP**. The difference between **PP** and **RP** also provides a measure of the unused capacity of the modified apps represented as the opportunities foregone to provide value to the user groups. Using the example of the modified social media apps adoption rate, we can generate a table of values that will allow a yield estimate based on results per Table 4.

Table 4. Modified Social Media Apps Example

Year	Potential User Groups	Actual Adoption of Apps by User Groups	PF (from Table 2)	PP	RP
2010	100	40	168	16800	6720
2011	120	60	284	34080	17040
2012	125	90	248	31000	22320
Totals	**345**	**190**	**700**	**81880**	**46080**
				Yield =	56%

Comparing total PP with RP provides a simple yield ratio of 46,080/81,880 or 56% yield for the three-year period. This value yield could be compared with industry averages for this kind of modified social media app as well as for other IT acquisitions cases. These yields comparisons might be very useful for acquisition leaders, as well as user group leaders, in tracking the conversion PP to RP performance.

In this example, the acquisition leadership wanted to estimate how rapidly the potential social media apps, modified for use by service member organizations, would be adopted. The estimate included the number of new social media services that are rolled out to the potential user groups. Included in the estimate is the relative mass (measured in terms of relative complexity on a 1–10 scale) of the app modifications. The velocity (i.e., change in forecasted adoption rate) of the modified apps is presumed to be a reasonable surrogate for predicting the future adoption rate of these apps by potential user organizations. The total potential proto-value is estimated in terms of the total potential energy field times the number of potential user organizations that might adopt the modified social media apps. For a more traditional developmental program that provides either a product or service, the potential proto-value would equate to the number of requirements in terms of products or services expected to be used by the user times the total number of capabilities being delivered by the respective contractor.

Table 5 is a summary of the potential adoption rate example and reflects the kind of data this simplified model would generate. It is based on the expectations of the planned acquisitions of these modified apps over a three-year period. In this example, the expectation is that the number of modified apps for two of the social media apps (i.e., T and S) will diminish in Year 3. This reduction in introduction of new modifications directly affects the potential adoption of these apps even though the number of user groups is expected to grow. After rising from Year 1 to Year 2, this drop in new modifications is reflected in Figure 2, which indicates that the adoption rate velocity of the modified social media apps should be falling precipitously from Year 2 to Year 3.

Table 5. Potential Adoption Rate Example

Year	Facebook (FB) **8*N**	Twitter (T) **6*N**	Snapchat (S) **4* N**	PF **FB+T+S**	V	Potential Users Organizations (PU)	Total Potential Energy Field **PF*PU**
2015	56	12	100	168	90	100	16800
2016	80	24	180	284	116	120	34080
2017	80	18	150	248	-36	125	31000
Totals	216	54	430	700		345	81880

Mass per service weightings
FB = 8 T = 6 S = 4

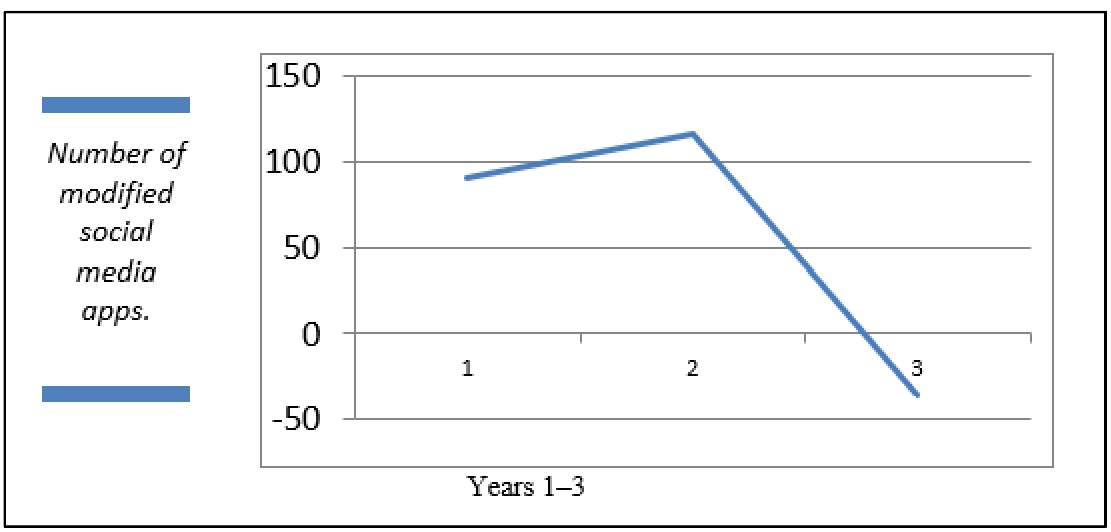


Figure 2. Potential Adoption Rate Velocity

This drop in potential proto-value, even with the increase in potential user groups, should provide a rationale to advocate for continued increases in development of modified social media apps or set expectations that there will be a potential reduction in the proto-value of these apps due to the reduction in investments in the modifications of the apps. One implication from the use of this simplified model for the adoption rate velocity estimate is that there is a correlation between the velocity of potential adoption of new social media apps and the proto-value of these apps. Increasing the velocity of introduction of modified apps would represent an increasingly larger potential field for customers, while decreasing velocity would represent an overall reduction of potential proto-value due to the decreasing number of modified apps being offered to the potential user groups. One can see that increasing the number of apps is only one way that the potential proto-value can be increased. It would also be possible to increase potential proto-value in a given year by offering the modified apps to a larger number of potential user groups. The goal of this example is to demonstrate a simplified way to forecast the potential proto-value of information technology investments.

Defense Acquisition Framework

While the preceding example begins to explain the relationships between econophysics and proto-value with regard to services-based applications, a more rigorous explanation of how these principles relate to more established developmental program business processes is necessary. We will introduce the concept of risk and probability of success to the model and show how significant these concepts are in predicting program performance. By introducing proto-value and risk, we will show how program performance prediction is significantly more reliable than traditional methods using forms of cost as the sole metrics.

Table 6 relates the econophysics terms defined for the simplified program example with a more generic defense acquisition program. Risk is introduced with regard to the probability of success (Ps) of meeting specified requirements defined by the operational user and articulated in the Capabilities Development Document (CDD).

Table 6. Framework for Simple Model to Estimate the Proto-Value of a Contract Pre-Award Modified for Standard DoD Acquisition Program

mass {**m**} = relative complexity of key performance parameters (KPP) as defined in the Capabilities Development Document (CDD) multiplied by its Technology Readiness Level (TRL) (Scale of 1–9, as defined in Figure 1)
Number of capabilities (**N**) = number of completed capability solutions in the contract relative to the proposed contractor schedule that supports meeting the KPPs
Potential Field (**PF**) = (**m*N**) the number of capabilities of a given complexity (m) offered by the contractor in the contract proposal
Velocity = Change in rate of **PF** over time period of contract performance
Probability of Success (P_s) = (1-%risk) of completing a specified requirement defined in the CDD where %risk is shown as; **r.** Therefore **P_s = (1-r)**
(PP) - Number of potential requirements **(R)** multiplied by the probability of Success – **[R*(P_s)]**
Realized Proto-value (RP) – number of requirements actually accomplished in the contract relative to the CDD and proposed schedule

Understanding risk is necessary for determining the probability of success for a particular program and, subsequently, the proto-value. Risk is the principle indicator as to whether a program will succeed. Program managers and decision-makers must make informed decisions prior to contract award based upon TRL and the overall risk of accomplishing the various requirements for the program. While risk is considered in current source selection processes, it is not integrated into a probability of success calculation that reflects the potential program's return on investment. Risk is typically managed as a separate entity concurrently with cost and schedule. While risk is derived from the same data by which cost information is collected, the integration of risk into the program performance calculations is not well developed. Consequently, risk is simply characterized as a qualitative function based upon subjective methods in determining the potential cost and schedule impacts a given contractor might experience throughout the program life cycle. The risk matrix shown in Figure 3 is a standard model that is explained in the DoD Risk Management Guide and is typically used in most programs within the DoD and industry.

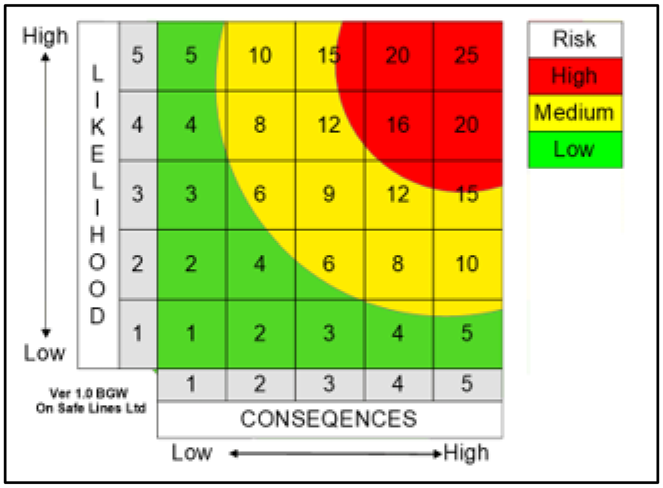


Figure 3. Standard Risk Matrix
(DoD Risk Management Guide)

This process determines the likelihood and consequence of realizing a risk and is reported to the program manager on a regular basis. Done correctly, potential risks are identified through the requirements analysis process, during which the requirements are decomposed into subordinate tasks. This process allows the program manager to allocate a cost and schedule risk to the individual requirements and subsequently to the overall program. The problem with this method, however, lies in the absence of translating risk into potential success and s-ROI. Intuitively, program managers feel that if they sufficiently mitigate the risk at the predetermined time identified in the risk management process, then this will result in a lower likelihood of cost and schedule creep. This says nothing about potential for actually succeeding and maximizing the surrogate financial return on investment relative to the operational utility of the system being developed. The goal of this research is to tie the potential for program success to operational utility by showing how s-ROI is a better measure of program performance than traditional cost methods. For the purpose of this research, we are using a surrogate measure for ROI derived from proto-value.

Using risk as a basis for understanding the potential for success, we have redefined the traditional risk matrix in terms of the probability of either meeting or not a meeting the specified requirements defined in the CDD. While these percentages are debatable, they simply reflect the logic of the argument. Table 7 reflects the likelihood and consequence of not realizing the completion of a particular defined requirement listed in the CDD, which is important in determining the overall value of the program.

Table 7. Percent Risk of *Not* Completing an Individual Requirement Defined in the CDD and the Relative Consequence of Not Completing the Requirement

High	60	70	80	90	100
High	50	60	70	80	90
Medium	40	50	60	70	80
Medium	30	40	50	60	70
Low	10	20	40	50	60
Low	Low	Medium	Medium	High	High

Return on Investment Performance Index (RPI) Comparison With Earned Value Cost and Schedule Indices (CPI/SPI)

Major defense programs and large commercial programs typically use EVM metrics to measure their performance. These data are generally historical in nature and require the program manager to extrapolate future performance based on program risk and other mitigating factors. While this is a good measure of tracking pre-contract award cost to work relationships, it does not provide an early assessment of program value relative to the potential for program success. Consequently, programs tend to get into trouble earlier than program managers are able to observe through traditional measures, and program managers are unable to ascertain the relative program performance based upon investments. If there were a way to inform the program manager on how a program was performing relative to the investment, decision-makers would be able to make decisions as to the program net value rather than simply falling victim to making cost and performance trades based upon increasing cost and schedule.

Using the principles of econophysics and basic EVM methods described previously, we are able to show that s-ROI is a better predictor of program performance than traditional EVM metrics alone and is referred to as s-ROI Performance Indicator (RPI) in subsequent discussions. By way of summary and explanation, the following equations show how each of the variables in Table 8 were derived for a notional developmental program with a 36-month expected period of performance.

BCWS – Performance Measurement Baseline and Cumulative Program Cost over the period of performance

BCWP – Budgeted Cost of Work Performed is the cost per unit of work budgeted at the start of the program

ACWP – Actual Cost of Work Performed is the actual cost charged by the contractor

R – Specified requirements that are identified in the CDD

N – Number of capabilities completed by the contractor over time

P_s – Probability of Success – (1-%risk) = (1-r); r = f(cost, schedule, TRL)

PF – Potential Field – (m*N)

PV – Proto-value (surrogate term for revenue). This term is non-dimensional for the purpose of our calculation of RPI.

PP – the number of potential user specified requirements multiplied by P_s

With this as a summary, the relevant equations follow:

$$CPI = BCWP/ACWP \quad (5)$$

$$SPI = BCWP/BCWS \quad (6)$$

$$RPI = [(PV)(BCWS)] - ACWP/ACWP \quad (7)$$

Where

$$PV = PP * PF \quad (8)$$

$$PV = (R * P_s)(m * N) = ([R * (1 - r)])(m * N) \quad (9)$$

Table 8 shows the contractor is expected to perform $10 worth of work every month for 36 months with the overall PMB reflected in the BCWS column. This baseline is developed using typical EVM methods, the process of which is defined in standard EVM textbooks.

The data in Table 8 reflects a program with some amount of anticipated risk with regard to developmental maturity. The risk is informed by the TRL level of the program and is considered in the calculation of the monthly and overall potential field (PF) (that also includes mass per requirement number) for the program. Generally, the program reflects a user requirement for 10 “needs” at a cost of $10/month for 36 months. The data in Table 8 reflects a delta between the Budget at Complete and the Actual at Complete to be $43, representing an overall cost variance of 11%. By Month 21, the program seems to be costing more than expected, and by Month 23, the program seems to be producing less output (i.e., value) per unit cost than expected as shown by the increase to an ACWP of $11 from an expected ACWP of $10 and decrease from $10 BCWP to $9 BCWP, indicating that there is less output than expected for that point in the schedule.

Typically, a program begins to suffer technical problems before these would be reflected in EVM cost reports. EVM does not provide an early warning signal of technical issues because of the lagging nature of EVM data. Using the econophysics model, this early indication of a technical problem is seen in the decrease in PF from 10 to 8 and a monthly decrease in PV from 90 to 72 at Month 19. This is realistic in that technical issues generally reveal themselves earlier in the process than they are reflected in the lagging indicators of EVM data. Using the equations defined previously for PV and RPI, a plot of PV relative to EVM data is shown in Figures 4 and 5. The cumulative PV shows a rate change as early as six months prior to the first significant indicator of a problem using EVM data. The first sign of trouble in EVM is the CPI at Month 24 and the second is SPI at Month 28, whereas RPI begins to inform the situation as early as Month 19.

Table 8. Notional Program EVM and Proto-Value Data

Month	Cost Est/ Mo	BCWS	BCWP/mo	BCWP	ACWP/mo	ACWP	R	Ps	PF	PV per Month	Cum PV	RPI	CPI	SPI
1	10	10	10	10	10	10	10	0.9	10	90	90	8.9	1	1
2	10	20	10	20	10	20	10	0.9	10	90	180	8.9	1	1
3	10	30	10	30	10	30	10	0.9	10	90	270	8.9	1	1
4	10	40	10	40	10	40	10	0.9	10	90	360	8.9	1	1
5	10	50	10	50	10	50	10	0.9	10	90	450	8.9	1	1
6	10	60	10	60	10	60	10	0.9	10	90	540	8.9	1	1
7	10	70	10	70	10	70	10	0.9	10	90	630	8.9	1	1
8	10	80	10	80	10	80	10	0.9	10	90	720	8.9	1	1
9	10	90	10	90	10	90	10	0.9	10	90	810	8.9	1	1
10	10	100	10	100	10	100	10	0.9	10	90	900	8.9	1	1
11	10	110	10	110	10	110	10	0.9	10	90	990	8.9	1	1
12	10	120	10	120	10	120	10	0.9	10	90	1080	8.9	1	1
13	10	130	10	130	10	130	10	0.9	10	90	1170	8.9	1	1
14	10	140	10	140	10	140	10	0.9	10	90	1260	8.9	1	1
15	10	150	10	150	10	150	10	0.9	10	90	1350	8.9	1	1
16	10	160	10	160	10	160	10	0.9	10	90	1440	8.9	1	1
17	10	170	10	170	10	170	10	0.9	10	90	1530	8.9	1	1
18	10	180	10	180	10	180	10	0.9	10	90	1620	8.9	1	1
19	10	190	10	190	10	190	10	0.9	8	72	1692	7.1	1	1
20	10	200	10	200	10	200	10	0.8	8	64	1756	6.3	1	1
21	10	210	10	210	10	210	10	0.8	8	64	1820	6.3	1	1
22	10	220	10	220	11	221	10	0.8	8	64	1884	5.7181818	0.909	1
23	10	230	10	230	11	232	10	0.8	8	64	1948	5.7181818	0.909	1
24	10	240	9	239	11	243	10	0.7	8	56	2004	4.9909091	0.818	0.9
25	10	250	9	248	11	254	10	0.7	7	49	2053	4.3549455	0.818	0.9
26	10	260	9	257	12	266	10	0.7	7	49	2102	3.9833333	0.75	0.9
27	10	270	8	265	12	278	10	0.7	7	49	2151	3.9833333	0.667	0.8
28	10	280	8	273	12	290	10	0.7	7	49	2200	3.9833333	0.667	0.8
29	10	290	8	281	12	302	10	0.7	7	49	2249	3.9833333	0.667	0.8
30	10	300	8	289	14	316	10	0.7	7	49	2298	3.4	0.571	0.8
31	10	310	7	296	14	330	10	0.7	7	49	2347	3.4	0.5	0.7
32	10	320	7	303	14	344	10	0.6	7	42	2389	2.9	0.5	0.7
33	10	330	7	310	14	358	10	0.6	7	42	2431	2.9	0.5	0.7
34	10	340	7	317	15	373	10	0.6	7	42	2473	2.7	0.467	0.7
35	10	350	7	324	15	388	10	0.6	7	42	2515	2.7	0.467	0.7
36	10	360	7	331	15	403	10	0.5	6	30	2545	1.9	0.467	0.7

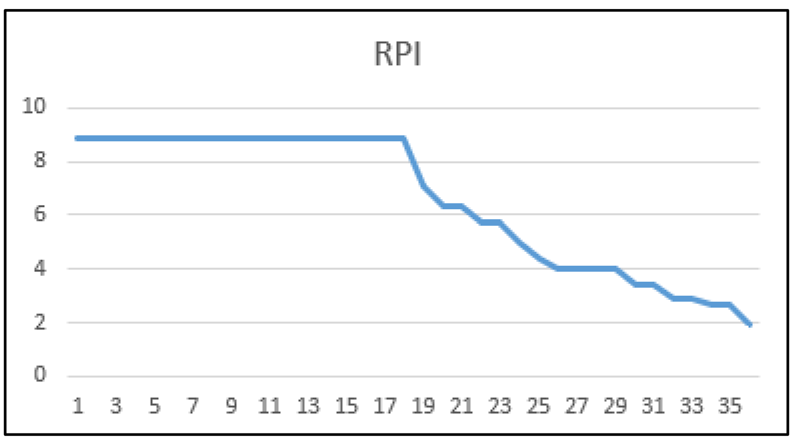


Figure 4. Program s-ROI Performance Index (RPI)

Figure 5 is another view of the same data using CPI and SPI as the performance indices. Comparing Figures 4 and 5, RMI begins to fall off much earlier than CPI and SPI. This is explained by the fact that risk and probability of success are incorporated into the PV calculation. Additionally, PF impacts the overall PV in that we are assuming in this basic example that mass does not increase significantly and N begins to drop by Month 19. This is fairly typical in programs in that contractor performance issues are first observed in technical performance, indicating a schedule impact. The value N is a function of schedule, leading us to conclude that N would be an early indicator of performance as the contractor begins to fall behind in completing tasks, followed quickly by cost (ACWP).

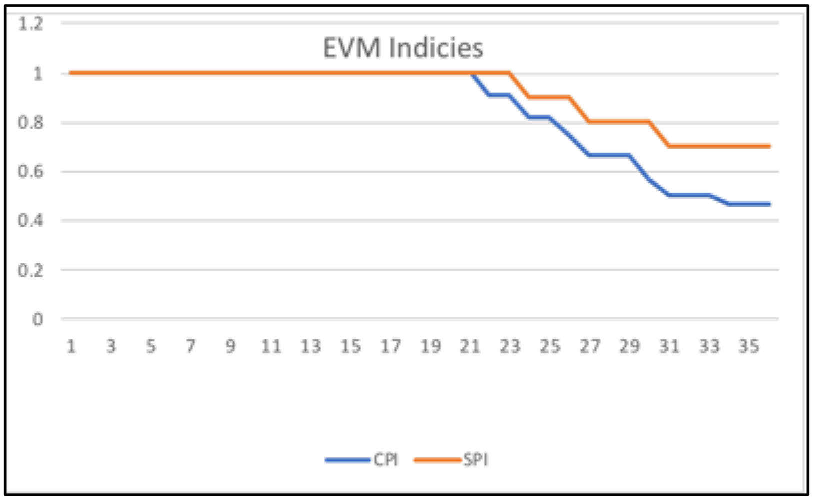


Figure 5. EVM CPI/SPI Indices

The data shows that establishing a measure for value based upon revenue will inform the decision-maker when a program ROI is decreasing. This decrease in ROI, as reflected in the RPI, can be an early indicator of program issues. Since the RPI is directly influenced by risk, the lag typically associated with EVM data is mitigated. Knowing that a program is attaining less value for its investment is a powerful measure by which leaders can make informed decisions regarding the viability of a program.

Potential Benefits

The results of this study provide a methodology for estimating a surrogate for financial value of a given technology at the pre-contract review stage of an acquisition program. Current methods used to predict program performance are based upon techniques such as EVM, which helps project managers to measure project performance. It is a systematic project management process used to find variances in projects based on the comparison of work performed and work planned. The EVM process establishes a Performance Measurement Baseline (PMB) which provides a baseline by which the contractor is measured. The PMB is a time–phased schedule of all the work to be performed, the budgeted cost for this work, and the organizational elements that produce the deliverables from this work. This baseline is agreed upon prior to contract award by the government and subsequently included in the statement of work for the contract.

While the PMB is an attempt to estimate cost over time, it provides no assessment of the financial value of the program and subsequent ROI. Furthermore, the cost estimates used to determine the PMB are typically based on incomplete information due to the program risk uncertainty. Development programs typically use cost reimbursable type contracts which attempt to account for unknowns due to technology immaturity and overall program risk.

