SYM-AM-21-041

Proceedings
of the
Eighteenth Annual Acquisition Research Symposium

Wednesday, May 12, 2021 Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 11–13, 2021

Published: May 5, 2021

Approved for public release; distribution is unlimited.

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 Defense Management at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Table of Contents

SYM-AM-21-041

Proceedings of the Eighteenth Annual Acquisition Research Symposium

Wednesday, May 12, 2021 Sessions
Volume II

Acquisition Research:
Creating Synergy for Informed Change

May 11-13, 2021

Published: May 5, 2021

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School

THIS PAGE INTENTIONALLY LEFT BLANK

WELCOME: DAVID H. LEWIS, VICE ADMIRAL, U.S. NAVY (RET), ACQUISITION CHAIR, ACQUISITION RESEARCH PROGRAM

Vice Admiral David H. Lewis, USN (Ret.) took the helm as the Naval Postgraduate School Chair of Acquisition. As chair, he will lead the Acquisition Research Program (ARP) in the Graduate School of Defense Management and connect NPS with leaders and policymakers in the acquisition community.

Lewis graduated from NPS in 1988 with a Master of Science in Computer Science, and we're pleased to welcome him back to campus in this leadership role. Lewis is replacing the founding Chair of Acquisition, Rear Admiral, USN (Ret.) Jim Greene, who retired this June.

Most recently, Lewis served as Director of the Defense Contract Management Agency, managing over $7 trillion in defense contracts. In this role, he oversaw the agency's efforts to ensure that supplies and services contracted for by the Department of Defense are delivered on time and in line with contract performance requirements.

During his career at sea, Lewis served as a communications officer, fire control and missile battery officer, and combat systems officer aboard destroyers and guided-missile cruisers.

Upon selection to flag rank in 2009, Lewis served as Vice Commander, Naval Sea Systems Command and then served four years as Program Executive Officer, Ships, where he directed the delivery of 18 ships and procurement of another 51 ships. From 2014-2017 he served as Commander, Space and Naval Warfare Systems Command where he led a global workforce of 10,300 civilian and military personnel who design, develop and deploy advanced communications and information capabilities.

Lewis's extensive experience in shipbuilding has given him a unique understanding of the full acquisition lifecycle. He has delivered ships as a program manager and program executive officer, then later sustained and modernized them as a fleet engineer and systems commander. He will bring valuable perspective to NPS students and faculty, as well as the broader acquisition innovation community working to get superior capabilities into the hands of our warfighters.

Lewis's expertise in product delivery will amplify ARP's ability to execute its mission of delivering the real-time information and analytical capabilities needed by today's acquisition professionals and policymakers. Adding VADM Lewis to the team also demonstrates NPS's continued commitment to providing world-class defense-focused education and research...

KEYNOTE SPEAKER: FREDERICK J. (JAY) STEFANY, ASSISTANT SECRETARY OF THE NAVY (RESEARCH, DEVELOPMENT AND ACQUISITION)

Frederick J. (Jay) Stefany serves as acting assistant secretary of the Navy (Research, Development and Acquisition) (Acting) as of Jan. 20, 2021, having served as Deputy Assistant Secretary of the Navy for Ship Programs since April 2018. In this role, he is responsible for executive oversight of all naval shipbuilding programs, major ship conversions, and the modernization and disposal of inservice ships. He is also responsible for executive oversight of cost, schedule and performance of surface ship, submarine, and Marine Corps combat systems, electronic warfare systems, shipboard radars, and Navy missile defense programs.

Previously, Stefany served as executive director, Amphibious, Auxiliary and Sealift Office, Program Executive Office, Ships. He provided executive leadership to 200 personnel and oversaw one of the broadest acquisition portfolios in the Navy, including more than $30 billion in complex shipbuilding procurements. His responsibilities spanned four major program offices: LPD 17 Program (PMS 317), Amphibious Warfare Program (PMS 377), Strategic and Theater Sealift Program (PMS 385), and the Auxiliary Ships / Small Boats and Crafts Program (PMS 325).

Additionally, he oversaw several active major programs, to include LHA 6, LPD 17, EPF, ESB, T-AKE, T-AO(X), and Heavy Icebreaker ship classes, as well as ship-to-shore connectors, landing craft, research ships, and service craft boats.

Stefany entered the Senior Executive Service in March 2012, and has been in Civil Service for more than 27 years. Serving in a variety of key leadership positions throughout his career, including program manager and deputy program manager for the LPD 17 class Amphibious Transport Dock ship program (2004-2012). During his tenure, the first six ships of the San Antonio class were delivered and construction started on four additional hulls. He also assumed responsibilities for management of the initial concept work on a replacement for the Navy's Command & Control Ships and later, the replacement for the LSD 41 and 49 class ships.

Previous assignments include director of Naval and Commercial Construction (2002-2004), responsible for oversight of the Navy's portfolio of Amphibious, Auxiliary and Special Mission ships and craft for the assistant secretary of the Navy for Research, Development and Acquisition (ASN RD&A); assistant program manager in PMS 377 for LCAC (2000-2002) and for Amphibious Ship Combat / C4I Systems (1998-2000); and project engineer for both the LHD 5-7 (1992-1998) and LHD 1-4 (1987-1991) ship acquisition programs as PMS 377 delivered LHD 1-6 and LSD 52 to the Fleet.

Stefany received his bachelor of science degree in mechanical engineering from Lehigh University, Bethlehem, Pennsylvania, in 1985, and his master of science degree in management from the Florida Institute of Technology, Melbourne, Florida (National Capital Campus), in 1998. He is also a 1996 graduate of the Defense Systems Management College Advanced Program Management Course.

During his distinguished federal service career, Stefany has received the Navy Civilian Meritorious Service Award and two Navy Civilian Superior Service Awards.

PANEL 8. ACQUISITION MODERNIZATION

Wednesday, May 12, 2021	
7:45 a.m. – 9:00 am.	**Chair: Brigadier General Michael E. Sloane, USA,** Program Executive Officer for Intelligence, Electronic Warfare and Sensors (IEW&S) **Panelists:** **Lieutenant General Thomas H. Todd III, USA,** Deputy Commanding General, Acquisition and Systems Management, United States Army Futures Command **Lieutenant General Michael Williamson, USA (Retired),** Vice President for Missiles, Lockheed Martin **Lieutenant General Neil Thurgood, USA,** Director for Hypersonics, Directed Energy, Space & Rapid Acquisition

Lieutenant General Thomas H. Todd III, USA—serves as the Program Executive Officer, Aviation. He accepted this post on 11 January 2017. Todd previously served as the deputy commanding general of Research, Development and Engineering Command (RDECOM) and senior commander at Natick Soldier Systems Center. Todd earned a bachelor of science in business administration from The Citadel in 1989 and received his commission as a second lieutenant upon graduation. He is a graduate of the Army Aviation Officer Basic Course and Initial Entry Rotary Wing training. Todd is also a graduate of the OH-58 and H-60 Maintenance Test Pilot Course, the Army Aviation Officer Advanced Course, as well as the Command and General Staff Officer Course. He holds masters of science degrees in contract management and strategic studies from the Florida Institute of Technology and U.S. Air War College, respectively.

Lieutenant General Michael Williamson, USA (Retired)—Lieutenant General Michael E. Williamson retires after 34 years of service to the U.S. Army in a Special General Officer Retirement ceremony hosted by General Daniel B. Allyn, 35th Vice-Chief of Staff of the Army, in Conmy Hall, Joint Base Myer-Henderson Hall, Va., March 2, 2017. Lieutenant General Williamson retired from his career with his final position as the Principal Military Deputy to the Assistant Secretary of the Army for Acquisition Logistics and Technology (ASA(ALT)).

Lieutenant General Neil Thurgood, USA— Lieutenant General L. Neil Thurgood is the Director for Hypersonics, Directed Energy, Space and Rapid Acquisition, Office of the Assistant Secretary of the Army (Acquisition, Logistics and Technology), Redstone Arsenal, Alabama. He assumed duties in March 2019. In this position, LTG Thurgood is responsible for the rapid fielding of select capabilities to deter and defeat rapidly modernizing adversaries, including overseeing development of an Army Long Range Hypersonic Weapon. He leads the Army Rapid Capabilities and Critical Technologies Office mission to rapidly and efficiently research, develop, prototype, test, evaluate, procure and field critical enabling technologies and capabilities that address immediate, near-term, and mid-term threats, consistent with the Army's modernization priorities.

PANEL 9. TRENDS IN DEFENSE ACQUISITION: ANALYSIS FROM CENTER FOR STRATEGIC AND INTERNATIONAL STUDIES (CSIS)

Wednesday, May 12, 2021	
9:15 a.m. – 10:00 a.m.	**Chair: Phil Rodgers,** Director, Acquisition Management and Approaches, OUSD (A&S) ***Assessing the Role of Congress in Defense Acquisition Program Instability*** Seamus Daniels, Center for Strategic and International Studies Todd Harrison, Director, Center for Strategic and International Studies ***Defense Acquisition Trends 2021*** Greg Sanders, Center for Strategic and International Studies Rhys McCormick, Center for Strategic and International Studies ***Trends in OTA Usage Across the Federal Government*** Rhys McCormick, Center for Strategic and International Studies Greg Sanders, Center for Strategic and International Studies

Phil Rodgers— is the Director of Acquisition Approaches and Management. He joined the Department of Defense in 1982 serving as a Naval Officer in a variety of positions. His ashore assignments included tours on the Chief of Naval Operations staff as a manpower analyst and program analyst concentrating on Logistics and Mobility analyses, as well as long range program planning.

After leaving active duty, Mr. Rodgers accepted a position with the Naval Center for Cost Analysis where he conducted independent cost estimates and financial analysis of major Defense contractors; and industrial base assessments of interest to the Navy. He then assumed duties as the Program Manager for the Navy Visibility and Management of Operating and Support Costs (VAMOSC) system.

Mr. Rodgers joined the staff of the Under Secretary of Defense (Acquisition, Technology & Logistics) (USD(AT&L)) in 1995, and was selected into the Senior Executive Service in 2000. From 2000 to May 2007 he served as the Deputy Director, Resource Analysis within the Directorate for Acquisition Resources and Analysis (ARA) leading acquisition and technology planning, programming, and budgeting process activities for the USD(AT&L). He also conducted a wide range of assessments of major resource issues for use in Defense Acquisition Board reviews and in the Planning, Programming, Budgeting and Execution (PPBE) process.

In May 2007, he assumed duties as the Principal Deputy Director, Acquisition Resources and Analysis; responsible for integrating all USD(AT&L) planning, programming and budgeting activities, as well as for overseeing the efficient functioning of the Department's formal weapons systems acquisition process. He managed the governance process and systems providing Department-wide access to authoritative and timely information supporting major acquisition oversight and decision making. He was the functional leader of the 7000 members of the Business community of the defense acquisition workforce, setting workforce policies and guiding training through the Defense Acquisition University. He also provided management oversight and guidance for the DoD FFRDCs and University Affiliated Research Centers.

In August 2018, he was reassigned to become the first Director, Acquisition Approaches & Management responsible for leading Defense acquisition policy formulation and implementation for various aspects of the Defense Acquisition System; to include the Middle Tier of Acquisition, Business Systems, and Software. He also leads policy development and implementation for the acquisition and licensing of Intellectual Property. He is the DoD lead for data collection and statutory reporting to Congress on $2T of investments in Major Defense Acquisition Programs.

He received his MS in Operations Research Analysis from the Naval Postgraduate School. His honors and awards include the Defense Medal for Exceptional Civilian Service (three awards) as well as two Presidential Rank Awards.

Assessing the Role of Congress in Defense Acquisition Program Instability

Seamus P. Daniels—is an Associate Fellow and Associate Director for Defense Budget Analysis in the International Security Program at the Center for Strategic and International Studies (CSIS), where he contributes to the center's research on issues related to defense funding, force structure, and military readiness. He has authored publications on trends in the overall defense budget, the legislative process surrounding defense appropriations, Navy readiness funding, and NATO burden sharing. He holds an A.B. from Princeton University's School of Public and International Affairs with minors in Near Eastern studies and Arabic language and culture. [sdaniels@csis.org]

Todd Harrison—is the Director of Defense Budget Analysis and the director of the Aerospace Security Project at CSIS. As a Senior Fellow in the International Security Program, he leads the center's efforts to provide in-depth, nonpartisan research and analysis of space security, air power, and defense funding issues. He has authored publications on trends in the overall defense budget, military space systems, defense acquisitions, military compensation, military readiness, nuclear forces, and the cost of overseas military operations. He graduated from the Massachusetts Institute of Technology with both a BS and an MS in aeronautics and astronautics. [tharrison@csis.org]

Abstract

This paper seeks to evaluate the impact of Congress on funding instability for defense acquisition programs. While the Department of Defense (DoD) requests specific funding levels for procurement and research, development, test, and evaluation (RDT&E) programs each fiscal year in the president's budget submission, Congress ultimately determines the funds that will be made available to those programs via the appropriations process. While lawmakers constitutionally hold the power of the purse and oversight authority over the DoD, changes between the requested and actual level of funding for programs can force defense officials and industry partners to adjust program schedules and funding requirements. At a macro level, changes can also disrupt wider defense planning and strategic priorities for the Department and the executive branch. This study assesses trends in funding for acquisition accounts and programs relative to the requested level of funding from Fiscal Year (FY) 2010 to FY 2020 to determine if Congress regularly funded programs above or below administrations' budget requests over that time frame. It identifies specific procurement and RDT&E accounts that typically have their funding adjusted by lawmakers and trends in the differences between requested funding and the actual level of funding.

Introduction

The Department of Defense (DoD) outlines the priorities of the administration in the budget request submitted to Congress for the upcoming fiscal year. Along with its request for funding for the next fiscal year, the Department submits thousands of pages of budget information justifying the funds required for its programs and outlining its plans in detail. For acquisition programs (primarily funded through the procurement and research, development, test, and evaluation [RDT&E] accounts), the DoD also provides lawmakers with detailed information at the line item and program element levels on program schedules and requirements as well as projected future funding requirements in the Future Years Defense Program (FYDP).

While the executive branch articulates its own strategic priorities, plans its defense acquisition agenda, and distributes contracts to private sector partners, Congress retains the power of the purse and ultimately has the final say in deciding which acquisition programs receive funding and how much they receive. Led by each chamber's respective appropriations subcommittee on defense, the legislative branch can choose to match, modify, or eliminate the DoD's requested funding levels for procurement and RDT&E programs as well as alter the quantity of systems or platforms procured. Congressional adjustments to the budget can also be made after funding is appropriated via rescissions that cancel some or all of the budget authority prior to its obligation (Saturno et al., 2016, p. 20). Likewise, the DoD can reprogram funding to move it among accounts as needed within the constraints set by appropriators (Saturno et al., 2016, p. 12).

Congressional decisions on funding for acquisition programs in the appropriations process can have a significant impact on an administration's defense plans. Adjustments to the program of record for acquisition projects can force the program management teams in the DoD to alter the program's schedule and contracting actions. In addition to affecting the performance of acquisition programs, these adjustments can have secondary effects on private sector partners, particularly the prime contractors tasked with developing and building systems and their suppliers. At the macro level, disruptions to acquisition plans in the appropriations process can also affect an administration's ability to operationalize its defense strategy. Ultimately, the power of Congress to appropriate money gives the legislative branch an important role in overseeing how defense dollars are spent and the execution of defense strategy.

This analysis seeks to assess trends in congressional action on funding for defense acquisition programs relative to the requested level. It will compare the actual funding level for procurement and RDT&E accounts with the original level proposed in the administration's budget request and identify specific accounts which are regularly adjusted by Congress. Ultimately, this analysis aims to inform defense planners, acquisition officials and program managers, and industry partners of trends in congressional appropriations for defense so they can better plan for how the congressional budget process is likely to affect the defense budget request on a more granular level using historical data.

Methodology

This study assesses trends in congressional action on defense acquisition funding from Fiscal Year (FY) 2010 through FY 2020. The time frame was selected to determine the impact of the Budget Control Act (BCA) of 2011 on funding for procurement and RDT&E accounts. In an effort to reduce the federal deficit, the BCA imposed caps on discretionary funding levels for defense and non-defense programs between FY 2012 and FY 2021, among other efforts. These budget caps were then increased by Congress in a series of budget agreements over that time frame (Harrison & Daniels, 2020, p. 6). However, the inability of lawmakers to identify an alternative to the deficit reduction plan outlined in the BCA following its passage in 2011 led to sequestration, a budgetary mechanism that reduces discretionary spending in excess of the budget caps. A sequester was triggered in March 2013 that led to cuts of 6.7% and 8.1% to procurement and RDT&E accounts, respectively (Daniels, 2019, p. 3).

In its approach to measuring the role of Congress in acquisition funding, this analysis compares the *requested* level of funding from the presidential budget request with the *actual* level of funding for procurement and RDT&E accounts. Budget data was compiled from the procurement programs' (P-1s) and RDT&E programs' (R-1s) justification books published by the Office of the Comptroller with each year's budget request.

The *actual* level of funding for acquisition programs is calculated approximately 2 years after it is originally requested (for example, actual funding levels for FY 2019 are published with the FY 2021 request). While the *enacted* level of funding passed by lawmakers would be a more direct comparison to illustrate congressional action on the budget request (since lawmakers can make adjustments to funding after it is appropriated), this data is not consistently captured in budget justification documents due to delays in enacting appropriations. It is also not captured in legislative text in a machine-readable format at the program level. The *actual* funding level will not typically differ dramatically from the *enacted* level of funding, with the exception of FY 2013, when sequestration occurred. Since the *actual* funding level for FY 2020 is not yet available, the *enacted* funding level is used because it is available and accurately reflects what was appropriated.

Occasionally, administrations may submit supplemental budget requests or amendments to their original budget requests. The requested level data analyzed in this study incorporates adjustments submitted by the administration after the fact and is also inclusive of Overseas Contingency Operations (OCO) funding. One exception is FY 2017, in which the Trump administration submitted an updated budget request after taking office, which amended the request submitted by the Obama administration. The data for that year's request in this study represents the Obama administration's request and any changes requested by that administration, but it does not include changes requested by the Trump administration that occurred during the middle of FY 2017.

Congressional adjustments to the requested level of funding for procurement and RDT&E accounts are measured by calculating the percent change between the actual funding level and the requested level for each given fiscal year. An average percent change for the FY 2010 to FY 2020 time frame is then calculated to allow for more direct comparison. Analysis is conducted at the topline procurement and RDT&E level, the military department level, and account level with further analysis at more granular levels to explain trends in the data.

Analysis of Congressional Action on Procurement Funding

Topline Procurement Funding

If assessed at the topline level, congressional funding for procurement accounts relative to the presidential request has fluctuated considerably over the period of analysis. Figure 1 presents the percent change from the actual level of procurement to the requested level. While funding is 2.9% higher than the requested topline on average from FY 2010 through FY 2020, actual funding for procurement accounts was below the requested amount for 4 years between FY 2011 and FY 2014. This was due in no small part to concerns over the federal deficit and the impact of the BCA in constraining the defense budget. The Obama administration repeatedly requested funding for defense programs above the level permitted by the BCA budget caps, yet congressional adjustments to the caps did not always match the level of increase requested by the Obama administration. Sequestration also contributed to lower actual levels of funding for procurement relative to the request in FY 2013.

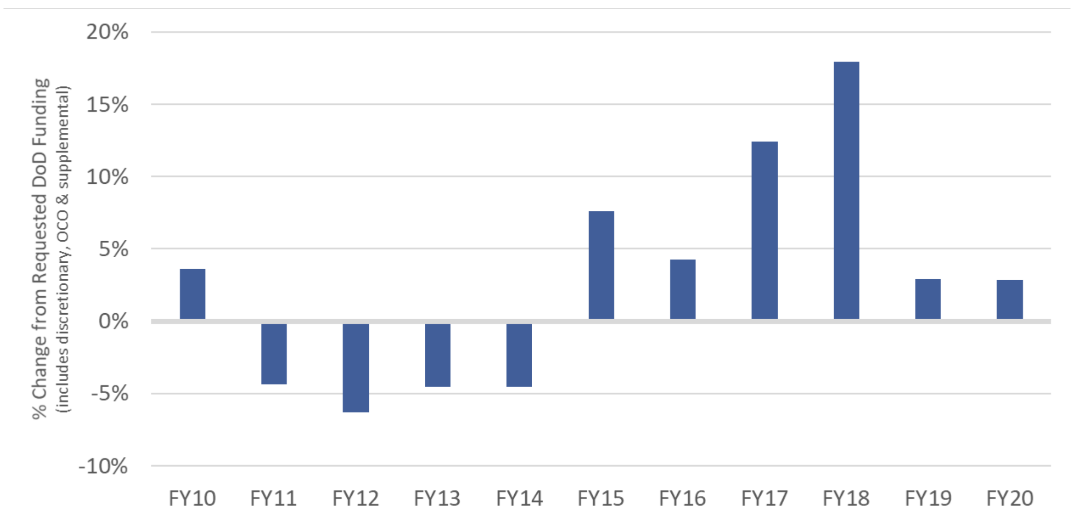


Figure 1. Actual Total Procurement Funding vs. Requested Level

Actual procurement funding surpassed the requested level by almost 18% in FY 2018 as a result of the Bipartisan Budget Act of 2018 (BBA 2018). The 2-year budget deal raised the spending caps for defense higher than the Trump administration had requested and was significantly larger than previous 2-year agreements that increased the caps from FY 2014 to FY 2015 and FY 2016 to FY 2017 (Daniels & Harrison, 2020).

Procurement Funding by Military Department

When assessed by military departments and defense-wide, or "Fourth Estate," accounts, the procurement funding data yields similar trends to the topline analysis, as shown in Figure 2. However, defense-wide procurement funding shows significant fluctuations and differences relative to the request. This is largely due to the small amount of procurement funds requested for defense-wide accounts compared to the military departments; adjustments above or below that requested level will appear more drastic when represented as a percentage change because of the smaller amounts of funding involved. At times, the DoD may also request funds for defense-wide accounts, which Congress then cuts from defense-wide and appropriates directly into other service accounts.

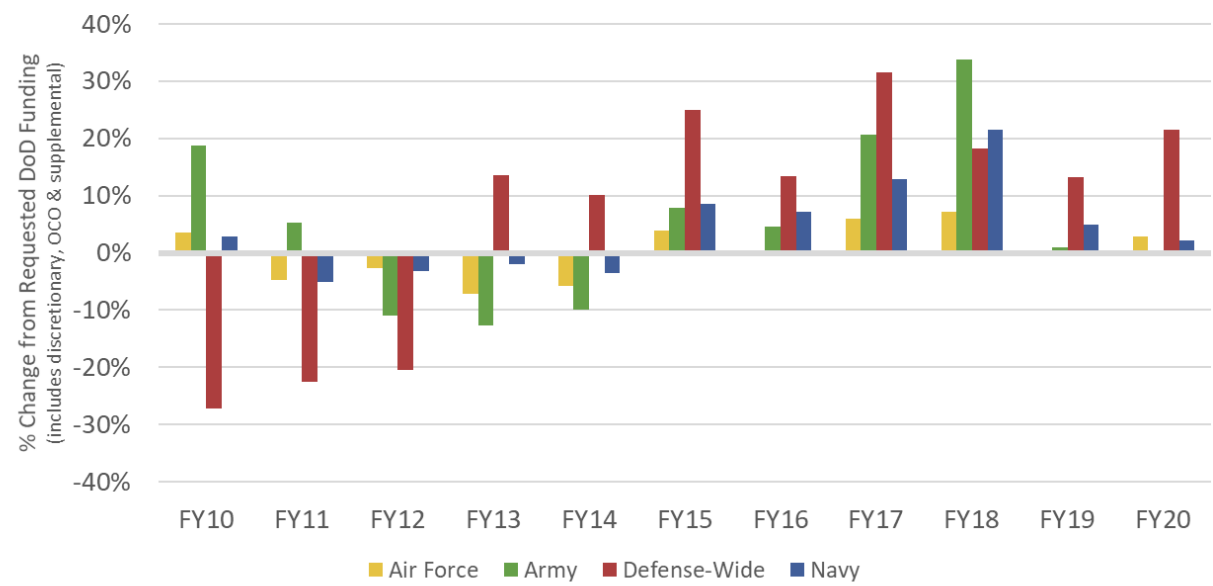


Figure 2: Actual Procurement Funding vs. Requested Level by Military Department

Compared across the FY 2010 to FY 2020 time frame, actual procurement funding for the defense-wide accounts was an average of 6.9% higher than the request, compared to 2.9% for overall procurement funding. Of the military departments, the Army received the greatest increase above the request at 5.3% on average, in comparison to a 4.2% average increase for the Navy and 0.3% for the Air Force. While Army accounts received significantly more funding than requested in FY 2017 and FY 2018, Congress was less generous in FY 2019 and appropriated slightly less funding than requested in FY 2020.

Procurement Funding by Account

A comparison of the requested and actual levels of funding at the account level provides a better idea of the factors driving trends at the military department level. Table 1 ranks procurement accounts by the average percent change between the actual level of funding and the requested level over the FY 2010 to FY 2020 time frame.[1] Three Army accounts—weapons and tracked combat vehicles (W&TCV), aircraft, and missiles—received the greatest percentage increase over the budget request. As Figure 3 shows, funding for those three accounts exceeded the requested level for at least 7 of the 11 years assessed. Funding for W&TCV was higher than the request every year until FY 2019 and exceeded the requested level by nearly 60% in FY 2018.

Table 1: Average Difference Between Requested and Actual Funding Levels by Procurement Account

Account	Average Percent Change
Procurement of Weapons and Tracked Combat Vehicles, Army	15.1%
Aircraft Procurement, Army	7.5%
Missile Procurement, Army	7.0%
Shipbuilding and Conversion, Navy	6.8%
Procurement, Defense-wide	5.8%
Aircraft Procurement, Navy	5.1%
Aircraft Procurement, Air Force	2.2%
Other Procurement, Army	2.0%
Procurement of Ammunition, Air Force	0.9%
Procurement of Ammunition, Army	0.9%
Procurement, Marine Corps	0.7%
Other Procurement, Air Force	0.5%
Weapons Procurement, Navy	-1.8%
Other Procurement, Navy	-2.8%
Space Procurement, Air Force	-2.9%
Chemical Agents and Munitions Destruction, Defense	-3.0%
Procurement of Ammunition, Navy and Marine Corps	-5.5%

[1] Only certain defense-wide accounts are included in this ranking given the irregular nature of some accounts and the fact that others only had several years' worth of data compared to the 11 years of data for most other accounts assessed.

Missile Procurement, Air Force	-5.7%

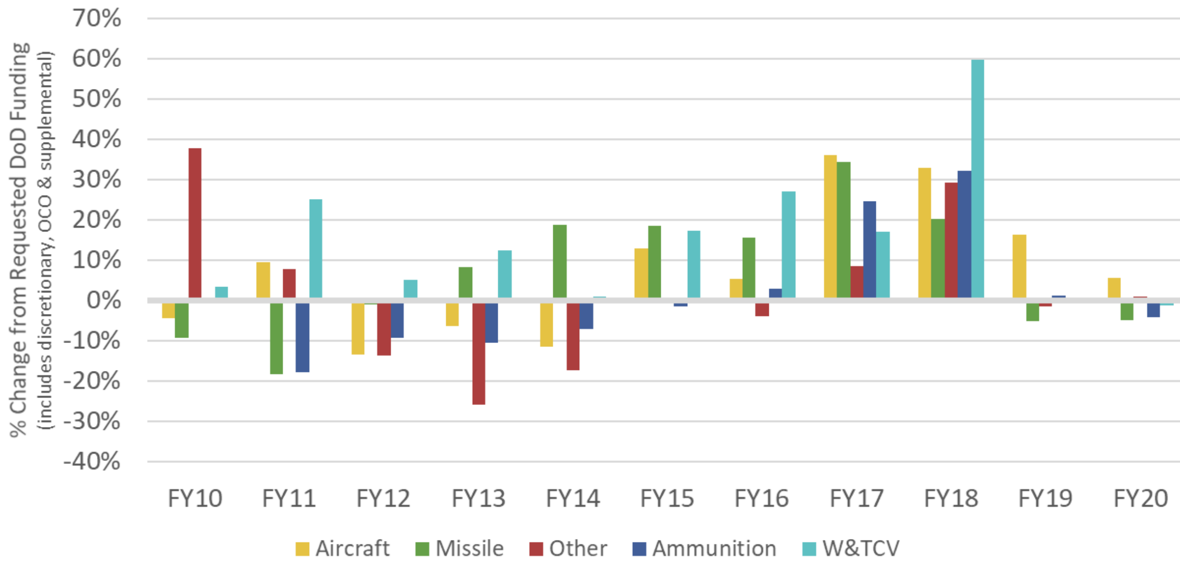


Figure 3. Army Actual Procurement Funding vs. Requested Level by Account

As previously discussed, the Navy received the second largest average increase above its requested level compared to the other two military departments. This was driven by funding for the shipbuilding and conversion account, which on average received 6.8% more funds than requested. As Figure 4 shows, Congress also added to the Navy's aircraft procurement account from FY 2015 through FY 2020 for an average increase of 5.1% above the requested level. However, Congress regularly appropriated less funding for the Navy's weapons, ammunition, and other accounts than requested over the selected time frame.

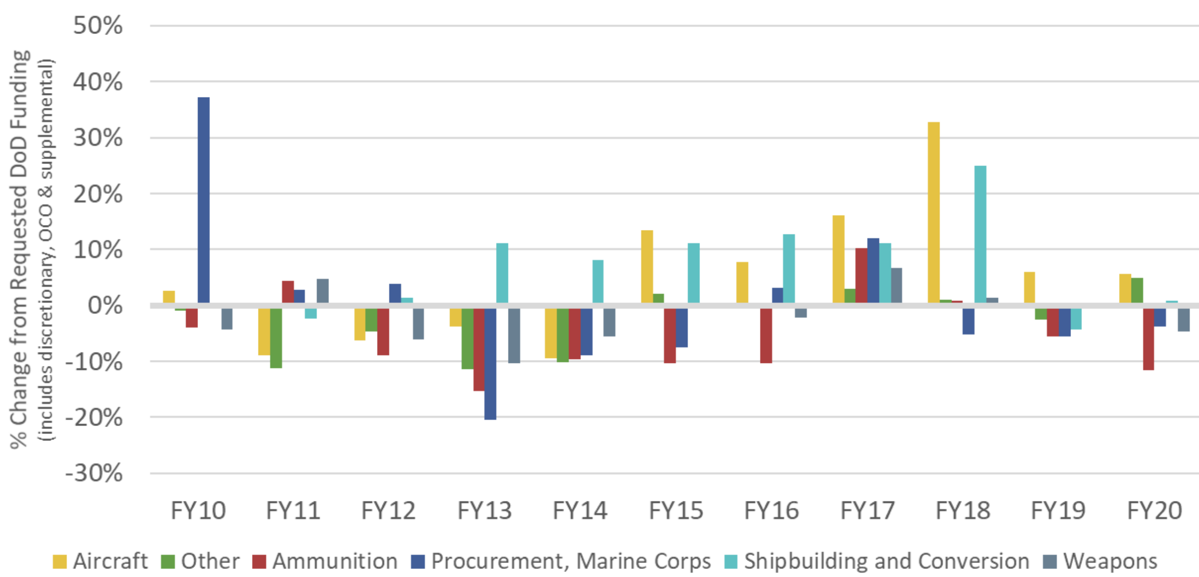


Figure 4. Navy Actual Procurement Funding vs. Requested Level by Account

Air Force procurement accounts received relatively smaller plus ups, if any, compared to their Army and Navy counterparts, as shown in Table 1. On average, Congress increased funding for the Air Force aircraft procurement account by an average of 2.2%, less than the increases for both the Army and Navy's aircraft accounts. After initially receiving nearly 9% more funding than requested when the account was created in FY 2016, Air Force space procurement has received 2.9% less funding than the requested level, on average. The Air Force's missile procurement account was the account cut most by Congress over the selected time frame, receiving an average of 5.7% less funding than requested. Congressional funding exceeded the requested level only once over the 11-year period, in FY 2015, as depicted in Figure 5.

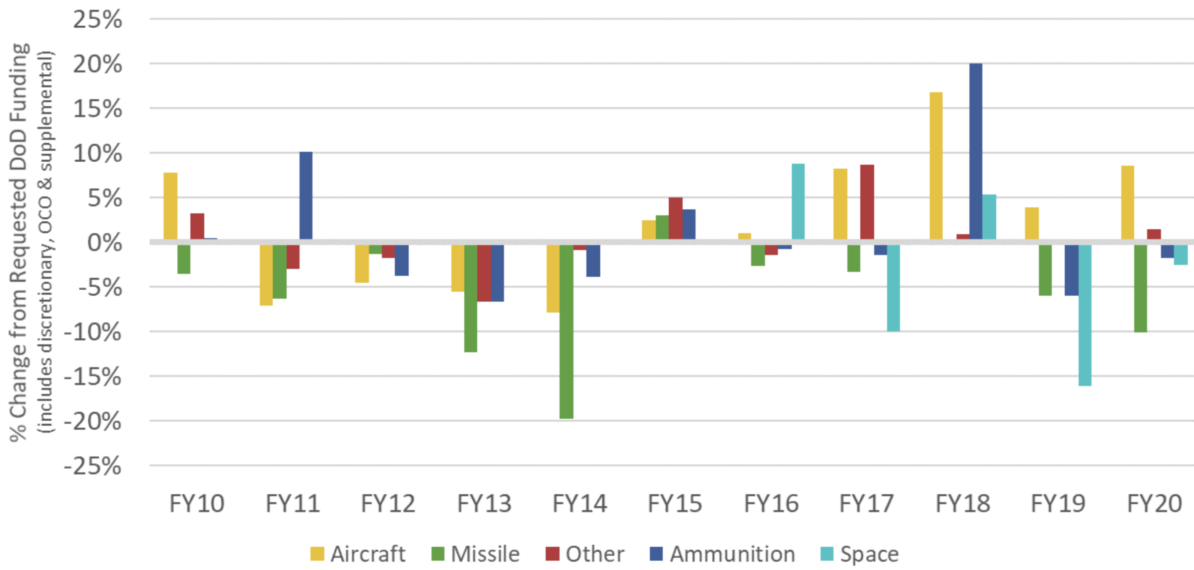


Figure 5. Air Force Actual Procurement Funding vs. Requested Level by Account

Procurement Funding by Category

Analyzing the difference between the requested and actual level of funding by category type provides a better impression of Congress's procurement priorities. Table 2 ranks the average percent change between the actual level of funding and the requested level over the FY 2010 to FY 2020 time frame for 10 distinct categories of procurements assigned by the authors. Based on this data, Congress has regularly increased funding for missile defense programs more than any other category at an average of nearly 18% over the requested level. Funding for missile defense exceeded the requested level in 10 out of the 11 years assessed and was almost 55% higher than what was requested in FY 2017 (see Figure 6).

Shipbuilding programs received the second largest increase on average of any distinct category, which could be due to the strong support from representatives for shipyard constituencies in Congress. Lawmakers also increased funding for aircraft and ground systems at an average of 4.6% and 3.6% above requested levels, respectively. The addition to ground systems was driven by a significant plus to the Army W&TCV account in FY 2018, as shown in the previous section. Four procurement categories received less than the requested level on average between FY 2010 and FY 2020: missiles and munitions; space systems; communications, sensors, and electronics; and defense-wide programs.

Table 2. Average Difference Between Requested and Actual Funding Levels by Procurement Category

Category	Average Percent Change
Missile Defense	17.8%
Support and Other	10.5%
Shipbuilding	8.9%
Aircraft	4.6%
Ground Systems	3.6%
Classified	0.5%
Missiles and Munitions	-1.1%
Space Systems	-3.7%
Comms, Sensors, and Electronics	-4.9%
Defense-Wide	-6.5%

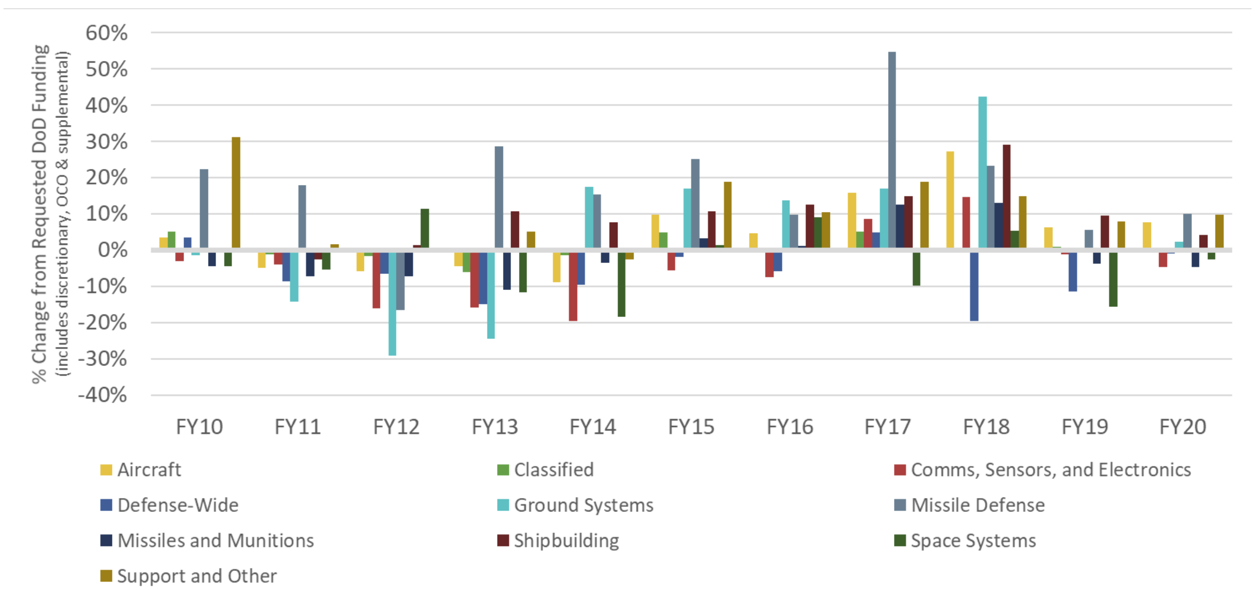


Figure 6. Actual Procurement Funding vs. Requested Level by Category

Analysis of Congressional Action on Research, Development, Test, and Evaluation Funding

Topline Research, Development, Test, and Evaluation Funding

As Figure 7 shows, topline RDT&E funding largely follows a similar pattern to trends in topline procurement in the FY 2010 to FY 2020 period, with exceptions in FY 2011 and FY 2020. Similar to procurement, RDT&E accounts also received a generous boost in FY 2018 as a result of the budget deal reached that year. Over the 11 fiscal years analyzed in this study, RDT&E funding was an average of 1.2% higher than the requested level.

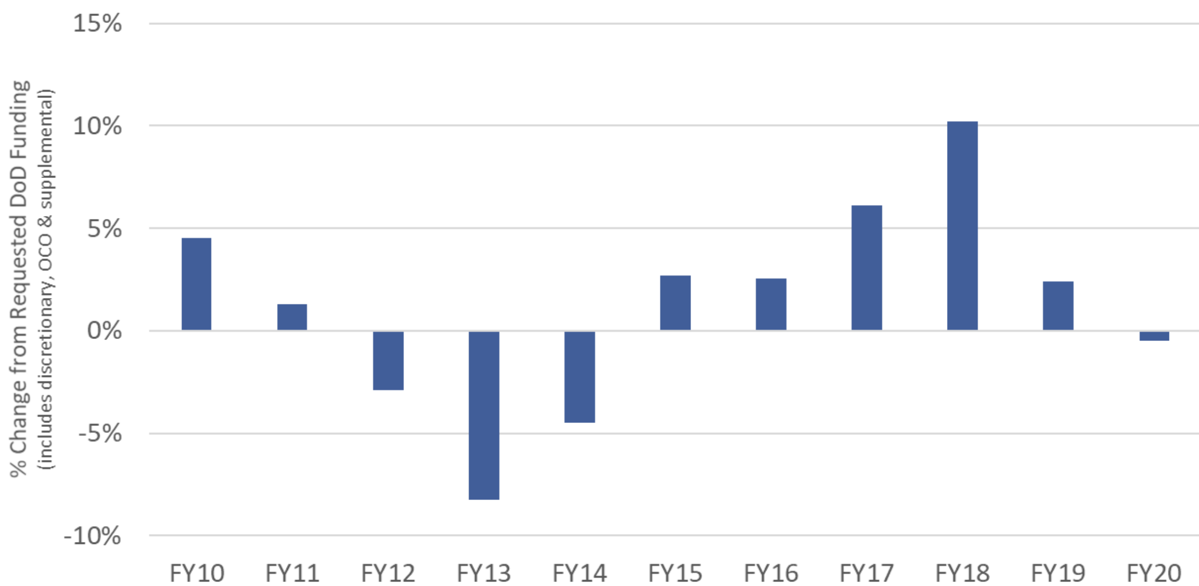


Figure 7. Actual Total RDT&E Funding vs. Requested Level

Research, Development, Test, and Evaluation Funding by Budget Activity

Funding for RDT&E is organized into different budget activities that “correspond to different phases of the development process” (Harrison and Daniels, 2020, p. 9). They include

- Basic Research (6.1)
- Applied Research (6.2)
- Advanced Technology Development (6.3)
- Advanced Component Development and Prototypes (6.4)
- System Development and Demonstration (6.5)
- Management Support (6.6)
- Operational Systems Development (6.7)

Funding for the first three budget activities is collectively referred to as Science and Technology (S&T) funds. In the FY 2021 request, the DoD requested funds for a new budget activity, Software & Digital Technology Pilot Programs (6.8), but because it did not appear until FY 2021, it does not fall within the range of data analyzed for this study.

Table 3 shows the average percent change between the requested and actual funding levels for each RDT&E budget activity between FY 2010 and FY 2020. On average, Management Support received 28.9% more funding than requested, while Applied Research received 16.5% more. Congressional support for Management Support funding was so strong that it received more funding than requested in all 11 years during the period of analysis, as shown in Figure 8. Lawmakers also provided overall S&T accounts an average of 9.1% more funding than requested over that time frame.

Table 3. Average Difference Between Requested and Actual Funding Levels by RDT&E Budget Activity

Budget Activity	**Average Percent Change**
Basic Research (6.1)	2.0%
Applied Research (6.2)	16.5%
Advanced Technology Development (6.3)	4.9%
Aggregate S&T Funding	*9.1%*
Advanced Component Development and Prototypes (6.4)	0.6%
System Development and Demonstration (6.5)	-6.6%
Management Support (6.6)	28.9%
Operational Systems Development (6.7)	-2.5%
Total RDT&E Average	*1.2%*

The System Development and Demonstration and Operational Systems Development budget activities both received less funding than the requested level on average over the FY 2010 to FY 2020 period. Congress provided an average of 6.6% less funding than requested for System Development and Demonstration, while Operational Systems Development received 2.5% less than the requested level, on average.[2] Advanced Component Development and Prototypes only received an average of 0.6% more than the request.

[2] In the FY 2010 R-1 justification spreadsheets, the DoD requested approximately $17.7 billion in RDT&E funding for "Other Programs" categorized under budget activity 99. The R-1 justification book in PDF form specified that amount as funding for classified programs. While the spreadsheets assigned that funding to budget activity 99, calculations based on the data in the justification book showed that just over $17.5 billion of the classified amount fell under budget activity 6.7, and the remainder fell under budget activity 6.6. The previous analysis incorporates the classified amount under 6.7 and 6.6 based on the justification book data.

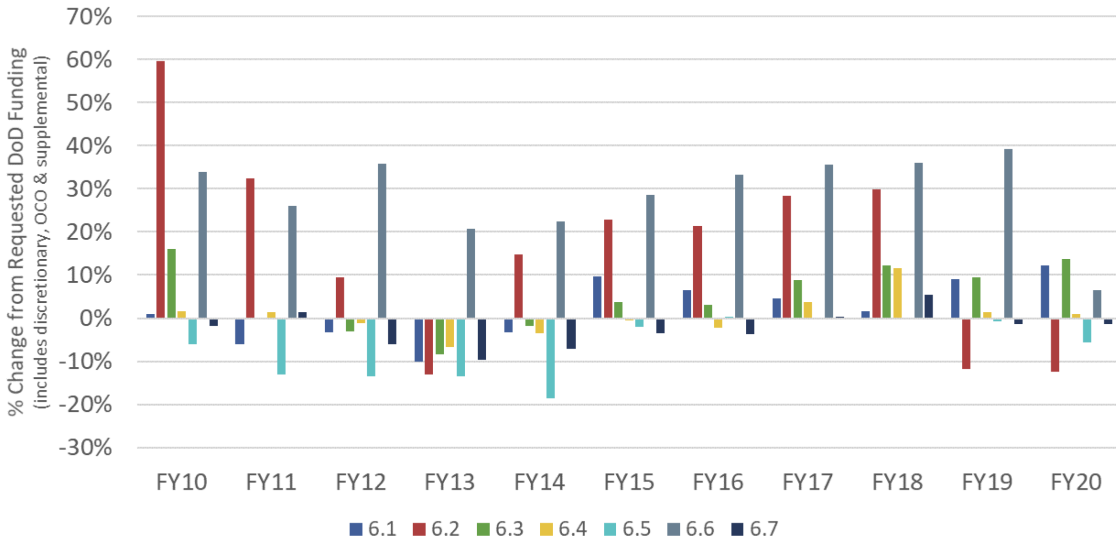


Figure 8. Actual RDT&E Funding vs. Requested Level by Budget Activity

Research, Development, Test, and Evaluation Funding by Military Department

When assessed by military department, defense-wide RDT&E programs received the greatest increase relative to the request at an average of 5.5%, followed by Army programs at 3.4%. As Figure 9 shows, Army RDT&E was over 10% higher than the requested level from FY 2016 to FY 2018 and surpassed 20% in FY 2018. Funding for the Navy and Air Force's RDT&E programs fell below the requested level on average between FY 2010 and FY 2020, at 0.4% and 1.5% lower than the request for each department, respectively.

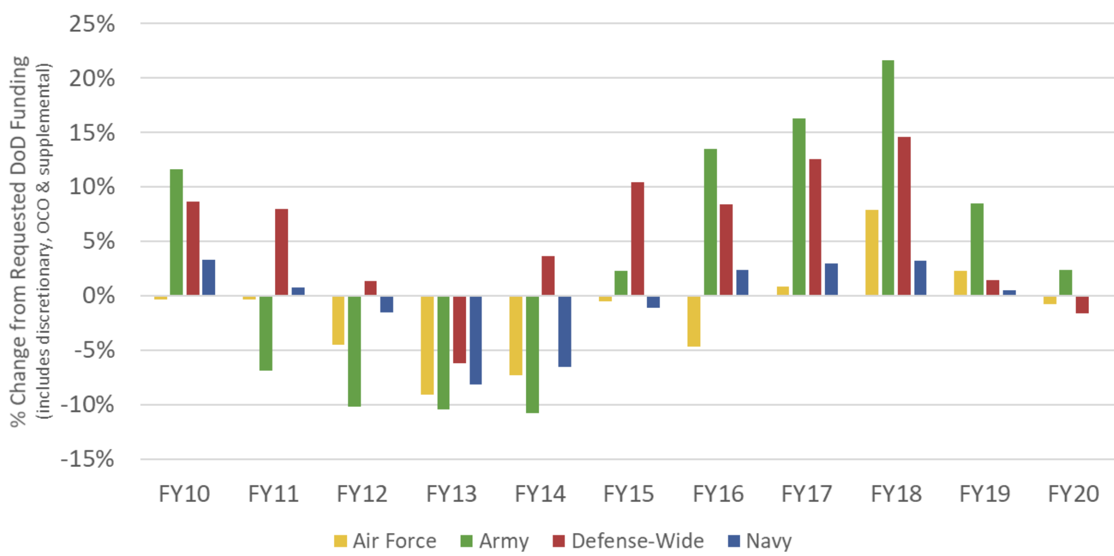


Figure 9. Actual RDT&E Funding vs. Requested Level by Military Department

Conclusion

The appropriations process is one of Congress's primary tools in exercising its oversight authority on the executive branch's defense policy and, more specifically, its defense acquisition plans. As the preceding analysis has illustrated, the process enables lawmakers to signal their priorities to the administration by increasing or decreasing the funding levels for programs in the annual budget request. This study reaches several findings based on that analysis:

- Congressional action on procurement and RDT&E accounts largely followed similar trends over the past decade, due in no small part to the impact of the BCA and subsequent budget deals. With minor exceptions, Congress underfunded (relative to the request) both procurement and RDT&E accounts following the passage of the BCA in FY 2011, and it appropriated more than requested beginning in FY 2015. The budget deal reached in 2018 (BBA 2018) led to a notable increase above the request for procurement and RDT&E accounts in that same year.
- Congress has clear favorites among procurement accounts. Programs for missile defense, shipbuilding, aircraft, and ground systems all received increases above the requested level on average over the FY 2010 to FY 2020 period.
- Lawmakers regularly increase RDT&E funding for Management Support (6.6) and S&T (6.1, 6.2, and 6.3). Support for other budget activities is not as strong.

While Congress is able to enact its own defense priorities via the appropriations process, concerns over strategy are not the only drivers of congressional preferences for some programs over others. The appropriations process also serves as a political tool for lawmakers to serve their constituencies, which may include defense factories that produce aircraft or shipyards constructing future vessels.

For the executive branch, the budget request can similarly serve a political purpose for enacting the administration's priorities. With the knowledge that Congress regularly increases funding for some accounts above the request, the DoD can be strategic in signaling its own plans to the Hill. For example, it could cut funding from some accounts in the budget request if it feels confident that Congress is likely to restore that money later in the process. Understanding these trends in congressional action can enable policy-makers, program managers, and industry leaders alike to improve planning and efficiency in the overall acquisition process.

References

Daniels, S. P. (2019). *How would sequestration impact DoD in FY 2020?* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/190515_Daniels_DBA_layout_FINAL_v2.pdf

Daniels, S. P., & Harrison, T. (2018, February 20). *Making sense of the Bipartisan Budget Act of 2018 and what it means for defense*. Center for Strategic and International Studies. https://www.csis.org/analysis/making-sense-bipartisan-budget-act-2018-and-what-it-means-defense

Harrison, T., & Daniels, S. P. (2020). *Analysis of the FY 2020 defense budget and its implications for FY 2021 and beyond*. Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/HarrisonDaniels_FY20DBA_v4.pdf?FA48YSbbC3Vj.aH7pKhpdIbU56OQ9QUZ

Harrison, T., & Daniels, S. P. (2020). *Analysis of the FY 2021 defense budget*. Center for Strategic and International Studies. http://defense360.csis.org/wp-content/uploads/2020/08/Analysis-of-the-FY-2021-Defense-Budget.pdf

Saturno, J., Heniff, Jr., B., & Lynch, M. S. (2016). *The congressional appropriations process: An introduction* (CRS Report No. R42388). Congressional Research Service. https://www.senate.gov/CRSpubs/8013e37d-4a09-46f0-b1e2-c14915d498a6.pdf

Defense Acquisition Trends 2021: A Preliminary Look

Rhys McCormick—is a Fellow with the Defense-Industrial Initiatives Group (DIIG) at the Center for Strategic and International Studies (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. [rmccormick@csis.org]

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. [gsanders@csis.org]

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 for Acquisition, Technology, and Logistics 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. [ahunter@csis.org]

Abstract

This report, the latest in an annual series, examines the trends in what the U.S. Department of Defense (DoD) is buying, how the DoD is buying it, and from whom the DoD is buying. This year's study focuses on the acquisition system's response to the 2018 National Defense Strategy's emphasis on peer and near-peer competition and the emergence of a new research and development (R&D) paradigm. This report looks at whether there has been a significant shift in the DoD's investment posture between platform portfolios and the composition of R&D spending between contracts and Other Transaction Authority (OTA) agreements. Additionally, this report includes analysis of the topline DoD contracting trends.