Once the contract is awarded, actual performance is measured against the PMB. With near certainty, all DoD programs tend to breach the PMB, leading to either a rebaseline or termination. A better measure of program performance is ROI. By establishing an ROI baseline, the desired ROI is measured over time, allowing decision-makers to focus their decisions on how to optimize program performance by balancing risk and proto-value. Rather than chasing costs, which inevitably increase due to risk and other programmatic influencers, increasing costs become less critical if they are measured against value and subsequent ROI. If the ROI of a program remains within predetermined thresholds, the PM can allow cost to "float," within reason, and offset this with increased efficiency, resulting in higher ROI. Essentially, the program manager can set cost threshold and objective limits in order to establish budget constraints but will manage to the ROI baseline vice the cost baseline. This method would allow the program manger more flexibility in developing innovating strategies and managing risk that are based upon value rather than simply focusing on cost. Cost as an independent variable (CAIV) would be replaced with ROI and an independent variable (RAIV).

Acquisition leadership should find the simplified econophysics and more complex model useful in the pre-award acquisition phases in estimating whether an IT investment has promise based on its potential value (i.e., proto-value) compared with other options. Continuous estimates of the proto-value, after an acquisition, should prove useful in attempting to improve the fitness and reduce the distance of the acquired IT. For these reasons, the econophysics models should help improve acquisition investment portfolios. Use of these models should also provide the acquisition leadership a way to track the use of their investments to avoid costly mistakes.

Conclusions

These examples of how the econophysics approach can be used to model the potential value of new or mature products or services demonstrated that (when the data values can be verified) it is possible to predict the potential value of the acquisition of a new or mature product or service. The purpose of this study was to demonstrate that it is possible to use econophysics formalisms to model the potential proto-value of new products and services before their acquisition in a pre-award phase. These estimates can be routinely updated during the product/service adoption rate life cycle, as well as when modified or discontinued. The econophysics approach can be combined with existing investment tools and approaches to create more accurate potential value estimates before services or products are acquired.

References

Assistant Secretary of Defense for Research and Engineering (ASD[R&E]). (2011). *Department of Defense technology readiness assessment guidance*. Washington, DC: Author

Baer, W., & Housel, T. (2016). An econo-physics theory of value: The case of micro-finance in Africa. IC12: *World Conference on Intellectual Capital for Communities in the Knowledge Economy*. Paris, France.

Beinhocker, E. D. (2006). *The origin of wealth: The radical remaking of economics and what it means for business and society*. Boston, MA: Harvard Business School Press.

Housel, T., Baer, W., & Mun, J. (2015). A new theory of value: The new invisible hand of altruism. In P. Ordonez de Pablos & L. Edvinsson (Eds.), *Intellectual capital in organizations: Nonfinancial reports and accounts* (pp. 16–52). New York, NY: Routledge.

Housel, T. J., & Bell, A. (2001). *Measuring and managing knowledge*. Boston, MA: McGraw Hill.

Housel, T. J., & Kanevsky, V. (1995). Reengineering business processes: A complexity theory approach to value added. *INFOR, 33*, 248–262.

Koller, T. (1994). What is value based management. *McKinsey Quarterly*.

Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.

Mirowski, P. (1989). *More heat than light: Economics as social physics, physics as nature's economics.* Cambridge, England: Cambridge University Press.

Mun, J., Housel, T., & Wessman, M. D. (2010, April 30). *PEO-IWS ACB insertion portfolio optimization* (NPS-AM-10-069). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Office of the Deputy Assistant Secretary of Defense for Systems Engineering (ODASD[SE]). (2017). *Department of Defense risk, issue, and opportunity management guide for defense acquisition programs*. Washington, DC: Author.

Pietronero, L., & Cader, M. (2016). Presentation on economic complexity at a conference dedicated to OECD, Paris, France. Retrieved from http://www.lucianopietronero.it/presentations/

Return on investment—ROI. (n.d.). In *Investopedia*. Retrieved January 8, 2013, from https://www.investopedia.com/terms/r/returnoninvestment.asp

Panel 16. Bid Protests: Reality and Perceptions

Wednesday, May 9, 2018	
3:30 p.m. – 5:00 p.m.	**Chair: Douglas Brook,** Visiting Professor, Duke University ***Cost-Benefit Analysis of Bid Protests: A Representative Bidder Model*** Francois Melese, Naval Postgraduate School ***An Analysis of DoD's Use of the Lowest Price Technically Acceptable Acquisition Strategy and Recommendations for Improvement*** Patricia Donahoe, Duke University ***A Closer Look at Bid Protests in the Department of Defense*** Jack Chutchian, Duke University

Douglas Brook—is Visiting Professor in the Sanford School of Public Policy at Duke University. He is also Emeritus Professor at the Naval Postgraduate School in Monterey, CA. At NPS he was Dean of the Graduate School of Business & Public Policy, Professor of Public Policy, and Director of the Center for Defense Management Research.

Brook has served in four presidentially-appointed positions. While on leave from NPS, he served a 14-month tour in the Pentagon serving first as Assistant Secretary of the Navy (Financial Management & Comptroller) and later as Acting Under Secretary of Defense (Comptroller)/Chief Financial Officer. From 1990 to 1992, he served as Assistant Secretary of the Army (Financial Management). In 1992 he was Acting Director of the U. S. Office of Personnel Management (OPM).

Brook holds a BA degree in political science and a Master of Public Administration degree from the University of Michigan. In 2001 he earned his PhD in Public Policy at George Mason University.

Cost-Benefit Analysis of Bid Protests: A Representative Bidder Model

"If those affected by the breach of rules cannot protest … rules have no teeth, and competition is stifled. Without the constraints of bid protests, government contracts will be let based on favoritism … and bribery—as they were before the system was initiated."

(Weckstein & Love, 1995)

Francois Melese—is a Professor at the Defense Resources Management Institute (DRMI) at NPS. He received his BA in Economics from the University of California at Berkeley in 1977, MA in Economics from the University of British Columbia, Canada in 1979, and PhD from the University of Louvain, Belgium in 1982. He was previously a Research Fellow at the Institut de Recherches Economiques et Sociale (IRES), University of Louvain, Belgium, and Assistant Professor of Economics at Auburn University. He has papers published in the *Quarterly Journal of Economics*, the *Southern Economic Journal, Energy Economics*, and the *International Trade Journal and Defense Analysis*. He has presented papers at meetings of the American Economic Association, the European Economic Association, the Southern Economic Association, and the World Econometric Society, as well as meetings of the International Association of Science and Technology for Development. He is a member of the American Economic Association, Southern Economic Association, Operations Research Society, and Research Society of American Scientists–Sigma XI. Professor Melese joined the faculty of DRMI in June 1987. His research interests include budgeting with incentives, pricing issues in revolving funds, the role of benefits and pay in compensation, defense industrial base issues, and the integration of cost and effectiveness.

Abstract

Most countries allow disappointed bidders[1] to protest public procurements. The dual goal is to reduce *favoritism*, reduce *fraud and errors*, and increase *competition*. The legal literature that underpins protest systems for the U.S. Federal Government and European Union generally reflects these two goals. The hypothesis is that allowing disappointed bidders to protest public procurements serves as a decentralized oversight mechanism that increases transparency and accountability, which encourages vendor participation. This study offers a cautionary tale for any government agency, country, or international institution that relies on, and/or promotes, bid protests to improve public procurement outcomes. The goal is to explore costs and benefits of bid protests for governments and taxpayers. As a first step, a probabilistic, micro-economic, partial equilibrium, representative bidder model is developed to help evaluate protest systems. The model reveals multiple unintended

[1] Other terms besides *disappointed bidder* found in the literature include disappointed offeror, unsuccessful offeror, excluded offeror, and interested party. For the purposes of this study, these terms are used interchangeably and refer to a company that has standing, or is allowed to protest the solicitation or award of a contract. "Interested party … means an actual or prospective offeror whose direct economic interest would be affected by the award of a government contract" (FAR 33.101, Definitions).

consequences of protest systems and suggests alternative approaches to improve public procurement outcomes.[2]

Introduction

Two key claims appear in the legal literature in support of bid protests. First, protests play an important role as a decentralized oversight mechanism to ensure "fairness" of the procurement process. The claim is that allowing vendors to protest public procurements reduces the risk of "crony capitalism," and helps deter favoritism, fraud, and errors. Military procurement offers an important illustration.[3]

The second claim is that allowing losing bidders to protest makes vendors more willing to compete,[4] that is, delivering benefits of competitive markets to improve performance, costs, and schedules (Arrowsmith et al., 2000). Experience from major defense acquisitions tends to undermine both claims: that protests deter favoritism, and that they increase competition.

According to the U.S. Federal Acquisition Regulations (FAR), a protest is defined as

> a written objection by an interested party to any of the following: (1) A solicitation … by an agency … for a contract for the procurement of property or services, (2) The cancellation of the solicitation … (3) ***An award … of the contract*** [emphasis added], or (4) A termination or cancellation of an award of the contract. (FAR, 2005, 33.101; see also U.S.C. 31 § 3551[1])

The relative frequency of protests over these four categories reveals the majority involve (3) "An award … of the contract," which is the primary focus of this paper. According to the Congressional Research Service, since 2008 the annual rate of protests of government procurements has increased by nearly 50% (CRS). In 2014 alone, the Government Accountability Office (GAO) received over 2,500 protests.[5]

---

[2] Whereas the most common term, and the term used in this study, is *bid protest*, the United Nations Commission on International Trade Law (UNCITRAL) refers to reviews, while the World Trade Organization's (WTO's) *Agreement on Government* Procurement uses the term challenges (see Gordon, 2006).

[3] (Fraction of Military Contracts/Total Government Contracts in U.S. & EU? Protest data in both?)

[4] "Fundamentally, bid protest systems, like audit systems, serve a procurement oversight function. They provide a means of monitoring the activities of government procurement officials, enforcing compliance with procurement laws and regulations, and correcting incidents of improper government action. … Enforcing compliance with procurement laws implicates not just high standards of integrity, but also … the maximization of competition." (Troff, 2005, pp. 118, 120)

[5] A key pillar of the U.S. Federal Government's protest process, the Competition in Contracting Act (CICA of 1984, Title 31 of the U.S. Code, §§ 3551–3556) claims bid protests improve procurement outcomes by reducing risks of fraud and errors, and increasing competition. The CICA gives the GAO authority over bid protests as a less expensive alternative to judicial proceedings. Congress directed "to the maximum extent practicable, the Comptroller General (at GAO) shall provide for the inexpensive and expeditious resolution of protests" (31 U.S.C. § 3554(a)). Since the majority of protests are filed with the GAO, that is the primary focus of this study. (Note: From FY2003–2007, nearly 7,000 cases were filed with the GAO, and only 328 with the Court of Federal Claims; see Schaengold, Guiffré, & Gill, 2009, p. 255.)

Although bid protests are relatively rare in low-cost procurements, vendor selection decisions in major (high-dollar) defense purchases appear to be routinely and strategically protested.[6] As former head of the Office of Federal Procurement Policy, Dan Gordon, observed,

> It is … true that very high-dollar procurements are much more likely to be protested: the higher the dollar value, the greater the likelihood of a protest. … For a company that loses the competition … with all the bid and proposal costs ["bidding costs"] …, the additional cost of filing a protest ["filing costs"] may seem minimal, so that filing a protest can be very tempting. (Clark, 2013)

The model introduced in this paper focuses on a representative bidder competing for a government contract. The bidder is assumed to be a strategic, profit-maximizing firm responsible to shareholders. Under this assumption, it is demonstrated that well-intentioned protest systems can inadvertently motivate inefficient (and potentially fraudulent) behavior on the part of bidders and public procurement officials, and may or may not increase competition. Some preliminary observations appear below.

Do Protests Minimize Fraud and Errors?

Strategic bidders can use the threat of protests to extract concessions from well-intentioned procurement officials unwilling to risk shortages of critical equipment, services, or supplies ("Fedmail").[7] Meanwhile, risk-averse procurement officials may have incentives to pre-emptively offer concessions to bidders, to ensure protest-proof procurements that avoid delays in acquiring critical equipment, services, and supplies ("Buy-offs").[8] For example, Reuters news service recently reported, "Lockheed Martin (LMT) is getting offered a multiyear block buy for its F-35 aircraft in exchange for not objecting to its rival Boeing (BA) getting new orders from the Navy for the F/A-18 fighter" (Reuters, 2016).

Especially troublesome is a measure of successful protests developed and routinely reported by the Government Accountability Office (GAO), misleadingly called the *effectiveness rate*. This measure captures "the percentage of protesters obtaining relief—either through a protest being sustained, or through *voluntary action* [*emphasis added*] taken by the agency" (Kepplinger, 2008).

---

[6] (GAO data reference; Aerial Tanker, Air Force Bomber examples; etc.)

[7] The Congressional Research Service (CRS) reports, "Many … acquisition professionals are concerned that bid protests can delay contract awards … costing millions of dollars [and] preventing government from getting the goods and services it needs when it needs them" (Schwartz & Manuel, 2009, p. 8). Government's incentive to avoid the risk of significant transaction costs from bid protests is revealed in an August 2007 memo by then Acting Under Secretary of Defense for Acquisition, Technology, & Logistics John Young, who stated, "Protests are extremely detrimental to the warfighter and the taxpayer. These protest actions consume vast amounts of time of acquisition, legal, and requirements team members; [and] delay program initiation and the delivery of capability" (Schwartz & Manuel, 2009, p. 8).

[8] A survey by the American Bar Association (1989) found that half of all federal agencies had settled protests to "simply move forward with the procurement … POs [procurement officials] often settle by enhancing the terms of other contracts that the protester currently has with the procuring agency" (Marshall et al., 1994, p. 300).

The Congressional Research Service (CRS) is on record stating, “The *effectiveness rate* may be a good way to measure the number of protests that have actual or potential merit” (Schwartz & Manuel, 2009, p. 5). At first glance, it might appear a greater effectiveness rate reflects positively on the dual goals of a protest system. Unfortunately, this is not necessarily the case. So-called “voluntary actions” can also involve inefficient and potentially fraudulent “Fed-mail” or “Buy-off” settlements. Increases in the “effectiveness rate” could inadvertently reflect government agencies over-generously engaged in Fed-Mail or Buy-off settlements with taxpayer dollars to keep procurements on schedule, minimize delays, or simply avoid negative publicity.

If so, then this clearly contradicts the conclusion drawn by the CRS that the effectiveness rate reflects protests that have merit. The risk of Fed-mail and Buy-off settlements warrants serious rethinking by the GAO, CRS, and others of the protest *effectiveness rate*. It also cautions departments, agencies, and Congress against implementing any analysis or recommendations that utilize this measure.

Do Protests Increase Competition?

A guiding principle of the Competition in Contracting Act and the Federal Acquisition Regulations (FAR) is to promote competition for government contracts. The implicit assumption woven through these documents, and in the legal and regulatory literature, is that the “second-chance” offered by bid protests to address fraud or errors in the procurement process makes prospective *losing* bidder types *more* inclined to participate (thereby increasing competition).

But this ignores prospective *winning* bidder types! The risk they face is that a winning bid will be delayed and disputed, increasing transaction costs, which reduces expected values of winning a government contract. On the margin, this makes prospective winning bidder types *less* inclined to participate (thereby reducing competition). This observation yields the counterintuitive result that reducing protests could actually increase competition (i.e., if reducing the risk of disputes motivates more winning bidder types to participate than losing bidder types drop out). In contradiction to the existing literature, it is therefore an empirical question whether or not a protest system increases the number of suppliers willing to participate.

Unfortunately, even if (on net) a bid protest system succeeds in attracting more vendors, insights from “Transaction Cost Economics” remind us ex-ante competition often leads to ex-post monopoly (Williamson). The risk is that a winning “foot-in-the-door” bidding strategy results in a “hold-up,” where the winning bidder more than covers its losses from high prices for change orders, etc. (see Melese et al., 2007). Therefore, attracting more vendors ex-ante does not guarantee better ex-post public procurement outcomes. Similarly, regardless of how slight the probability a protest will be sustained, a losing incumbent on a re-competed contract has a strong incentive to strategically protest to artificially extend the contract.[9]

To achieve desirable competitive market outcomes, instead of bid protests, the "contestable markets" literature (Baumol, Panzar, & Willig, 1982, etc.) urges lowering entry barriers as a more cost-effective strategy—for example, by reducing military specifications; excessive rules and regulations; complexity (e.g., "bundling"); unique government accounting/reporting and other regulatory requirements; the degree of asset specificity; or the ability of incumbents to raise entry barriers through strategic bid protests.

What Interventions and Alternatives Exist to Achieve the Goals of a Protest System?

If profit-maximizing strategic bidders can undermine government's goals of a bid protest system, then it pays to investigate ways to modify bidder behavior, and to explore alternatives, that is, more cost-effective governance mechanisms. Risks posed by significant transaction costs and unintended consequences from bid protests should encourage public officials to review costs and benefits of their protest system, and seek alternatives. The model developed in this paper offers a starting point.

The comparative statics results of the model reveal how several key government decision variables could impact a profit-maximizing representative bidder. Recognizing costs as well as benefits of a protest system, this study invites a review of alternative portfolios of governance mechanisms to improve procurement outcomes that could substitute for, or complement, bid protests (e.g., internal audits, external audits, independent investigations, alternative dispute resolution, integrity pacts, and other incentive mechanisms). Results of the model suggest there may be significant returns from another critical investment that impact the protest system—education, training, motivation (incentive alignment), and retention of experienced public procurement officials.

If it is determined the burden of protests outweighs the benefits, then reducing the rate of protests is appropriate, and can be accomplished in two ways: by *reducing expected benefits* of a protest to a "disappointed bidder" (including enabling the protester to achieve desired outcomes through other means), or by *increasing expected costs*. Options include the following: narrowing standing (eligibility), setting stricter time limits for filing and deciding protests, encouraging alternative dispute resolution (ADR), raising filing fees, setting fines for frivolous protests, instituting new rules or reputation assessments to restrict frequent or repeated protestors, or making losers pay as in the UK (see Appendix 1).

A major concern expressed in the legal and regulatory literature is that limiting protests will inhibit competition and result in higher costs.[10] However, the literature is mostly silent regarding the *strategic* behavior of bidders and procurement officials. It also ignores potential benefits of more timely delivery of projects, products, and services, and lower

---

[9] "Federal statutes and regulations … [require] GAO to … [resolve] protests within 65 to 100 days after they are filed." "[Automatic] stays triggered by GAO protests [can] encourage contractors to 'game the system … [where] contractors knowingly file … protests with GAO in order to harass their competitors and delay awards … or in the hopes of obtaining short-term contracts … during the pendency of the GAO protest" (Manuel & Schwartz, 2016, pp. 7, 11).

[10] "Attempts to disincentivize protests … may have, on balance, the unintended consequence of harming the federal procurement system by discouraging participation in federal contracting and, in turn, limiting competition" (Kepplinger, 2009, p. 12).

transaction costs, and potentially lower prices, from fewer protests. Finally, the legal and regulatory literature mostly neglects how the growing burden and complexity of regulations to address past procurement problems complicates the task of procurement officials. The more complex the regulatory environment, the more likely errors are made in the procurement process, raising the probability of bid protests and the probability those protests are sustained, which in turn increases risks of Fed-mail and Buy-off settlements.[11]

The next section leverages these observations to develop a probabilistic, micro-economic, partial equilibrium, representative bidder model. The following section summarizes and interprets results of the model. The concluding section offers policy recommendations and important avenues for future research.

The Model

The literature generally focuses on two players: a disappointed bidder and the government. In the United States, the "government" consists of several distinct players. Disappointed bidders have the option to challenge any of three key players: government procurement officials (POs) and their agency (department or activity); the quasi-judicial GAO; and the Court of Federal Claims (COFC). For simplicity, we restrict our representative bidder to a single protest (e.g., either with the agency, the GAO, or the COFC).[12]

Other key stakeholders are often overlooked in the protest literature. Besides a "disappointed bidder," it is critical to consider other bidders (especially the "winning bidder," eager to defend the award); those that ultimately depend on procurement outcomes (e.g., our troops and/or citizens); and taxpayers who foot the bill.[13] The goal of this paper is to represent the best interests of the last two players, in the case of military contracts, troops and taxpayers. This section develops a probabilistic, micro-economic, partial equilibrium representative bidder model to help identify opportunities to enhance the efficiency and effectiveness of government procurements, in order to obtain the greatest (troop) value for (taxpayer) money.[14]

---

[11] In fact, Wong & Gerras (2015) conclude U.S. Army officers became comfortable lying about complying with regulations, partly as a result of the challenge of compliance with conflicting regulations.

[12] Note GAO issues preliminary and final decisions on protests. Again for simplicity, the model assumes a single decision is taken by the Agency, GAO, or COFC.

[13] Gordon (2006) focuses on four principal parties: the disappointed offeror who is denied a contract award or the potential offeror who is excluded from competition, the acquiring agency, the public at large and their elected representatives, and an intervening offeror or successful awardee. Each has a different objective in resolving the protest. The unsuccessful offeror seeks a forum to air complaints, to learn as much information as possible about the denial or exclusion of their offer, and, ultimately, to obtain some type of meaningful relief. The acquiring agency seeks to resolve the protest in a way that minimizes the impact on the efficiency and effectiveness of the acquisition process. The public seeks a resolution that promotes the integrity, transparency, and accountability of the acquisition system. The successful awardee (or intervening offeror) seeks a resolution that supports the original award (Gordon, 2006, p. 4).

[14] "The federal procurement system was designed by Congress to leverage maximum public benefit from scarce taxpayer funds through three guiding principles: competition, integrity, and transparency. The aim of allowing bid protests is to "play an important role in ensuring integrity in the federal procurement system while … enhancing transparency and accountability" (JAT, 2009, p. 1).

Assuming a representative bidder's goal is to maximize expected profits, the objective function for any bidder/offeror entering a competition for a government contract consists of three scenarios (or "states of nature"): (1) expected returns from winning the competition, E(W); (2) expected returns from winning a protest given they lose the competition (i.e., protest is "sustained," E(W/L); and (3) expected returns from losing the competition and losing the protest (i.e., protest is not sustained, E(L/L).

Our representative bidder's problem is illustrated in Figure 1. The probability the bidder wins the competition is Pw, and the probability a protest is sustained is Ps.[15] Expected payoffs at the end of each branch (E[W]; E[W/L]; E[L/L]) are explained in detail below.