Introduction

This paper examines the notable trends in what, how, and from whom the Department of Defense (DoD) has been buying. These trends provide critical insights into the DoD's priorities and the industrial base's response to those priorities. These trends provide vital information describing the timely situation of defense acquisition. Since Fiscal Year (FY) 2020 was one of the first years in which the budget was developed following and in accordance with the release of the 2018 National Defense Strategy (NDS), the trends for FY2020 are particularly interesting. Identifying and discussing the shifts in the defense industrial base and acquisition system in response to the NDS are of particular intrigue given that previous Naval Postgraduate School (NPS)–funded research showed that it can take the acquisition system years to respond to changes in priorities.

This report uses the methodology used in Center for Strategic and International Studies (CSIS) reports on federal contracting. For over a decade, the Defense-Industrial Initiatives Group (DIIG) has issued a series of analytical reports on federal contract spending for national security by the government. These reports are built on Federal Procurement Data System (FPDS) data, which is downloaded in bulk from https://www.usaspending.gov/. DIIG now maintains its own database of federal spending, which includes data from 1990 to 2020. This database is a composite of

FPDS and DD350 data. For this report, the study team relied on FY2000 to FY2020 data. All dollar figures are in constant FY2020 dollars, using Office of Management and Budget (OMB) deflators. For additional information about the CSIS contracting data analysis methodology, see https://github.com/CSISdefense/Lookup-Tables.

For this paper, CSIS focused on the following research questions:

- **Area:** Has there been a significant shift in DoD investment between and within the areas of products, services, and research and development (R&D) to reflect the 2018 National Defense Strategy priorities?
- **Platform Portfolio:** Have there been significant changes across the different sectors of the defense industrial base?
- **Other Transaction Authorities (OTA):** What are the significant trends in OTA usage across the DoD and how does the growth of OTAs affect the DoD's technology development efforts?
- **Components:** Have there been significant shifts in defense contracting trends between the major DoD components?

DoD Contract Spending in a Budgetary Context

Defense contract spending continued to grow in FY2020, at nearly twice the rate of growth of overall defense spending. As shown in Figure 1, total defense contract obligations increased from $391.5 billion to $421.3 billion, an 8% increase. Comparatively, defense spending increased less than 4%, rising from approximately $707 billion to $732 billion. Defense contract spending accounted for 57.5% of defense spending in FY2020, the highest level this century. This continues the ongoing trends over the course of the defense contracting rebound where defense contract spending has grown at rates faster than the overall rate of growth in defense spending. Between FY2015 and FY20202, overall defense spending increased 18% compared to the 41% growth in defense contract obligations. Given that the defense budget is likely to remain relatively flat in the coming years, it will be difficult for the DoD to maintain this level of defense contract spending.

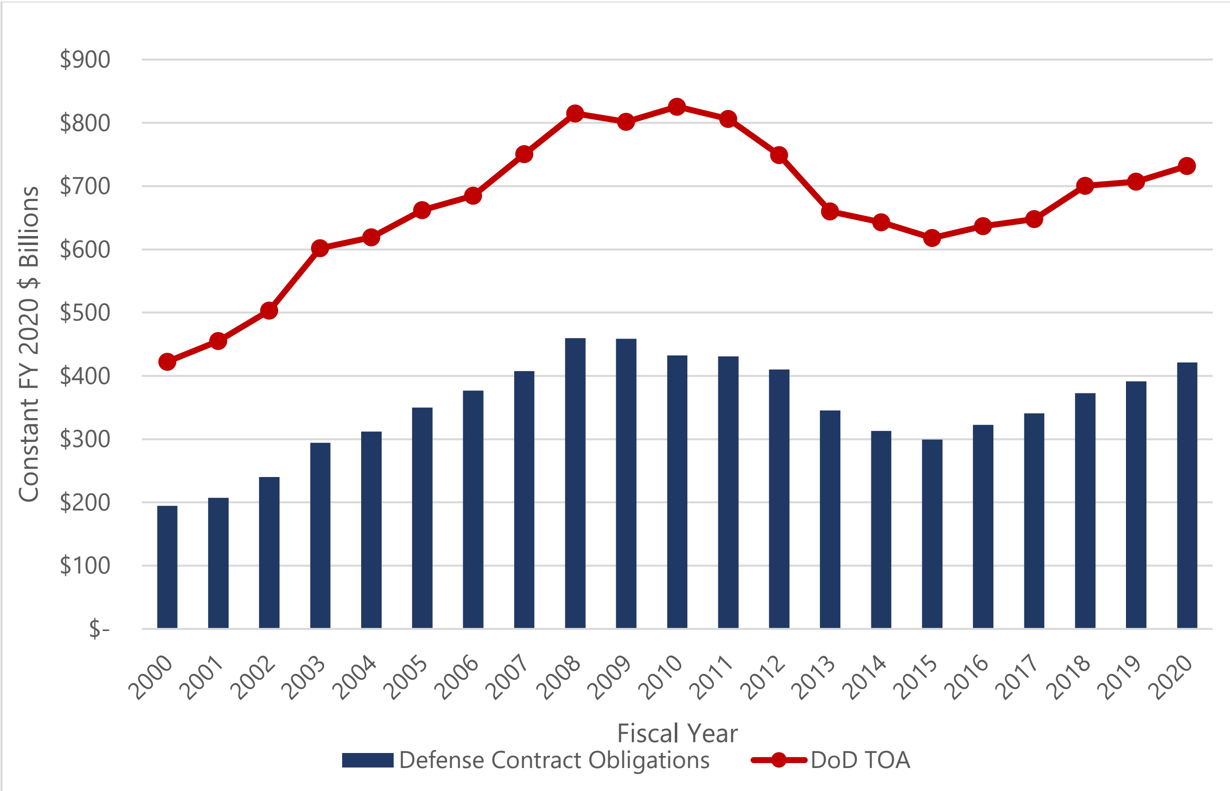


Figure 1. Defense Contract Obligations Versus Budget Authority, 2000–2020

Source: FPDS; Department of Defense, "National Defense Budget Estimates for Fiscal Year 2021 (Green Book)," Office of the Under Secretary of Defense (Comptroller), April 2020; Department of Defense, "Analysis of the FY 2021 Defense Budget" Todd Harrison and Seamus Daniel, CSIS (Washington, DC), August 2020; CSIS analysis

In addition to the growth in defense contract spending, OTA obligations have continued to grow as the DoD increasingly uses them in response to the FY2016 legislative changes in the National Defense Authorization Act (NDAA), aimed at incentivizing their usage. DoD OTA obligations increased 113% in FY2020, rising from $6.6 billion in FY2019 to $14.1 billion in FY2020. Between FY2015 and FY2020, DoD OTA obligations increased 2030%. Of note, while the sum of OTA dollars obligated increased 113% last year, the Sum of Base and All Options Value or potential total contract value of DoD OTA obligations only increased 1%. This could suggest that while OTAs are likely to continue to rise in future years, we might not see the same level of year-over-year growth that we've seen in recent years.

Figure 2 shows defense OTA obligations from FY2015 to FY2020.

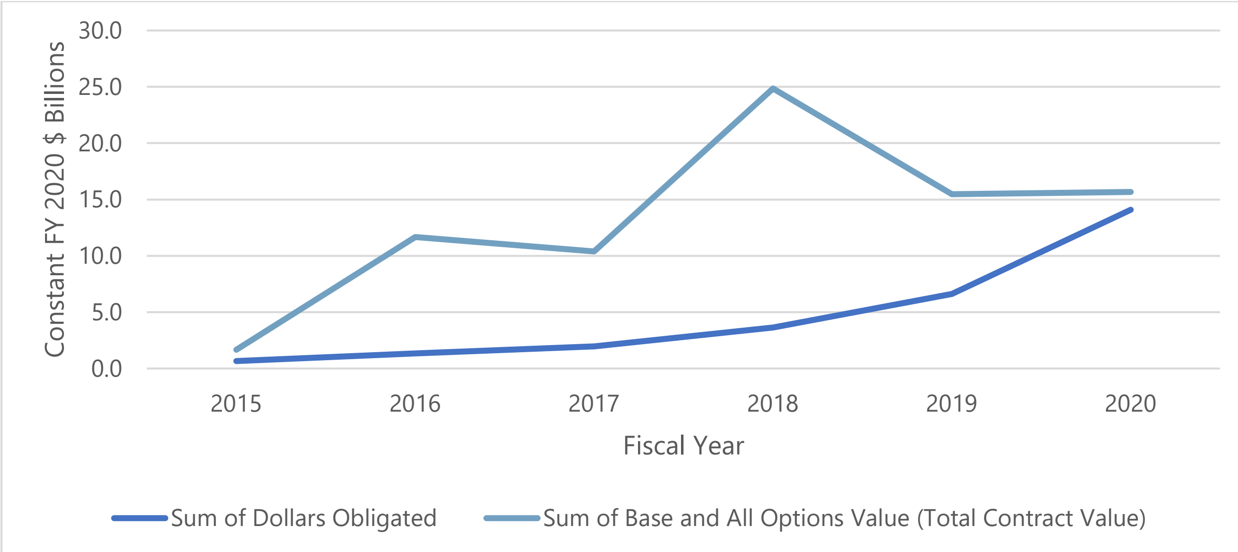


Figure 2. Defense OTA Obligations, 2015–2020

Source: FPDS; CSIS analysis

What Is the DoD Buying?

As previous CSIS research showed, you started to see some of the "emergent shifts in the composition of the DOD's investment portfolio" in the FY2019 contract data (McCormick, 2020). You continue to see some of these shifts in the DoD's investment portfolio, but there are also some outliers. Defense Products, which had started to slow down in FY2019, rebounded strongly in FY2020. Defense Products contract obligations increased 11% in FY2020, rising from $197.2 billion to $218 billion. Defense Services contract obligations continued to grow in FY2020, but at not at the same rate as before. Defense Services contract obligations increased from $164.0 billion in FY2019 to $173.1 billion, a 6% increase. Defense R&D contract obligations fell slightly in FY2020, falling from $30.3 billion to $30.2 billion, a 0.5% decline. However, that does not tell the full story as OTAs are increasingly supplementing contracts, particularly for R&D activities. Overall, if you include both R&D Contract and OTA Obligations, defense R&D spending increased 22% in FY2020, rising from $37.0 billion to $45.0 billion.

Figure 3 shows defense contract obligations by area from FY2000 to FY2020.

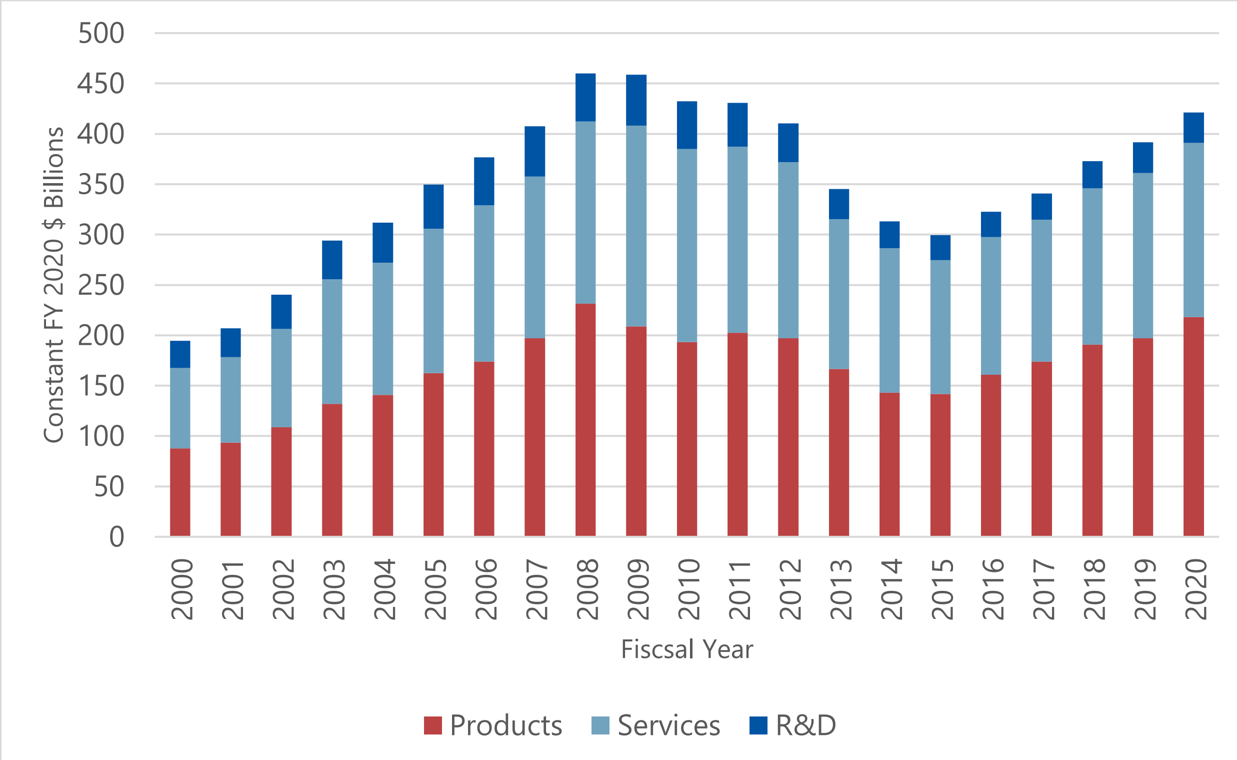


Figure 3. Defense Contract Obligations by Area, 2000–2020

Source: FPDS; CSIS analysis

Defense Contract Obligations by Stage of R&D

Previous CSIS analysis last year showed evidence that there is a "paradigm shift ongoing in the DoD as OTAs have become a core element in the DoD's approach to technology acquisition over the last five years." While we have seen some recovery in defense contracting in "the early and mid-stages of the development pipeline for major weapon systems" (McCormick, 2020a), that has not yet proven out for the later stages. However, there was evidence to suggest that "OTAs are partially supplanting contracts in the mid-to-late stages of the development pipeline for major weapon systems" (McCormick, 2020b).

There were mixed fortunes for the two early-stage R&D activities, Basic Research (6.1) and Applied Research (6.2) in FY2020. Defense Basic Research defense contract obligations declined 5% in FY2020, falling from $4.0 billion to $3.8 billion. Applied Research contract obligations increased from $7.9 billion to $8.0 billion, a 1% increase. Between FY2015 and FY2020, Basic Research and Applied Research contract obligations increased 10% and 18% respectively.

Similar to the early-stage R&D activities, there were divergent fortunes for the two mid-stage R&D activities, Advanced Technology Development (6.3) and Advanced Component Development & Prototypes (6.4), after several years of growth in both. Advanced Technology Development contract obligations declined slightly in FY2020, falling from $6.2 billion in FY2019 to $6.0 billion in FY2020, a 3% decline. Advanced Component Development & Prototypes contract obligations increased 8%, rising from $7.3 billion in FY2019 to $7.9 billion in FY2020. Between FY2015 and FY2020, Advanced Technology Development (6.3) and Advanced

Component Development & Prototypes (6.4), contract obligations have increased 40% and 84% respectively.

The data continue to suggest that “OTAs are partially supplanting contracts in the mid- to-late stages of the development pipeline for major weapon systems” (McCormick, 2020b). Last year, System Development & Demonstration (6.5) contract obligations declined 8% and have declined 18% between FY2015 and FY2020. Similarly, Operational Systems Development (6.7) contract obligations declined 7% in FY2020 and are down 39% from where they were in FY2020. Meanwhile, defense OTA R&D obligations went from $6.7 billion in FY2019 to $14.8 billion in FY2020, a 122% increase. Between FY2015 and FY2020, defense OTA R&D obligations have increased 1,850%.

Figure 4 shows defense contract obligations by stage of R&D from FY2000 to FY2020.

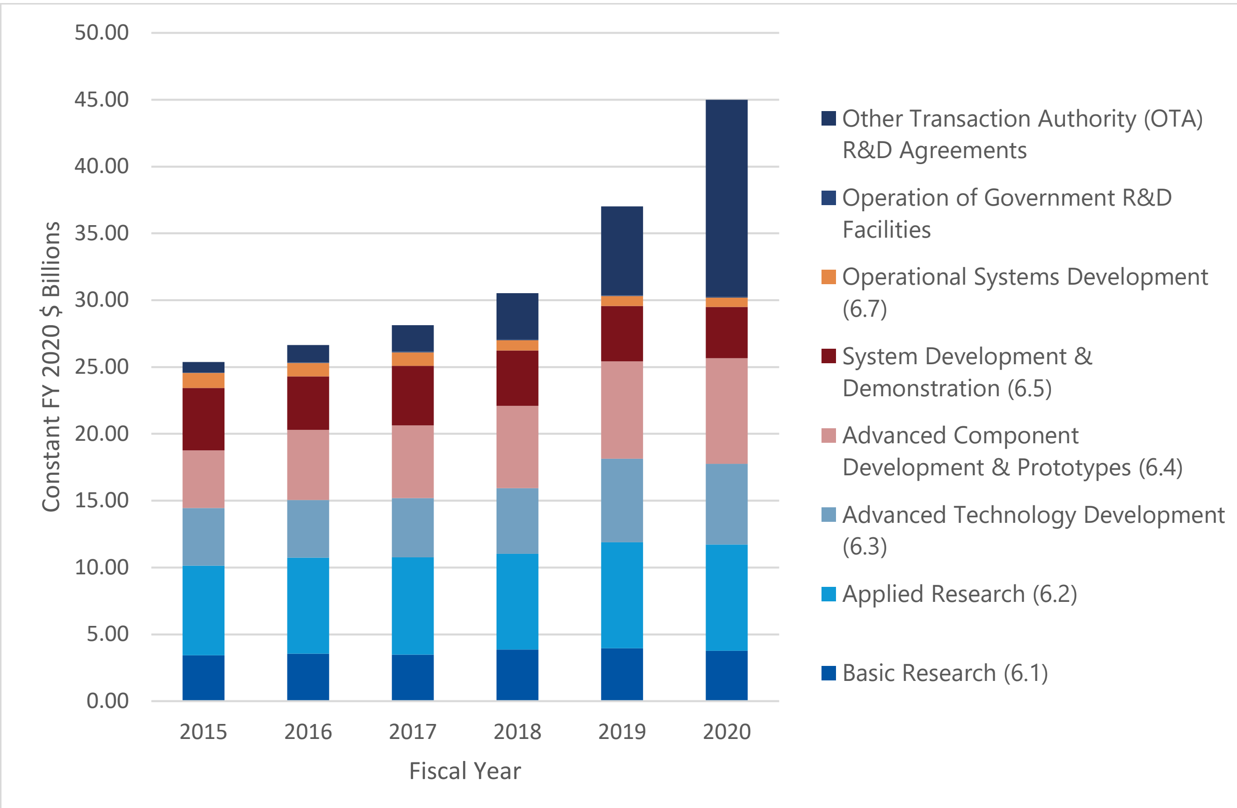


Figure 4. Defense R&D Obligations by Stage of R&D, 2015–2020

Source: FPDS; CSIS analysis

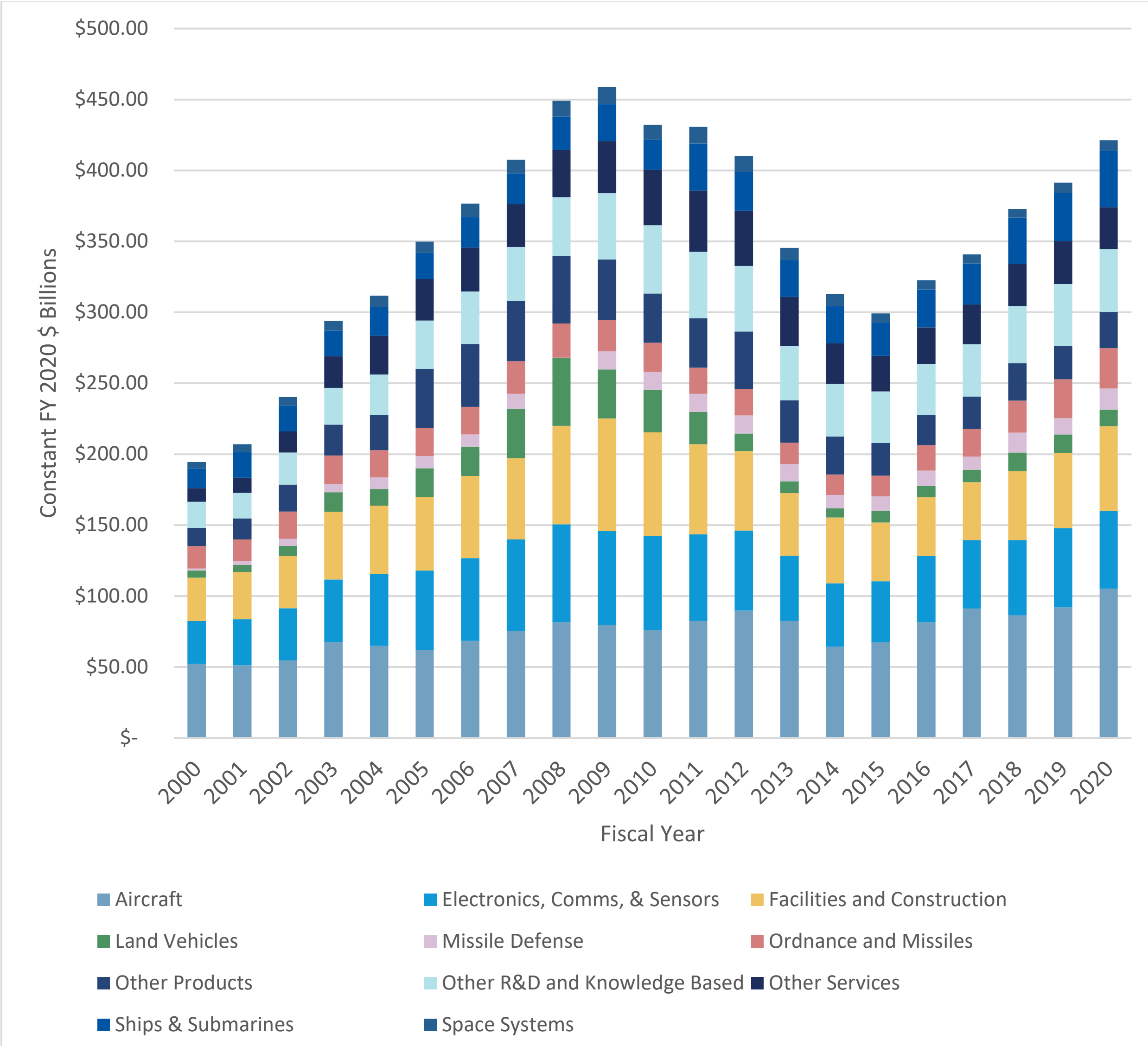


Figure 5. Defense Contract Obligations by Platform Portfolio, 2000–2020

Source: FPDS; CSIS analysis

Similar to previous years, the data show mixed trends for the platform portfolios emphasized in the NDS: Air and Missile Defense, Nuclear, Space, Cyberspace, and C4ISR (Department of Defense, 2018).

Over the course of the defense contracting rebound, Air and Missile Defense contract obligations have whipsawed back and forth between growth and declines, and FY2020 was no exception to that trend. After declining 18% between FY2018 and FY2019, Air and Missile Defense contract obligations rebounded in FY2020. Air and Missile Defense contract obligations increased from $11.5 billion in FY2019 to $14.8 billion in FY2020, a 29% increase. Between FY2015 and FY2020, Air and Missile Defense contract obligations have increased 45%.

After seeing substantial growth in FY2019, defense Space Systems contract obligations slightly came back down to Earth last year. Defense Space Systems contract obligations fell 1% last year, falling from $7.3 billion in FY2019 to $7.2 billion in FY2020. Despite this decline, defense Space Systems contract obligations are up 12% between FY2015 and FY2020.

Electronics, Communications, and Sensors (EC&S) had seen steady growth at the start of the defense contracting rebound, but that trend came to a halt in FY2020. EC&S contract obligations declined from $55.8 billion in FY2019 to $54.9 billion in FY2020, a 2% decline. EC&S contract obligations increased 28% between FY2015 and FY2020.

Ordnance and Missiles contract obligations continued to increase in FY2020, but not at the level higher than the overall rate of growth like they had done so in previous year. Last year, defense Ordnance and Missile contract obligations increased 4%, less than half the overall rate of growth. While Ordnance and Missiles grew at a rate less than the overall growth rate last year, Ordnance and Missiles contract obligations grew 95% between FY2015 and FY2020, over twice the overall rate of growth in defense contract obligations.

Outside of the platform portfolios emphasized in the NDS, there were interesting trends in the three major weapon system sectors: Aircraft; Ships & Submarines; and Land Vehicles.

After notable whiplash between growth and declines at the start of the defense contracting rebound, defense Aircraft contract obligations have continued their steady growth over the last 2 years. Defense Aircraft contract obligations increased from $92.0 billion in FY2019 to $105.1 billion in FY2020, a 14% increase. Between FY2015 and FY2020, defense Aircraft contract obligations have increased 56%.