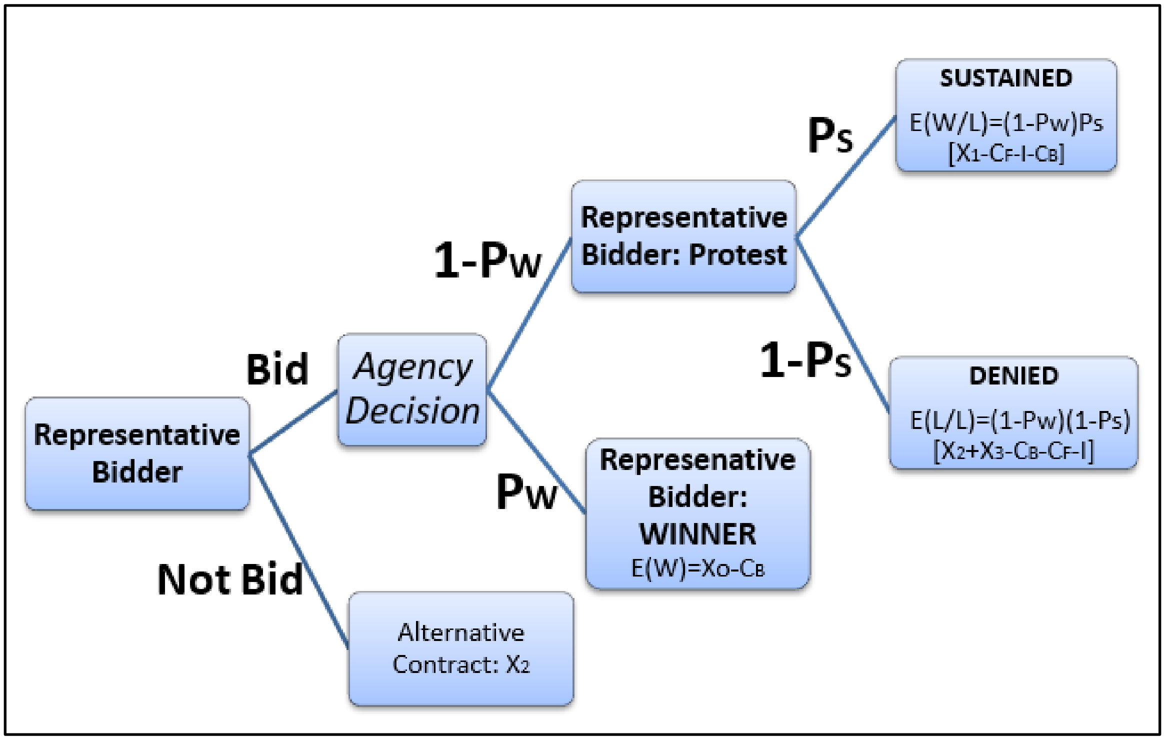


Figure 1. Representative Bidder Decision Problem

The two key decision variables controlled by our representative bidder are the bid price, P≥0, for the contracted quantity, Q≥0; and the investment, I≥0, to sustain a protest in the event the bidder loses the competition. The bidder's problem is to select an optimal combination of bid price and protest investment (P*, I*) to maximize overall expected profits:

$$\mathbf{Max\ V(P,I) = E(W) + E(W/L) + E(L/L).} \qquad (1)$$

---

[15] Note the sum of the probabilities of the three possible states of nature (winning the competition E[W]; losing but winning the protest E[L/W]; losing and losing the protest E[L/L]) are mutually exclusive and collectively exhaustive, and therefore must sum to one: i.e., $P_w+(1-P_w)[P_s+(1-P_s)]=1$.

The first term is the expected return from winning the competition given by

$$E(W) = P_W(X_0 - C_B). \qquad (2)$$

Profits from a winning bid are: $X_0 = PQ - C(Q,R)$, where $C(Q,R)$ is the winning bidder's cost function, and R represents regulatory complexity, such that $\delta C/\delta Q>0$, and $\delta C/\delta R>0$, that is, a more complex and burdensome regulatory environment raises production costs.

To allow the possibility other bidders might protest a winning bid, we introduce the possibility of "split buys" (see Coughlan & Gates, 2012). The contracted quantity is therefore given by $Q=Q(I_0)$, where $I_0\geq 0$ represents cumulative protest investments of other losing bidders, such that $\delta Q/\delta I_0<0$, that is, the greater the cumulative protest investment, the smaller the quantity allocated to a winning bidder.

To simplify the model, bid and proposal costs, C_B, act as a proxy for quality of the project, product, or service, and capture any other non-price variables of interest to the government. We assume these costs are directly related to the measure of effectiveness (MOE) of the bidder's proposal (i.e., not including price).

Therefore, higher bid and proposal costs, C_B, incurred by our bidder (ceteris paribus), increase the probability they win the competition, given by $P_W=P_W(P,N,C_B)$, such that $\delta P_W/\delta C_B>0$ and $\delta^2 P_W/(\delta C_B)^2<0$. Conversely, the higher the price bid, P, and the more bidders, $N\geq 2$, the lower the probability of winning, such that $\delta P_W/\delta N<0$ and $\delta P_W/\delta P<0$,[16] where $\delta^2 P_W/(\delta P)^2<0$, $\delta^2 P_W/\delta P\delta N\leq 0$, and $\delta^2 Pw/\delta P\delta C_B\geq 0$.[17]

The second term in Equation 1 represents expected returns from losing the competition, but winning the protest (i.e., protest is "sustained"):

$$E(W/L)= (1-P_W)P_S[X_1-C_F-I-C_B], \qquad (3)$$

where $(1-P_W)$ is the probability of losing the competition; C_F, are exogenous government-set filing fees, and for simplicity, $X_1\leq X_o$ is the award or "prize" in the event the protest is sustained.

The probability a protest is sustained is given by $P_s=P_s(P,I,N,T,I_0,R, C_B)$. The higher the bid price, P>0, the lower the probability a protest is sustained, such that $\delta P_s/\delta P<0$,

[16] We further assume the absolute value of the elasticity of the probability of winning the competition with respect to the bid price is less than one, that is, the elasticity is the %reduction in P_w for a given %increase in P, or $|(\delta P_w/\delta P)(P/P_w)|<1$

[17] The greater the number of bidders, N, then for any given bid price, P, the lower the probability of winning, P_W. Conversely, the greater a representative bidder's investment in the quality of their proposal reflected in bid and proposal costs, C_B, then for any given bid, P, the greater the probability of winning, P_W.

where $\delta^2P_s/(\delta P)^2<0$.[18] However, the greater the representative bidder's investment in the protest process, I>0, the greater the likelihood a protest is sustained, such that $\delta P_s/\delta I>0$, where $\delta^2P_s/(\delta I)^2<0$, that is, bidder protest investments increase the probability a protest will be sustained, but at a decreasing rate, and $\delta^2P_s/\delta P\delta I\leq 0$ (the larger the protest investment, the smaller the impact of a high bid price on the probability the protest is sustained). Also, for any given protest investment, the greater the quality of the proposal (reflected in higher bid and proposal costs, C_B), the greater the probability the protest is sustained, or $\delta^2P_s/\delta I\delta C_B>0$.

The proxy variable, T, represents the training/education/experience of government procurement officials. The greater T, the lower the risk of errors in the acquisition process, and the better communication, documentation, contract specifications, quality of debriefings, and so forth. Thus the greater T, the less likely a protest will be sustained, or $\delta P_s/\delta T<0$, where $\delta^2P_s/\delta I\delta T<0$, and $\delta^2P_s/\delta P\delta T\leq 0$.

The reverse is true for regulatory complexity, represented by the parameter, R. The more complex and burdensome the regulatory environment, not only does this increase production costs, but it leads to a greater risk of missteps and errors by procurement officials, which increases the probability a protest is sustained, or $\delta P_s/\delta R>0$, where $\delta^2P_s/\delta I\delta R>0$ and $\delta^2P_s/\delta P\delta R\geq 0$.

Since bid and proposal costs are a proxy for quality, the greater C_B, the greater the probability a protest is sustained, $\delta P_s/\delta C_B>0$, where $\delta^2P_s/\delta C_B)^2<0$, $\delta^2P_s/\delta P\delta C_B\geq 0$, and $\delta^2P_s/\delta I\delta C_B\geq 0$, that is, the greater a representative bidder's investment in their proposal, C_B, then for any given bid, P, or protest investment, I, the greater the probability a protest is sustained, P_s.

Data reported in Maser & Thompson (2010) suggests increasing the number of bidders (ceteris paribus) increases the probability a protest is sustained, $\delta P_s/\delta N>0$. However, the sign on $\delta^2P_s/\delta I\delta N$ is an empirical question: positive (negative) depending if more bidders increases (decreases) the likelihood a representative bidder's protest is sustained, for any given protest investment.

Finally, it is also an empirical question whether greater cumulative protest investments by other bidders, I_o, raises or lowers the probability a given bidder's protest is sustained (i.e., $\delta P_s/\delta I_o=?$). It is also unclear if greater cumulative protest investment makes it more or less likely an individual bidder's protest will be sustained for any given bid price (i.e., $\delta^2P_s/\delta P\delta I_o=?$), and more or less likely any individual bidder's protest is sustained for a given protest investment (i.e., $\delta^2P_s/\delta I\delta I_o=?$).

---

[18] Note that it is likely the percentage difference in price relative to the low price bid, $(P-P_L)/P_L$, is more likely to influence the probability a bid is sustained, than the absolute price bid, P. This can be accommodated by constraining the functional form of the relationship given by $\delta P_s/\delta P<0$ and $\delta^2Ps/(\delta P)^2<0$, so that the Price is bounded between the low price bid, P_L (where $P_s=1$) and $P_{Max}=(1+X\%)P_L$ (where $P_s=0$), that is, where X reflects how far the price in percentage terms can reasonably be above the low bid before there is no chance a protest will be sustained. In this case, the comparative static results will be the same if we use either the bid price, or the percentage difference between the bid price and the low bid.

The final term in our representative bidder's profit Equation 1 is the expected return from losing the competition, and losing the protest:

$$\mathbf{E(L/L)= (1\text{-}P_w)(1\text{-}P_s)[X_2+X_3\text{-}C_F\text{-}I\text{-}C_B],} \quad (4)$$

where the variable, $X_2<X_o$, represents the bidder's "opportunity cost," or value of the next best alternative project available if they lose the competition (or decide not to participate; see Figure 1). If a representative bidder's protest is denied, then the variable, X_3, can represent two possibilities: $X_3>0$ represents compensation that might be offered a losing bidder (i.e., possibly reflecting "Fed Mail" or "Buy Offs," or perhaps valuable information obtained about competitors), while $X_3<0$ represents a penalty for losing the protest (e.g., "loser pays").

Maximizing the representative bidder's expected profits (given by (1), (2), (3), and (4)) to solve for the optimal bid price and protest investment (P*,I*) yields the following First Order Necessary Conditions for an Optimum:

$$\mathbf{V_1= \delta V/ \delta P} = P_wQ+(\delta P_w/ \delta P)[X_o\text{-}(X_2+X_3)\text{-}C_F\text{-}I]+(X_1\text{-}X_2\text{-}X_3)[(1\text{-}P_w)(\delta P_s/ \delta P)–Ps(\delta P_w/\delta P)]=0; \quad (5)$$

and

$$\mathbf{V_2= \delta V/ \delta I} = (1\text{-}P_w)[(\delta P_s/\delta I)(X_1\text{-}X_2\text{-}X_3)\text{-}1]=0. \quad (6)$$

Conditions required to ensure the Second Order Sufficient Conditions are satisfied at the optimum (or that, $V_{11}V_{22}\text{-}V_{12}^2>0$), include the following: $\delta^2 P_s/(\delta I)^2<0$; $\delta^2 P_s/(\delta P)^2\leq 0$; $X_1>(X_2+X_3)$; and $X_o>[(P_sX_1+(1\text{-}P_s)(X_2+X_3))\text{-}C_F\text{-}I]$.

From the Implicit Function Theorem, the first order necessary, and second order sufficient, conditions for a maximum yield a set of comparative statics results for the two decision variables: the optimal bid price, P*, and protest investment, I*. Applying the Envelope Theorem[19] further reveals the impact of changes in the exogenous variables on a representative bidder's expected profits, V*. A summary of the results appears in Table 1.

Table 1. Comparative Statics Results

	X_1	X_2	$X_3>0$ ($X_3<0$)	N	I_o	C_B	C_F	T (R)
P*	+?	+	+ (-)	-?	-?	+	-	- (+)
I*	+	-	- (+)	-?	-	+	0	- (+)
V*	+	+	+ (-)	?	-	?	-	- (?)

Note. "?" indicates: given certain conditions.

Results

To interpret the results in Table 1, we work our way from left to right across the top row, and discuss each model parameter in turn. The bigger the expected "prize" from a protest, $\mathbf{X_1}$, the greater the optimal bid price, P*. When a protest offers a bigger prize/award, it is optimal to increase bids, taking a greater risk of losing the competition, because of the

[19] From the Envelope Theorem (Silberberg, 1978, pp. 168-71), taking partial derivatives of the objective function with respect to any parameter, k, yields the change in the overall value function at the optimum, **V***, with respect to a change in k. Detailed calculations are available upon request.

greater expected value from the "second chance" provided by a bid protest. Not surprisingly, a bigger protest prize also justifies a bigger investment, I*, to increase chances of winning the protest. Naturally, a bigger protest prize also boosts overall expected profits at the optimum, V*.

While a greater value of a representative bidder's next best alternative project, $\mathbf{X_2}$, justifies a higher bid price for the government contract, P* (i.e., there is a lower opportunity cost to losing the competition), it also reduces the incentive to invest in a protest, I*. Of course, an increase in the value of any alternative to the government contract increases overall expected profits, V*.

Now consider the possibility of "Fed Mail" or "Buy Offs" so that losing a protest still offers a consolation prize, $\mathbf{X_3>0}$. Then any increase in such benefits (presumably made by a government agency to reduce future disruptions from unhappy bidders), *besides directly raising agency costs and indirectly other costs, involves a serious negative externality—an unintended consequence is higher optimal bid prices, P*, which increases overall government procurement costs*. However, since a bigger consolation prize means the same expected value of a protest can now be achieved with a lower probability of winning the protest, this has the effect of lowering incentives to invest in bid protests, I*. Of course, a higher consolation prize increases overall expected profits, V*.

Instead of receiving a consolation prize for losing a protest, now suppose penalties apply, or that $\mathbf{X_3<0}$. In this case, increasing penalties yields the opposite results: The optimal bid price, P*, will be lower to try to win the competition, since there is now greater risk in protesting. The greater risk of punishment from *losing* a protest also means it pays to invest *more in winning the protest*, I*. Finally, the added risk (expected cost) of a possible penalty reduces overall expected profits, V*.

The greater the number of bidders, N, the lower any individual competitor's optimal bid price, P* (a public benefit of increased competition), and protest investment, I* (if $\delta^2 Ps/\delta I \delta N<0$). It also lowers the expected overall profits of any individual bidder, V*.[20]

Knowing there is an increase in protest expenditures by other bidders, I_0, will reduce the optimal bid price of a representative bidder, P* (if the absolute value of elasticity of P_w with respect to P is less than one), and the representative bidder's own protest investment, I*. It also reduces the bidder's expected profits, V*.[21]

An increase in filing fees for a protest, C_F, lowers the optimal bid price, P*, but has no impact on protest investment, I*, since they are essentially "sunk costs." Of course, higher filing fees will lower overall expected profits, V*.

---

[20] Condition for P* is satisfied if $\delta^2 P_w/\delta P \delta N=0$ (or small enough). Condition for I* is satisfied if $\delta^2 P_s/\delta I \delta N<0$. Condition for V* is satisfied if positive impact of N on P_s ($\delta P_s/\delta N>0$) is small enough, and/or if the absolute value of the impact of N on P_w ($\delta P_w/\delta N<0$) is big enough. Higher bid costs increase the probability of winning the competition, but the extra costs lower profits from winning, requiring a higher price to "break even."

[21] Condition for P* holds if absolute value of impact of cumulative protest expenditures by other bidders, I_o, on a representative bidder's contract quantity ($\delta Q/\delta I_o<0$) is small enough, or the probability the representative bidder's protest is sustained ($\delta P_s/\delta I_o<0$) is big enough, and/or that ($X_1-X_2-X_3$) is big enough. Condition for I* holds since $\delta^2 P_s/\delta I \delta I_o<0$.

Conversely, since an increase in bid and proposal costs, C_B, reflects an increase in the quality (MOE) of the proposal, this increases the optimal bid price, P*, and the optimal protest investment, I*.[22] The impact on overall expected profits, V*, is indeterminate, that is, positive (negative) if benefits from increasing the probability of winning and sustaining a protest are bigger (smaller) than the higher investment costs of preparing the bid.

Paradoxically, adding well-intentioned rules and regulations that inadvertently increase regulatory complexity, R, can have perverse effects. It raises optimal bid prices, P, increasing the costs of public projects, products, and services. Increased regulatory complexity also contributes to higher bid protest investments, I*, increasing transaction costs and possibly triggering other unintended consequences*. The impact on overall expected profits, V*, is negative (positive) if added expected production costs from regulation, $P_w(\delta C/\delta R)$, are bigger (smaller) than the expected increase in profits from a protest, given the marginal increase in probability of winning a protest from greater errors, etc., resulting from increased regulatory complexity, $(1-P_w)(\delta P_s/\delta R)[X_1-(X_2+X_3)]$.

Finally, boosting government investments in education/training/experience of public procurement officials, T, has multiple payoffs. It lowers optimal bid prices, P, cutting the costs of public projects, products and services. It also reduces the optimal amount invested in bid protests, I*, lowering transaction costs, and possibly limiting other unintended consequences*. The impact of increasing the competency of procurement officials in reducing optimal prices bid, P*, and protest investments, I*, is reflected in lower overall expected profits for bidders, V*.[23]

Conclusion

The goal of a public procurement system is to obtain the best "value for money." To help achieve this goal, countries around the world have adopted bid protest systems. The legal and regulatory literature that underpins protest systems in the United States and European Union claim allowing disappointed bidders to protest public procurements reduces *favoritism, fraud*, and *errors*, and encourages *competition*. This study offers a cautionary tale for any government agency, country, or international institution that relies on, and/or promotes, bid protests to improve public procurement outcomes.

The paper explores costs and benefits of bid protests. As a first step, a probabilistic, micro-economic, partial equilibrium, representative bidder model is developed to evaluate protest systems. The bidder for a government contract is assumed to be a strategic, profit-maximizing firm responsible to shareholders. Under this assumption, it is demonstrated that well-intentioned protest systems can inadvertently motivate inefficient (and potentially fraudulent) behavior on the part of bidders and public procurement officials, and may or may not increase competition.

Risks posed by significant transaction costs and unintended consequences from bid protests should encourage public officials to review protest systems and consider alternatives. If the burden of protests outweighs the benefits, then reducing protests is appropriate and can be accomplished in two ways: reducing expected benefits of a protest

[22] Condition on P* is satisfied if $\delta^2P_w/\delta P\delta C_B\geq0$, and $\delta^2P_s/\delta P\delta C_B\geq0$. Condition on I* is satisfied if $\delta^2P_w/\delta I\delta C_B>0$.
[23] Condition on P* is satisfied since $\delta^2P_s/\delta P\delta T\leq0$. Condition on I* is satisfied since $\delta^2P_s/\delta I\delta T<0$.

to a “disappointed bidder” (including enabling the protester to achieve desired outcomes through other means) or increasing expected costs.

The comparative statics results of the model reveal how several key government decision variables could impact a profit-maximizing representative bidder. Our bidder controls the bid price, P*, and any investment they make to sustain a protest, I*. The government controls the variables: X_1, X_3, C_F, T, and R. Reducing the protest prize (X_1) and unnecessary regulatory burdens (R), and increasing investments in human capital (T), all reduce expected benefits of a protest. Alternatively, governments can raise expected costs by increasing filing fees (C_F) and/or introducing penalties for losing a protest ($X_3<0$).

Reducing the protest award, (X_1), can be accomplished by (i) unbundling the contract vertically, in terms of different stages of production, or horizontally, in terms of quantities; (ii) sharing the award (split buys); or (iii) keeping records of protests by firms and using this information in future competitions (i.e., using a company’s reputation to establish contract quantities).

Constructive ways of reducing the probability of a successful protest include (i) investing in training and experience, (ii) initiatives to build integrity (e.g., codes of conduct, ethics training, etc.), (iii) aligning incentives for procurement officials to improve procurement outcomes, such as linking pay and promotions to successful procurement outcomes, (iv) ensuring transparency of assessment criteria, (v) ensuring the transparency and accountability of the evaluation and selection process, (vi) making companies aware of the low probability of awards being overturned, and (vii) substituting protests for alternatives such as random (internal and external) audits and investigations, encouraging alternative dispute resolution (ADR), etc. Alternatively, the probability of successful protests would drop if the GAO narrowed standing (i.e., eligibility) or raised the threshold required for a protest to have merit.

An important avenue for future research is to review alternative portfolios of governance mechanisms to improve procurement outcomes that could substitute for, or complement, bid protests. This study offers a starting point.

References

Arrowsmith, S., Linarelli, J., & Wallace, D., Jr. (2000). *Regulating public procurement: National and international perspectives.* The Hague, Netherlands: Kluwer International Law.

Baumol, W., Panzar, J., & Willig, R. (1982). Contestable markets and the theory of industry structure. New York, NY: Harcourt Brace Jovanivich.

Benishek, P., Sheinman, B., Kidalov, M., & Angelis, D. (2011). *Better acquisition management through alternative dispute resolution (ADR) and other best practices for preventing and resolving bid protests.* Retrievable from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Biery, F. (1992). The effectiveness of weapon system acquisition reform efforts. *Journal of Policy Analysis and Management, 11*(4), 637–664.

Brown, T., & Potoski, M. (2003). Transaction costs and institutional explanations for government service and production decisions. *Journal of Public Administration Research and Theory, 13*(4), 441–468.

Camm, F., et al. (2009, October). *GAO bid protests in Air Force source selections*. RAND Corp., Unpublished Manuscript for USAF.

Clark, C. (2013, March 12). *Bid protests are worth their costs, ex-procurement chief says. Government Executive, 1*(12), 17.

Competition in Contracting Act of 1984, Pub. L. No. 98-369, 98 Stat. 1175-1203 (1984) (codified in various sections of 10 U.S.C., 31 U.S.C., 40 U.S.C., and 41 U.S.C.).

Coughlan, P., & Gates, W. (2012, July 18). *Endogenous split buys as a bid protest management tool* (NPS-CM-12-180). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Court of Federal Claims Rules. (n.d.). Retrieved from http://law.justia.com/us/codes/title28a/28a_7_.htmlSAF

Crean, S. (2008, January 8). *Improving communication during competitive source selections* [Memorandum for heads of contracting activities]. Retrieved from https://acquisition.navy.mil/content/download/5263/23838/file/enhancing%2520competition%25201-18-2008.pdf.

Defense Acquisition University (DAU). (2009). *Interim defense acquisition guidebook.* Retrieved from https://acc.dau.mil/dag

Demski, J., & Feltham, G. (1978). Economic incentives in budgetary control systems. *Accounting Review, 53*, 336–359.

Department of Justice. (2008, withdrawn on May 11, 2009). *Competition and monopoly: Single-firm conduct under section 2 of the Sherman Act,* 144. Washington, DC: Author. Retrieved from http://www.usdoj.gov/atr/public/reports/236681.htm

Dillard, J., Franck, R., & Melese, F. (2006). A transaction costs economics approach to defense acquisition management. In *Proceedings of the Third Annual Acquisition Research Symposium.* Monterey, CA: Naval Postgraduate School.

DoD. (2010, February 1). *2010 QDR fact sheet*.

DoD. (2009). *Consolidated acquisition reporting system (CARS) user's guide*. Retrieved from http://www.acq.osd.mil/cars/Downloads/CARS%20Users%20Guide.doc. [Note: CARS has been replaced by DAMIR: http://www.acq.osd.mil/damir]

Flyvbjerg, B., Holm, M., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association, 68*(3), 279–295.

FMTV 2010-2015: Oshkosh wins the re-compete. (2010, November 21). *Defense Industry Daily.* Retrieved from http://www.defenseindustrydaily.com/FMTV-2010-2015-Oshkosh-Wins-The-Re-Compete-05744/

Fox, R. (1974). *Arming America: How the U.S. buys weapons*. Boston, MA: Harvard University Press.

Fox, R., & Field, J. (1988). *The defense management challenge: Weapons acquisition*. Boston, MA: Harvard Business School Press.

Francis, P., Golden, M., & Woods, W. (2010). *Defense acquisitions: Managing risk to achieve better outcomes* (GAO-10-374T). Washington, DC: GAO.

Gansler, J., Lucyshyn, W., & Arendt, M. (2009, September). *Bid protests in the Defense Department: An analysis of recent trends (*Center for Public Policy and Private Enterprise Technical Report, UMD-CM-09-135). College Park, MD: CPPPE.

GAO. (1986, January 31). *GAO statutory report on FY1985 bid protest activity to the Honorable George Bush, President of the Senate* (B-158766). Washington, DC: Author.

GAO. (1990, March 30). *ADP bid protests: Better disclosure and accountability of settlements needed* (GAO/GGD-90-13). Washington, DC: Author.

GAO. (2005). *Better support of weapon systems managers needed to improve outcomes* (GAO-06-11). Washington, DC: Author.

GAO. (2008). *Aerial refueling tanker protes*t (GAO-08-991T). Washington, DC: Author.

GAO. (2009). *GAO bid protest overview*. Retrieved from http://www.gao.gov/new.items/d10534sp.pdf

GAO, Office of General Counsel. (2008). *Bid protests at the GAO: Descriptive guide* (9th ed., GAO Special Publication GAO-09-471SP). Washington, DC: Author.

Gates, D. (2008, June 19). Boeing tanker "back in the game" after GAO backs company's protest. *Seattle Times*. Retrieved from Congressman Norm Dicks website: http://www.house.gov/apps/list/speech/wa06_dicks/seattletimestanker.shtml

Gordon, D. I. (2006). Constructing a bid protest process: Choices every procurement challenge system must make. *Public Contract Law Journal, 3*, 1–18.

Graham, R. (2003, Summer). The transformation of contract incentive structures. Acquisition *Review Quarterly*. Retrieved from http://www.dau.mil/pubscats/PubsCats/arq2003.aspx#summer

Grossman, S., & Hart, O. (1983). An analysis of the principal-agent problem. *Econometrica, 51*(1), 7–45.

Holmstom, B., & Milgrom, P. (1991, Spring). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization, 7*, 24–52.

Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics, 3,* 305–360.

Johnsson, J. (2009, July 19). Split could seal Air Force aerial-refueling tanker contract. *Chicago Tribune*. Retrieved from http://www.chicagotribune.com

Kepplinger, G. L. (2008). *Bid protest report for fiscal year 2008 and bid protest statistics for fiscal years 2004-2008* (GAO-09-251R, B-158766) [Letter to the Honorable Nancy Pelosi, Speaker of the House]. Washington, DC: GAO.

Kepplinger, G. L. (2009a, January 19). Commentary: GAO's bid protest role. *Federal Times.* Retrieved from http://www.federaltimes.com

Kepplinger, G. L. (2009b, April 9). *Report to Congress on bid protests involving defense procurements* (GAO B-401197). Washington, DC: GAO.

Klein, B., Crawford, R., & Alchian, A. (1978). Vertical integration: Appropriable rents and the competitive contracting process. *Journal of Law and Economics, 21*, 279–326.

Kogut, B., & Kulatilaka, N. (2001). Capabilities as real options. *Organization Science, 12*(6), 744–758.

Kovacic, W. E. (1995). Procurement reform and the choice of forum in bid protest disputes. *Administrative Law Journal of the American University, 9*, 461, 486.

Lieberman, R. D., & Morgan, J. D. (2008). *The 100 worst mistakes in government contracting*. Ashburn, VA: National Contract Management Association.

Manuel, K., & Schwartz, M. (2016, January 19). *GAO bid protests: An overview of time frames and procedures* (Report 7-5700: R40228). Washington, DC: Congressional Research Service.

Marshall, R. C., Meurer, M., & Richard, J.-F. (1991). The private attorney general meets public contract law: Procurement oversight by protest. *Hofstra Law Review, 20*(1), 1–71.

Maser, S., & Thompson, F. (2010). *Understanding and mitigating protests of DoD acquisition contracts* (Technical Report WIL-CM-10-164, 1-83). Salem, OR: Willamette University, Atkinson Graduate School of Management.

Melese, F., Franck, R., Angelis, D., & Dillard, J. (2007). Applying insights from transaction cost economics to improve cost estimates for public sector purchases: The case of U.S. military acquisition. *International Public Management Journal, 10*(4), 357–385.

Melese, F. et al. (2010). *A new paradigm to address bid protests* (NPS-CM-10-159, 1-129). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Menard, C., & Saussier, S. (2003). La délégation de service public, un mode organisationnel efficace? Le cas de la distribution d'eau en France. *Economie Publique, 1,* 99–129.

Metzger, R., & Lyons, D. (2007). A critical reassessment of the GAO bid protest mechanism. *Wisconsin Law Review,* (6), 1225.

Office of the General Counsel, GAO. (2004, January). *Principles of federal appropriations law* (GAO-04-261SP, 3rd ed., Vol. 1). Washington, DC: GAO.

Pint, E., & Baldwin, L. (1997). *Strategic sourcing: Theory and evidence from economic and business management* (Mr-865-Af). Santa Monica, CA: RAND.

Potter, M. (2009, December 21). FMTV contract award and protest raises industrial policy issues. Retrieved from http://industry.bnet.com/government/10004609/fmtv-contract-award-and-protest-raises-industrial-policy-issues/?tag=untagged

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature, 37*(1), 7–63.

Ross, S. (1973). The economic theory of agency: The principal's problem. American *Economic Review, 63*, 134–139.

Sappington, D. (1991). Incentives in principal-agent relationships. *Journal of Economic Perspectives, 5*(2), 45–66.

Schaengold, M. J., Guiffré, T. M., & Gill, E. M. (2009). Choice of forum for federal government contract bid protests. *Federal Circuit Bar Journal, 18*.

Scherer, F. (1964). The weapons acquisition process: *Economic incentives*. Boston, MA: Harvard University Press.

Schooner, S. L. (2001). Fear of oversight: The fundamental failure of businesslike government. *American University Law Review, 50*(3), 627–639.

Schwartz, M., & Manuel, K. M. (2009). *GAO bid protests: Trends, analysis, and options for Congress* (CRS R40227). Washington, DC: Congressional Research Service.

Shalal-Esa, A. (2009, November 2). Pentagon concerned about routine protests. Retrieved from http://reuters.com

Sherman Act, 15 U.S.C. § 2 (2010).

Source Selection Joint Analysis Team. (2009, April 24). Best practices FINAL product.doc (pp. 9–10).

Spring, B. (2002, May). *Don't let politics or bureaucracy hobble missile defense* (Executive Memorandum 817). Retrieved from Heritage Foundation website: https://www.policyarchive.org/bitstream/handle/10207/8296/em_817.pdf

S. Rep. No. 103-258 (1994). Report of the Senate Committee on Governmental Affairs on the Federal Acquisition Streamlining Act of 1994. Washington, DC: Government Printing Office.

Troff, E. A. (2005). The United States agency-level bid protest mechanism: A model for bid challenge procedures in developing nations. *Air Force Law Review, 57*, 113.

Tucker Act, 28 U.S.C. § 1491 (2010).

United Nations Commission on International Trade Law (UNCITRAL). (1994). Model law on procurement of goods, construction and services.

Waterman, R., & Meier, K. (1998, April). Principal-agent models: An expansion? *Journal of Public Administration Research and Theory, 8*(2), 173–202.

Weckstein, K., & Love, M. (1995, June 12). Bid protest system under review [Special report]. *Legal Times.* Retrieved from http://www.law.com/jsp/nlj/legaltimes/index.jsp

Williamson, O. (1971, May). The vertical integration of production: Market failure considerations. *American Economic Review, 61*, 112–123.

Williamson, O. (1979). Transaction-cost economics: The governance of contractual relations. *Journal of Law and Economics, 22*, 233–261.

Williamson, O. (1999). Public and private bureaucracies: A transaction cost economics perspective. *Journal of Law, Economics and Organization, 15,* 306–342.

Wong & Gerras. (2015). *Lying to ourselves*. U.S. Army War College Press.

Worthington, M. M., & Goldsman, L. P. (1998). *Contracting with the federal government*. New York, NY: John Wiley.

Acknowledgment

I am grateful to my co-principal investigator, Troy Terronez, for helpful comments and suggestions.

Appendix

Note that in the United States, restricting the number of protests may be unconstitutional on First and Fifth Amendment grounds. To limit the number of non-frivolous protests would violate the First Amendment right to petition the government for the redress of grievances, and the Fifth Amendment right to due process. Federal courts tend not to favor broad limitations on access to the legal process. For example, the Supreme Court held in *Bill Johnson's Restaurants, Inc. v. National Labor Relations Board* (461 U.S. 731 [1983]), that a Federal agency cannot halt lawsuits brought even for improper motives unless those lawsuits are based on "intentional falsehoods or on knowingly frivolous claims," or otherwise lack a reasonable basis. In another case, *California Motor Transport Co. v. Trucking Unlimited* (404 U.S. 508 [1972]), the Supreme Court held that Federal antitrust laws may penalize businesses bringing lawsuits and petitions to Federal agencies only if such petitions and lawsuits are "a mere sham to cover what is actually nothing more than an attempt to interfere directly with a business relationship of a competitor." Federal appellate courts also identified two limited ways which can render a legal action frivolous:

> First, a legal action is considered "frivolous as filed" when a plaintiff or appellant grounds its case on arguments or issues "that are beyond the reasonable contemplation of fair-minded people, and no basis for [the party's position] in law or fact can be or is even arguably shown." … Second, a legal action is considered "frivolous as argued" when a plaintiff or appellant has not dealt fairly with the court, has significantly misrepresented the law or facts, or has abused the judicial process by repeatedly litigating the same issue in the same court. (GAO, 2009, p. 11)

However, options include agency policies requiring mandatory consideration of stay overrides, requiring vigorous objections, setting stricter time limits for deciding or resolving protests, mandating alternative dispute resolution (ADR) as the default resolution mechanism, or other approaches such as replicating sanctions for frivolous protests available at the Court of Federal Claims in GAO protests, or instituting rules such as the posting of bonds for the expenses of delays resulting from stays of protests that are ultimately not sustained. In addition, the standard of review at the GAO may be adjusted from the more relaxed and subjective "reasonableness" standard to the "arbitrary, capricious, abuse of discretion, or otherwise not in accordance with law" standard used by the COFC under the Administrative Procedures Act (*Choice of Forum for Federal Government Contracts Bid Protests,* at 298, 2009). Further, agencies can be encouraged not to allow post-award bid protests challenging the evaluation and the conduct of source selection to result as a matter of course in pre-award corrective actions, such as total cancellation of solicitation and full re-competition. (The Competition in Contracting Act [CICA of 1984, Title 31 of the U.S. Code, §§ 3551–3556]) is a key pillar of the U.S. protest process, together with the Federal Acquisition Regulation [FAR Parts 5, 10, 12–15, and 33], the Tucker Act, Title 28, Section 1491 of the U.S. Code, Executive Order No. 12979, *Agency Procurement Protests*, and various case law precedents. Legal insights provided by former NPS colleague, Max Kidalov, in Melese et al., 2010.)

An Analysis of DoD's Use of the Lowest Price Technically Acceptable Acquisition Strategy and Recommendations for Improvement

Patricia Donahoe—is a second-year Master of Public Policy (MPP) candidate at Duke University's Sanford School of Public Policy, where she is specializing in national security policy. Donahoe previously taught English in rural Brazil as a Fulbright Scholar. Most recently, Donahoe interned for the Section 809 Panel in Arlington, VA. Donahoe holds a bachelor's degree from the University of New Hampshire where she completed a double major in Political Science and International Affairs and graduated summa cum laude. She is a member of Phi Beta Kappa. Donahoe is excited to join L.M.I. as a Graduate Fellow after graduating from Duke. [patricia.donahoe@duke.edu]

Abstract

The DoD's use of the lowest price technically acceptable (LPTA) source selection method is a source of concern for many in the defense acquisition community. Some argue that the DoD has increasingly misused LPTA to procure complex goods and services that are difficult to define. Using data collected from the Federal Business Opportunities (FBO) website and the Federal Procurement Data System (FPDS), this report seems to test claims that the DoD has increased its use of LPTA and that its usage has yielded poor results (as measured by contract cancellation rates and vendor re-award rates). The results from this data query are mixed and show that LPTA usage has increased for all types of procurements and that there is some dissatisfaction associated with LPTA. The results, however, are questionable due to data validity concerns. This report concludes with recommendations for improving data collection for DoD source selection methods and contract cancellation rates.

Introduction

The United States Department of Defense (DoD) spends approximately $300 billion on goods and services contracts each year (Section 809 Panel, 2017). Specifically, during fiscal year (FY) 2015, the DoD "obligated more money on federal contracts ($274 billion) than all other federal agencies combined" (Schwartz & Manuel, 2015, p. 2). DoD acquisitions amounted to 7% of the federal government's total discretionary and mandatory spending and 62% of all federal contract obligations in FY 2015 (Schwartz & Manuel, 2015, p. 3). Ensuring efficiency within the defense acquisition system is paramount given the significant portion of taxpayer dollars the U.S. government commits annually.

This report's central research question is the following: Is the DoD using the lowest price technically acceptable (LPTA) source selection method to achieve its mission? If the DoD is not using LPTA effectively, what should the Department do to mitigate this problem?

DoD's Mission

The mission of the U.S. Department of Defense is to "provide the military forces needed to deter war and protect the security of our country" (DoD, 2017).

What Is LPTA?

The Federal Acquisition Regulation (FAR) governs executive federal agency acquisition processes. First written in 1984, the FAR ensures that executive federal agencies “deliver on a timely basis the best value product or service … while maintaining the public’s trust and fulfilling public policy objectives” (FAR Foreword; FAR 1.102). While the DoD has its own internal acquisition policy guidelines and its own supplement to the FAR, the Defense Federal Acquisition Regulation Supplement (DFARS), DoD acquisitions are subject to all of the rules contained in the FAR, unless explicitly exempt (Manuel et al., 2015, p. 33).

FAR Part 15 broadly regulates the processes for “competitive and non-competitive negotiated acquisitions” or contracts. FAR Part 15.1 establishes and governs the various source selection practices (i.e., the processes by which federal agencies may legally select bidders or vendors in a competitive bidding environment) that executive federal agencies may use to acquire goods and services. The FAR determines that federal agencies “can obtain [the] best value in negotiated acquisitions by using any one or a combination of source selection approaches” from the “Best Value Continuum” (FAR 15.101). The Best Value Continuum is a spectrum of source selection methods differentiated by the degree to which such methods prioritize cost factors over non-cost factors in a contract award process (see Figure 1). There are at least three key source selection methods along the Best Value Continuum: Subjective Tradeoff, Value Adjusted Total Evaluated Price (VATEP), and Lowest-Price, Technically-Acceptable (LPTA; DoD, 2016, pp. 2–3).

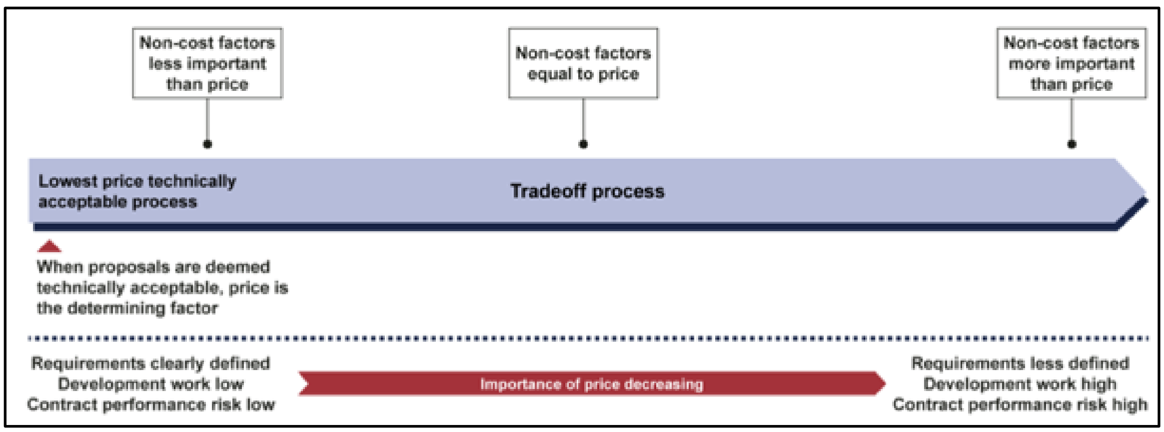


Figure 1. Best Value Continuum for Source Selection Methods
(DiNapoli, 2014, p. 4)

FAR 15.101-2 briefly describes the LPTA source selection process. The complete description contained in the FAR is as follows:

> (a) The lowest price technically acceptable source selection process is appropriate when best value is expected to result from selection of the technically acceptable proposal with the lowest evaluated price.
>
> (b) When using the lowest price technically acceptable process, the following apply:
>
> > (1) The evaluation factors and significant subfactors that establish the requirements of acceptability shall be set forth in the solicitation. Solicitations shall specify that award will be made on the basis of the lowest evaluated price of proposals meeting or exceeding the

acceptability standards for non-cost factors. If the contracting officer documents the file pursuant to 15.304(c)(3)(iii), past performance need not be an evaluation factor in lowest price technically acceptable source selections. If the contracting officer elects to consider past performance as an evaluation factor, it shall be evaluated in accordance with 15.305. However, the comparative assessment in 15.305(a)(2)(i) does not apply. If the contracting officer determines that a small business' past performance is not acceptable, the matter shall be referred to the Small Business Administration for a Certificate of Competency determination, in accordance with the procedures contained in Subpart 19.6 and 15 U.S.C. 637(b)(7)).

(2) Tradeoffs are not permitted.

(3) Proposals are evaluated for acceptability but not ranked using the non-cost/price factors.

(4) Exchanges may occur (see 15.306).

Unlike the FAR, the DFARS does not offer such a description of LPTA, nor does it offer a department-specific definition of LPTA.

Advantages and Disadvantages of LPTA

There are several key advantages and disadvantages associated with LPTA source selection procedures. One of the core advantages of LPTA is that it generally results in procurements with the lowest overall price for products and/or services. Additionally, LPTA provides a clear basis for decision-making because it is a less subjective award process (Gansler, Harrington, & Lucyshyn, 2013, p. 2). As such, it is also one of the quickest ways to equip warfighters with the products and services they need (Kendall, 2015, p. 1). Finally, LPTA is believed to diminish the probability of encountering a bid protest because of the reduced subjectivity in evaluating bidders (Gansler et al., 2013, p. 6).

One of the main concerns surrounding the DoD's use of LPTA is that, in using this source selection procedure, the DoD is pushing vendors or contractors to design their products so cheaply that they cannot afford to design products or plan their services in a way that is outside-of-the-box and potentially more efficient than previous products or service modes. "The downward pressure on price [caused by LPTA] reduces industry's incentive to innovate and may drive quality suppliers entirely out of the defense marketplace as they look for more lucrative opportunities" (Goodman, 2015).

Others in the acquisition community contend that the DoD has increased its use of LPTA for complex procurements and risky acquisitions (Gansler et al., 2013, p. 22). A prime example of this is the Department of the Navy's 2009 LPTA contract award to Hewlett Packard Enterprise Services to replace the Navy's prior network system, Navy/Marine Corps Intranet (NMCI), with a new system called Next Generation Enterprise Network (NGen). The Navy's transition to NGen is a significant and highly complicated undertaking. The Navy has now "delay[ed] the previously scheduled contract award" for the NGen project, "leading one to question the wisdom of using LPTA as a source selection criterion" (Gansler et al., 2013, p. 22).

Another salient concern regarding the DoD's use of LPTA is that LPTA yields poor quality products and services because the DoD is using LPTA more frequently to acquire "higher risk" goods and services (Gansler et al., 2013, p. 22). For example, in one instance in which the DoD issued an LPTA contract for procuring "network equipment for military bases across the country"—a complex project—the "technical evaluators" for the project were forced to select a vendor from a small and less preferable pool of bidders that met the minimum qualifications (Gansler et al., 2013, p. 22). The technical evaluators would have preferred to use non-cost factors to select a vendor, yet they were "required to choose the lowest priced option over one they believed to be a superior proposal that would provide the best value to the government" (Gansler et al., 2013, p. 22).