The Ships & Submarines had been steadily growing throughout the defense contracting rebound but saw sizable growth in the last year. In FY2020, Ships & Submarines contract obligations increased 18%, rising from $34.0 billion to $40.0 billion, a new record high level. Ships & Submarines contract obligations increased 70% between FY2015 and FY2020.

The Land Vehicles sector had been on the rebound after experiencing the brunt of the cuts of sequestration and the defense drawdown but experienced a setback in FY2020. Defense Land Vehicles contract obligations fell from $13.2 billion in FY2019 to $11.7 billion in FY2020, an 11% decline. Despite this 1-year decline, Land Vehicles contract obligations increased 43% between FY2015 and FY2020, a rate slightly than the overall rate of growth.

Defense Components

Navy contracting obligations, which had been on the decline at the start of the defense contracting rebound, have continued their bounce back over the last 2 years. Navy contracting obligations increased 20% last year, rising from $124.5 billion in FY2019 to $150.0 billion in FY2020. As a share of total defense contract obligations, the Navy went from 31.8% to 35.6%, a 20-year high. Between FY2015 and FY2020, Navy contract obligations have increased 62%, the second-largest growth among DoD components only behind the Missile Defense Agency.

The Army continued its slow but steady growth path that it has been on over the course of most of the defense contracting rebound. Army contract obligations increased from $96.7 billion in FY2019 to $100.1 billion in FY2019, a 3% increase. As a share of total defense contract obligations, the Army continued to fall slightly, from 24.7% to 23.8%. Between FY2015 and FY2020, Army contract obligations have increased 26%, the lowest among the three military services and well below the overall growth in defense contract obligations (41%).

After 2 years of steady growth between FY2018 and FY2019, Air Force contract obligations stalled out in FY2020. Air Force contract obligations went from $77.5 billion in FY2019 to $77.8 billion in FY2020, a 0.5% increase. As a share of total defense contract obligations, the Air Force fell from 19.8% to 18.5%. Between FY2015 and FY2020, Air Force contract obligations increased 26%.

The Missile Defense Agency (MDA) continued on its ascendent path that it has been on the last few years. MDA contract obligations increased from $9.0 billion in FY2019 to $12.3

billion in FY2020, a 37% increase. Between FY2015 and FY2020, MDA contract obligations increased 141%, the highest level among all DoD components and nearly 3.5 times the rate of overall growth in defense contracting obligations.

The Defense Logistics Agency (DLA) continued to decline, as it has done so over the past few years. DLA contract obligations declined 7% last year, falling from $45.1 billion in FY2019 to $41.9 billion in FY2020. Despite the declines in recent years, DLA contract obligations are still up 24% in FY2020 from where they were in FY2015.

Figure 6 shows defense contract obligations by component from FY2000 to FY2020.

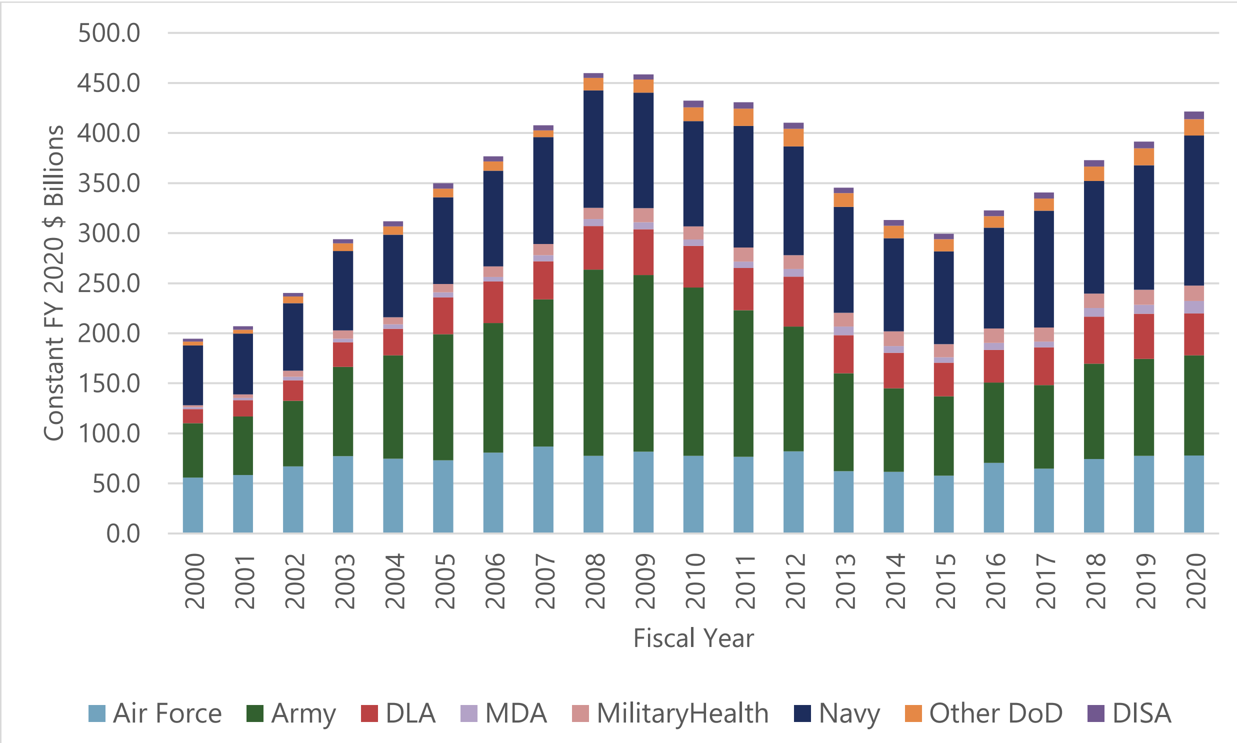


Figure 6. Defense Contract Obligations by Component, 2000–2020

Source: FPDS; CSIS analysis

Conclusion

Defense Contract Obligations Continued to Grow Even as Defense Budget Leveled Off

Defense contract spending continues to grow in FY2020 even as defense spending has started to level off some. In FY2020, defense contract spending grew at nearly twice the overall rate of growth in defense spending. Defense contract spending accounted for 57.5% of defense spending in FY2020, the highest level this century. If the defense budget remains relatively flat in the coming years as currently expected, it will be difficult for the DoD to maintain this level of defense contract spending.

Mixed Trends in the 2018 National Defense Strategy Priority Platform Portfolios

The data show mixed trends for the platform portfolios emphasized in the NDS (Air and Missile Defense, Nuclear, Space, Cyberspace, and C4ISR).

EC&S declined 2% in FY2020 after several years of growth. Despite this decline, EC&S contract obligations increased 28% between FY2015 and FY2020.

Space Systems fell 1% in FY2020 after experiencing growth the previous year.

Air and Missile Defense contract obligations continued to whipsaw over the course of the defense contracting rebound, rebounding again in FY2020, seeing a 29% increase after a sizeable decrease the preceding year.

Mixed Growth in Early and Mid-Stage of the Weapon Systems Pipeline

The data show mixed trends in both the early and mid-stages of the weapon systems pipeline.

In the early R&D stages, Basic Research (6.1) contract obligations declined 5% while Applied Research (6.2) increased 1%. In the mid-stage, Advanced Technology Development (6.3) and contract obligations declined 3% while Advanced Component Development & Prototypes (6.4) contract obligations increased 8%.

OTA Usage Continues Increasing, Supplementing Contracts in the Mid-to-Late Stages of the Development Pipeline for Major Weapon Systems

OTA usage across the DoD continues to surge in response to the FY 2016 NDAA legislative changes that aimed to incentivize their usage. In FY2020, defense R&D OTA obligations increased from $6.6 billion to $14.1 billion, a 113% increase. Since FY2015, defense R&D OTA obligations have increased 2,030%.

The impact of R&D OTAs is most notable in the late stages of the development pipeline for major weapon systems, where contracting has largely fallen off a cliff. System Development & Demonstration (6.5) and Operational Systems Development (6.7) contract obligations declined 8% and 7% respectively last year. However, comparatively, defense OTA R&D obligations went from $6.7 billion in FY2019 to $14.8 billion in FY2020, a 122% increase. While not all of that $14 billion goes to late-stage weapon systems development, a sizable percentage does.

As highlighted in seven recent CSIS reports, the growth in OTAs is massive, and it is increasingly clear that OTAs are supplementing contracts in certain traditional defense acquisition activities and that a new R&D paradigm is emerging. However, the full implications of that final paradigm shift for both government and industry remains unknown.

Navy and Air Force Bounce Back; Army Slows Down, but Continues Growing

Navy contract obligations continued their rebound into FY2020, increasing 20%. As a share of total defense contract obligations, the Navy rose from 31.8% to 35.6%.

Air Force contracting continued its whipsaw over the course of defense contracting rebound, stalling out in FY2020 after 2 years of steady growth. Air Force contract obligations went from $77.5 billion in FY2019 to $77.8 billion in FY2020, a 0.5% increase.

The Army continued its slow but steady growth path that it has been on over the course of most of the defense contracting rebound, increasing 3% in FY2020. Between FY 2015 and FY2020, Army contract obligations have increased 26%, the lowest among the three military services and well below the overall growth in defense contract obligations (41%).

Final Thoughts

The overall picture that emerges from a detailed examination of the 2020 data reveals important signals for the future of the defense acquisition system. First, it appears that contract spending may either be approaching or has reached a local peak in the typical boom and bust cycle of defense acquisition. Second, the 5-year rebound in the acquisition system made some progress in addressing the priorities of the NDS, but few if any of the NDS priority areas really stand out in their performance compared to areas of the defense acquisition system that were not so prioritized. These two factors create a challenge for the department as resources tighten, which may compel more profound shifts in support of various sectors of the acquisition system. Lastly, the growth of OTAs has substantially reshaped the system development pipeline for the DoD, altering how the department ingests new technology and who participates and controls that process. This shift will likely reshape relationships in the defense industry as well over time, but the nature of that reshaping and what new centers of control emerge is very much open to debate.

References

DoD. (2018). *Summary of the 2018 National Defense Strategy of the United States of America: Sharpening the American military's competitive edge*. https://csis-prod.s3.amazonaws.com/s3fs-public/publication/181120_NewEntrantsandSmallBusiness_WEB.pdf

McCormick, R. (2020a). *Defense acquisition trends, 2020: Topline DoD trends.* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/201007_McCormick_Defense_Acquisition_Trends_2020.pdf

McCormick, R. (2020b). *Department of Defense other transaction authority trends: A new R&D paradigm?* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/201207_McCormick_DoD_OTA.pdf

Trends in Department of Defense Other Transaction Authority Usage: A Preliminary Look

Rhys McCormick—is a Fellow with the Defense-Industrial Initiatives Group (DIIG) at the Center for Strategic and International Studies (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. [rmccormick@csis.org]

Abstract

The federal government's use of Other Transaction Authority (OTA) agreements has exploded in recent years, thanks in large part due to a surge in popularity within the Department of Defense (DoD). Neither a contract, grant, or cooperative agreement, OTAs are an acquisition approach that enable certain federal agencies to access goods and services outside of the traditional acquisition system. This research examines the trends in OTA usage across the DoD to provide insights into what the DoD is using OTAs for, how they are spending under an OTA, and to whom the majority of OTA obligations go.

Introduction

Other Transaction Authorities (OTAs) have become an increasingly popular tool across the DoD as senior Pentagon officials and congressional leadership seek ways to guide the defense acquisition enterprise as it seeks to maintain continued U.S. technological superiority against global competitors like China and Russia. Subsequently, DoD OTA obligations increased from $0.76 billion to $16.18 billion between Fiscal Year (FY) 2015 and FY2020. Neither contracts, grants, nor cooperative agreements, OTAs are a more flexible acquisition approach that enables specific federal agencies to access goods and services outside of traditional acquisition processes.[1] These authorities give these agencies greater flexibility and customization in designing appropriate acquisition approaches, but they are not without risk. OTAs are often more restricted to a specific set of activities, largely centered around research and development (R&D), and require a more skilled acquisition workforce to design and execute these activities that may lack the necessary familiarity and training amongst the broader community.

The DoD has had some form of OTA authority since 1989 (when the Defense Advanced Research Projects Agency [DARPA] was given the authority to enter into OTAs), so what explains its increased popularity in recent years? The DoD's recent interest in OTAs is heavily driven by the FY2015 and FY2016 National Defense Authorization Act (NDAA) expanding what the DoD can use OTAs to accomplish. Section 812 of the FY2015 NDAA expanded the range of what types of prototypes could be perused under an OTA, while Section 815 of the FY2015 NDAA "expanded DoD's OTA authority by making DoD's OTA authority permanent, modifying the definition of nontraditional defense contractor, and allowing DoD to issue follow-on production contracts for OTA prototypes" (McCormick, 2019). In the FY2016 NDAA conference report, House and Senate conferees noted that the expansion of the DoD's OTA authorities was designed to "support Department of Defense efforts to access new source of technical

[1] Besides the DoD, the following 10 federal agencies have some form of OTA authority: Advanced Research Projects Agency–Energy, Department of Energy, Department of Health and Human Services, Department of Homeland Security, Department of Transportation, Federal Aviation Administration, Domestic Nuclear Detection Office, National Aeronautics and Space Administration, National Institute of Health, and the Transportation Security Agency.

innovation" by making OTAs "attractive to firms and organizations that do not usually participate in government contracting due to the typical overhead burden and 'one size fits all' rules" (H.R. Rep. No. 114-270, 2015).

The following paper examines the notable trends in the DoD OTA authorities since the FY2015 and FY2016 NDAA statuary changes expanded the DoD's OTA authorities and seeks to answer the following research questions:

- What are the topline trends in the DoD's OTA usage?
- What is the DoD procuring using OTAs?
- How are the different DoD components using OTAs?
- What is the extent of competition for DoD OTA awards?
- From whom is the DoD procuring using OTAs?

This brief builds and expands on the methodology used in other CSIS reports that employ data from the Federal Procurement Data System (FPDS). Unlike other Defense-Industrial Initiatives Group reports on federal contracting, which rely on bulk data downloaded from https://www.usaspending.gov/, this brief relies on the data downloaded directly from https://beta.sam.gov/ and https://www.fpds.gov/. All dollar figures are reported in constant FY2020 dollars, using Office of Management and Budget (OMB) deflators.

Topline Trends

The data show that the rapid growth in the DoD's usage of OTAs did not slow down in FY2020. DoD OTA obligations increased 113% last year, rising from $7.6 billion in FY2019 to $16.2 billion in FY2020. Between FY2015 and FY2020, DoD OTA obligations have increased from $0.76 billion to $16.2 billion, a 2,030% increase (see Figure 1). Of note, while the sum of OTA dollars obligated increased 113% last year, the Sum of Base and All Options Value or potential total contract value of DoD OTA obligations only increased 1%. This could suggest that while OTAs are likely to continue to rise in future years, we might not see the same level of year-over-year growth that we have seen in recent years.

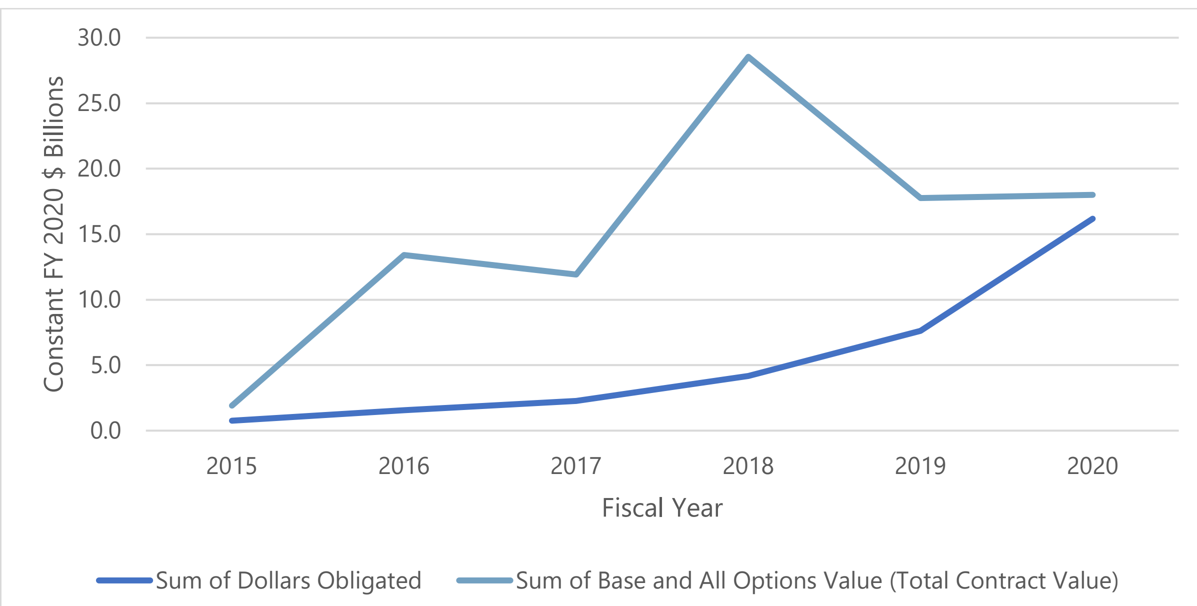


Figure 1. Defense OTA Obligations, 2015–2020 (Source: FPDS; CSIS analysis)

What Is the DoD Buying With OTAs?

Given the purpose of OTAs, it is not surprising that the DoD has predominantly used OTAs for R&D activities, but OTAs are not unique to R&D. Between FY2015 and FY2020, 89% of total DoD OTA obligations were awarded for R&D compared to 8% for Products and 3% for Services.

OTA R&D obligations increased from $6.7 billion in FY2019 to $14.8 billion in FY2020, a 122% increase. Between FY2015 and FY2020, DoD OTA R&D obligations increased 1,850%.

OTA Products contract obligations increased from $0.6 billion in FY2019 to $0.95 billion in FY2020, a 59% increase. Between FY2015 and FY2020, DoD OTA Products obligations increased 43,654%.

OTA Services contract obligations increased from $0.4 billion to $0.5 billion last year, a 29% increase. DoD OTA services obligations are up 58,761% between FY2015 and FY2020.

Figure 2 shows defense OTA obligations by area from FY2015 to FY2020.

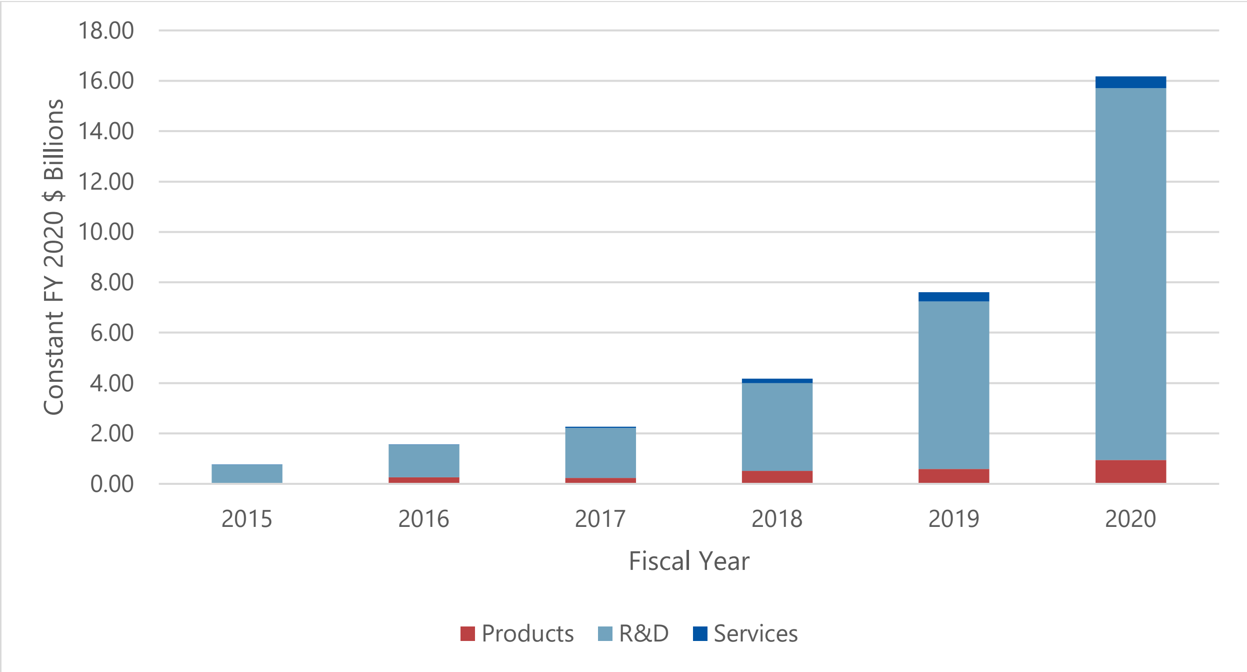


Figure 2. Defense OTA Obligations by Area, 2015–2020 (Source: FPDS; CSIS analysis)

As shown in Figure 3, unsurprisingly the predominance of DoD OTA obligations in recent years have gone to prototypes efforts. It is only in recent years that the DoD has received the authority to award follow-on production OTA agreements, so it is not too surprising to see that production OTAs are still in their infancy. While there is not much to this data at this point in time, this will be an important area that CSIS will continue to monitor into the future as the DoD evolves its approach to the emerging new R&D funding paradigm.

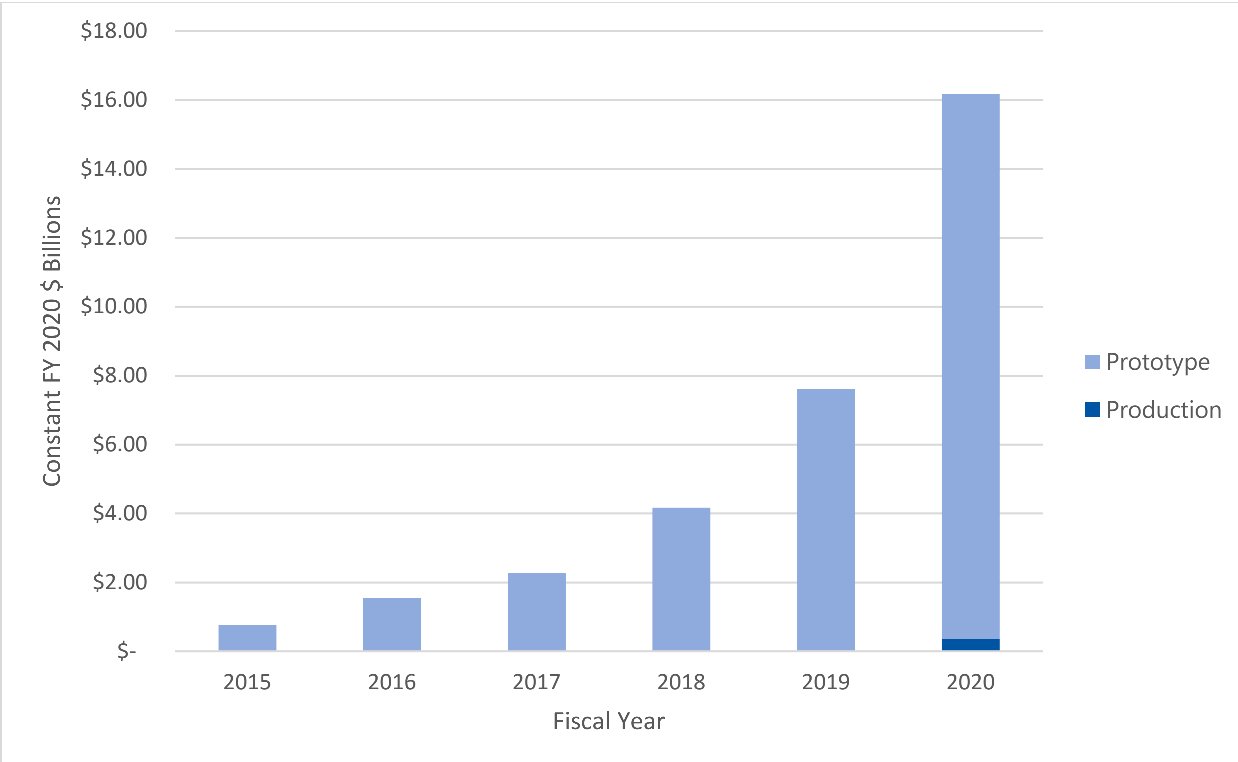


Figure 3. Defense OTA Obligations by Type of Agreement, 2015–2020 (Source: FPDS; CSIS analysis)

How Is the DoD Using OTAs for R&D?

Previous CSIS research showed that "OTAs are taking on a more major role in the mid-to-late stages of the development pipeline for major weapon systems" (McCormick, 2000a). While this largely held true into FY2020, there were several notable developments and shifts in the composition of the DoD's OTA R&D portfolio.

In the mid-stage R&D activities, there was significant growth in Advanced Technology Development (6.3), while Advanced Component Development & Prototypes (6.4) actually declined slightly. Advanced Technology Development OTA obligations increased from $0.6 billion in FY2019 to $8.0 billion, a 1,196% increase. Meanwhile, Advanced Component Development & Prototypes OTA obligations declined 1% in FY2020, falling from $3.9 billion to $3.8 billion.

In the later-stages of the weapon-systems development pipeline, there was actually a drop-off from previous levels. System Development & Demonstration (6.5) OTA obligations declined 37%, totaling $0.5 billion in FY2020 compared to $0.8 billion in FY2019. This decline was somewhat offset by the gains in OTA obligations Operational Systems Development (6.7), but Operational Systems Development still accounts for less than 1% of all DoD OTA obligations.

Finally, both Basic Research (6.1) and Applied Research (6.1) saw increased OTA obligations in 2020, but the two early-stage R&D activities both fell as a share of overall defense OTA spending. Basic Research OTA obligations increased from $0.3 billion to $0.5 billion, a 50% increase. However, Basic Research fell as a share of overall defense obligations from 5% to 3%. Applied Research saw an 87% increase in OTA obligations in FY2020 from FY2019 but fell as a share of overall defense obligations from 15% to 13%.

Figure 4 shows defense OTA obligations by stage of R&D from FY2015 to FY2020.

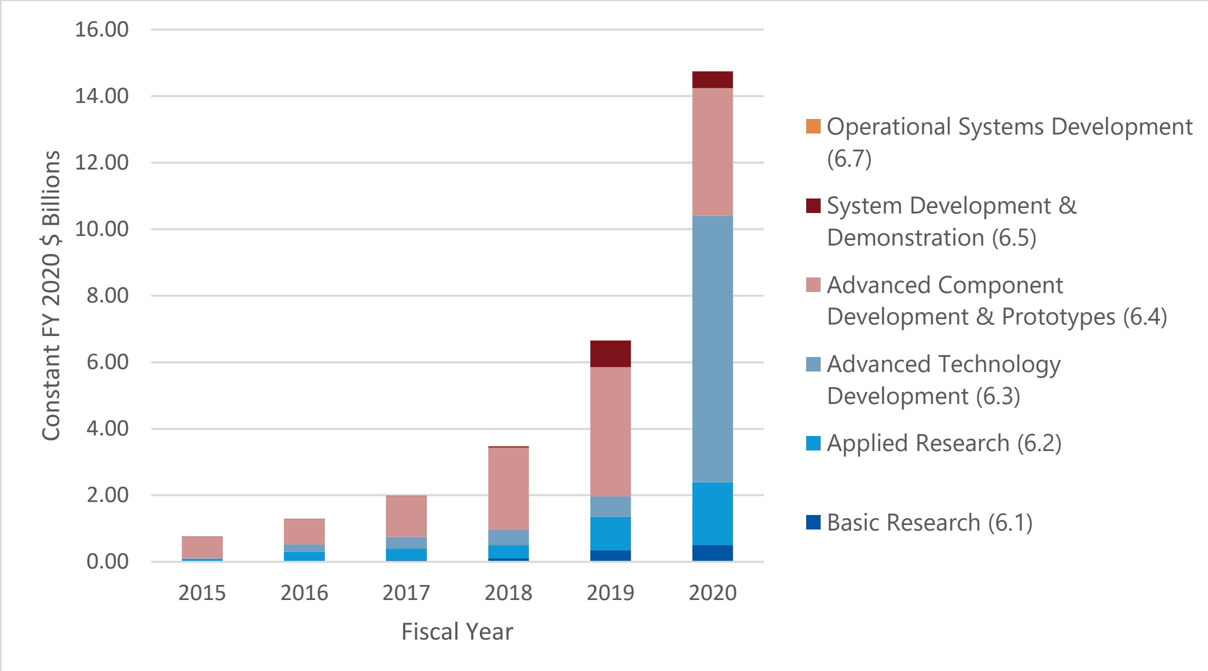


Figure 4. Defense OTA Obligations by Stage of R&D, 2015–2020 (Source: FPDS; CSIS analysis)

DoD OTA Awards by Platform Portfolio

As shown in Figure 5, several trends emerge in analyzing DoD OTA obligations by platform portfolio.

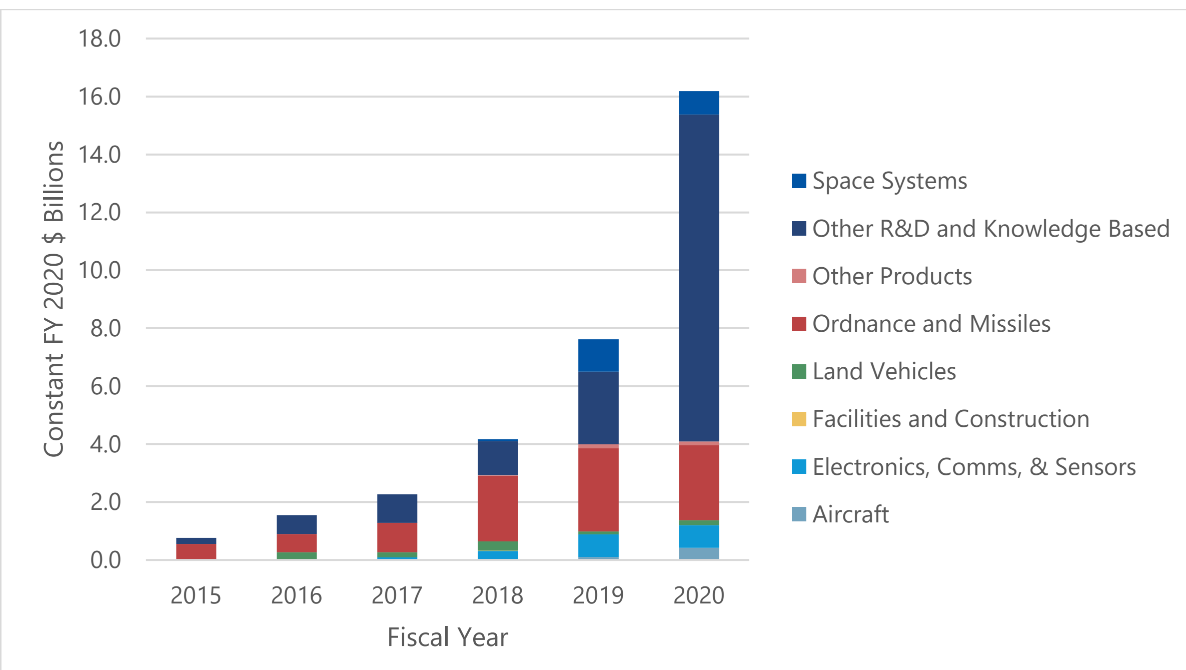


Figure 5. Defense OTA Obligations by Platform Portfolio, 2015–2020 (Source: FPDS; CSIS analysis)

Aircraft OTA obligations increased from $0.1 billion in FY2019 to $0.4 billion in FY2020, a 3,365% increase.

Space Systems, which had been on an uptick in recent years, saw a decline in OTA obligations last year. Defense Space Systems OTA obligations declined 27% in FY2020, falling from $1.1 billion to $0.8 billion.

Ordnance and Missiles, the predominant OTA platform portfolio prior to the recent statutory changes, saw a decline in OTA obligations in FY2020 but remains the second largest platform portfolio. Ordnance and Missile OTA obligations declined from $2.9 billion in FY2019 to $2.6 billion in FY2020, a 10% decline. However, Ordnance and Missiles OTA obligations are still up 373% between FY2020. As a share of overall defense OTA obligations, Ordnance and Missiles fell from 72% in FY2015 to 16% in FY2020.

Other R&D and Knowledge Based, previously the second-largest platform, succeeded Ordnance and Missiles as the largest OTA platform portfolio in FY2020.[2] Other R&D and Knowledge Based contract obligations increased a staggering 350% last year. Total Other R&D and Knowledge Based OTA obligations increased from $2.5 billion to $11.3 billion. This increase was primarily driven by R&D- DEFENSE OTHER: SERVICES (ADVANCED DEVELOPMENT), which saw an increase in OTA obligations from $0.14 billion in FY2019 to $7.2 billion in FY2020, a 5,013% increase. Of note, the following product or service codes comprised the top five Other R&D and Knowledge Based accounts ordered by total OTA obligations between FY2015 and FY2020:

1. R&D- DEFENSE OTHER: SERVICES (ADVANCED DEVELOPMENT)
2. R&D- DEFENSE OTHER: OTHER (ENGINEERING DEVELOPMENT)
3. EDUCATION/TRAINING- COMBAT
4. R&D- MEDICAL: BIOMEDICAL (APPLIED RESEARCH/EXPLORATORY DEVELOPMENT)
5. R&D- MEDICAL: BIOMEDICAL (ADVANCED DEVELOPMENT)

How Are the Different DoD Components Using OTAs?

The Army remains the leader in OTA usage across DoD components, but other components have also seen substantial increases in recent years. In FY2020, Army OTA obligations increased from $5.1 billion to $13.2 billion, a 161% increase. Army OTA obligations have increased 1,840% since FY2015. After seeing an uptick in OTA obligations in FY2019, Air Force OTA obligations declined last year. Air Force OTA obligations declined 20% last year, falling from $1.7 billion in FY2019 to $1.3 billion in FY2020. After a slow start to OTA usage, the Navy has seen a significant increase in OTA usage over the last 2 years. Navy OTA obligations increased from $0.2 billion in FY2019 to $0.6 billion in FY2020, a 253% increase. Between FY2015 and FY2020, Navy OTA obligations increased 24,633%. There was a notable increase in OTA obligations last year for "Other DoD," largely driven by the Washington Headquarters Services (WHS).

Between FY2015 and FY2020, the Army accounted for 76% of total defense OTA obligations compared to the Air Force and DARPA, which both accounted for 12% while the Navy accounted for approximately 3%. In FY2020, the Army accounted for 82% of defense OTA obligations; the Air Force accounted for 8% of defense OTA obligations last year after accounting for 22% the previous year; DARPA fell to 2%; and the Navy rose slightly to 4%.

---

[2] *Other R&D and Knowledge Based* serves as a catch-all category that doesn't fit into platform portfolios but includes a wide range of activities that include biomedical, technical services, and other R&D activities.

Figure 6 shows defense OTA obligations by customer from FY2015 to FY2020.

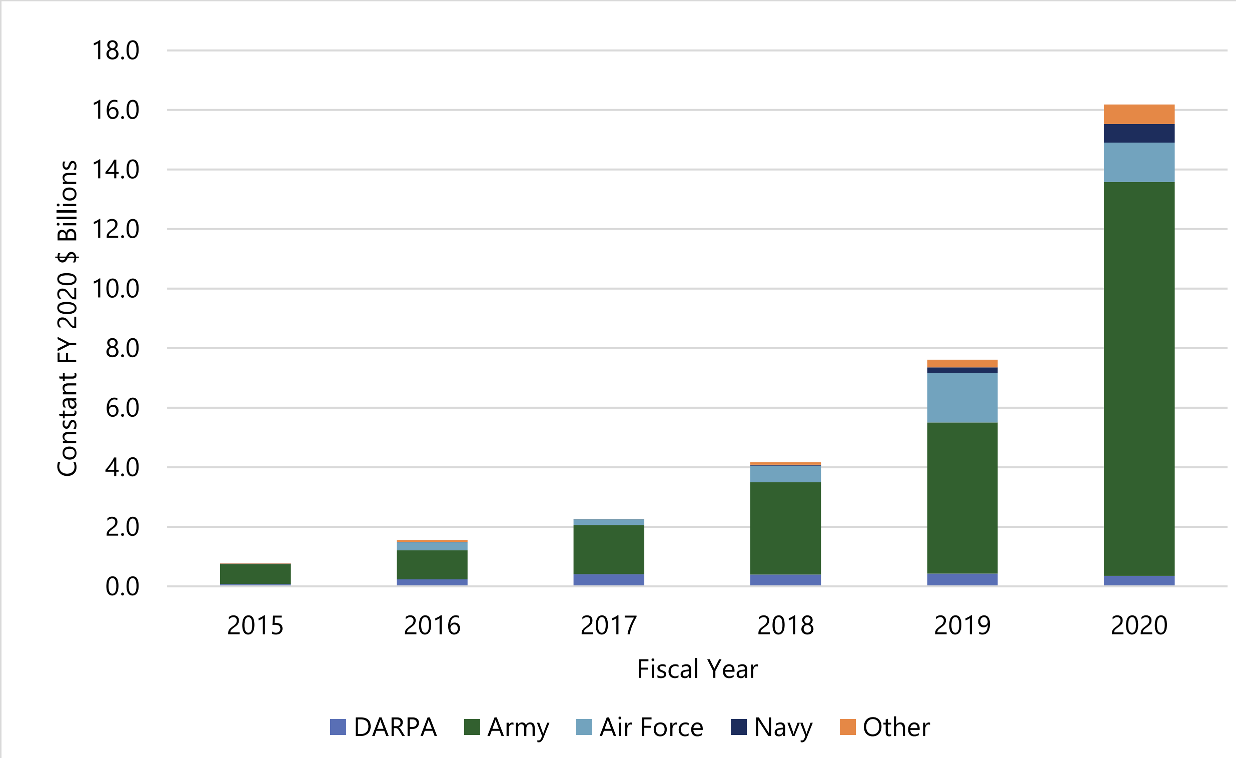


Figure 6. Defense OTA Obligations by Customer, 2015–2020 (Source: FPDS; CSIS analysis)

Army Contracting Command New Jersey (ACC–NJ), headquartered out of Picatinny Arsenal, once again remains as the largest contracting office awarding OTAs across all of the DoD. In FY2020, ACC–NJ accounted for 62% of all DoD OTA obligations and has accounted for 60% of all DoD OTA obligations between FY2015 and FY2020. Outside of ACC–NJ, the Army continues to retain several contracting offices executing OTAs, accounting for five of the top 10 DoD OTA contracting offices between FY2015 and FY2020. Outside of the Army, two Air Force contracting offices remained in the top 10—Launch Enterprise Systems Directorate and Space Development & Test Wing—but the Air Force Life Cycle Management Center (HNK C3IN), fell out of the top 10 and was replaced by Joint Munitions Command.

Table 1 shows the top 10 defense OTA contracting offices between FY2015 and FY2020.

Table 1. Top 10 DoD OTA Contracting Offices, 2015–2020 (Source: FPDS; CSIS analysis)

Contracting Office Rank	Contracting Office	Component	Total Obligations 2015–2020 ($ Billions)
1	ACC–Picatinny, NJ	Army	19.5
2	DARPA	DARPA	1.9
3	Launch Enterprise Systems Directorate	Air Force	1.8
4	ACC–Aberdeen Proving Grounds	Army	1.7
5	ACC–Redstone Arsenal	Army	1.3
Top 5 Total			**26.2**
6	Space Development & Test Wing	Air Force	0.8
7	WHS	Other DoD	0.6
8	TACOM	Army	0.5
9	Joint Munitions Command	Army	0.5
10	ACC–Orlando	Army	0.4
Top 10 Total			**28.9**
Overall DoD Total			**32.5**

Competition for DoD OTA Awards

As shown in Figure 7, the data continue to show positive trends in the rates of competition for DoD OTA obligations. Just 10% of DoD OTA obligations were competed in FY2015, but that share has been rising every year since. In FY2020, 92% of DoD OTA obligations were awarded after competition.

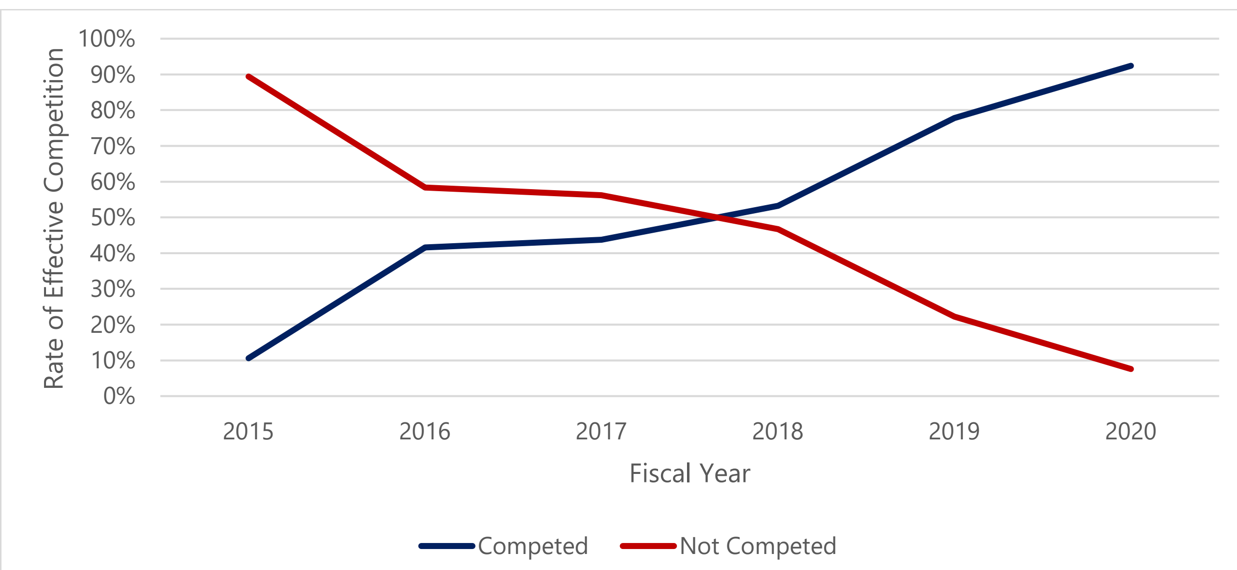


Figure 7. Competition for DoD OTA Obligations, 2015–2020 (Source: FPDS; CSIS analysis)

From Whom is the DoD Buying?

As shown in Figure 8, the rise in the vast majority of DoD OTA obligations in recent years were awarded to vendors categorized as having nontraditional significant participation.[3] Between FY2018 and FY2019, it appeared that there might be an emerging trend showing an increased share of DoD OTA obligations being awarded with cost sharing, but that trend halted in FY2020. In FY2020, defense OTA obligations awarded with cost sharing fell from $1.1 billion to $0.9 billion, a 14% decline, and subsequently fell as a share of DoD OTA obligations from 15% to 6%. Of note, as highlighted in previous CSIS reports (McCormick, 2000), although the data show that nearly 96% of DoD OTA obligations were awarded to nontraditional significant participation, consortia were awarded the majority of OTA obligations in recent years.

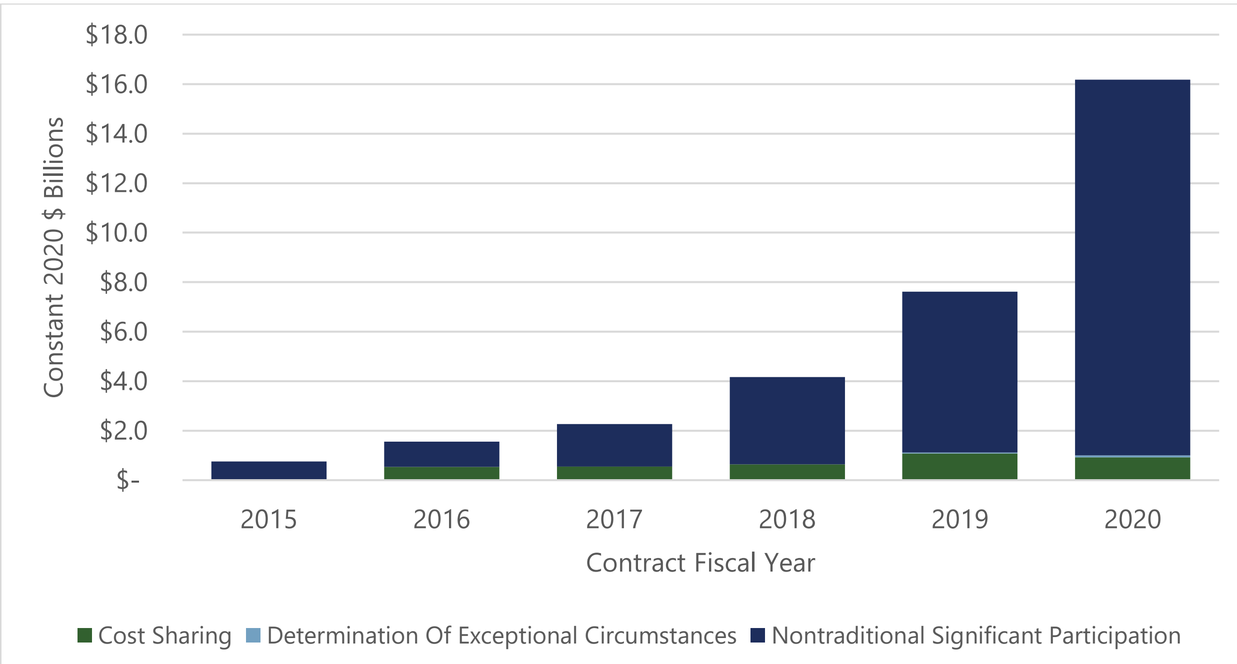


Figure 8. Defense OTA Obligations by Nontraditional Government Contractor Participation, 2015–2019 (Source: FPDS; CSIS analysis)

Conclusion

Defense OTA Obligations Continued to Grow at Staggering Rates

The data show that the rapid growth in the DoD's usage of OTAs did not slowdown in FY2020. DoD OTA obligations increased 113% last year, rising from $7.6 billion in FY2019 to $16.2 billion in FY2020. However, the Sum of Base and All Options Value or potential total contract value of DoD OTA obligations only increased 1% last year, suggesting we could see some slowdown in the same level of year-over-year growth that we've seen in recent years.

R&D Remains the Majority of DoD OTA Obligations

Defense R&D OTA obligations increased 122% between FY2019 and FY2020, compared to the 59% increase and 29% increase in Products and Services respectively.

---

[3] Nontraditional vendors are often used as a shorthand for major Silicon Valley firms, other commercial technology leaders, or individual startups with breakthrough technology.

Between FY2015 and FY2020, 89% of total DoD OTA obligations were awarded for R&D compared to 8% for Products and 3% for Services.

Mid-Stage R&D Continues Growing While Later-Stage R&D Falls Off

Although there was a slight decline in Advanced Component Development & Prototypes (6.4) OTA obligations in FY2020, those losses were more than offset by the 1,196% increase in Advanced Technology Development (6.3) OTA obligations. However, the later-stages of the weapon-systems development pipeline saw a drop off where the decline in System Development & Demonstration (6.5) was not nearly close to being offset by the relatively small total increase in Operational Systems Development (6.7).

The Army Remains the Predominant User of OTAs Across the DoD

The Army remains the predominant user of OTAs across all of the DoD, but other components, notably the Navy, have made more extensive use of OTAs in recent years than they previously did. Army OTA obligations increased 161% in FY2020 and are up 1,840,416% since FY2015. Navy OTA obligations increased from $0.6 billion in FY2019 to $0.8 billion in FY2020, a 253% increase.

Nontraditional Significant Participation Remains Dominant as Cost-Sharing Declines

For a few years, it seemed that there might be an emerging trend showing that cost-sharing was gaining a foothold for defense OTA obligations. However, this trend halted in FY2020 as OTA obligations awarded with cost sharing declined 14% and fell as a share of OTA obligations to 6% from 14%.

References

H.R. Rep. No. 114-270 (2015). https://www.govinfo.gov/content/pkg/CRPT-114hrpt270/pdf/CRPT-114hrpt270.pdf

McCormick, R. (2019). *Defense acquisition trends, 2019: Topline DoD trends. Bounces Back*. Center for Strategic and International Studies. https://csis-prod.s3.amazonaws.com/s3fs-public/publication/191011_McCormick_AcquisitionTrendsTopline_v4.pdf?Sqww2x572efEl3rUndsaZ2ikRvYdF35R

McCormick, R. (2020a). *Department of Defense other transaction authority trends: A new R&D paradigm?* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/201207_McCormick_DoD_OTA.pdf

McCormick, R. (2020b). *Defense acquisition trends, 2020: Topline DoD trends.* Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/201007_McCormick_Defense_Acquisition_Trends_2020.pdf

National Defense Authorization Act for Fiscal Year 2015, Pub. L. No. 113-291, 128 Stat. 3292 (2014). https://www.congress.gov/113/plaws/publ291/PLAW-113publ291.pdf

National Defense Authorization Act for Fiscal Year 2016, Pub. L. No. 114-92, 129 Stat. 726 (2015). https://www.congress.gov/114/plaws/publ92/PLAW-114publ92.pdf

PANEL 10. IMPROVING ACQUISITION PERFORMANCE WITH SYSTEMS ENGINEERING

Wednesday, May 12, 2021	
9:15 a.m. – 10:30 a.m.	**Chair: Thomas A. McDermott Jr.,** Deputy Director/Chief Technology Officer, Systems Engineering Research Center (SERC), Stevens Institute of Technology ***Contractual Safety of Model-Based Requirements: Preliminary Results of an Experimental Study*** Alejandro Salado, Virginia Tech Niloofar Shadab, Virginia Tech Kaitlin Henderson, Virginia Tech ***Adapting Systems Engineering Leading Indicators to the Digital Engineering & Management Paradigm*** Donna Rhodes, Massachusetts Institute of Technology Eric Rebentisch, Sociotechnical Systems Research Center ***Defining a Model-Based Systems Engineering Approach for Technical Reviews*** Ronald Carlson, Naval Postgraduate School Warren Vaneman, Naval Postgraduate School

Thomas A. McDermott Jr.—is a leader, educator, and innovator in multiple technology fields. He currently serves as Deputy Director of the Systems Engineering Research Center (SERC) at Stevens Institute of Technology in Hoboken, NJ, as well as a consultant specializing in strategic planning for uncertain environments. He studies systems engineering, systems thinking, organizational dynamics, and the nature of complex human socio-technical systems. He teaches system architecture concepts, systems thinking and decision making, and the composite skills required at the intersection of leadership and engineering. He has over 30 years of background and experience in technical and management disciplines, including over 15 years at the Georgia Institute of Technology and 18 years with Lockheed Martin.

Tom's professional accomplishments in this position come from a combination of servant leadership, systems thinking, and heuristic knowledge of complex system architectures. His long-term research goal is to develop methods and tools that support better systems thinking in the management and engineering domains and enable more rapid development of system knowledge. His current research activities focus on innovation models, strategic foresight techniques, system data analytics, and modeling and simulation of policy implications in current and future complex systems.

Tom is a graduate of the Georgia Institute of Technology, with degrees in Physics and Electrical Engineering. He developed his career in the defense electronics industry, culminating in a leadership position with Lockheed Martin as Chief Engineer and Program Manager for the F-22 Raptor Avionics Team. Tom was GTRI Director of Research and interim Director from 2007-2013. During his tenure the impact of GTRI significantly expanded, research awards doubled to over $300M, faculty research positions increased by 60%, and the organization was recognized as one of Atlanta's best places to work as well as one of the nation's leaders in employee development. He also has a visiting appointment in the Georgia Tech Sam Nunn School of International Affairs. Tom is one of the creators of Georgia Tech's Professional Masters degree in Applied Systems Engineering and lead instructor of the "Leading Systems Engineering Teams" course.