Background

DoD Policy Guidance on LPTA

Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]), Frank Kendall's publication of Better Buying Power (BBP) 1.0 in June 2010 may have sparked the DoD's increased usage of LPTA (DoD, n.d.). "The common view is that the first version of Better Buying Power's emphasis on lowering costs led the acquisition workforce to interpret the guidance as a preference for LPTA contracts whenever possible" (Serbu, 2017). BBP 1.0 emphasized fiscal austerity; its objectives included "do more without more" and "restore affordability to defense goods and services" (Carter, 2010). The policy guidance does not explicitly advocate for the use of LPTA; however, the efficiency-minded objectives likely steered defense contracting officers in the direction of LPTA (Serbu, 2017).

In March 2015, USD(AT&L) Frank Kendall issued a memorandum entitled *Appropriate Usage of Lowest Priced Technically Acceptable Source Selection Process and Associated Contract Type*. The memo established greater guidance for the DoD's use of LPTA. According to Kendall (2015),

> LPTA is the appropriate source selection process to apply only when there are well-defined requirements, the risk of unsuccessful contract performance is minimal, price is a significant factor in the source selection, and there is neither value, need, nor willingness to pay for higher performance. … LPTA is most appropriate when best value is expected to result from the selection of the technically acceptable proposal with the lowest evaluated price. … [LPTA] has a clear, but limited place in the source selection "best value" continuum. … Whenever the warfighter is willing to pay more for above threshold requirements or performance standards and may benefit from an innovative and technologically superior solution to meet their mission needs, a tradeoff source selection process between cost or price and non-cost factors is optimal. (pp. 1–2)

Much of Kendall's guidance on how the DoD should apply LPTA is reflected in the recent regulatory changes to LPTA.

Recent Regulatory Changes to LPTA

Congress mandated regulations around the DoD's LPTA usage via the National Defense Authorization Acts (NDAAs) for FY 2017 and FY 2018. Specifically, sections 813, 814, 885, and 892 of the 2017 NDAA establish circumstances under which the DoD may use the LPTA source selection procedure. Section 813 requires the Secretary of Defense to revise the DFAR to limit LPTA usage to the following contracting scenarios:

1. When the DoD can clearly articulate criteria for "performance objectives, measures, and standards that will be used to determine acceptability of offers" in a request for proposals (RFP);

2. When the DoD does not realize any additional advantage by “exceeding the minimum technical or performance requirements set forth in the request for proposal”;
3. When the technical requirements of the contract do not call for “subjective judgement … as to the desirability of one offeror’s proposal versus a competing proposal”;
4. When the “source selection authority has a high degree of confidence that a review of technical proposals of offerors other than the lowest bidder would not result in the identification of factors that could provide value or benefit to the Department”;
5. The DoD contracting officer must provide a written justification for their use of LPTA source selection; and
6. The DoD must conclude that the lowest-price proposal “reflects full life-cycle costs, including for operations and support.” (NDAA for FY 2017, 2016, §§ 2270–2271)

Sections 813 further encourages DoD contracting officers to avoid the use of LPTA in the following contracting scenarios:

1. Information technology services, cybersecurity services, systems engineering and technical assistance services, advanced electronic testing, audit or audit readiness services, or other knowledge-based professional services;
2. Personal protective equipment;
3. Knowledge-based training or logistics services in contingency operations or other operations outside the United States, including in Afghanistan or Iraq. (NDAA for FY 2017, §§ 2270–2271)

Finally, Section 813 requires that the Comptroller General report to Congress documenting LPTA usage for contracts with a value greater than $10 million by December 2017 (NDAA for FY 2017, §§ 2270–2271).

Section 814 prohibits the use of LPTA for personal protective equipment (PPE) contracts (NDAA for FY 2017, § 2271). Section 885 calls for an assessment of the bid protest system for DoD contracts and requires the Secretary of Defense to assess and provide data regarding the extent to which the existing bid protest system affects the “decision to use lowest price technically acceptable procurement methods” (NDAA for FY 2017, § 2319). Section 892 of the 2017 NDAA requires the DoD to award audit services and audit readiness service contracts to bidders only using the tradeoff source selection method, not LPTA (NDAA for FY 2017, § 2324).

The 2018 NDAA contained three new provisions governing the DoD’s LPTA usage. Specifically, Sections 822, 832, and part of 874 further regulate how the DoD may use LPTA. Section 822 amends Section 813 of the 2017 NDAA by adding the following paragraphs:

(7) the Department of Defense would realize no, or minimal, additional innovation or future technological advantage by using a different methodology; and

(8) with respect to a contract for procurement of goods, the goods procured are predominantly expendable in nature, nontechnical, or have a short life expectancy or short shelf life. (U.S. House of Representatives, 2017)

Section 822 also reduces the reporting requirement for LPTA threshold from $10 million to $5 million (U.S. House of Representatives, 2017).

Section 832 amends Title X, Chapter 42 of United States Code to include § 2442. Subsection 2442 proscribes the DoD from using LPTA to procure "engineering and manufacturing development contract [for] major defense acquisition program[s]" (MDAPs; U.S. House of Representatives, 2017). Finally, Section 874, subsection (g)(2), limits the use of LPTA for software development and agile acquisitions (U.S. House of Representatives, 2017).

These recent legislative changes underscore the importance and relevance of the LPTA issue. Further, these new laws demonstrate Congress' awareness of the problems associated with LPTA. In the House Armed Services Committee (HASC) Report for the 2017 NDAA, the committee expressed concern that DoD contracting officers have frequently used LPTA source selection inappropriately for procurements such as "electronic test equipment that are very technical in nature and require calibration, repair, and software updates during their life cycle" (U.S. House Armed Services Committee, 2016, p. 183). The report further stated, "These anecdotal examples suggest a more widespread over-use of LPTA processes and contracts that may be having substantial unintended consequences" (U.S. House Armed Services Committee, 2016, p. 183). The Senate Armed Services Committee (SASC) Report for the 2017 NDAA expressed similar concerns and disagreement with the use of LPTA for personal protective equipment. The Senate Committee report also said,

> While LPTA and reverse auction contracting techniques are appropriate for some types of purchases, the committee believes that lowest price is not always the best strategy when quality and innovation are needed. In these cases, the committee believes a best value acquisition approach is more appropriate. (U.S. House Armed Services Committee, 2016, p. 215)

The congressional committees continued to express their concern for better defining LPTA in the 2018 NDAA. In the HASC Report for the 2018 NDAA, the committee wrote that LPTA is a valid source selection criterion for "acquisitions with well-defined and non-complex requirements that are not expected to evolve over the life of a contract." The HASC report further noted its concern that the DoD "continues to use LPTA criteria for other acquisitions, including those for innovative professional services and high-performance technologies" (U.S. House of Representatives, 2017).

GAO Reporting on LPTA

In November 2017, the Government Accountability Office (GAO) released a report at the end of 2017 assessing the DoD's LPTA practices. The report found that, during the first half of 2017, the three military departments, the Air Force, Army, and Navy, "rarely used LPTA source selection procedures for IT and support services contracts valued at $10 million or more" (GAO, 2017). The GAO pulled 781 contracts valued at $10 million or more and identified 133 contract awards within this larger pool of contracts that were for IT and support services. The GAO found that only nine of the 133 IT and support services contracts valued at $10 million or more were awarded on an LPTA basis. The GAO also found that for seven of the nine LPTA contracts identified, contracting officers "determined that the government would not receive a benefit for paying more than the lowest price" and that "LPTA procedures were used, in part, because the requirements were well-defined, non-complex, or reoccurring" (GAO, 2017, p. 10).

In 2014, the GAO found that the DoD's use of LPTA had grown between 2009 and 2013 for contracts valued at $25 million or more. Specifically, between 2009 and 2013, the DoD's LPTA usage had grown from 26% to 36%, respectively (see Figure 2).

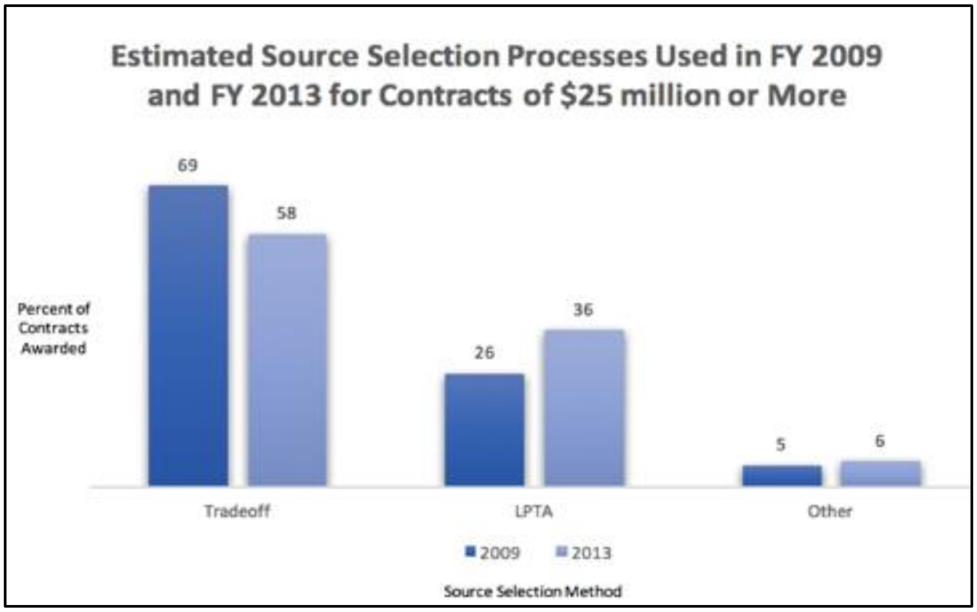


Figure 2. Source Selection Method Frequency in FY 2009 and FY 2013
(DiNapoli, 2014)

The GAO (2017) also found that 45% of contracts with obligation values between $1 million and $25 million were granted on an LPTA basis. These awards were for both "products and services" (p. 8). The GAO (2017) wrote, "We identified relatively few uses of LPTA to acquire high dollar services" and that contracting officials chose source selection methods based on their "knowledge about the requirements and contractors" (p. 8).

In sum, the GAO's 2014 and 2017 reports on the DoD's LPTA usage suggest that the Department has used LPTA marginally to procure complex services. Moreover, the GAO found that in those instances in which the DoD used LPTA to procure complex services, its usage was appropriate. These findings contradict the broader literature's anecdotal findings, which suggest that the DoD has misused LPTA.

Methodology

To answer the central research question, I used a two-phased, mixed methodology approach. The first phase was a simple data collection and comparison of LPTA contracts. In the second phase, I used the Delphi method to collect qualitative data from experts in the Defense acquisition field. For the purposes of this paper and forum, I will focus only on Phase 1, the FBO-FPDS data query.

Phase 1: FBO-FPDS Data Query

I gathered information on LPTA contracts using the Federal Business Opportunities website (FBO.gov) and the Federal Procurement Data System website (FPDS.gov). For this phase, I operationalized the central research question by asking the following subsidiary questions:

1. Has the DoD increased its usage of LPTA contracts over time?
    a. If so, has the DoD increased its usage of LPTA source selection for complex or non-complex procurements?
2. Does LPTA source selection yield poor outcomes for DoD contracts?

a. Is there high dissatisfaction (as measured by contract cancellation rates) for LPTA contracts? If so, is this dissatisfaction greater for LPTA contracts used for complex or non-complex procurements?
b. To what extent is the DoD's usage of LPTA source selection associated with contractor non-performance (as measured by failure to re-award to a vendor after awarding an LPTA contract)?

A "Yes" answer to these questions would indicate that the DoD has not been effectively using LPTA to fulfill its mission.

As I collected data from FBO (See Appendix for further details on FBO data collection), I identified complex and non-complex LPTA contracts. I originally intended to collect data on non-LPTA contracts as well; however, collecting such data yielded tens of thousands of observations (i.e., individual contract award announcements). Due to time and resource limitations, I was unable to collect information on non-LPTA contracts. I identified complex and non-complex LPTA contracts by using the Product and Service Codes (PSCs) search filter in FBO. I collected contract information using the following PSCs:

- **Simple Good:** (51) Hand tools, (74) Office machines, text processing systems & visible record equipment, (75) Office supplies and devices.
- **Complex Good:** (10) Weapons, (17) Aircraft launching, landing & ground handling equipment, and (18) Space vehicles.
- **Simple Service:** (C) Architect and engineering services, (S) Utilities and housekeeping services, and (Z) Maintenance, repair, and alteration of real property
- **Complex Service:** (A) Research and development, (D) Information technology services, including telecommunications services, and (H) Quality control, testing & inspection services.

Using these search criteria and steps in FBO, I constructed four discrete samples based on the complexity of the procurement (i.e., complex vs. non-complex, or high-risk vs. low-risk). Using this strategy allowed me to test whether or not LPTA is increasingly applied for inappropriate types of acquisitions. I collected most of the contract data from FBO and used FPDS (1) to corroborate information found on FBO, (2) to assess whether the contract subsequently won an award with the same buying entity (i.e., from the same service department), and (3) to find out whether the contract was eventually terminated.

Contract terminations and re-awards are appropriate proxy variables for contractor performance because termination and failure to re-award are indicators that the contractor was no longer able to meet the government's needs at the estimated price—they are proxies for "contractor non-performance" (Staff member, Personal communication, October 13, 2017). For example, if a contractor underestimates the cost of labor in their bid, and is unable to find labor at the estimated price after winning a contract, the DoD may terminate the contract as it no longer meets the LPTA standard (Staff member, Personal communication, October 13, 2017).

To the best of my knowledge, there is no other research publicly available that tracks the misuse of LPTA source selection by procurement category. The 2014 GAO tracked LPTA usage by dollar value, which may be a proxy variable for procurement complexity (DiNapoli, 2014). I noted contract award value for all contracts in my data collection, but the focus of my data collection was to identify contracts through a set of pre-determined procurement categories.

Limitations

While FBO is a very useful tool for collecting basic information on government contracts, it is an imperfect data collection system with some noteworthy limitations. First, data collected from FBO does not account for contract renewal fatigue, wherein the government automatically decides not to re-award a contract to the same supplier (Staff member, Personal communication, October 13, 2017). It is also possible that some vendors simply did not seek re-award with the DoD. Second, while my total sample size included over 400 award notices across 12 PSCs, the sample itself represents a small portion of total DoD LPTA contracts and, therefore, may not be generalizable to all LPTA contracts across all procurement categories. Third, it is likely that within each PSC category, there is an inherent range of complexity. For example, with an IT service PSC, IT services may range from front desk help to a new cutting-edge technology. Unfortunately, with the Federal Business Opportunities website, it is not possible to filter potential complexity within PSCs. Finally, because data is manually entered into FBO, there is the potential for human error and inconsistency. In other words, contracting personnel who are entering data into FBO may have done so incorrectly or not thoroughly, and their level of accuracy and thoroughness may vary from year to year. We therefore must assume that the personnel who are entering data for LPTA contracts and non-LPTA contracts are doing so with the same level of accuracy (or inaccuracy) and the same level of thoroughness (Staff member, Personal communication, November 3, 2017).

As stated previously, the data collected for the FBO-FPDS data query portion of this research was derived from award announcements found on FBO.gov using a predetermined set of search criteria. While FBO allows users to search for award announcements that contain the term "LPTA," the results do not necessarily mean that an award was made on an LPTA basis. The award announcement results that FBO generates when using the "LPTA" search term may be a mixture of (1) awards that the service departments granted by using LPTA source selection procedures, or (2) awards that the service departments granted in which the solicitation referenced the LPTA evaluation criteria at one point or another (but didn't necessarily stay that way). The only way to verify whether a contract was solicited and awarded on an LPTA basis is to look within Section M (Evaluation Factors for Award) of a solicitation (Staff member, Personal communication, February 8, 2018). Unfortunately, FBO does not include copies of solicitations used for all contract awards. Further, I was unable to collect and verify the available solicitations in this sample due to time constraints. Therefore, this report assumes that each of the contracts identified was, during at least one point, solicited on an LPTA basis, but may not have been awarded as such.

Data & Analysis

I operationalized the quantitative data collection and analysis into sub-questions. The subsidiary questions and the answers gleaned from the FBO-FPDS Data Query are summarized in Table 1.

Table 2. Summary of FBO-FPDS Data Findings

Subsidiary Question	Key Findings	Effective Usage?
Has DoD increased its usage of LPTA contracts over time? If so, has DoD increased its usage of LPTA for complex or non-complex procurements?	• DoD's LPTA usage for complex/non-complex goods and services rose rapidly after 2011. • LPTA contracts since 2002 have been marginal. • LPTA usage for simple goods/services has generally exceeded LPTA usage for complex goods/services. • DoD's use of LPTA appears to be declining	Yes and No
Is there higher dissatisfaction (as measured by contract cancellation rates) for LPTA? If so, is this dissatisfaction greater for LPTA contracts used for complex or non-complex procurements?	• Of the 467 contracts in the sample, only 13 contract cancellations noted in FPDS. • 10 out of the 13 cancellations were for service procurements.	Yes and No
To what extent is DoD's usage of LPTA source selection associated with contractor non-performance (as measured by failure to re-award to a vendor after awarding an LPTA contract)?	• 261 of the 373 individual awardees in the sample (70%) were subsequently re-awarded a contract with the same department. • Re-award rates were 68% and 75% for simple services and complex services, respectively.	Yes

Sub-Question #1: Has the DoD increased its usage of LPTA contracts over time? If so, has the DoD increased its usage of LPTA for complex or non-complex procurements?

The data collected from FBO suggests that the number of award notices linked to solicitations that evaluated bidders on an LPTA basis at least once during the source selection process has increased over time. The total increase in LPTA criteria usage over time, however, has been marginal. Figure 3 shows that, since 2002, the percent of award notices and solicitations that referenced LPTA at least once represented less than 2.25% of those award notices and solicitations that did not.

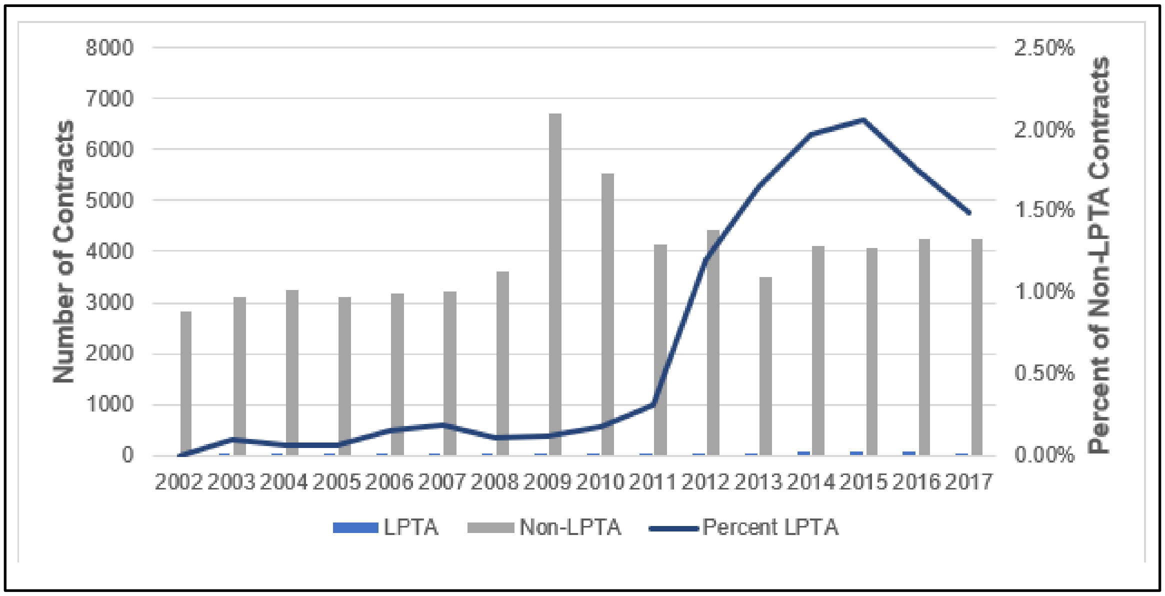


Figure 3. Total Sample of LPTA and Non-LPTA Contracts (2002–2017)
(Federal Business Opportunities website)

While the use of the LPTA evaluation criteria has marginally increased since 2002, it also appears to be declining in recent years. Figure 4 shows that the frequency of award notices and corresponding solicitations that referenced an LPTA evaluation criteria sharply increased between 2011 and 2012, peaked in 2015, and has been declining ever since.

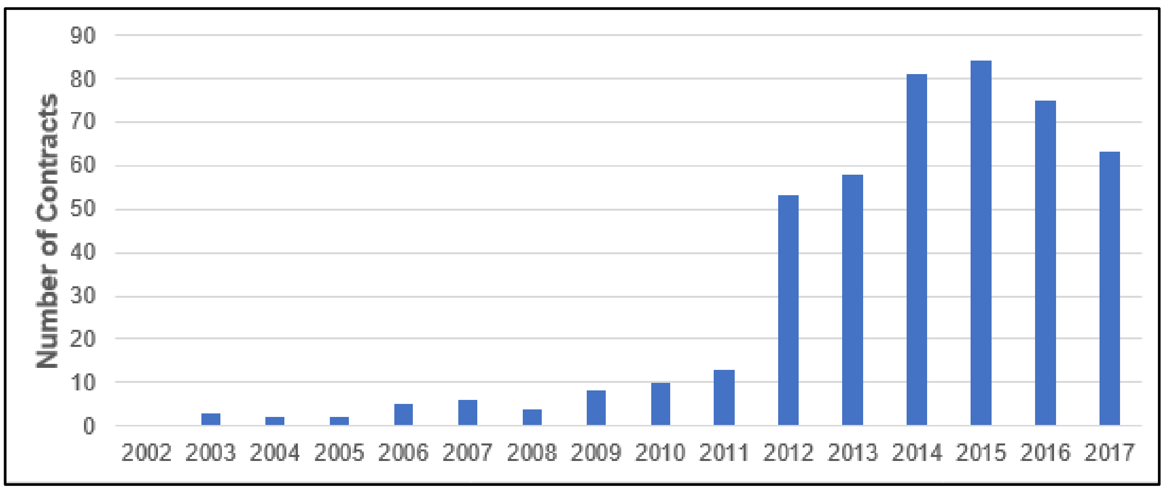


Figure 4. Total Sample LPTA Frequency (2002–2017)
(Federal Business Opportunities website)

There is no discernable trend when comparing the frequency of LPTA awards for simple goods with the frequency of LPTA awards for complex goods. Figure 5 shows that, in total, the service departments awarded only four more contracts referencing an LPTA evaluation criteria for simple goods than for complex goods. In 2015, the service departments granted awards four times more often for complex goods than for simple goods. The year 2015, however, was the only one in which the frequency of LPTA award announcements for complex goods exceeded that of simple goods. By 2017, none of the service departments granted an LPTA-based award for complex goods

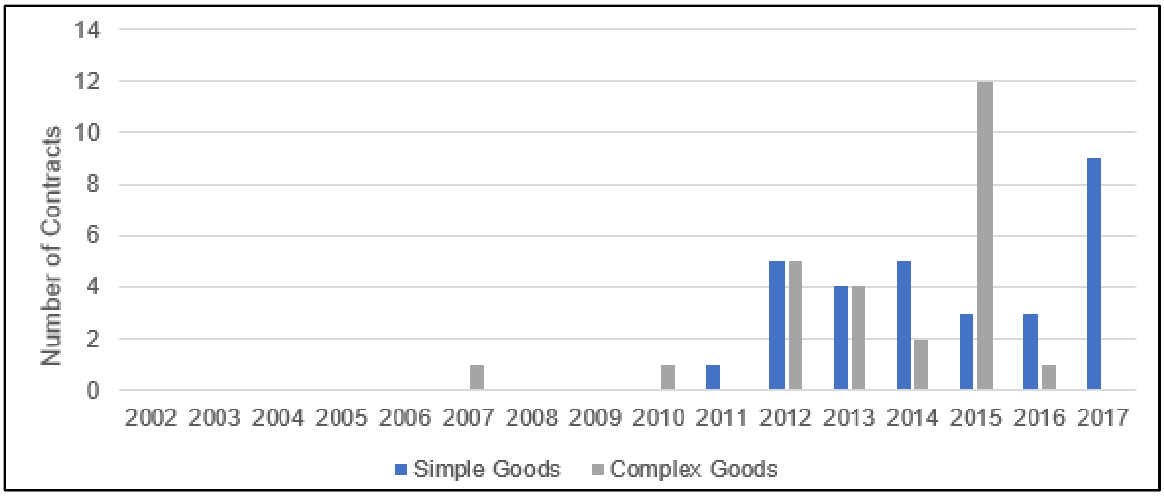


Figure 5. LPTA Award Announcements for Simple Goods vs. Complex Goods (2002–2017)

(Federal Business Opportunities website)

Since 2002, the frequency of award notices for simple services has typically surpassed the frequency of award notices for complex services (see Figure 6). The frequency gap grew sharply in 2012 when the number of award notices for complex services was eight, and the number of award notices for simple services was 35. That said, the number of award notices for complex services gradually increased from four in 2006 to 14 in 2017. It is important to note that the number of award announcements for complex services peaked in 2016 at 29 award notices.

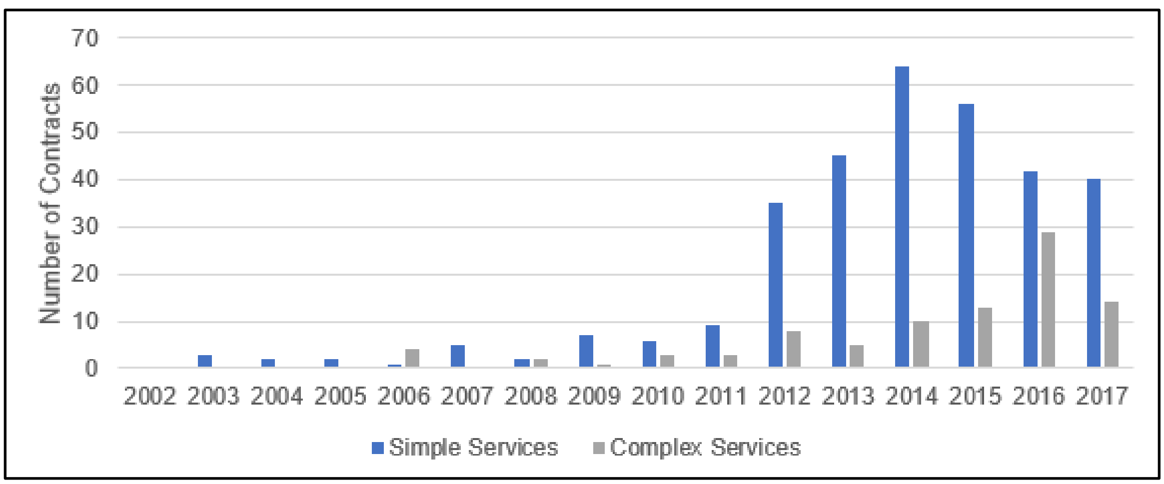


Figure 6. LPTA Award Announcements for Simple Services vs. Complex Services (2002–2017)

(Federal Business Opportunities website)

In sum, the DoD has increased its LPTA usage over time. LPTA usage increased sharply between 2011 and 2012. This is unsurprising given that this occurred shortly after Frank Kendall issued BBP 1.0 in 2010. The DoD's use of LPTA to procure complex goods and complex services also rose markedly after 2010 but has generally been lower than the DoD's use of LPTA to procure non-complex goods and services. With the understanding that using LPTA to procure complex goods and services is an ineffective or inappropriate usage of LPTA, the data so far suggests that the DoD has not been using LPTA effectively. However, it appears that the DoD typically uses LPTA more frequently for appropriate procurements than inappropriate procurements. Further, the DoD's inappropriate usage of LPTA appears to be declining.

Sub-Question #2a: Is there higher dissatisfaction (as measured by contract cancellation rates) for LPTA? If so, is this dissatisfaction greater for LPTA contracts used for complex or non-complex procurements?

From the total sample of 467 discrete award announcements collected from FBO, there were only 13 contract cancellations noted in FPDS. This represents 2.7% of the entire sample of LPTA award announcements. The sub-sample with the greatest frequency of contract cancellations was the simple services sample, which had seven contract cancellations. The sub-sample with the second highest frequency of contract cancellations was the complex services sub-sample, which included four contract cancellations (see Figure 7). Ten of the total contract cancellations identified were terminated for convenience: One was terminated for default, one was a legal contract cancellation, and one was a terminate for cause.

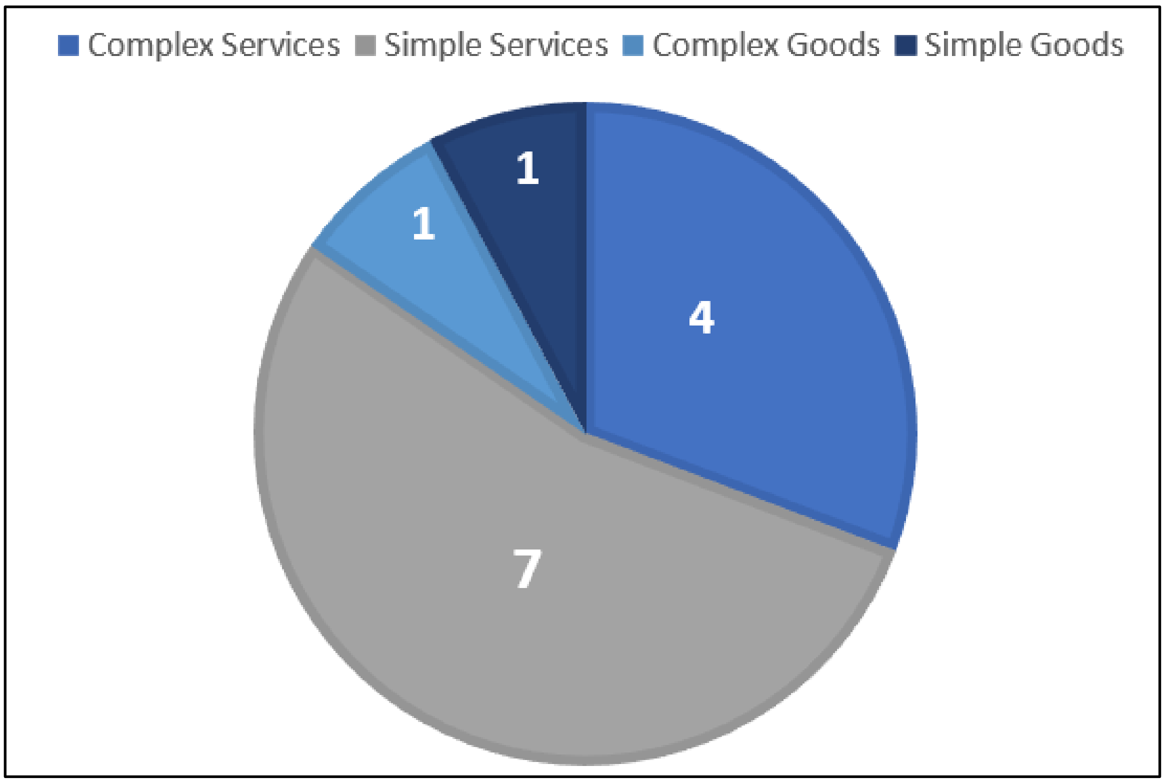


Figure 7. Frequency of Contract Cancellations by Sub-Sample
(Federal Procurement Data System)

The two sub-samples with the highest contract cancellation frequency were for service procurements. Because the number of contract cancellations captured in this sample was so small, this data only marginally supports the notion that LPTA, when used to procure services, yields greater dissatisfaction than when the DoD uses LPTA for goods. Therefore, when the DoD uses LPTA for service procurement, it is not rapidly equipping warfighters, and therefore, is compromising its ability to fulfill its mission.

Sub-Question #2b: To what extent is the DoD's usage of LPTA source selection associated with contractor non-performance (as measured by failure to re-award to a vendor after awarding an LPTA contract)?

Of the 373 individual awardees in the sample, 261 of the awardees were subsequently re-awarded a contract with the same service department (see Figure 8). This represents approximately 70% of the total sample. Meanwhile, 112 of the awardees were not re-awarded a contract with the same service department, which equates to roughly 30% of the total sample. Re-award rates for vendors who delivered a complex service or a complex good on an LPTA basis were higher than that of the total sample. Re-award rates were 75% and 77% for complex services and complex goods, respectively. Re-award rates were lower for simple services and simple goods, 68% and 69%, respectively.

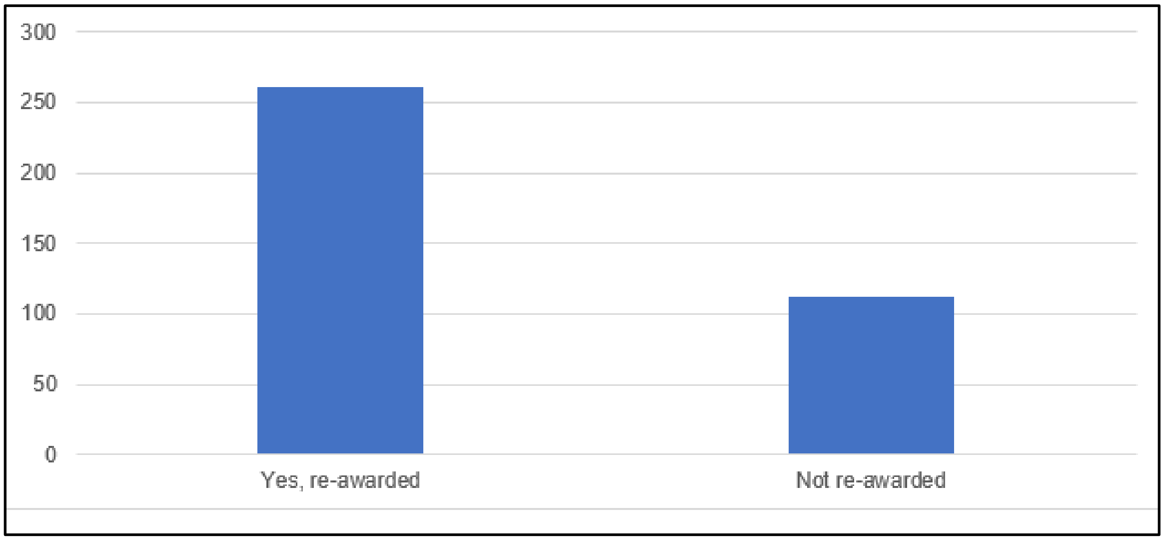


Figure 8. Frequency of Re-Award Post-LPTA Award
(Federal Procurement Data System)

These figures may overestimate the re-award and non-re-award rates because different contract awards were sometimes granted to the same awardee. Also, because the sample lacks information on contract re-awards for non-LPTA contracts, it is unclear whether these re-award rates are normal or abnormal. Objectively, however, the re-award rate for LPTA contracts appears to be moderate, and the re-award rates for vendors who delivered services under an LPTA-based contract were high. This could suggest that the DoD is satisfied with contractor performance after awarding on an LPTA basis.

In sum, the answers to the subsidiary questions are mixed. The data suggests that the answer to sub-question 1 is both yes and no because the DoD has increased its usage of LPTA over time for both complex and non-complex services, but it has generally used LPTA more to acquire simple goods and services. Further, the Department's LPTA usage is marginal when comparing the number of award notices that reference LPTA to those that do not. The answer to sub-question 2a is also mixed because the number of contract cancellations represents less than 3% of the entire sample. At the same time, however, 10 of the 13 contract cancellations were for service procurements. Finally, the answer to sub-question 2b is definitively yes. Seventy percent of the LPTA awardees in the sample were subsequently re-awarded a contract with the same service department. While it is unclear what contract re-award rates are for non-LPTA contracts, the re-award rates observed in this study are objectively high.

Subsidiary Findings: Data Issues

A subsidiary but salient theme emerged through this research concerning the dearth of concrete, publicly accessible data on the DoD's use of LPTA. As discussed within the methodology section, monitoring the DoD's usage of LPTA through platforms such as FBO and FPDS presents its own challenges because data is oftentimes unclear and is subject to variability. Because many contracting officers and contracting personnel are entering information into FBO, researchers must assume that contracting personnel are doing so with different rates of thoroughness and accuracy (Staff member, Personal communication, November 3, 2017).

Assumptions about the accuracy of data entered into FPDS seem to be warranted in light of a 2003 GAO report highlighting incomplete data and accuracy issues within FPDS. Specifically, the GAO wrote,

> FPDS has been the federal government's central database of information on federal procurement actions since 1978. Congress and executive branch agencies rely on FPDS to assess the impact that governmentwide acquisition policies and processes are having on the system generally, as well as with respect to specific geographical areas, markets, and socio-economic goals. Yet despite the importance of the data, we continue to find that FPDS data are inaccurate and incomplete. Although we have not fully assessed the extent of reporting errors, we have found sufficient problems to warrant concern about the current reliability of FPDS information. (Woods, 2003, p. 1)

In 2009, the GAO reported that the accuracy of FPDS has improved due to the rise in electronic data submissions. The GAO noted that "the quality of some FPDS-NG data remains a concern" (Woods, 2009).

FPDS does not track data on source selection procedures. FBO is one of the only public interfaces that allows the public to use a general search filter to specifically identify large amounts of LPTA contract award announcements. Even still, the reliability of the results is questionable. Oftentimes, when closely evaluating a single contract resulting from an LPTA search in FBO, the evaluation criteria portion of the solicitation, including the LPTA criteria, was crossed out without explanation. This made it unclear to the researcher whether the award was granted on an LPTA basis or not.

Another issue concerning LPTA data was the dearth of contract cancellation information. Having such information could allow researchers and government officials to have a much better understanding about whether LPTA contracts yield successful results for the government. However, locating contract cancellation data in either FBO or FPDS is exceedingly difficult. In 2016, the GAO noted that "the FAR does not require contracting officials to publicize notices of canceled solicitations"; however, "officials may post cancelations [at will] on FBO" (Woods, 2016). This indicates that the data collected for this report significantly underestimates the number of contract cancellations. Similarly, the GAO wrote, "No information was available [in FPDS] on canceled solicitations as only awarded contract actions are recorded in the system" (Woods, 2016). FPDS does have a feature that allows users to track the value of obligated dollars associated with contract terminations. Further, FPDS allows users to track these values by termination category (i.e., Legal Contract Cancellation, Terminate for Cause, Terminate for Convenience [Complete or Partial], and Terminate for Default [Complete or Partial]). Aside from this, though, FPDS does not offer a simple, mechanized way to track cancellations by contract. The only apparent way to find information on contract cancellations in FPDS (i.e., the method used in this research project) is to manually open each of the "modifications" associated with a contract to check for termination. This method can be challenging, however, because some contracts have 400 or more modifications, creating an undue time and resource burden for the researcher or user.

Recommendations

The FBO-FPDS data query revealed that the DoD's LPTA usage for complex and non-complex products and services has risen significantly since 2011. The overall number of LPTA contracts since 2002 has been marginal, and the DoD's LPTA usage for simple goods and services has generally exceeded LPTA usage for complex goods and services. Ten out of the 13 contract terminations identified were for service procurements, and 67%of the awardees who were granted an LPTA contract were subsequently re-awarded a contract by the same DoD service department. The lack of reliable data in FBO and FPDS not only calls some of these results into question, but it also underscores the need for greater attention on improving the DoD's data collection and monitoring.

Based on the information gleaned through the FBO-FPDS data query, as well as the subsidiary findings, the need for data collection and data management reform is evident. While anecdotal information and qualitative literature on the DoD's LPTA usage acknowledges that the Department's use of LPTA has been inappropriate, testing or corroborating those findings is virtually impossible due to the lack of concrete, reliable data on the DoD's source selection practices. Therefore, I recommend that the United States Congress, DoD, Office of Defense Procurement Acquisition Policy (DPAP), and Defense Acquisition Regulation Council (DAR Council) work jointly to take the following actions:

- **Recommendation 1: Update the DFARS to require defense contracting officers to enter information on cancelled contracts into FPDS.**
- **Recommendation 2: Update the DFARS to require defense contracting officers to enter information on source selection methods into FPDS.**

Because the FAR does not require agencies to track contract termination data, it is likely that this report has severely underestimated the number of contract terminations associated with LPTA-based awards. Therefore, Congress must mandate regulations—and DPAP and the DAR Council must help implement those regulations—that require DoD contracting officers to record and monitor contract terminations (and the associated reasons) in FPDS. Additionally, the FAR does not require agencies to track source selection data. Concrete source selection data is needed in order to better understand the DoD's source selection practices. Having such data would allow the DoD to definitively test or corroborate the wealth of anecdotal information suggesting that the Department's use of LPTA has been inappropriate, and whether its usage of LPTA has supported the DoD's mission and acquisition interests. Any ongoing or further LPTA reform efforts will continue to be of questionable value until the information in FBO or FPDS is more reliable.

In its 2003 report on FPDS data, the GAO aptly wrote, "Reliable information is critical to informed decision making and to oversight of the procurement system" (Woods, 2003, p. 1). Effecting necessary change is impossible without access to the right metrics and accurate data. The lack of clear and consistent data on the DoD's source selection practices, including contract cancellations, point to a transparency crisis. Having such data could help policymakers better understand the context around the DoD's LPTA usage and the extent to which the DoD has misused LPTA, if at all.

Conclusion

Further research is needed to further confirm or supplement the data findings presented in this report. One way to do this would be to interview defense acquisition experts to identify more examples of LPTA misuse and to continue to build a consensus around how the DoD could improve its LPTA practices. Another step that could improve the validity of the data used in this report would be to locate the solicitations of the contracts gathered from FBO to verify that they were awarded on an LPTA basis.

The DoD is currently facing a transparency crisis with respect to its source selection practices. Scholars, members of Congress, and members of the defense industrial base acknowledge that the DoD's use of LPTA is harmful toward industry and threatens to undermine the Department's mission. Concurrently, however, there is very little data to support their findings. Consequently, defense acquisition stakeholders are not on the same page regarding the breadth and depth of the LPTA problem. Mandating more and better data collection could improve our collective understanding of the DoD's use of LPTA, and it could help mitigate the potential negative effects of LPTA on the DoD and industry.

References

Carter, A. B. (2010, June 28). *Better Buying Power: Mandate for restoring affordability and productivity in defense spending.* Retrieved from http://bbp.dau.mil/docs/Better%20Buying%20Power--Mandate%20for%20Restoring%20Affordability%20and%20Productivity%20in%20Defense%20Spending.pdf

DiNapoli, T. (2014). *Defense contracting: Factors DOD considers when choosing best value processes are consistent with guidance for selected acquisitions* (GAO-14-584). Retrieved from https://www.gao.gov/products/GAO-14-584

DoD. (n.d.). *Better Buying Power*. Retrieved March 18, 2018, from http://bbp.dau.mil/background.html

DoD. (2016, April 1). *Department of Defense source selection procedures*. Retrieved from https://www.acq.osd.mil/dpap/policy/policyvault/USA004370-14-DPAP.pdf

DoD. (2017, January 27). About the Department of Defense (DoD). Retrieved March 17, 2018, from https://www.defense.gov/About/

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2018).

Gansler, J. S., Harrington, L. H., & Lucyshyn, W. (2013). *The DoD's use of lowest price technically acceptable (LPTA) price selection.* College Park, MD: UMD Center for Public Policy and Private Enterprise. Retrieved from http://www.cpppe.umd.edu/publications/dod%E2%80%99s-use-lowest-price-technically-acceptable-lpta-price-selection

GAO. (2017). *Defense contracting: DOD's use of lowest price technically acceptable source selection procedures to acquire selected services* (GAO-18-139). Retrieved from https://www.gao.gov/products/GAO-18-139

Goodman, W. (2015). Lowest price technically acceptable: Overrated, overused? *Defense AT&L, 44*(2). Retrieved from http://link.galegroup.com.proxy.lib.duke.edu/apps/doc/A409048975/ITOF?u=duke_perkins&sid=ITOF&xid=1dea4bda

Kendall, F. (2015, March 4). Appropriate use of the lowest priced technically acceptable source selection process and associated contract type. Retrieved from http://acqnotes.com/wp-content/uploads/2015/01/Appropriate-Use-of-LPTA-4-Mar-2015.pdf

Manuel, K. M., Halchin, L. E., Lunder, E. K., & Christensen, M. D. (2015). *The Federal Acquisition Regulation (FAR): Answers to frequently asked questions.* Washington, DC: Congressional Research Service.

National Defense Authorization Act for Fiscal Year 2017, Pub. L. No. 114–328, 130 Stat. 2000 (2016).