Contractual Safety of Model-Based Requirements: Preliminary Results of an Experimental Study

Niloofar Shadab—is a PhD student in systems engineering in the Department of Industrial and Systems Engineering at Virginia Tech, Blacksburg, VA. Her research track is on using systems theory to align systems engineering practices for Engineering Intelligent Systems. Shadab received her MSc in model-based systems engineering (MBSE) from the University of Maryland, College Park, MD, in 2016; she focused on Unmanned Aerial Vehicles' safety in civil applications. She has used MBSE languages such as SysML and UML. She worked as a Systems Analyst for AvMet Applications, a consulting company in the aviation industry. She received her BSc in aerospace engineering at the Sharif University of Technology, specializing in flight dynamics and controls. [nshadab@vt.edu]

Alejandro Salado—is an assistant professor in the Grado Department of Industrial and Systems Engineering and the director of its systems engineering program at Virginia Tech. He conducts research in problem formulation, design of verification and validation strategies, model-based systems engineering, and engineering education. Salado holds a BS/MS in electrical and computer engineering from the Polytechnic University of Valencia, an MS in project management and an MS in electronics engineering from the Polytechnic University of Catalonia, the SpaceTech MEng in space systems engineering from the Technical University of Delft, and a PhD in systems engineering from the Stevens Institute of Technology. [asalado@vt.edu]

Abstract

Requirements form the backbone of contracting in acquisition programs. Requirements define the problem boundaries within which contractors try to find acceptable solutions (design systems). At the same time, requirements are the criteria by which a customer measures the extent that their contract has been fulfilled by the supplier. Therefore, requirements are instrumental in the success of acquisition programs. In this context, the quality of a requirement set is determined by the level of contractual safety that it yields. From a technical perspective, contractual safety is driven by the accuracy, precision, and level of completeness of the requirement set. Unfortunately, textual requirements do not provide acceptable levels of contractual safety, as they remain a major source of problems in acquisition programs. This is partly caused by the inherent limitations of natural language to statically capture written statements with precision and accuracy. In addition, natural language is difficult (often impossible) to parse into consistent logical or mathematical statements, which limits the use of systematic and/or automated tools to explore completeness. Model-based requirements have been proposed as an alternative to textual requirements, with the promise of enabling higher accuracy, precision, and completeness when eliciting requirements. However, this promise has not been demonstrated yet. Therefore, research is needed to understand the contractual impacts of using model-based requirements instead of textual requirements before model-based requirements can be widely adopted to support acquisition programs. This paper presents preliminary results of a research project that measures the contractual safety yielded by model-based requirements. Specifically, the research addresses the main question of whether using model-based requirements improves the contractual safety of acquisition programs compared to using textual requirements. The accuracy, precision, and completeness achieved by model-based requirements are empirically measured using an experimental study. We employ a notional airborne solution to a surveillance and detection problem.

Introduction

Requirements form the backbone of contracting in acquisition programs. Requirements define the problem boundaries within which contractors try to find acceptable solutions (design systems; Salado et al., 2017). At the same time, requirements are the criteria by which a customer measures the extent that their contract has been fulfilled by the supplier (INCOSE, 2015). Hence, it is not surprising that some authors consider requirements the cornerstone of systems engineering (Buede & Miller, 2016).

Within an acquisition context, the quality of a requirement set is determined by the level of contractual safety that it yields. In the experience of the second author in acquisition programs (leaving contractual mechanisms aside and focusing only on the technical side of acquisition) contractual safety is driven by the accuracy, precision, and level of completeness of the requirement set:

- Achieving accuracy is necessary to guarantee that the requirements capture the real needs of the customer (Salado & Nilchiani, 2017c).
- Achieving precision is necessary to guarantee that the supplier interprets the requirements exactly as the customer intended when writing them (Salado & Wach, 2019b).
- Achieving completeness is necessary to avoid gaps in the problem formulation (Salado et al., 2017). If requirements are missing, a supplier may reach contractually acceptable solutions that do not fulfill the needs of the customer.

Unfortunately, textual requirements do not provide acceptable levels of contractual safety, as they remain a major source of problems in acquisition programs (GAO, 2016; Gilmore, 2011). This is partly caused by the inherent limitations of natural language to statically capture written statements with precision and accuracy (Pennock & Wade, 2015). In addition, natural language is difficult (often impossible) to parse into consistent logical or mathematical statements (Fockel & Holtmann, 2014; Gervasi & Zowghi, 2005; Tjong et al., 2006), which limits the use of systematic and/or automated tools to explore completeness (Carson et al., 2004; Salado & Nilchiani, 2017b; Salado et al., 2017).

Model-based requirements have been proposed as an alternative to textual requirements, with the promise of enabling higher accuracy, precision, and completeness when eliciting requirements (Salado & Wach, 2019b). Hence, we suggest that model-based requirements will improve contractual safety in acquisition programs. However, this statement remains to be proven. Although prior work has provided some indication in this direction (Salado & Wach, 2019b; Wach & Salado, 2021a, 2021b), we currently do not completely understand how engineers will interact with and interpret model-based requirements.

In this paper, we present preliminary results of an experimental study that evaluates the contractual safety of model-based requirements by studying their precision, accuracy, and potential for completeness compared to textual requirements. Particularly, we show some initial results that measure the impact of both types of approaches in the elicitation of requirements that are properly bounded and free of unnecessary constraints.

Literature Review

Model-based Requirements

Most of the literature in model-based requirements deals with aspects related to requirements management (e.g., requirements traceability and allocation (Badreddin et al., 2014; Borgne et al., 2016; Holder et al., 2017; Holt et al., 2012; Marschall & Schoemnakers, 2003; Mordecai & Dori, 2017; Ribeiro, 2018; Schmitz et al., 2010) or requirements engineering and management processes (Holt et al., 2012; Holt et al., 2015; Holt et al., 2012; Holt et al., 2015). Modeling the actual requirements is generally accomplished with one of two approaches. The first approach defines a specific type of model object that encapsulates the requirement, which is formulated using a textual statement. For example, SysML uses elements called requirement element and requirement diagram (Friedenthal et al., 2015). Given the inherent vagueness of natural language to formulate requirements (Salado & Wach, 2019b), such modeling approach provides minimal improvement with respect to working with textual documents (from the perspective of enabling computational assessment of requirement completeness). While parsing textual requirement statements into a set of properties or constraints associated to objects has been shown to be feasible in software systems (Lu et al., 2008), since parsing protocols rely on the structure of natural language, and not on meaning, the resulting requirements model inherits the vagueness of natural language.

In the second approach, system models are directly flagged as requirements. They often use behavioral models and/or state machines to capture functional requirements (Aceituna et al., 2011; Aceituna et al., 2014; Adedjouma et al., 2011; Soares & Vrancken, 2008; Ouchani & Lenzinia, 2014; Pandian et al., 2017; Siegl, 2010), and non-functional requirements are captured as properties or attributes of the system (Reza, 2017; Saadatmand, 2012). For example, in SysML, this is achieved by defining values for the physical block that represents the system for which requirements are being formulated (Fockel & Holtmann, 2014; Holt et al., 2015). However, this second approach presents two weaknesses, which are discussed at length in Salado and Wach (2019b). First, the separation between functional and non-functional requirements is ambiguous (Salado & Nilchiani, 2014), since, from a systems-theoretic standpoint, such a distinction does not really exist (Salado & Wach, 2019b). Requirements modeled in such a way may therefore inaccurately capture the requirement of concern. Second, since directly using behavioral models of the system of interest imposes a solution as the requirement (INCOSE, 2012), such requirements may also unnecessarily constrain the solution space (Salado et al., 2017).

To overcome these problems, formal requirement models that prescribe a requirement structure without relying on pre-existing textual statement have been proposed (Borgne et al., 2016; Micouin, 2008). For example, Micouin defines a requirement as a combination of a condition (e.g., when flying), a carrier (e.g., the system), a property (e.g., power consumption), and a domain (e.g., less than 100 W) (Micouin, 2008). While internally consistent, these structures do not prescribe the type of property that may be defined. As a result, they allow for imposing a system solution by defining design-dependent requirements, which are considered a poor practice in requirements engineering (INCOSE, 2012) because they unnecessarily constrain the solution space (Salado et al., 2017).

Alternatively, requirements have also been modeled as exchanges in which the system of interest participates. Three approaches are predominant: model requirements as data exchanges (Teufl, 2013), as exchanges between actors (Miotto, 2014), and as input/output transformations through physical interfaces at the system boundary (Salado & Wach, 2019b). The first approach is insufficient to model space system requirements because it is only capable of modeling data exchanges, not physical aspects of the problem space. While the second approach (modeling requirements as exchanges between actors) may be promising to model stakeholder needs, it is not adequate for system requirements. This is because the requirement remains unbounded (dependent on the actions of external systems), which is considered a poor practice in requirements engineering (INCOSE, 2012). The third approach, which is the basis of True Model-Based Requirements (TMBR), is consistent with the principles of systems theory and the guidelines for writing good requirements (Salado & Wach, 2019b). It was used as the basis to develop the TMBR approach used in this paper and described in the next section.

The True Model-Based Requirements Approach

In this paper, TMBR is implemented as an extension of behavioral and structural model elements of SysML (Friedenthal et al., 2015). The usage of the different model elements relies on semantics that differ from those corresponding to the original model elements in SysML. Specifically, the models presented in this paper capture solution spaces (sets of solutions), not systems (single solutions). While SysML models are used for diagrammatic purposes, their *meaning* differs from the traditional SysML specification. In particular, TMBR's implementation in SysML is architected as follows:

1. An extended sequence diagram captures the required logical transformation required to the system.
2. Signals capture required logical inputs and outputs with their required attributes.
3. Ports in block elements capture the required physical interfaces and their required properties through which inputs and outputs are conveyed.
4. An extended state machine diagram is used to capture mode requirements, which capture the simultaneity aspects of requirements applicability.

A visual representation of the basic construct of a requirement modeled as an input/output transformation is shown in Figure 1. *Blocks* are used to represent the system of interest for which requirements are being defined and *sequence diagrams* are used to capture the required flow of inputs (and outputs) to (and from) the system. In this way, the system remains a *solid line*, preventing the modeler from defining design-dependent or inner aspect of the system; only the system's behavior at its boundary in the form of external inputs and outputs is allowed. In (and out)-flows to (and from) the system are defined as items (i.e., energy, information, or material) not as actions, hence guaranteeing consistency with systems theoretic principles for system requirements.

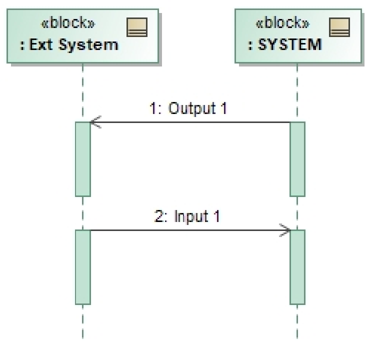


Figure 1. Input/Output Transformation Sequence Between a System and an External System

The *signal* element is used to capture the required input and output characteristics. Attributes of the signal are used to capture the required characteristics of the inputs and outputs. Examples of an input and an output are shown in Figure 2. The required interfaces through which the required inputs and outputs must be exchanged are captured using *ports*. An example of this is shown in Figure 3. The required properties of the interfaces are captured using *InterfaceBlocks*. Properties and values are used to capture requirements on the physical and transport (data) layers of the interface. An example of a modeled interface is shown in Figure 4.

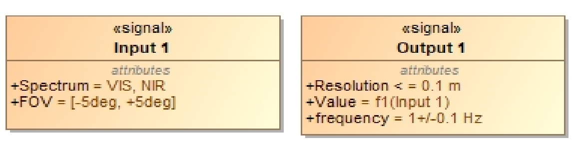


Figure 2. Example of Input/Output Signals

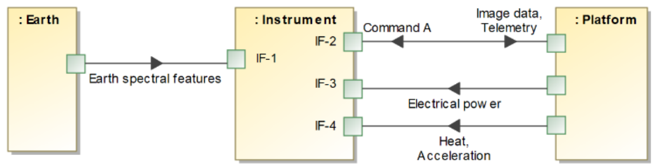


Figure 3. Example of Logical Capture of Interfaces

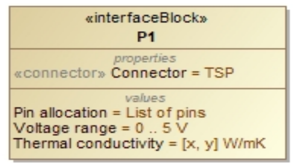


Figure 4. Example of Use of an InterfaceBlock to Capture the Characteristics of an Interface

Simultaneity of requirements applicability is captured by extending the use of SysML state machine diagrams rather than capturing all requirements in one large sequence diagram. This is defined as a *mode requirement*, which captures all requirements that “do not have conflicting requirements and that must be fulfilled simultaneously” (Salado & Wach, 2019b). An example of this is shown in Figure 5. In this example, the *Accept external energy* requirement and the *Compute tasks* requirements need to be fulfilled simultaneously.

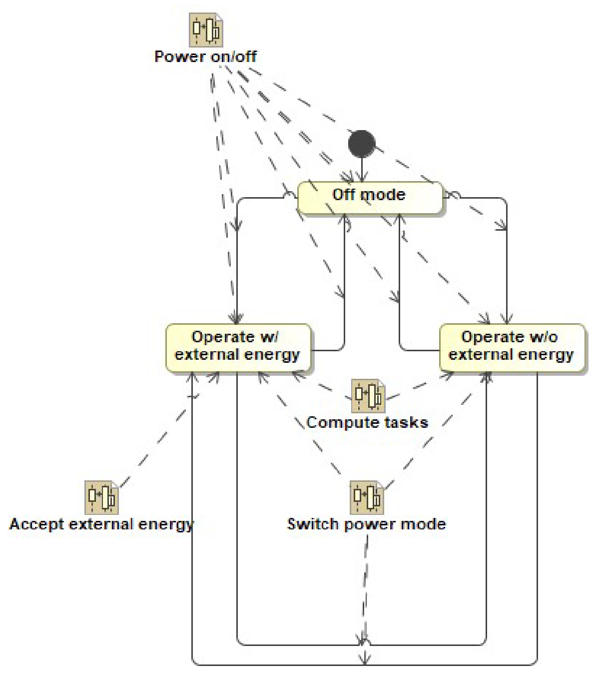


Figure 5. Example of a Mode Requirement Used to Capture Requirement Simultaneity

It is critical to note that the state objects in the state diagram do not represent *states* in the traditional sense of SysML. They are only used in this implementation of TMBR to capture operational scenarios under which different requirements are expected to be fulfilled at the same time. The model does not impose any design constraint for the system, such as what states it will have; such design decision is left open. Eventually, a *real* state machine diagram that captures the actual behavior of a potential system (not the required one) may have a completely different set of state elements (in the diagram) than the mode requirement diagram contains for indicating requirements applicability (i.e., as used in this paper). This is because in a solution model, state elements capture system states, whereas, as indicated, state elements in TMBR capture operational conditions that differ in the requirements that need to be fulfilled simultaneously. This choice avoids unnecessarily constraining the solution space (Wach & Salado, 2020).

Research Method

Participants

A total of 44 participants participated in the study, and 40 participants finished the study and turned in their artifacts. Participation was voluntary and participants received a compensation of $15 per hour, up to a total of $225, for participating in the study. Only adults participated in the study. The following inclusion criteria were used when selecting participants:

- Undergraduate or graduate student in systems engineering at Virginia Tech.
- Undergraduate or graduate student in aerospace engineering at Virginia Tech.

The following exclusion criteria were used when selecting participants:

- Not meeting the inclusion criteria.
- Students registered in a course taught by the authors during the study.
- Minors, prisoners, and adults incapable of consenting on their own behalf.

These criteria were considered appropriate because of three reasons. First, students are easier to recruit than professional engineers and can devote a significant amount of time to the study on short notice. Second, we could control the base knowledge of all participants more easily, factoring out the effects of prior experiences or individual preferences on the results of the study. For example, whereas a professional engineer may confront a conflict between applying a newly learned approach (i.e., model-based requirements) with their experience using a different approach, a student is embedded in a natural dynamic of learning and applying new methods. Third, we did not consider the inexperience of students to be a factor in this study. In fact, such inexperience was a necessary condition for this study because of the difficulty in controlling for the experience of engineers when eliciting requirements.

Determination of compliance of participants to selection criteria was performed by the authors at the start of the study.

Students were assigned to each group randomly. To do this, we separated aerospace engineering students and systems engineering students. Then we randomly split them into two groups. To ensure that all the participants meet the experiment criteria and to avoid any conflicts of interests, we conducted a survey to gather demographic information.

Instruments

Five instruments were developed for use in the experiment (see Table 1). All instruments were developed before the study was initiated. The table is organized sequentially, i.e., the order indicates the sequence in which the different instruments were provided to the participants.

Table 1. Research Instruments Used in the Study

Instrument	Description
Consent form	This instrument was used to inform the participants about the conditions of the study.
Survey	This instrument was used to gather demographic information of the participants.
Training material textual requirements	This instrument consisted of a slide deck and synchronous online presentations by an instructor.
Training material model-based requirements	This instrument consisted of a slide deck, research papers, and synchronous online presentations by an instructor.
Problem description	This instrument was used as a problem statement. It also listed the expected behavior of the participants during the study. It is provided in the Appendix.

Design

In this section, the study design is discussed. This discussion includes the statement of the hypotheses, factors in the design, a discussion on the validity and reliability of the design, detailed procedures for executing the study, and a summary of the data analysis methods.

Hypotheses

The study was designed to test the following three hypotheses:

H1. Model-based requirements yield fewer unbounded requirements than textual-based requirements.

H2. Model-based requirements yield fewer unnecessary constraints than textual-based requirements.

H3. Model-based requirements achieve higher completeness than textual-based requirements.

All hypotheses focus on the performance of the groups. It was expected that the results would confirm the three hypotheses.

Experimental Design

Two groups of engineers were asked to elicit requirements from potential users of a surveillance and detection system. One group acted as the control group and the other group acted as the experimental group. The control group employed textual requirements, while the experimental group employed model-based requirements.

Each group consisted of 22 students. Eleven students were in the aerospace engineering major and the other 11 were in the systems engineering major. Aerospace engineering students participated to both bring subject matter expertise on the problem statement and prevent investigators' biases on the subject. Each aerospace engineering student was teamed up with another study participant (i.e., a system engineering student). The one-to-one allocation was intended to avoid coupling effects between the study participants, which eases the factorial analyses necessary to test the hypotheses of this study. However, later in the experiment, four students from the model-based requirements engineering dropped out of the study before they finished their artifacts. Therefore, the study consisted of 11 groups of textual requirements and 9 groups of model-based requirements.

After splitting participants into two groups, each group was trained in just one of the methodologies, either model-based requirements or textual requirements. Group 1 received 10-hour training on textual requirements. Group 2 received 10-hour training on Model-based Requirements. This split helps avoiding confounding effects between knowing both methods but applying only one of them. The 10-hour training was divided into two blocks of 5 hours apiece.

Training was not provided by the researchers, but by independent instructors. Training in model-based requirements was provided by an author of a seminal Model-Based Requirement paper. In this way, we could control for adequate learning and application of the model-based requirements framework, by having an instructor who developed such a framework. Training in using textual requirements was provided based on material in (Buede & Miller, 2016; Lee et al., 2009; Wasson, 2016). In this way, we mitigated potential biases that the researchers might have introduced if provided the training of both methods. The instructor had over 5 years of experience in eliciting requirements for large-scale engineered systems and prior experience in conducting this type of training.

An important observation was that during or after the training sessions for Model-Based Requirements (i.e., Group 2), three students dropped out of the experiment. They cited difficulties in understanding Model-Based Requirements Engineering as the main reason for their decision. Therefore, we started with 22 students for each group and ended up with 22 students in Group 1 and 18 students in Group 2.

The one-on-one study was conducted in five 1-hour sessions. Each session was separated by 1 week to allow the participant to reflect and process the insights and data collected during the elicitation session.

Each team performed the elicitation sessions in isolation from other participants. The teams performed their task only with the knowledge they gained from the training sessions. That is, the elicitation sessions were conducted sequentially and not for the entire sample at once. No outsider source was used during the elicitation process. To ensure that the teams worked in isolation with no outside help, all the sessions were video recorded.

The surveillance and detection need of the case study defined in Larson et al. (2009) was used as the problem statement. The hypothesis was tested using stakeholder needs for surveillance and detection of fire over the U.S. map. In the study, participants in both groups developed a requirement set for an Earth Observation Satellite. The stakeholder need was to build a system that could detect and monitor potentially dangerous wildfires throughout the United States. This satellite would survey the United States daily to give the Forest Service a means for earlier detection to increase the probability of containment and to save lives and property.

A survey was employed to gather demographic information of the participants.

Factors in the Design

The independent variable in this investigation is the *requirements approach*. There are two alternatives: (1) textual requirements and (2) model-based requirements as defined in Salado & Wach, 2019b.

Four dependent variables were measured:

- Number of inapplicable requirements. This variable provides a measure of the actual effectiveness of both the control method and the experimental method to elicit inapplicable-free requirements.
- Number of unnecessary constraints. This variable provides a measure of the actual effectiveness of both the control method and the experimental method to elicit unnecessary requirements, such as solution-dependent ones.
- Number of unbounded requirements. This variable provides a measure of the actual effectiveness of both the control method and the experimental method to elicit adequately bounded requirements.
- Level of completeness of the requirement set. This variable provides a measure of the completeness of the resulting requirement sets when using both the method employed by the control group and the method employed by the experimental group.

Effects related to experience, competence, and specific knowledge were controlled by the inclusion and exclusion criteria of the participants.

In designing the experiment, several constraints were imposed that could have restricted the ways in which the independent variables could be manipulated. Three primary factors constrained the experiment:

- Time. The elicitation problem was limited to 5 hours per problem. This is considered much lower than what would be allocated in a real-life development for the given system of interest. Therefore, this limitation poses a threat to completeness in the elicitation effort. However, since all participants are subjected to the same limitation, we suggest that the effectiveness of the method can still be measured.
- Participants. First, the elicitation activity was performed in isolation (that is, one analyst and one stakeholder) and not in teams of analysts. This is not necessarily representative of a real-life development for the given system of interest. Therefore, this limitation poses a threat to correctness in the elicitation effort due to potential lack of domain knowledge. However, since all participants are subjected to the same limitation, we suggest that the effectiveness of the method can still be measured.
- Single domain. The problems only address one type of system, a satellite. This poses a threat to generality of the results.

Threats to Internal Validity

In an internally valid experiment, the relationships between observed differences on the independent variable are a direct result of the manipulation of the independent variable, not some other variable. Table 2 lists internal threats to validity and their potential interference with the experimental design, if any.

Table 2. Threats to Internal Validity

Threat	Factor	Justification
Ambiguous temporal precedence	No	The cause variable (requirements formulation method) was used as an input to create the effect variable (formulated requirements).
Confounding	No	Prior knowledge and experience in the field of systems engineering was controlled through recruiting.
History	No	The study was conducted in a short time and no extraneous event was recorded during the study.
Maturation	Yes (to be mitigated)	The study was conducted in several sessions separated in time. Although participants were instructed to not read or learn anything on the topics relevant to the study until their responses were delivered, the researchers had no mechanism to control maturation. However, the video recording of the different sessions could indicate if maturation happened. This assessment has not been done for this paper.
Repeated testing	No	No pre-test was given.
Instrumentality	Potentially	Factor: The researchers could link the artifact under evaluation to the different groups. Mitigations: The experiment did not use pre-test instruments in conjunction with post-test instruments.
Statistical regression	No	Random allocation of participants to groups.
Selection bias	No	(1) Groups were randomly created. (2) Pre-testing was not performed.
Mortality	No	All participants that conducted the first 1-to-1 session completed the study.
Selection-Maturation interaction	No	Random allocation of participants to groups.
Diffusion	Yes	Although participants were instructed to not exchange any information or opinion about the experiment with other participants until cleared out by the researcher, the researcher had no mechanism to control it.
Compensatory rivalry/resentful demoralization	No	(1) The experiment did not have intermediate results gates. (2) Participants did not have access to the results of the other group.
Experimenter bias	Yes (mitigated)	No verbal interaction between the researcher and the participants regarding the experiment besides pre-produced instruments and minor clarifications about the expected deliverables.

Threats to External Validity

In an externally valid experiment, the results are generalizable to groups and environments outside of the experimental setting. Table 3 lists internal threats to validity and their potential interference with the experimental design, if any.

Table 3. Threats to External Validity

Threat	Factor	Justification
Pre-test treatment interaction	No	No pre-test was given.
Multiple treatment interference	No	There was a single treatment in the study (one training session before the executing of the requirements elicitation activity).
Selection-treatment interaction	Yes	All participants were students from controlled departments.
Reactivity and Situation	Yes	
Rosenthal effects	No	(1) Instruments prepared before the experiment and provided in written form. (2) Participants were randomly assigned to groups. (3) Administration of treatment (instruction) was not performed by the researchers. (4) Evaluation criteria not defined by the researchers.

Data Analysis

For textual requirements, requirements were de-categorized and compiled as a single list for each participant. For model-based requirements, requirements were transformed into individual statements using the template described in Salado and Wach (2019a) and consolidated as a single list for each participant.