Schwartz, M., & Manuel, K. M. (2015). *GAO bid protests: Trends and analysis*. Washington, DC: Congressional Research Service. Retrieved from https://fas.org/sgp/crs/misc/R40227.pdf

Section 809 Panel. (2017). *Advisory Panel on Streamlining and Codifying Acquisition Regulations: Section 809 Panel interim report*. Arlington, VA: Author. Retrieved from http://docs.house.gov/meetings/AS/AS00/20170517/105973/HHRG-115-AS00-Wstate-LeeM-20170517.pdf

Serbu, J. (2017, February 28). The legacy of Better Buying Power: DoD's gambit to reform acquisition from within. Retrieved from https://federalnewsradio.com/defense/2017/02/bbpndaa-special-report-part-1/

U.S. House Armed Services Committee. (2016). *Committee report for the National Defense Authorization Act for Fiscal Year 2017*. Retrieved from https://www.congress.gov/114/crpt/hrpt537/CRPT-114hrpt537.pdf

U.S. House Armed Services Committee. (2017). *Committee report for the National Defense Authorization Act for Fiscal Year 2018*. Retrieved from https://www.congress.gov/115/crpt/hrpt200/CRPT-115hrpt200.pdf

U.S House of Representatives. (2017). *FY 2018 NDAA conference report to accompany H.R. 2810.* Retrieved from https://www.congress.gov/115/crpt/hrpt404/CRPT-115hrpt404.pdf

U.S. Senate Armed Services Committee. (2016). *Committee report for the National Defense Authorization Act for Fiscal Year 2017.* Retrieved from https://www.congress.gov/114/crpt/srpt255/CRPT-114srpt255.pdf

Woods, W. T. (2003). *Reliability of Federal Procurement Data* (GAO-04-295R). Retrieved from https://www.gao.gov/products/GAO-04-295R

Woods, W. T. (2009). *Federal contracting: Observations on the government's contracting data systems* (GAO-09-1032T). Retrieved from https://www.gao.gov/assets/130/123442.pdf

Woods, W. T. (2016, June 29). *Defense contracting: Complete historical data not available on canceled DOD solicitations*. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/1013843.pdf

Appendix: FBO Data Collection Steps

The steps I took to collect data from FBO were as follows:

1. Open the “Advanced Search” form on FBO.
2. Select “Archived” under the “Documents to Search” field.
3. Select “Award Notice” under the “Opportunity/Procurement Type” field.
4. Select “Specific Agency/Office/Locations” under the “Agency/Office/Locations” field.
5. Select the Department of the Air Force, Department of the Navy, and Department of the Army under the “Specific Agencies/Offices” field.
6. Type “LPTA” into the “Keywords or SOL#” field.
7. Select the relevant, pre-determined Product and Service Codes (PSCs).
8. Select a search date range in the “Posted Date Range” field. The date range selected for all searches was from the earliest record through December 31, 2017.
9. Click “Search.”
10. Repeat steps #1–#9 and change PSCs.

Disclaimer

A Closer Look at Bid Protests in the Department of Defense

Jack Chutchian—is a second year Master of Public Policy candidate at the Sanford School of Public Policy at Duke University. He graduated from Northeastern University with a BA in political science and international affairs before spending two years in Huntsville, AL, with Teach for America. Upon graduation in May 2018, he will use his policy analysis education to work for a consulting firm in Washington, DC. [jzc6@duke.edu]

Abstract

Congress tasked the Section 809 Panel with streamlining the acquisition process at the Department of Defense (DoD; Section 809 Panel, 2018). Streamlining the acquisition process should make government more efficient and attract new business partners. The notion of large private contractors excessively protesting at the Government Accountability Office (GAO) is a potential barrier to entry for new businesses.

This paper will explore congressional attempts to limit protests filed at the GAO through reforms in the National Defense Authorization Act for Fiscal Year 2018. The reform instituted a three year “losers pay” pilot program that requires contractors with unsuccessful protests to reimburse the DoD for costs incurred in processing certain protests. Based on my analysis of the data collected and a concurrent study by the RAND Corporation, this reform will not achieve the desired result of streamlining the acquisition process.

Instead, the data and information gathered from the GAO, Federal Business Opportunities, and stakeholder interviews suggest that condensing protests into a singular review would streamline the acquisition process more effectively. Companies sometimes file multiple protests at the GAO regarding a single solicitation to obtain information about why they lost the bid. This policy alternative will make the federal government transparent for current contractors and potential partners.

Introduction

The Section 809 Panel was tasked by Congress in the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2016 to streamline and improve the defense acquisition process. The panel is divided into 10 teams. Team Four addresses “Barriers to Entry.” This team evaluates and removes regulatory, cultural, or bureaucratic barriers to entry at the Department of Defense (DoD) marketplace (Section 809 Panel, 2018).

The Section 809 Panel proposes that the current bid-protesting environment may prevent new companies from conducting business with the DoD. Subsequently, bid-protest reforms became a policy initiative of Team Four “Barriers to Entry.” Administrative costs and the loss of time connected to contract protests may outweigh the potential benefits of new companies bidding on and securing DoD contracts. This paper will analyze the protesting landscape at the Government Accountability Office (GAO) regarding DoD contracts and make appropriate recommendations based on the findings of this and other scholarly research.

Policy Question

This paper addresses the following policy questions: Considering the passage of Government Accountability Office (GAO) Bid Protest Reforms in the FY 2018 NDAA, what is the potential effect on protests of DoD contract awards? Furthermore, are there other policy options that the Section 809 Panel should consider outside of the FY 2018 NDAA recommendations? And if so, how would these alternatives alter the protest process to make it easier for private contractors to do business with the federal government?

Background

In this section, I will address the original intent of contract protests and current trends in the contract protest at the GAO, and I will summarize the RAND Corporation findings. This section will not cover court systems' protest history nor current trends because there is no uniform procedure in which the courts adjudicate bid protests.

Original Intent of the Contract Protest

Bid protesting to the federal government began in the 1920s shortly after the creation of the GAO. Losing companies wrote to the agency complaining that government contracts had been unjustly awarded to competitors and petitioned the GAO to review and mitigate the dispute (Gordon, 2013). The GAO delivered its first decision in 1926 when the agency intervened in a trucking contract regarding the Panama Canal. The Autocar Sales and Service Company petitioned the GAO to review a contract awarded to Federal Motor Truck Company. The contract stipulated certain truck features unique to Federal Motor Truck Company vehicles and the Autocar Sale and Service Company sought remedies for the unfair favoritism (GAO, 2017). Federal Motor Truck Company kept the contract, yet the GAO's review and decision prevented future contracts from containing specific modifications that favored one private company.

Since the original Panama Canal bid protest presented to the GAO in 1926, several laws have attempted to funnel the protest process into a single government agency or system. However, these laws always included a sunset clause or were replaced by new laws with different methods of bid protest. While the GAO has accepted and reviewed bid protests since the 1920s, the current definition of a bid protest was not formalized until the Competition in Contracting Act (CICA) of 1984. Federal Agencies amended the Federal Acquisition Regulation (FAR) in 1995 to reflect CICA. FAR Part 33 Section 101 Definitions contains the four conditions in which a losing bidder may provide a written objection:

(1) A solicitation or other request by an agency for offers for a contract for the procurement of property or services.

(2) The cancellation of the solicitation or other request.

(3) An award or proposed award of the contract.

(4) A termination or cancellation of an award of the contract, if the written objection contains an allegation that the termination or cancellation is based in whole or in part on improprieties concerning the award of the contract. ("Rules and Regulations," 1995)

Protests provide businesses a valuable mechanism that serves two purposes. First, protests allow entities doing business with the government to air their grievances about government contracting processes and seek relief. Without this mechanism, private entities may be less inclined to do business with the federal government. Second, the protest mechanisms hold procurement officials and government agencies accountable by highlighting and correcting inaccuracies in the bid process (Manuel & Schwartz, 2016).

CICA established the GAO as the venue for contractors to address perceived errors in the bid process by all federal agencies. Upon a protest receipt to the GAO, the law stipulated that activities pertaining to the awarded contract and within the scope of the legislation must be suspended until the GAO rules on the case. Federal agencies may override the stay if agencies determine urgent and compelling circumstances will not permit waiting for the GAO's decision, or if performance of the contract is in the best interests of the United States (Manuel & Schwartz, 2016). Originally, the law established a deadline of 90 working days for the GAO to issue a ruling on a protest. Newer laws extended the deadline to 100 days. Furthermore, the awarding agency (DoD, Energy, etc.) may take up to 60 days to administer the GAO's recommendations (Manuel, 2011).

Current Trends

The White House recently confirmed the need for acquisition reform in the President's Management Agenda for 2018. The agenda indicated a problem with acquisition managers who are more concerned with compliance than best business practices.

> Major acquisitions (over $50 million) often fail to achieve their goals because many federal managers lack the program management and acquisition skills required to successfully manage and integrate large and complex acquisitions into their projects. These shortcomings are compounded by complex acquisition rules that reward compliance over creativity and results. (President's Management Council, 2018)

This sentiment will be discussed later while analyzing stakeholder interviews from government officials and private companies.

Prior to FY 2018, reports from the Congressional Research Service, the GAO, and others described trends in the *number* of protest cases brought to the GAO by the DoD and other government agencies. However, little information existed about the cost of these protests. The sheer number of protests in recent years has doubled, specifically for the DoD. Protests regarding DoD contracts increased from approximately 600 in FY 2001 to 1,200 in FY 2015 (Manuel, 2011). The GAO did not necessarily issue a decision on each DoD bid-protest. However, each protest carried the possibility of an operations delay which could negatively impact the business and the warfighter.

Some analysts within the government and defense industries suggest that the mere threat of a bid protest affects both the bidding business and the awarding agency business practices and workflow. Maser and Thompson (2010) assert that "rejected offerors have incentives to threaten to protest as a way to obtain a percentage of the award as a subcontractor to the winner, or to obtain a settlement payment from the agency to avoid a protest." However, little statistical evidence supports the threat of protesting claim, and it will, therefore, receive limited attention in the following research. Some reports suggest that "contractors knowingly file frivolous protests with GAO in order to harass their competitors and delay awards, or in the hopes of obtaining short-term contracts from the government while the GAO is reviewing the protest" (Defense Industry Daily, 2010).

The U.S. Air Force's CSAR-X contract awarded to Boeing to supply its H-47 Helicopter in 2006 illustrates some of the cost and problems associated with excessive protests. Two of the losing bidders, Sikorsky and Lockheed Martin, filed protests with the GAO over the Boeing award. The GAO sustained the protests and recommended a re-compete of the CSAR-X program. More protests followed this contract over the next two years. Protests challenged the Air Force's compliance with the GAO's recommendations and for fuel-cost specifications in the Request for Proposal, among others. The Pentagon's Inspector General began an audit into "key performance parameters" in the CSAR-X RFP in

March 2008. Due to the Pentagon audit's adverse findings, Defense Secretary Gates decided to cancel the program "for convenience" in April 2009. Commander of the Air Force Material Command, General Bruce Carlson estimated that the GAO protest process cost the USAF $800 million in the CSAR-X RFP case (Defense Industry Daily, 2010). It is unclear how General Carlson calculated his estimate, and this paper challenges his estimate later on.

RAND Corporation Findings

The RAND Corporation published a study in February 2018 regarding bid-protest reform in the federal government titled *Assessing Bid Protests of U.S. Department of Defense Procurements* (Arena, 2018). Their research into protests was much broader than my narrow GAO focus. Their four research questions were as follows:

1. When bid protests are filed, what is the nature and value of these contracts, and what is their share of total defense procurement contract dollars?
2. What are the outcomes of bid protests?
3. How do protesters perceive post-award debriefings in which the reasons for the contract award are explained?
4. When a protester is successful, how often is voluntary corrective action taken by the DoD contracting agency? (Arena, 2018)

The report produced three key findings. First, RAND discovered that despite a steady increase in bid protests filed by DoD contractors, the overall number of protests remains small. Second, RAND found that DoD agencies and the private sector had differing views on the bid protest process. Last, RAND noted the different trends between the GAO and court-filed protests (Arena, 2018). I intentionally omitted court-filed protests in my research. The GAO has nearly a century of policies in place that govern their bid-protest process, and their data is relatively accessible to the public. Omitting court filed bid protests will not take away from the research question nor the solutions provided. Furthermore, courts have multiple appeals systems which companies can file protests within. Retrieving data from multiple court systems was too large in scope for this project.

RAND recommended six courses of action for the federal government. Five of the recommendations are most applicable to my research:

- Enhance the quality of post-award debriefings. Improved debriefings will give disappointed bidders a better understanding of the evaluation and award process and help them better analyze potential protest grounds before filing a protest.
- Be careful in considering reductions to the timeline for resolving bid protests filed with the U.S. Government Accountability Office (GAO) from 100 to 65 days. Seventy percent of protest cases are resolved within 60 days, but complex decisions typically take close to the 100-day limit.
- Be careful in considering restrictions on task-order protests. These protests are more likely to be sustained or involve corrective action, so they may fill an important role in improving the fairness of DoD procurements.
- Consider implementing an expedited process for adjudicating protests involving contracts valued under $0.1 million. The costs to adjudicate

> these protests under the current system may exceed the value of the procurements. (Arena, 2018)

Analysis of FY 2018 NDAA

A *Washington Post* article on October 10, 2017, reported on a new provision of the Senate's version of FY 2018 NDAA that would institute penalties for unsuccessful protests on companies with annual revenues over $100 million (Davenport, 2017). The law would encompass every DoD bid and subsequent protest filed through the GAO. The DoD is by far the largest contracting agency within the federal government which is why the Senate included the protest reform targeting the DoD within its version of the FY 2018 NDAA. The penalties were intended to dissuade large companies with annual revenues over $100 million from excessive protesting at the GAO. The theory was that if a large company was fined for losing every protest it filed, companies would be more selective when filing future protests. The GAO would subsequently spend less time and money reviewing the smaller protests.

The proposed Senate reform may have caught the attention of defense companies like Boeing, Lockheed Martin, Raytheon, Northrop Grumman, General Dynamics, and other defense contractors whose annual revenues exceed $100 million. Senate staffers and others working on protest reform believed that these companies' wealth would encourage them to protest more frequently in absence of any penalty (Contractor, personal communication, March 19, 2018). The relevant wording of Section 821 of the Senate's version of FY 2018 NDAA is presented below:

> (a) Payment Of Costs For Denied Protests.—
>
> (1) IN GENERAL.—A contractor who files a protest described under paragraph (2) with the GAO on a contract with DoD shall pay to DoD costs incurred for processing a protest at the GAO and the DoD.
>
> (2) COVERED PROTESTS.—A protest described under this paragraph is a protest—
>
> (A) all of the elements of which are denied in an opinion issued by the GAO; and
>
> (B) filed by a party with revenues exceeding $100 mil during the previous year.
>
> (b) Withholding Of Payments Above Incurred Costs Of Incumbent Contractors.—
>
> (1) IN GENERAL.—Contractors who file a protest on a contract on which they are the incumbent contractor shall have all payments above incurred costs withheld on any bridge contracts or temporary contract extensions awarded to the contractor as a result of a delay in award resulting from the filing of such protest.
>
> (2) DISPOSITION OF WITHHELD PAYMENTS ABOVE INCURRED COSTS.—
>
> (A) RELEASE TO INCUMBENT CONTRACTOR.—All payments above incurred costs of a protesting incumbent contractor withheld pursuant to paragraph (1) shall be released to the protesting incumbent contractor if—

(i) the solicitation that is the subject of the protest is cancelled and no subsequent request for proposal is released or planned for release; or

(ii) if the Government Accountability Office issues an opinion that upholds any of the protest grounds filed under the protest.

(B) RELEASE TO AWARDEE.—Except for the exceptions set forth in subparagraph (A), all payments above incurred costs of a protesting incumbent contractor withheld pursuant to paragraph (1) shall be released to the contractor that was awarded the protested contract prior to the protest.

(C) RELEASE TO DoD IN EVENT OF NO CONTRACT AWARD.—Except for the exceptions set forth in subparagraph (A), if a protested contract for which payments above incurred costs are withheld under paragraph (1) is not awarded to a contractor, the withheld payments shall be released to the DoD and deposited into an account that can be used by the Department to offset costs associated with GAO bid protests. (S. 115-125, 2017)

Section 821 was added by the Senate after the House passed its own NDAA without the protest reform language.

The GAO, Congress, and the DoD repeatedly called for protest reform through speeches and reports over the last decade. However, most defense contractors criticized the proposed legislative change while acknowledging only some companies would be affected.

Steven Koprince is a managing partner at Koprince Law LLC and primary contributor to SmallGovCon, a legal news and notes outlet for small government contractors. He identified potential pitfalls with the Senate reform in a July 2017 article. Koprince questioned whether a problem of "frivolous" protest existed. He noted that over 99% of contracts are not protested and that even though protesters have the burden of proof, overall GAO protests are successful roughly 40% of the time (Koprince, 2017). However, Koprince failed to separate the DoD from other government agency protests which could illustrate a different environment of protesting.

Koprince also failed to elaborate on what constitutes a "successful" protest at the GAO. There are five decisions the GAO can issue for a protest: Granted, Sustained, Withdrawn, Dismissed, or Denied. If he considers "successful" protests as those which the GAO issued a final ruling in favor of the protestor (Sustained or Granted), then my data will show a significantly lower success rate than 40%. His definition of success is further muddled by the government's ability to take "corrective action." When the government takes corrective action to address an error or misstep, many protests which are then "Dismissed" or "Withdrawn" could be considered a success. "Denied" protests go through the full protest process and their grounds are rejected in the final decision by the GAO (Arena, 2018).

Koprince wrote the article during the Senate sessions on the NDAA in the summer of 2017. He identified a serious problem with the legislation; the bill would allow the protesting company to recover its profits under two circumstances: if the solicitation in question was cancelled, or if the GAO issued an opinion to uphold any of the protest grounds filed under the protest. These caveats could have allowed protestors to rescind their protests on the last day or recoup their money if just one small piece of the protest was upheld. While the reconciled bill eliminated withholding profits, Koprince's other reservations about the reform still hold true. Even considering the flaws highlighted by Koprince and others, Section 821 of the Senate's FY 2018 NDAA was the first attempt at fixing a perceived problem with bid protests in the federal government.

On November 16, 2017, the NDAA left conference committee with a diluted protest reform provision. The final version of the NDAA for FY 2018 submitted to the President for signature contained three major differences from Section 821 of the Senate version of the bill. First, the resolved bill changed the "loser pays" protest reform from a law affecting all companies with revenue exceeding $100 million into a three-year pilot program beginning in 2019 applicable only to companies with revenues over $250 million. Second, the reconciled bill no longer withheld profits from incumbent contractors who protest contract awards. Third, the bill eliminated the accelerated GAO decision time frame. The GAO's 100-day decision deadline remained intact (Lasky, 2017). Section 827: Pilot Program on Payment of Costs for Denied Government Accountability Office Bid Protests of the final FY 2018 NDAA is presented below:

> (a) PILOT PROGRAM REQUIRED.—The Secretary of Defense shall carry out a pilot program to determine the effectiveness of requiring contractors to reimburse the Department of Defense for costs incurred in processing covered protests.
>
> (b) DURATION.—The pilot program shall—
>
> > (1) begin on the date that is two years after the date of the enactment of this Act; and
> >
> > (2) end on the date that is five years after the date of the enactment of this Act.
>
> (c) REPORT.—Not later than 90 days after the date on which the pilot program under subsection (a) ends, the Secretary shall provide a report to the Committees on Armed Services of the House of Representatives and the Senate assessing the feasibility of making permanent such pilot program.
>
> (d) COVERED PROTEST DEFINED.—In this section, the term "covered protest" means a bid protest that was—
>
> > (1) denied in an opinion issued by the Government Accountability Office;
> >
> > (2) filed by a party with revenues in excess of $250,000,000 (based on fiscal year 2017 constant dollars) during the previous year; and
> >
> > (3) filed on or after October 1, 2019 and on or before September 30, 2022. (H.R. 2810, 2017)

This diluted reform still attempts to address the perceived problem of too many costly bid protests from DoD contractors. However, after the FY 2018 NDAA was signed into law, the RAND Corporation produced its report that dispelled the notion of excessive protesting. RAND found that the share of contracts protested remains very small—less than 0.3%. A significant number of these protests concerned smaller-value contracts ($0.1 million or less; Arena, 2018). Larger contractors are not prevented from bidding on smaller contracts but if RAND found that more protests stem from smaller solicitations, then Section 827 of the FY 2018 NDAA may have missed the mark.

Another problem with the compromise reform in FY 2018 NDAA is how the government will determine the cost of protesting to the government. The vague language in FY 2018 NDAA does not specify what constitutes reimbursement to “the Department of Defense for costs incurred in processing covered protests” (H.R. 2810, 2017). There is no payment to companies who won the protests but lose time working on it due to stays from the bid-protest review at the GAO. It remains unclear what “costs incurred” to the DoD are and how the pilot program plans to calculate this fine.

In FY 2018 NDAA, the federal government attempted to address the perceived problem of excessive protesting by large government contractors. Congress initially tried to impose fines and withhold profits from large protestors but the final reform in FY 2018 NDAA imposes unclear and highly targeted fines. The three-year pilot program that penalizes large companies for unsuccessful protests will undergo further analysis using GAO and Federal Business Opportunities (FBO) data later in this paper.

Data and Methods

Introduction

To operationalize my research question, I looked at data provided by the GAO and FBO’s Opportunities list. FBO is a free web-based portal which allows vendors to review federal procurement opportunities over $25,000 (Department of Commerce, 2012). The government website also retains all posted solicitations over the last calendar year in its archives. This data provided the average value of protests at time of award decided by the GAO, the decisions the GAO ruled regarding the protests, and the number of companies protesting DoD contracts. I linked high value at time of award with large companies because those businesses have greater resources to bid and perform on high value contracts. Trends in bid protest values and companies who file them provided by GAO and FBO data will help analyze the potential effects of GAO Bid Protest Reform in FY 2018 NDAA. The analysis will also provide useful information to propose new reforms that could help Section 809 fulfil its duty to streamline the acquisition process.

Patterns in Bid-Protesting Methodology

The GAO maintains a database that records every protest brought to the Office in recent years. I could not record the thousands of DoD subdivision protests due to time and resource constraints. Instead, I first noted all Department of Defense contracts and recognized that the subdivision of Defense Logistics Agency contained the largest number of protests. I proceeded to select the similar Army Material Command and Navy Supply Systems command to complete my three subdivision choices.