Each requirement in each list was evaluated and classified as inapplicable, unbounded, unnecessary constraint, or adequate. The following criteria, which were derived from industry guidelines (INCOSE, 2012) and used in prior research for the same purpose (Salado & Nilchiani, 2017a), were used to classify the requirements:

- *Inapplicable requirement*: A requirement that addresses a system external to the system of interest. Indications of this include statements where the subject of the requirement is not the system of interest or one of its parts or the requirement addresses aspects of the development process.
- *Unbounded requirement*: A requirement that the system of interest cannot fulfill on its own, but which fulfillment depends on the action of systems external to the system of interest. An indication of this is that the statement contains more than one system.
- *Unnecessary constraint*: A requirement that enforces a particular design solution. Indications of this include the use of terms such as *use*, *be composed of*, *consist of*, or *include*, among others, or the use of a system's part as the subject of the statement.
- *Adequate requirement*: Any requirement that is not classified in any of the other three.

Completeness was assessed in two steps. First, all generated requirements were aggregated, consolidating those that refer to the same aspect or mutually exclusive aspect that the system had to fulfill as a single requirement. Second, the coverage of each set of requirements from the participants was assessed against this aggregated list. While completeness of a requirement set cannot be proven (Carson, 1998; Carson et al., 2004; Carson & Shell, 2001), we suggest, as in Salado and Nilchiani (2017a), that this coverage provides valuable insights about how the different approaches might impact completeness.

These evaluations were performed independently by both authors and then consolidated. The authors knew the approach used to define the requirements when performing the evaluation, which leads to some of the biases described in the section Threats to Internal Validity.

Descriptive statistics were used to characterize the responses of the participants, as well as their demographics. Inapplicable requirements were removed from the comparisons of unbounded requirements, unnecessary constraints, and completeness to enable fair comparisons between the two approaches. While inferential statistics were initially planned to quantitatively compare both approaches, we found some problems to process the deliverables of the group using model-based requirements. Instead, a qualitative assessment was performed.

Results

In this section, we show preliminary results of a subset of the gathered data during the experiment. In particular, we randomly picked the deliverables of three participant pairs from the control group and three from the experiment group.

Group Composition

Figure 6 and Figure 7 show the distributions of prior experience using textual requirements and using MBSE, respectively. While experience in textual requirements among the two groups was a bit imbalanced, the group had, in general, little or no experience. A similar situation was given with respect to experience in MBSE. Comparing both the control and the experimental groups, the experience relevant to their approach was similar.

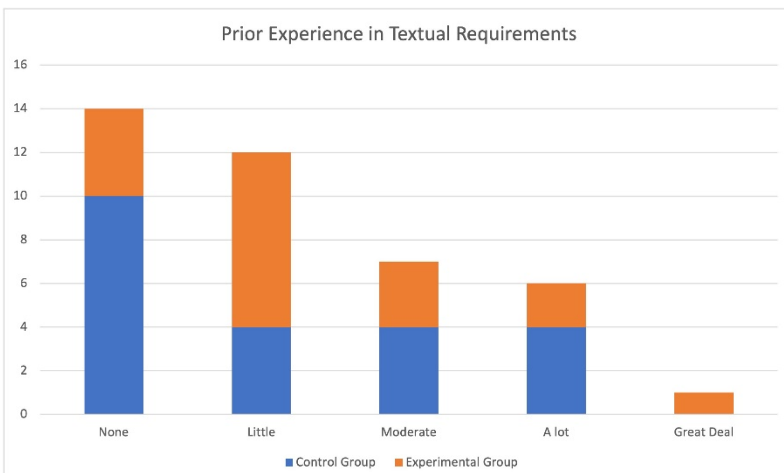


Figure 6. Prior Experience Using Textual Requirements

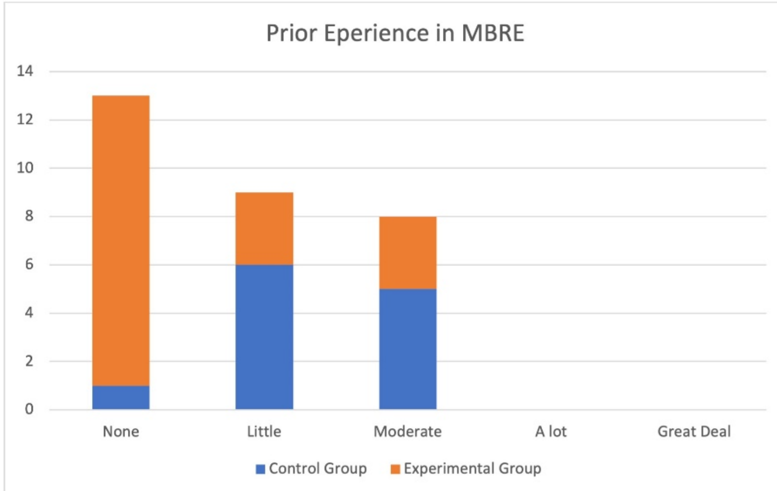


Figure 7. Prior Experience Using MBSE

Figure 8 shows the distribution of prior experience in designing or working with space systems among the different groups. The responses indicated more dispersion, which could be explained by the fact that around half of the participants were aerospace students and half not aerospace students. It should be restated that each pair of participants in each group were formed by one aerospace student and non-aerospace student.

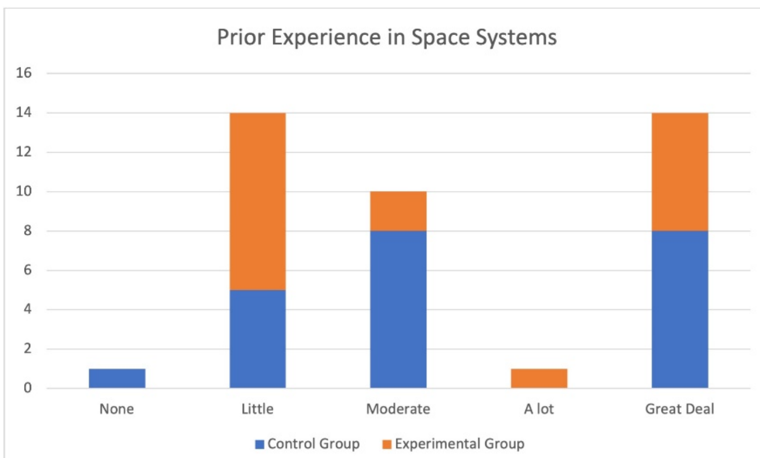


Figure 8. Prior Experience in Designing or Working with Space Systems

Evaluation of Textual Requirements

Table 4 shows the results of assessing the requirements delivered by three participant pairs in terms of the number of unbounded requirements, inapplicable requirements, and unnecessary constraints. Table 5 shows a few examples of such requirements directly taken from the participants' responses.

Table 4. Summary Assessment of Textual Requirements

Project	Total Req	Unbounded Req	Unnecessary constraint	Inapplicable Req	Adequate Req
1	45	16	10	16	4 (10%)
2	145	49	40	55	7 (5%)
3	98	27	41	1	32 (33%)

Table 5. Requirement Examples from the Participants' Responses

Unbounded requirements	The system shall provide space-based "fire-scouts" that survey the United States daily. The system shall provide space-based "fire-scouts" that survey the United States daily. The satellite shall be deployed in low Earth orbit.
Unnecessary constraints	The antenna shall allow the satellite to communicate with the ground. Propellant shall be an ionized thrust that can be recharged using solar cells. The satellite shall utilize GPS.
Inapplicable requirements	The rockets shall withstand temperatures from XXX-to-XXX degree Fahrenheit. Separation shall occur once satellite is in specified orbit. The satellite shall pass all the Vega Launch Vehicle Manual's quality inspection requirements.

To evaluate the fidelity of the requirements activity, we compare the results presented in Table 4 with those obtained in a similar study, which was conducted by one of the authors with professional engineers and is reported in Salado and Nilchiani (2017a). The comparison is presented in Table 6. While data are not available in Salado and Nilchiani (2017a) for unbounded requirements, participants employed textual requirements a bit less effectively (that is, with more unnecessary constraints and inapplicable requirements). We suggest that this difference is, however, not dramatic, and consider that the participants correctly used the training to formulate textual requirements.

Table 6. Comparison Against Performance of Practicing Engineers (Salado & Nilchiani, 2017a)

Variable		Practicing Engineers	This Experiment
Relative number of unbounded requirements	Mean	n/a	32%
	Median	n/a	34%
Relative number of unnecessary constraints	Mean	27%	31%
	Median	26%	28%
Relative number of inapplicable requirements	Mean	16%	25%
	Median	18%	36%

Evaluation of Model-Based Requirements

Two of the three responses we randomly selected for this pilot evaluation show significant misuse of TMBR. Two main issues were found. First, some participants used signal elements to represent actions instead of items that are exchanged between systems. An example of this, directly taken from a participant's response, is shown in Figure 9. Second, some participants decomposed the system of interest into its components. An example is shown Figure 10. Therefore, they cannot be used in the assessment.

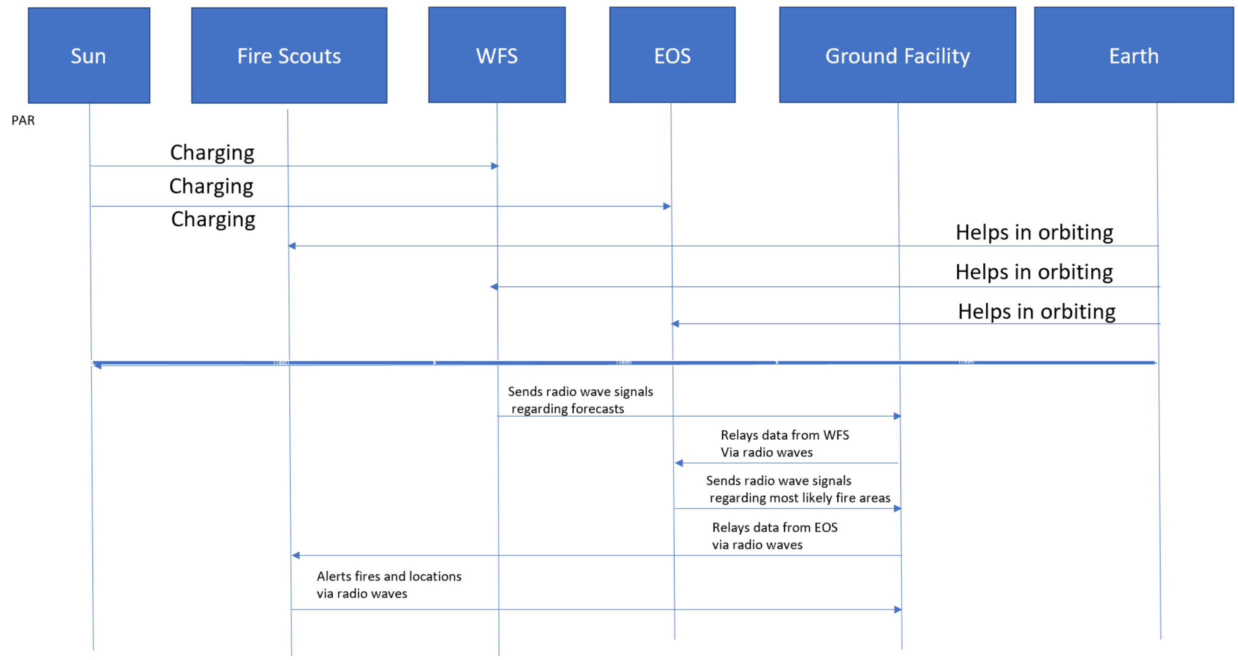


Figure 9. Example of TMBR Misuse: Use of Actions Instead of Signals in Exchanges Between Systems

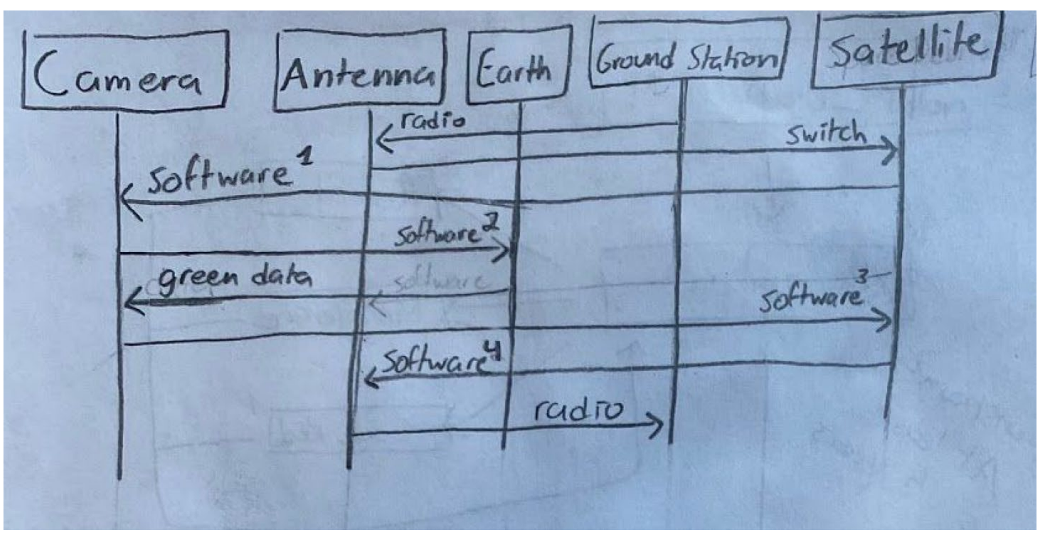


Figure 10. Example of TMBR Misuse: Decomposition of System of Interest into Components.

The third response, while adequately using TMBR, did not generate complete models (a partial example is shown in Figure 11). Specifically, the characteristics of the signals and the interfaces through which they ought to be exchanged were not modeled. Yet, we consider that this response is valid for the pilot assessment, since effects on the hypotheses listed in the section Hypotheses can be evaluated.

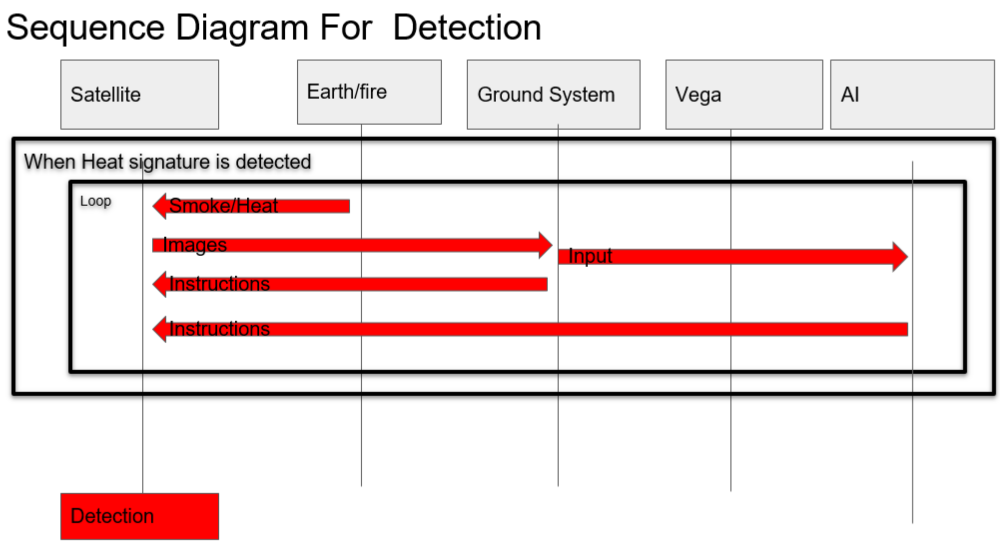


Figure 11. Example of Adequate Yet Incomplete Use of TMBR

The requirement models led to 31 unique requirement statements, none of which was an unbounded requirement, an unnecessary constraint, or an inapplicable requirement. A few examples are shown here:

R1. The satellite shall accept force of launch.
R2. The satellite shall accept instructions 1 when heat signature is detected.
R3. The satellite shall provide images when heat signature is detected.

As discussed, according to TMBR, the models should have incorporated required properties of the items and of the interfaces through which they are transferred. For example, the item “force of launch” in R1 should have been completed with the actual mechanical forces injected into the satellite (e.g., a random vibration profile). Similarly, the models should have been completed with the description of the satellite’s attachment/physical point through which such vibration profile was to be injected. These were not provided in the participant’s response.

Comparison

The performance of the control and experiment groups with respect to the hypotheses listed in Section 3.3.1 are compared in Table 7. While these results are preliminary, given the small size of the sample and the problems encountered with the responses in the experiment group, they provide some indication of the superiority of model-based requirements to textual requirements. In terms of unbounded requirements, inapplicable requirements, and unnecessary constraints, these results were expected because such types of requirements are avoided by design in TMBR (Salado & Wach, 2019b).

Table 7. Results Comparison Between Control and Experiment Groups

Variable		Control Group (textual reqs)* *(sample size = 3)*	Experiment Group (TMBR) *(sample size = 1)*
Relative number of unbounded requirements	Mean	32%	0%
	Median	34%	0%
Relative number of unnecessary constrains.	Mean	31%	0%
	Median	28%	0%
Relative number of inapplicable requirements	Mean	25%	0%
	Median	36%	0%
Adequate requirements	Mean	14	31
	Median	7	31
Coverage	Mean	26%	51%

*Note: As shown in Table 4, none of the responses in the sample achieved 0% performance in any of the variables, and the maximum number of adequate requirements obtained was 32 with a 54% coverage.

For completeness, in absolute terms, using model-based requirements led to more adequate requirements than using textual requirements. To assess coverage, the requirements from every participant were aggregated to a total of 57 adequate requirements. Model-based requirements achieved higher coverage than textual requirements, although one of the responses in the control group achieved a coverage of around 54%. However, it is important to note that the coverage achieved in the experiment group may not have a high fidelity due to the incompleteness in the models, as discussed earlier.

5. Conclusions and Future Research

We have presented in this paper preliminary results of an experimental study to evaluate the contractual safety of model-based requirements as compared to textual requirements. The preliminary results reported here have been limited to a subset of the all the data collected and to just four variables: number of unbounded requirements, inapplicable requirements, and unnecessary constraints, and coverage. The preliminary data support the claim that model-based requirements are superior in all variables to textual requirements.

However, some issues were encountered with the application of model-based requirements. Particularly, participants failed to use the modeling rules of TMBR, the modeling paradigm for requirements used in this study. We interviewed the TMBR instructor after reviewing the responses from the participants to better understand the results of the training prior to the start of the requirements definition activity. According to the instructor, the two issues described in the previous section were common misunderstandings he encountered during the whole training, and while he addressed them several times, he was not confident that the participants fully grasped them at the conclusion of the training.

We have reviewed the material that was used for the training and conjecture three potential causes. First, the material that was used for training relied heavily on academic papers, which might have been too hard to process for the participants, who were undergraduate students. Second, the participants in charge of modeling the requirements were primarily students in an industrial engineering program, being biased toward process flows over design. Third, TMBR may require more training time than initially scoped. The experience gained in this pilot study informs the development of dedicated training material that is more easily digested by participants, including an evaluation gate during the training and potentially increased training duration.

Finally, we note that future work is planned to complete the evaluation with the full set of responses and to include measures related to precision and accuracy with which the requirements were captured in textual forms and with models.

Appendix

We want to develop an Earth Observation Satellite. Our goal is to build a system that can detect and monitor potentially dangerous wildfires throughout the US. wildfires claim hundreds of lives, threat thousands more, and lay waste to millions of acres, causing losses in the billions of dollars. The system needs to provide space-based “fire-scouts” that would survey the US daily to give the Forest Service a means for earlier detection to increase the probability of containment and to save lives and property.

The following needs, among others you consider necessary, are to be taken into account:

1. Satellite will be placed in LEO.
2. Continuous monitoring.
3. Be operational for at least 5 years.
4. Launched onboard Vega. (Link to the actual user manual was provided.)
5. Use Space@VT ground station. (Link to a datasheet of the actual ground station was provided.)

If you need to quantify any value, use notional ones and specify your assumptions. No need for actual analyses.

Please derive the requirements for this system.

RULES:

1. You cannot use or read requirements related to similar satellite
2. Report all information that you have used for this activity
3. Ideally, you just use your own knowledge
4. No work in between sessions.
5. No learning about requirements engineering during the whole study. If required as part of formal coursework, let us know.
6. You cannot talk to anyone during the duration of the study about the study, not even your peer in the session.
7. You should brainstorm with your peer in each session.
8. You should hand in the final result of requirement derivation process after all the five sessions are finished.

References

Aceituna, D., Do, H., Walia, G. S., & Lee, S. (2011). *Evaluating the use of model-based requirements verification method: A feasibility study*. Paper presented at the Workshop on Empirical Requirements Engineering (EmpiRE 2011), Trento.

Aceituna, D., Walia, G., Do, H., & Lee, S.-W. (2014). Model-based requirements verification method: Conclusions from two controlled experiments. *Information and Software Technology*, *56*(3), 321–334. https://doi.org/10.1016/j.infsof.2013.11.004

Adedjouma, M., Dubois, H., & Terrier, F. (2011, April 27–29). *Requirements exchange: From specification documents to models.* Paper presented at the Proceedings of the 16th IEEE International Conference on Engineering of Complex Computer Systems, Las Vegas, NV.

Badreddin, O., Sturm, A., & Lethbridge, T. C. (2014). *Requirement traceability: A model-based approach*. Paper presented at the 2014 IEEE 4th International Model-Driven Requirements Engineering Workshop (MoDRE), Karlskrona. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6890829&isnumber=6890816

Borgne, A. L., Belloir, N., Bruel, J., & Nguyen, T. (2016). *Formal requirements engineering for smart industries: Toward a model-based graphical language*. Paper presented at the 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), Toulouse, France. http://ieeexplore.ieee.org.ezproxy.lib.vt.edu/stamp/stamp.jsp?tp=&arnumber=7816956&isnumber=7816796

Buede, D. M., & Miller, W. D. (2016). *The engineering design of systems: Models and methods* (3rd ed.). John Wiley & Sons.

Carson, R. S. (1998). Requirement completeness: A deterministic approach. *INCOSE International Symposium*, *8*, 738–746.

Carson, R. S., Aslaksen, E., & Gonzales, R. (2004). *Requirements completeness.* Paper presented at the INCOSE International Symposium.

Carson, R. S., & Shell, T. (2001). Requirements completeness: Absolute or relative? *Systems Engineering*, *4*, 230–231.

Fockel, M., & Holtmann, J. (2014). *A requirements engineering methodology combining models and controlled natural language*. Paper presented at the 2014 IEEE 4th International Model-Driven Requirements Engineering Workshop (MoDRE), Karlskrona. http://ieeexplore.ieee.org.ezproxy.lib.vt.edu/stamp/stamp.jsp?tp=&arnumber=6890827&isnumber=6890816

Friedenthal, S., Moore, A., & Steiner, R. (2015). *A practical guide to SysML—The systems modeling language* (3rd ed.). Morgan Kaufman.

GAO. (2016). *Defense acquisitions—Assessments of selected weapon programs* (GAO-16-329SP).

Gervasi, V., & Zowghi, D. (2005). Reasoning about inconsistencies in natural language requirements. *ACM Transactions on Software Engineering and Methodology*, *14*(3), 277–330.

Gilmore, J. M. (2011). Key issues causing program delays in defense acquisition. *ITEA Journal*, *32*, 389–391.

Holder, K., Zech, A., Ramsaier, M., Stetter, R., Niedermeier, H.-P., Rudolph, S., & Till, M. (2017). Model-based requirements management in gear systems design based on graph-based design languages. *Applied Sciences*, *7*(11), 1112. https://doi.org/10.3390/app7111112

Holt, J., Perry, S., Brownsword, M., Cancila, D., Hallerstede, S., & Hansen, F. O. (2012). *Model-based requirements engineering for system of systems*. Paper presented at the 2012 7th International Conference on System of Systems Engineering (SoSE), Genova, Italy. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6384145&isnumber=6355890

Holt, J., Perry, S., Payne, R., Bryans, J., Hallerstede, S., & Hansen, F. O. (2015). A model-based approach for requirements engineering for systems of systems. *IEEE Systems Journal*, *9*(1), 252–262. https://doi.org/10.1109/JSYST.2014.2312051

INCOSE. (2012). *Guide for writing requirements*.

INCOSE. (2015). *Systems engineering handbook: A guide for system life cycle processes and activities* (version 4.0). John Wiley & Sons.

Larson, W. J., Kirkpatrick, D., Sellers, J. J., Thomas, D., & Verma, D. (2009). *Applied space systems engineering.* McGraw Hills.

Lee, R. E., Hooks, I., & Barker, B. G. (2009). Engineering and managing system requirements. In W. J. Larson, D. Kirkpatrick, J. J. Sellers, L. D. Thomas, & D. Verma (Eds.), *Applied space systems engineering* (pp. 95–139). McGraw Hill.

Lu, C., Chang, C., Chu, W. C., Cheng, Y., & Chang, H. (2008). *A requirement tool to support model-based requirement engineering*. Paper presented at the 2008 32nd Annual IEEE International Computer Software and Applications Conference, Turku. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=4591654&isnumber=4591503

Marschall, F., & Schoemnakers, M. (2003). *Towards model-based requirements engineering for web-enabled B2B applications*. Paper presented at the 10th IEEE International Conference and Workshop on the Engineering of Computer-Based Systems, Huntsville, Alabama. http://ieeexplore.ieee.org.ezproxy.lib.vt.edu/stamp/stamp.jsp?tp=&arnumber=1194813&isnumber=26873

Micouin, P. (2008). Toward a property based requirements theory: System requirements structured as a semilattice. *Systems Engineering*, *11*(2).

Miotto, B. L., & a. P. (2014). *Model-based requirement generation*. Paper presented at the 2014 IEEE Aerospace Conference, Big Sky, MT.

Mordecai, Y., & Dori, D. (2017). *Model-based requirements engineering: Architecting for system requirements with stakeholders in mind*. Paper presented at the 2017 IEEE International Systems Engineering Symposium (ISSE), Vienna, Austria. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8088273&isnumber=8088244

Ouchani, S., & Lenzinia, G. (2014). *Attacks generation by detecting attack surfaces*. Paper presented at the 5th International Conference on Ambient Systems, Networks and Technologies.