After identifying the three subdivisions, I decided on the timeframe and type of protest to research. By restricting the search to Closed Docket DoD protests decided within the last calendar year, I could record the success rate of DoD protests. I initially selected FY 2016 because it would have contained the most updated record of GAO protests within a solid timeframe. However, once I began recording data, many of the older protests disappeared from the GAO server. To ensure a complete cache of bid protests, I changed my parameters to protests decided between January 17, 2017, through January 24, 2018, for the Defense Logistics Agency; January 26, 2017, through January 24, 2018, for Army Material Command; and February 8, 2017, through January 30, 2018, for Navy Supply Systems Command. These slightly different date parameters may have missed a small number of decided protests, but that number should not affect the findings. I recorded 472 closed bid protests from the GAO from the three DoD subdivisions.

I then populated 472 closed bid protests from the GAO with the “value at award” of each solicitation found in the Federal Business Opportunities (FBO) archived database. I searched other databases including the Federal Procurement Data System (FPDS), but the site contained less accessible information than FBO. Some incomplete or missing solicitation numbers from the GAO prevented me from obtaining the value at award information for some protests. However, many more protests on my GAO list displayed complete solicitation numbers but were indiscoverable on the FBO website. In total, I found value at award information on 159 out of 472 closed docket protests. The 33.7% completion rate may be caused by incorrect data entry by the GAO or FBO or more likely, the initial solicitation notice was over a year old. Per FBO’s support department, “some of the notices could have been archived more than a year ago, (and FBO’s) archive database only looks back 365 days. Also, fbo.gov does not communicate directly with the GAO site, (GAO is) independent. So, it is possible that some of the notices were not posted on FBO” (Federal Business Opportunities, 2018).

Once the matrix was completed, I analyzed the data to determine if larger contracts (which are likely to be bid upon by companies with revenues over $250 million) are in fact protested more than smaller contracts (under $250 million). Analyzing the closed-docket protests over the last year served to test a perception that the rate of protest varies with size. Government officials tend to believe that a major factor in determining a company’s protest strategy is the sheer size of the contract. A higher value contract results in more profit so larger companies will use their extra assets to protest the valuable contracts (Contractor, personal communication, March 19, 2018). Furthermore, the analysis of whether there is a higher rate of protest among larger companies will operationalize my research question. Answering the rate of protest question by comparing the rates side by side will determine whether the proposed policy change targeting only large companies is more effective in saving federal money than small companies.

Results

GAO and FBO Findings

I was unable to find the value at time of award for over 66% of the 472 closed docket protests between the three subdivisions. Part of the reason for the lack of dollar value information may be because companies filed protests during the solicitation process and before the contract was awarded. Additionally, the GAO and FBO regularly miscommunicate and share inconsistent information with each other according to a GAO official contacted for this research.

Nevertheless, there was still valuable information among the protests with incomplete information. The 159 completed closed-docket bid protests provided a valuable snapshot of the perceived problem regarding excessive protest filing at the GAO from DoD contractors. The average value at time of award for the bid-protest in this research was $188,703,404.28. Two DoD subdivisions reflected similar average values at time of award: Defense Logistics Agency averaged $23,222,383.76 among their 39 protests and Navy Supply Systems Command averaged $31,690,997.52 among their 24 protests. Army Material Command was a significant outlier given the average value at the time of award was $292,140,760.84 within the data set of 97 protests. Attributing to the inflated Army Material Command average values are multiple bid protests over the same high-value DoD solicitation. I will discuss the inflated Army Material Command average values in the analysis section.

As mentioned earlier, the GAO designates five possible outcomes for a protest: Dismissed, Withdrawn, Denied, Granted, and Sustained. The GAO generally sustains protests where it determines that the contracting agency violated procurement statutes or regulations, unless it concludes that the violation did not prejudice the protester (GAO, 2018). In the fully populated 472 protests, 381 protests did not go through the entire protest process. 293 were Dismissed and 88 were Withdrawn. Ninety-one protests were completed: 85 were Denied, five were Sustained, and one was Granted. This information is represented in Figure 1.

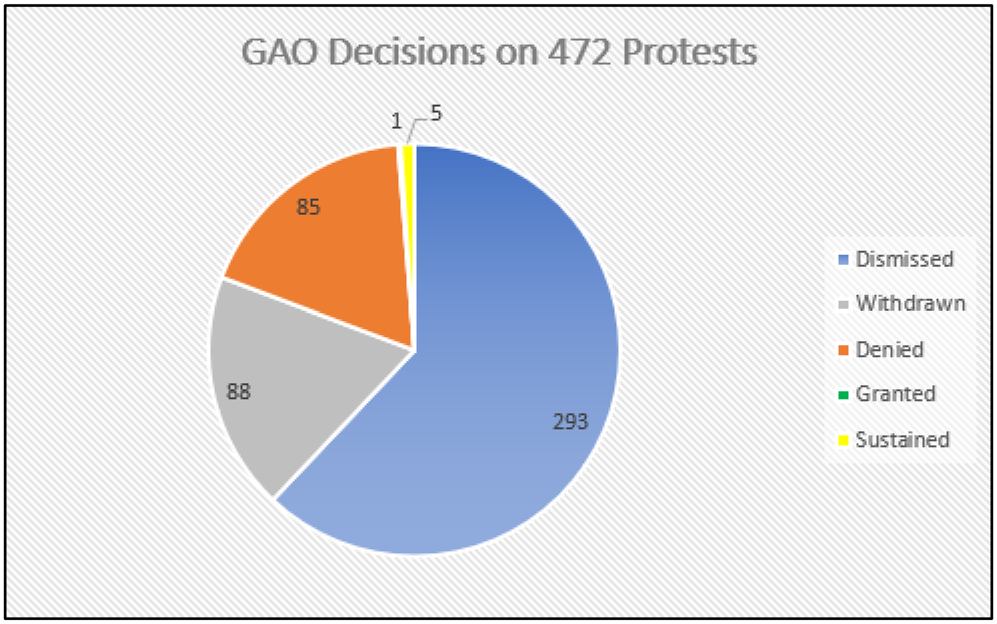


Figure 1. GAO Decisions on 472 Protests

In the 159 complete protests dataset, 127 protests did not go through the entire protests process. Ninety-nine were Dismissed, and 28 were Withdrawn. Thirty-two protests were completed with 31 Denied and only one protest Sustained (see Figure 2).

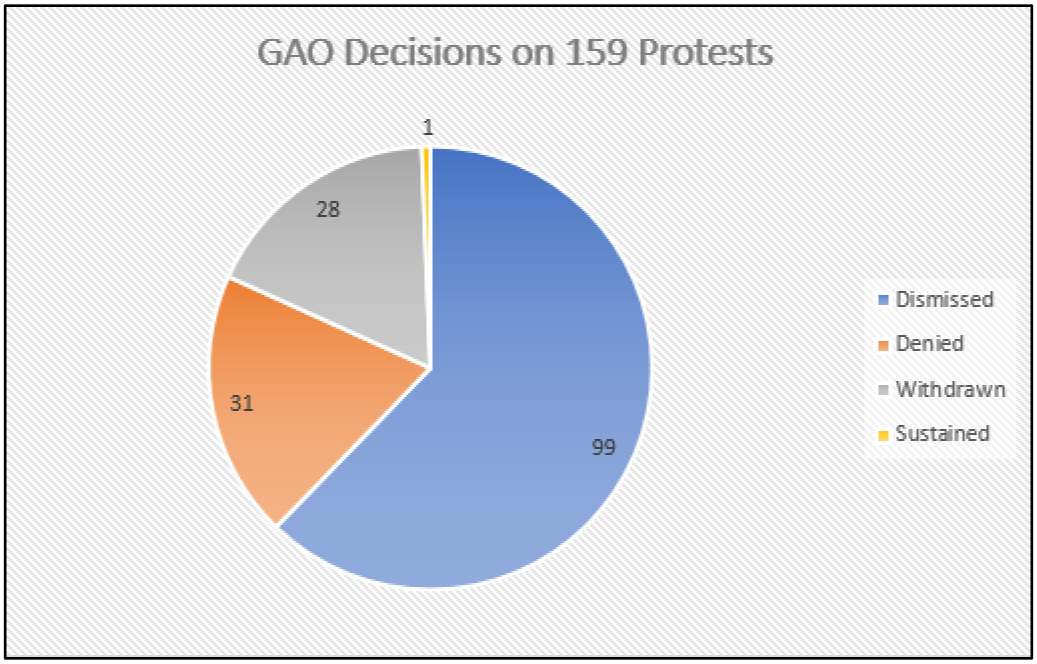


Figure 2. GAO Decisions on 159 Protests

The average value of the 127 Withdrawn and Dismissed protests was $219,999,788.66. The average value of the 31 Denied protests was $75,506,673.33; and the value of the singular Sustained protest was $3,161,460.00.

My data showed average values at time of award for Army Material Command nearly 10 times higher than Defense Logistics Agency or Navy Supply Systems Command. There were protests with a very high “Maximum Potential Contract” ceiling for Army Material Command solicitations. Fifty-four of the 97 closed docket bid protests were valued at over $100 million. However, those 54 protests represented only seven different solicitations. One $500 million solicitation for electronic computing technology, W52P1J-15-R-0122, generated 34 separate protests from 23 companies. One company alone submitted four protests for a $248 million engineering contract award, W900KK-15-R-0012. Table 1 displays the protests regarding high value solicitations.

Table 1. Protests for High Value Solicitations

Count of Company Name	Column Labels							
Row Labels	$110,000,000.00	$133,899,937.00	$248,000,000.00	$500,000,000.00	$580,200,000.00	$750,389,911.00	$975,000,000.00	Grand Total
W15QKN-15-R-0002					1			1
W52P1J-15-R-0122				34				34
W52P1J-16-R-0058		1						1
W58RGZ-15-R-0045						2		2
W900KK-15-R-0012			5					5
W900KK-17-R-0001							7	7
W911QY-17-R-0020	4							4
Grand Total	**4**	**1**	**5**	**34**	**1**	**2**	**7**	**54**

Multiple protests from one company was another pattern established in my data. The GAO processed and closed protests for 263 unique companies within the last calendar year for the three DoD subdivisions. Four companies alone accounted for 15.5% of the GAO processed protests and one company, Aerosage, accounted for 6.8% of the 472 closed docket protests. This trend is reflected in the entire dataset, not just the high value contracts displayed previously.

Stakeholders' Perspective

The RAND Corporation report provided robust information regarding stakeholders' views on the protest process. A key takeaway from their stakeholder analysis was that DoD personnel generally believe incumbent companies are more likely to protest when they have lost in a follow-on competition (Arena, 2018). In addition to the RAND findings, I conducted interviews with individuals who questioned the perceived problem of excessive protesting.

A lawyer for a major defense contractor cited the RAND study and his own work when he rejected the idea that companies were engaging in frivolous protesting. Furthermore, he claimed the majority of his company's protests, especially those of larger value, were protested not for substantive reasons. Instead, the bids were protested for more information. In such cases, protests were often withdrawn or dismissed once the company was satisfied with the information regarding their loss of a contract. RAND addressed this in their recommendation when they suggested the GAO improve their debriefings to losing companies. My interviewer acknowledged that debriefs have improved over the years but that there is still no standard method of delivering debriefs to companies and said this needs to change.

In general, the interviewer placed the responsibility of improving the protest process on the government. He suggested that the overwhelming majority of protests are due to a lack of information from the government or a government mistake. Protesting a bid is, therefore, a tool to keep the government accountable for intentional or unintentional harmful business practices. Furthermore, he questioned the ability of the federal government to accurately audit its own cost of processing a protest. FY 2018 NDAA's language remains vague in this regard, thus, leading to skepticism that the government will enact any penalty.

His final comments placed doubt on the effectiveness of the "loser pays" provision in the FY 2018 NDAA. Speaking candidly, he said the threat of paying the cost of GAO processing is essentially a non-factor when his company decided to protest. If his company lost a high value contract worth over $250 million, the opportunity cost of not protesting would be incredibly high. Discovering the relevant information to become more competitive in future bids outweighs any administrative penalties the federal government would impose (Contractor, personal communication, March 19, 2018). He also floated the idea of shifting all protests to the court system and eliminating the GAO to save the government money but acknowledged this will likely never happen.

After reviewing the results of the bid protest data provided by the GAO and FBO, the need for an alternative solution is clear. The following criteria and alternatives section may better reform the bid protest process than the solution in FY 2018 NDAA. I will analyze these alternatives against the criteria to determine if they satisfy my second research question: how the Section 809 Panel can make it easier for private companies to conduct business with the DoD.

Criteria for Policy Evaluation

Some criteria are required to evaluate potential solutions to GAO filed protests. Listed below are three criteria which, if satisfied, will indicate a potentially successful policy option to improve the protest process and fulfill the Section 809 Panel's goal to make it easier for private companies to conduct business with the DoD. They are derived from a GAO report on protest reform along with suggestions from former Director of Defense Procurement and Acquisition Policy, Deidre Lee. The purpose of the policies is to

- decrease the number of contract protests presented to the GAO,
- increase potential new business partners with the DoD, and
- save the government money and time (Murphy, 1995).

Policy Alternatives

The criteria for successful bid protest reform will be matched against the following three policy alternatives. The first policy alternative is a continuation of FY 2018 NDAA beyond the three-year pilot program. The next two policy alternatives were developed through an interview with a lawyer at a major defense contractor who is affected by the FY 2018 NDAA reforms.

1. **Make "losers pay" pilot program in FY 2018 NDAA permanent law**. This alternative would codify the three-year pilot program into permanent law extending beyond 2022.
2. **Funnel all DoD protests regarding a singular solicitation into one review**. Companies currently file one protest for each issue they identify with a single solicitation. This alternative would require the GAO to consider all points of contention regarding one solicitation in a single review and decision process.
3. **Eliminate the GAO's role in adjudicating bid protests**. The GAO established a procedure for reviewing bid protests nearly 100 years ago and the number of reviewed DoD protests has grown rapidly over the last decade. This policy alternative would eliminate the GAO's role in judging bid protests and move it to the court system.

Analysis

The three policy alternatives will be graded against each individual criterion to determine a final score. This score will demonstrate the likelihood that the policy alternative will solve part of the protest problem established in the policy question.

Table 2. Criteria Scoring System

Score	Explanation
0	Fails to satisfy criterion
1	Partially Satisfies criterion
2	Fully satisfies criterion

Table 3. Policy Alternative 1

Alternative	Decrease Contracts Protested	Increase DoD Business	Save Time and Money	Total Score
Losers Pay Law	0	0	1	**1**

As shown in my data and RAND's data, the highest volume of bid protests originates with contracts valued under $1,000,000 and companies affected by the pilot program are less likely to bid on such contracts. Furthermore, this alternative will not deter large companies from protesting bids on high value contracts. My interview subject asserted that the consensus of the contracting sector is if a large company decides to file a protest regarding a large solicitation, the penalties imposed by the GAO and the DoD are simply the cost of business (Contractor, personal communication, March 19, 2018). This policy alternative does not affect companies with annual revenues under $250 million and is therefore unlikely to decrease the number of protests filed at the GAO.

Turning the pilot program into permanent law is unlikely to attract new business to the DoD because few new companies would be large enough to be affected by the protest penalties. Additionally, because the number of protests is unlikely to decrease, new companies will not view the policy alternative as a new opportunity for business.

Turning the pilot program into permanent law will minimally impact the GAO or DoD budget or time spent on protests. The number of protests is likely to remain the same and the fines imposed on contractors is unclear. The DoD may recuperate some money from the fines imposed on contractors but only when the GAO fully denies the protest. The data shows this is a small portion of the protests and therefore will not save the government much time or money.

The RAND Corporation, my interview subject, and other stakeholders all highly doubt that the "losers pay" provision of the FY 2018 NDAA will generate dramatic changes to the federal procurement and acquisition system. Section 809 Panel is charged with making recommendations that will shape the DoD's acquisition system into one that is bold, simple, and effective (Section 809 Panel, 2018). Establishing this insignificant pilot program to an already complex acquisition system is counterproductive.

Policy Alternative 2: Funnel all DoD protests regarding a single solicitation into one review.

Table 4. Policy Alternative 2

Alternative	Decrease Contracts Protested	Increase DoD Business	Save Time and Money	**Total Score**
Funnel Protests	2	1	2	**5**

Forcing the GAO to consider all protests regarding one solicitation at the same time will decrease the number of protests filed at the GAO. Unlike ASRC Communications who filed four protests regarding one $248 million engineering contract award, this alternative will allow companies to air all their grievances in one protest.

The streamlining mechanism would improve the efficiency image of the federal government and potentially attract new businesses. However, this is only speculative, and therefore only partially satisfies the criterion of bringing new business partners to the DoD.

Currently, each protest is processed separately at the GAO and potentially by different GAO officers. If the bid was for a re-compete, a vast number of protests could significantly delay the process and keep the bridge-contractor working on the contract. By funneling all protests regarding one solicitation to a single review, the government will save time and money while addressing the concerns of business partners.

Policy Alternative 3: Eliminate the GAO's role in adjudicating bid protests.

Table 5. Policy Alternative 3

Alternative	Decrease Contracts Protested	Increase DoD Business	Save Time and Money	**Total Score**
Eliminate the GAO's role	0	0	1	**1**

Transferring protest review responsibility from the GAO to the courts will not decrease the number of protests within the federal government. This alternative would only impose a later shift of the protest responsibility.

If DoD contractors enjoy the courts protest review system more than the GAO, it could potentially attract new business to the DoD. However, there is no guarantee that the courts would perform better than the GAO. Therefore, this alternative only partially satisfies the second criterion.

The court system does not have a standard method of hearing protest cases. While the GAO's protest process has its flaws, there is a standard practice in place that has developed over nearly a century. Uncertainty in the court systems would cost the government more time and money in the larger scheme.

Conclusion

The federal government and private companies hold different views on the problems surrounding protests filed at the GAO. Government officials expressed their belief that large corporations are filing frivolous protests to hurt competitors at a high cost to the federal government. Private corporations and data from RAND, the GAO, and FBO assert that there is no such problem of frivolous protesting. Private corporations view protests as a mechanism to obtain information from the federal government about why they lost the bid, so they can improve in the future. However, the information often comes in separated by multiple protests regarding the same solicitation. Converting the GAO Bid Protest Reform section of FY 2018 NDAA into permanent law does not satisfy the mission of the Section 809 Panel to streamline acquisition. My analysis and scoring shows that the government should look into funneling all protests regarding a single solicitation into one protest review.

Recommendation

Table 6. Policy Alternative Evaluation Matrix

Alternative	Decrease Contracts Protested	Increase DoD Business	Save Time and Money	**Total Score**
Losers Pay Law	0	0	1	**1**
Funnel Protests	2	1	2	**5**
Eliminate GAO's role	0	0	1	**1**

The analysis from the data shows lower propensity of the GAO filed protests for low-value solicitations. The data also showed multiple protests regarding one solicitation even from the same company. Therefore, the Section 809 Panel should recommend instituting the RAND Corporation recommendations, abandoning the pilot program established in the FY 2018 NDAA, and funneling all protests regarding a singular solicitation into one review.

References

Arena, M. V. (2018). *Assessing bid protests of U.S. Department of Defense procurements: Identifying issues, trends, and drivers*. Santa Monica, CA: RAND Corporation.

Davenport, C. (2017, October 10). Senate bill tries to curb Pentagon contract protests. *Washington Post*.

Defense Industry Daily. (2010, April 22). How the US GAO's bid protest process works and why defense contractors abuse it. Retrieved from https://www.defenseindustrydaily.com/gao-protests-defense-programs-06269/

Department of Commerce. (2012, March 27). Facts you should know about federal business opportunities. Retrieved from Minority Business Development Agency website: https://www.mbda.gov/news/blog/2012/03/facts-you-should-know-about-federal-business-opportunities

Federal Business Opportunities. (2018, March 15). Unable to locate a notice. Federal Service Desk.

GAO. (2017, September 21). *A-11259, MARCH 9, 1926, 5 COMP. GEN. 712*. Retrieved from http://www.gao.gov/products/A-11259#mt=e-report

GAO. (2018, April 2). *Bid protests at GAO: A descriptive guide*. Retrieved from https://www.gao.gov/decisions/bidpro/bid/protest.html

Gordon, D. L. (2013). Bid protests: The costs are real, but the benefits outweigh them. *George Washington Law Faculty Publications & Other Works, 42*(3).

Koprince, S. (2017, July 28). Senate 2018 NDAA re-introduces flawed GAO bid protest "reforms." SmallGovCon.

Lasky, A. (2017, November 12). House and Senate strike deal on bid protest reforms in FY 2018 NDAA. The Procurement Playbook.

Manuel, K. M. (2011). *Competition in federal contracting: An overview of the legal requirements.* Washington, DC: Congressional Research Service.

Manuel, K. M., & Schwartz, M. (2016). *GAO bid protests: An overview of time frames and procedures.* Washington, DC: Congressional Research Service.

Maser, S., & Thompson, F. (2010). *Understanding and mitigating protests of Department of Defense acquisition contracts*. Salem, OR: Williamette University Center for Governance and Public Policy Research.

Murphy, R. (1995, April 6). Procurement reform: Opportunities for change. Retrieved from http://www.gao.gov/assets/110/105967.pdf

National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114-92, 129 Stat. 787 (2015). Retrieved from https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf

National Defense Authorization Act for Fiscal Year 2018, H.R. 2810, 115th Cong. (2017). Retrieved from https://www.congress.gov/bill/115th-congress/house-bill/2810/text#toc-H0A7CC8EC4BA448928FD234B8AD8FBE9D

National Defense Authorization Act for Fiscal Year 2018, S. 115-125, 115th Cong. (2017). Retrieved from https://www.congress.gov/bill/115th-congress/senate-bill/1519

President's Management Council. (2018). *President's management agenda*. Washington, DC: Executive Office of the President.

Rules and regulations. (1995, September 18). *Federal Register, 60*(180). Retrieved from https://www.gpo.gov/fdsys/pkg/FR-1995-09-18/pdf/95-22777.pdf#page=2

Section 809 Panel. (2018). About us. Retrieved from https://section809panel.org/about/

Disclaimer

This student paper was prepared in 2017 in partial completion of the requirements for the Master's Project, a major assignment for the Master of Public Policy Program at the Sanford School of Public Policy at Duke University. The research, analysis, and policy alternatives and recommendations contained in this paper are the work of the student who authored the document, and do not represent the official or unofficial views of the Sanford School of Public Policy or of Duke University. Without the specific permission of its author, this paper may not be used or cited for any purpose other than to inform the client organization about the subject matter.

Acquisition Research Program
Graduate School of Business & Public Policy
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net