Pandian, M. K. S., Suri, K., Cadavid, J., Barosan, I., Brand, M., Alférez, M., & Gérard, S. (2017, April 3–7). *Towards industry 4.0: Gap analysis between current automotive mes and industry standards using model-based requirement engineering.* Paper presented at the Proceedings of the 2017 IEEE International Conference on Software Architecture Workshops (ICSAW).

Pennock, M. J., & Wade, J. P. (2015). The top 10 illusions of systems engineering: A research agenda. *Procedia Computer Science, 44*, 147–154. https://doi.org/10.1016/j.procs.2015.03.033

Reza, H., Sehgal, R., Straub, J., & Alexander, N. (2017). *Toward model-based requirement engineering tool support*. Paper presented at the 2017 IEEE Aerospace Conference, Big Sky, MT. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7943647&isnumber=7943554

Ribeiro, F. G. C., Pereira, C. E., Rettberg, A., & Soares, M. S. (2018). Model-based requirements specification of real-time systems with UML, SysML and MARTE. *Software & Systems Modeling*, *17*(1), 343–361. https://link.springer.com/article/10.1007/s10270-016-0525-1

Saadatmand, M., Cicchetti, A., & Sjödin, M. (2012). *Toward model-based trade-off analysis of non-functional requirements.* Paper presented at the 2012 38th Euromicro Conference on Software Engineering and Advanced Applications, Cesme, Izmir. http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6328142&isnumber=6328119

Salado, A., & Nilchiani, R. (2014). A categorization model of requirements based on Max-Neef's model of human needs. *Systems Engineering*, *17*(3), 348–360. https://doi.org/10.1002/sys.21274

Salado, A., & Nilchiani, R. (2017a). Reducing excess requirements through orthogonal categorizations during problem formulation: Results of a factorial experiment. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *47*(3), 405–415. https://doi.org/10.1109/TSMC.2015.2502953

Salado, A., & Nilchiani, R. (2017b). The tension matrix and the concept of elemental decomposition: improving identification of conflicting requirements. *IEEE Systems Journal*, *11*(4), 2128–2139. https://doi.org/10.1109/JSYST.2015.2423658

Salado, A., & Nilchiani, R. (2017c). Understanding how problem formulation in systems engineering influences systems affordability: A systems thinking investigation. *International Journal of Applied Systemic Studies*, *7*(4).

Salado, A., Nilchiani, R., & Verma, D. (2017). A contribution to the scientific foundations of systems engineering: Solution spaces and requirements. *Journal of Systems Science and Systems Engineering*, *26*(5), 549–589. https://doi.org/10.1007/s11518-016-5315-3

Salado, A., & Wach, P. (2019a). *Automatic generation of contractual requirements from MBSE artifacts.* Paper presented at the 16th Annual Acquisition Research Symposium, Monterey, CA.

Salado, A., & Wach, P. (2019b). Constructing true model-based requirements in SysML. *Systems*, *7*(2), 19.

Schmitz, D., Nissen, H. W., Jarke, M., & Rose, T. (2010). *Relating domain model based requirements management and situational method engineering*. Paper presented at the 2010 Third International Workshop on Managing Requirements Knowledge, Sydney, NSW.

Siegl, S., Hielscher, K., & German, R. (2010). *Model based requirements analysis and testing of automotive systems with timed usage models*. Paper presented at the 2010 18th IEEE International Requirements Engineering Conference, Sydney, NSW. http://ieeexplore.ieee.org.ezproxy.lib.vt.edu/stamp/stamp.jsp?tp=&arnumber=5636645&isnumber=5636534

Soares, M. S., & Vrancken, J. (2008). Model-driven user requirements specification using SysML. *Journal of Software*, *3*(6), 57–68.

Teufl, S., Mou, D., & Ratiu, D. (2013). *MIRA: A tooling-framework to experiment with model-based requirements engineering*. Paper presented at the 2013 21st IEEE International Requirements Engineering Conference (RE), Rio de Janeiro.

Tjong, S. F., Hallam, N., & Hartley, M. (2006, September 20–22). *Improving the quality of natural language requirements specifications through natural language requirements patterns.* Paper presented at the Sixth IEEE International Conference on Computer and Information Technology.

Wach, P., & Salado, A. (2020). *The need for semantic extension of SysML to model the problem space*. Paper presented at the Conference on Systems Engineering Research (CSER), Redondo Beach, CA.

Wach, P., & Salado, A. (2021a). *True model-based requirements (TMBR) of a satellite TTC transponder.* Paper presented at the IEEE Aerospace Conference.

Wach, P., & Salado, A. (2021b). True model-based requirements (TMBR): Application to an Earth observation satellite. In *AIAA Scitech 2021 Forum*.

Wasson, C. S. (2016). *Systems engineering analysis, design, and development—Concepts, principles, and practices*. John Wiley & Sons.

Adapting Systems Engineering Leading Indicators to the Digital Engineering & Management Paradigm

Donna H. Rhodes—is a Principal Research Scientist at Massachusetts Institute of Technology Sociotechnical Systems Research Center, and director of the Systems Engineering Advancement Research Initiative (SEAri). She conducts research on innovative approaches and methods for architecting complex systems and enterprises, digital transformation, model curation, and human–model interaction. Previously, she held senior leadership positions at IBM, Lockheed Martin, and Lucent. Rhodes is a Past President and Fellow of the International Council on Systems Engineering (INCOSE), and INCOSE Founders Award recipient. She received her PhD in systems science from T. J. Watson School of Engineering at Binghamton University. [rhodes@mit.edu]

Eric Rebentisch—is a Research Associate at MIT's Sociotechnical Systems Research Center. His research focuses on implementation of digital engineering and model-based systems engineering in enterprises. He is Editor in Chief of *Integrating Program Management and Systems Engineering*, and teaches systems engineering in the MIT System Design and Management (SDM) program. Rebentisch received a doctorate in management of technological innovation from Sloan School of Management at the Massachusetts Institute of Technology, a master's degree in organizational behavior from Brigham Young University, and a Bachelor of Science degree in aerospace engineering from California State Polytechnic University, Pomona. [erebenti@mit.edu]

Allen Moulton—is a Research Scientist at MIT's Sociotechnical Systems Research Center. His research has addressed issues of capability-based defense acquisition requirements planning and cybersecurity in DoD; critical success factors in achieving economic benefits from information quality across enterprises; and applying abductive logic programming for automated reasoning to resolve semantic differences in the interchange of information among autonomous, heterogeneous data sources and receivers. Prior work included applications of information technology to substantive foreign policy analysis. He is co-author of *Managing International Conflict: from Theory to Policy* and the CASCON software for analyzing international conflict. [amoulton@mit.edu]

Abstract

Digital engineering transformation changes the practice of systems engineering, and drives the need to re-examine how engineering effectiveness is measured and assessed. Early engineering metrics were primarily lagging measures. More recently leading indicators have emerged that draw on trend information to allow for more predictive analysis of technical and programmatic performance of the engineering effort. By analyzing trends (e.g., requirements volatility) in context of the program's environment and known factors, predictions can be forecast on the outcomes of certain activities (e.g., probability of successfully passing a milestone review), thereby enabling preventative or corrective action during the program. This paper discusses continuing research on the adaptation of existing systems engineering leading indicators (developed under the assumptions of document-based engineering) for digital (model-based) engineering. Model-based implications identified in the research are discussed in support of the use of existing leading indicators in digital engineering programs. An illustrative example describes how measurement data can be extracted from a digital system model and composed into indicators. The importance of visualization and interactivity is discussed, especially the potential role of visual analytics and interactive dashboards. Several recommendations for future research are proposed based on interim research findings.

Introduction

Defense programs have long used engineering metrics to provide status and historical information, but implementation has been limited by the nature of the traditional, document-based engineering approach. Further, early systems engineering metrics were primarily lagging measures, providing information for the next program instead of the current one. Systems engineering leading indicators were subsequently developed to allow for more timely predictive analysis of the technical and programmatic performance of the engineering effort on a program. Leading indicators use an approach that draws on trend information to allow for more predictive insight (Rhodes et al., 2009). A systems engineering leading indicator is a measure for evaluating the effectiveness of how a specific program activity impacts engineering effectiveness, which may affect the system performance objectives.

Both lagging and leading indicators are found to be useful in many fields (e.g., economics, health, social science; Zheng et al., 2019). While lagging measures (e.g., system defects) continue to provide useful information over time for an enterprise, they are insufficient for real-time decisions during a program. Relatively little evidence exists on the application of leading indicators in the engineering of systems. The value of leading indicators comes from examining trends (e.g., requirements volatility) in context of the program's characteristics and known factors. This information can then be used to make predictions to forecast the outcomes of certain activities (for example, likelihood of successfully passing a milestone review). Leading indicators have provided some improved ability to assess ongoing engineering effort, and where necessary, take preventative or corrective action during the program.

Existing leading indicators were developed under the document-based engineering approach. The introduction of digital engineering practices can have a potentially radical or disruptive impact on the processes, tools, and time lines of engineering programs. Rapidly accelerating analytical and design capabilities will have limited impact on overall program pace and effectiveness if reviews and decision-making processes fail to adapt to the processes and cadence of digital engineering and management. Research is necessary in order to understand and adapt existing indicators for digital engineering and management practice. Additionally, the art of the possible needs to be explored including how digital system model information could be used to extract and compose base measures into indicators. Investigation is also needed to understand how newer sciences and technologies—such as data science, visual analytics, and interactive dashboards—could better inform timely leadership decisions in model-centric programs.

Background

Foundational work on systems engineering leading indicators was initiated in 2004. The early efforts produced a systems engineering leading indicators guide (Roedler & Rhodes, 2007) with 13 leading indicators defined using measurement specifications. This work was subsequently evolved through collaboration from organizations and individuals across the systems engineering community with over 20 organizations as contributors. The result was a second version of the guide (Roedler et al., 2010), with five new leading indicators and several appendices added. Additional studies and papers have also been published by various authors (Elm et al., 2008; Elm & Goldenson, 2013; Gerst & Rhodes, 2010; Gilbert et al., 2014; Knorr, 2012; Montgomery & Carlson, 2010; Orlowski, 2017; Orlowski et al., 2015; Rhodes et al., 2009; Shirley, 2016; Zheng et al., 2017; and Zheng et al., 2019).

Prior work on leading indicators was done under the assumption of traditional, document-based engineering practice. This paper shares interim findings of a continuing research effort to investigate adaptation of the systems engineering leading indicators for digital engineering and use in model-centric programs.

Motivation and Research Approach

The broad motivation for the work is to enable more timely and informed decisions on systems engineering activities and resources. The transformation to digital engineering has prompted a need to re-examine the systems engineering leading indicators for this new context. The investigation aims to provide findings for model-centric programs seeking to use the leading indicators, as well as contribute recommendations to inform the larger effort of the systems engineering community to establish the next generation of digital engineering effectiveness measurement.

Since each of the indicators requires some additional considerations under digital engineering, the first year in this research focused on identifying potential modifications and interpretation guidance (Rhodes, 2020). The digital engineering environment and newer technologies open new possibilities for providing program leaders with leading insights into the effectiveness of systems engineering efforts. Accordingly, in the second year of the research the focus has been on defining specific model-based implications for the 18 leading indicators, and possible enhancements through three areas of inquiry. First, the research explores how metrics data can be obtained from digital system models that are produced by the systems engineering team. Second, it explores how information from descriptive system models can be extracted and composed as composite leading indicators that give overall indication of effectiveness versus a set of separate indicators. An illustrative case is used to show how measurement data can be extracted directly from descriptive system models and composed as enhanced leading indicators that can provide insight into effectiveness of engineering on a model-centric program. (*Note: Integrated model-based tools were used in investigations during this research project. The selection of tools in the examples is not intended as an endorsement; the software used by the research team was selected based on pre-existing availability and case examples).* Third, the research seeks to understand how interactive dashboards can be used to extract and more effectively display measurement information to positively impact program reviews and decisions.

The justification for pursing this research approach extends from the DoD Digital Engineering Strategy (DoD, 2018), which discusses five goals. Goal 3, Incorporate Technological Innovation to Improve Engineering Practice, has a Subgoal 3.2 that discusses the use of technological innovations to improve digital engineering practice. There are many technological innovations of interest; some are specifically relevant to both the measures of digital engineering effectiveness on a program and the enabling technologies to support collection, analysis, and display of leading indicators of engineering effectiveness. As noted in the strategy, “data analytics can help gain great insights from existing model data” (DoD, 2018, p. 14). The strategy recommends that stakeholders use technological innovations to improve decision-making and performance of computationally intensive engineering activities.

The collection and analysis of systems engineering measurement data falls under that category of activities. Digital engineering tools are recognized as a means to increase engineering efficiency (DoD, 2018, p. 17) and to provide access to vast data. Leading indicators are especially important to monitoring effectiveness on a continuous basis, and also to ensure that effectiveness is not compromised for sake of efficiency. The strategy calls for leadership to “establish accountability to measure, foster, demonstrate, and improve tangible results across programs and the enterprise” (DoD, 2018, p. 22). Common enabling technologies used in digital environments to generate, analyze, and display measurement data will encourage a common foundation for cross-program comparison and learning.

Knowledge gathering from subject matter experts through technical exchanges and workshops provided insights regarding adaptation of leading indicators and potential new indicators of interest. This includes investigation of publications, studies, workshop reports and interim research findings from academic research groups, professional and industry societies, and cross-industry initiatives. Literature review is used to explore newer leading-edge techniques and approaches for collection and synthesis of measurement data, as enabled by digital engineering practices and environments.

Measurement Specifications

Standardizing leading indicators of engineering effectiveness across programs is facilitated through measurement specifications. The systems engineering community has been using measurement specifications for many years, based on foundational work of PSM in software and systems measurement (PSM, 2020). The systems engineering leading indicators initiative adopted the PSM measurement specification format. Accordingly, each of the systems engineering indicators is characterized using a measurement specification with detailed description, insights provided, interpretation guidance, and usage guidance. Detailed contents of the measurement specifications for leading indicators is described in Roedler et al. (2010) and summarized in Table 1.

Table 1. Systems Engineering Leading Indicator Specification Fields (Roedler et al., 2010, adapted by Zheng et al., 2019)

1.Information need description	
Information need	Specifies what the information need is that drives why we need this leading indicator to make decisions
Information category	Specifies what categories (as defined in the PSM) are applicable for this leading indicator(for example, schedule and progress, resources and cost, product size and stability, product quality, process performance, technology effectiveness, and customer satisfaction)
2. Measurable concept and leading insight	
Measurable concept	Defines specifically what is measurable
Leading insight provided	Specifies what specific insights that the leading indicator may provide in context of the Measurable concept - typically a list of several or more
3. Base measure specification	
Base measures	A list of the base measures that are used to compute one or more leading indicators - a base measure is a single attribute defined by a specified measurement method
Measurement methods	For each base measure, describes the method used to count the base measure,for example simple counting or counting then normalized
Unit of measurement	Describes the unit of measure for each of the base measures
4. Entities and attributes	
Relevant entities	Describes one or more particular entities relevant for this indicator - the object is to be measured (for example, requirement or interface)
Attributes	The function for computing the derived measure from the base measures
5. Derived measure specification	
Derived measure	Describes one or more measures that may be derived from base measures that will be used individually or in combination as leading indicators
Measurement function	The function for computing the derived measure from the base measures
6. Indicator specification	
Indicator description and sample	A detailed specific description and display of the leading indicator,including what base and/or derived measures are used
Thresholds and outliers	Would describe thresholds and outliers for the indicator; this information would be company (and possibly project) specific
Decision criteria	Provides basic guidance for triggers for investigation and when possible action to be taken
Indicator interpretation	Provides some insight into how the indicator should be interpreted, each organization would be expected to tailor this
7. Additional information	
Related processes	Lists related processes and sub-processes
Assumptions	Lists assumptions for the leading indicator to be used, for example, that a requirements database is maintained
Additional Analysis Guidance	Any additional guidance on implementing or using the indicators
Implementation Considerations	Considerations on how to implement the indicator (assume this expands with use by organization)
User of Information	Lists the role(s) that use the leading indicator information
Data Collection Procedure	Details the procedure for data collection
Data Analysis Procedure	Details the procedure for analyzing the data prior to interpretation

In the near term, the existing measurements specifications can be augmented with model-based implications. In the future, modified and new measurement specifications are envisioned in a new release of the leading indicators guide. Developing the next version of the guide necessitates a community effort extending from implications identified in this research, insights from practitioners, and results of ongoing investigations and initiatives on digital engineering metrics.

Model-Based Implications for Leading Indicator Implementation

The existing 18 leading indicators, as investigated through semi-structured interviews and technical exchange workshops, were shown to have varying implications related to model-based systems engineering. The implementation of a leading indicators in context of digital engineering will be based on many factors, such as nature of the program, processes used by the enterprise, model-based toolset selection and implementation, engineering culture of the enterprise, and maturity of digital engineering in the enterprise, as well as external influences (e.g., customer preferences, etc.).

Based on research findings, the leading indicators are grouped into three subsets: (1) leading indicators most likely to be implemented with direct use of a model-based toolset; (2) leading indicators most likely to be partially implemented with use of a model-based toolset; and (3) leading indicators less likely to be implemented with use of a model-based toolset. The three groups of leading indicators are then summarized in Tables 2, 3, and 4 to highlight model-based implications. Prior to the summary tables, two indicators, *requirements trends* and *facility and equipment trends,* are discussed in greater detail.

Requirements Trends Leading Indicator

The *Requirements Trends* leading indicator is used to evaluate the stability and adequacy of the requirements to understand the risks to other activities toward providing required capability, on time and within budget. This is done through an evaluation of trends in the growth, change, completeness, and correctness of the system requirements definition, as well as the quality of and consensus around the system operations concept. This indicator provides insight into rate of maturity of the system definition against the plan, and whether the system definition is maturing as expected. Additionally, it characterizes the stability and completeness of system requirements that could potentially impact design, production, operational utility, or support.

Requirements growth, changes, or impacts that exceed expectations or exhibit a lower closure rate of TBDs/TBRs than planned may indicate insufficient quality of architecture, design, implementation, verification, and validation efforts. This in turn could result in elevated schedule and cost risks, and/or a future need for different levels or types of resources/skills.

Near Term: The use of requirements management tools and databases is a mature practice in systems engineering. Tracking the growth trends and volatility of requirements is therefore a relatively straightforward matter of the compilation of data on the requirements within the database and the development of processes for regular review and action where implied. These functions could be incorporated into or added to existing requirements management tools within the MBSE environment to assist program decision-makers in assessing progress during the system development.

Longer Term: MBSE tools and methods introduce a number of new ways to assess and understand the quality of requirements and the degree to which they are being met over the course of the system development life cycle. A transition to primary use of an MBSE approach in system development could enable a broader range of analysis and model checking.

The expression of requirements as executable models has been demonstrated to improve the quality of requirements and decrease errors relating to poorly-defined requirements (Micoun et al., 2018). Model-based requirements provide the ability to validate that the system model is logically consistent, and the ability to answer questions such as the impact of a requirement or design change, or the assessment of how a failure could propagate through a system. Using this approach, it is possible to verify design models using a simulation-based verification process in order to detect and remove design errors. Model-based requirements may be included in a curated database for reuse in other development efforts, with the potential for savings in time and resources.

Model-based requirements may be used in early system analysis to assess requirements completeness and correctness through the identification of gaps, conflicts, or redundancies in the existing requirements set, prior to the development of more detailed engineering models and analysis. MBSE analysis using model-based requirements could validate the requirements themselves and ensure they don't contribute to undesirable emergent behaviors at the system level. A potential indicator of requirements quality in the MBSE environment might include the percentage of requirements that are formatted and expressed as models and the rate and total proportion of requirements validation through modeling and simulation at both the component and system level.

Model-based requirements may be archived and reused across multiple development projects. Any issues that are identified with the requirements in one project could potentially be traced to other projects that use the same models. The traceability inherent in using these archived requirements models enables enhanced root cause analysis and system refinement, triggering actions to correct and validate the originating requirement to prevent continuing propagation of errors. An indicator of requirements maturity in an MBSE environment might include the proportion of requirements models that include a validation pedigree. The presence of requirements models without a validation pedigree (at least to a specific standard defined by the enterprise) could indicate greater risk of potential future requirements changes and instability in the system baseline.

Facilities and Equipment Trends Leading Indicator

The *Facility and Equipment Availability Trends* leading indicator is used to determine the availability of critical facilities and equipment needed for systems engineering activities over the project life cycle. The indicator is composed of two metrics, measuring facility availability and equipment availability. The intent of this indicator is to provide a view of facility and equipment availability on the project over time. Facilities and equipment are of different types and may provide key capabilities to the program. The facility availability measurement provides insight into the difference between the planned need for a facility type and the existing inventory of available facilities that meets the need for the desired capability. Insufficient facilities—labs, test ranges, floor space, etc.—of various types may cause a project to be unable to meet its customer needs, create costly overruns, and inability to meet schedule targets. Similarly, a project requires various types of equipment that also may provide key capabilities for the program. Equipment availability measurement provides insight into the difference between the planned need for an equipment type and the existing inventory of available inventory that meets the need for the desired capability. Insufficient equipment (fabrication equipment, measurement equipment, cleanroom equipment, test equipment, software and systems applications, etc.) may cause a project to be unable to meet its customer needs, create costly overruns, and inability to meet schedule targets. Facility Availability and Equipment Availability as measurable concepts assess whether adequate facilities and equipment can be allocated to the project to meet life cycle milestones. This reveals differences between systems engineering needs on the project and available facilities and equipment based on projected needs. The leading insights provided to the project are potential shortfalls of systems engineering related facilities and equipment, and potential problems with the project's ability to meet desired milestones (Roedler et al., 2010).

Near Term: As an initial step in adapting the existing systems engineering (SE) leading indicators, the measurement specification can be augmented by adding model-based systems engineering implications to the Implementation Considerations within the Additional Information section of the measurement specification. Model-based programs necessitate personnel have (or have access to) computing “equipment,” including desktop/laptop computers or workstations with adequate performance, access to networks and/or intranet, data and model repositories, model libraries, computer services support, data/cloud storage, etc. Facilities may include the individual engineer’s workspace, as well as collaborative spaces. There is also a need to have access to the selected version of model-based toolset that is maintained. The facilities and equipment need to support any required upgrades of versions, which may have implications for the existing computing facilities. Another implication consideration is that facilities and equipment must accommodate any necessary collaboration with other internal groups and/or external organizations (such as a supplier or customer) as needed. The facilities and equipment must be adequate to support this. This includes necessary facilities and equipment to support tool interoperability, data/model exchange, version compatibility control, model sharing, model security, etc. Model-based programs need to have adequate budget allocated, as insufficient availability of the necessary facilities and equipment will have major impact on systems engineering effectiveness.

Longer Term: As we look to the future of digital engineering, the issue with using the existing Facilities and Equipment Availability leading indicator is that it takes a somewhat decoupled approach at these rather than as highly interconnected, as is the case for MBSE. In fact, with the transformation of traditional engineering to digital engineering, there is a need to look at this in context of the larger digital ecosystem. This includes interconnected digital environments that extend beyond the boundaries of the engineering organization. In the existing SE leading indicator guide published in 2010, the Facilities and Equipment Leading Indicator has relatively less substance than the other indicators given that it was not a major focus of the team. With digital engineering transformation, taking the perspective of the overall digital engineering ecosystems is necessary. The success of systems engineering on a program will be fully dependent upon the environment and infrastructure available to participate as part of the larger ecosystem. The supporting infrastructure required for digital engineering (Bone et al., 2018) necessitates that a new leading indicator be developed respective to the importance it has to system success and the dimensions and complexity of that infrastructure.

Leading Indicators Most Likely to Be Implemented with Direct Use of a Model-Based Toolset

The first subset of leading indicators, as shown in Table 2, are those that are most likely to be implemented with the direct use of the program’s MBSE toolset. In this case, the base measures as shown in the respective measurement specifications in the leading indicator guide (Roedler et al., 2010) are likely to be obtainable from the system model and composed into a leading indicator. Assuming an effective user interface and any required trend data, this provides the ability to obtain real-time leading indicator information to better inform and accelerate decisions based on this information.

Table 2. Leading Indicators Most Likely to Be Implemented With Direct Use of Model-Based Toolset
(Roedler et al., 2010)

Leading Indicator	Insight Provided	Model-Based Implications
Requirements Trends	Rate of maturity of the system definition against the plan. Additionally, characterizes the stability and completeness of the system requirements that could potentially impact design, production, operational utility, or support.	• See the section Requirements Trends Leading Indicator for a detailed discussion.
System Definition Change Backlog Trend	Change request backlog which, when excessive, could have adverse impact on the technical, cost, and schedule baselines.	• Model-based tools will enable collection and analysis of data • MBSE enables fixing defects earlier in time, where less effort is typically required. Accordingly, historical trends will vary from model-centric programs.
Interface Trends	Interface specification closure against plan. Lack of timely closure could pose adverse impact to system architecture, design, implementation, and/or V&V, any of which could pose technical, cost, and schedule impact.	• Similar to requirements trends; see the section Requirements Trends Leading Indicator for a detailed discussion.
Requirements Validation Trends	Progress against plan in assuring customer requirements are valid and properly understood. Adverse trends would pose impacts to system design activity with corresponding impacts to technical, cost, & schedule baselines and customer satisfaction.	• Similar to requirements trends; see the section Requirements Trends Leading Indicator for a detailed discussion. • Model-based tools may accelerate the pace of validation so historical data trend data may not be as useful.
Requirements Verification Trends	Progress against plan in verifying design meets the specified requirements. Adverse trends would indicate inadequate design and rework that could impact technical, cost, and schedule baselines. Also, potential adverse operational effectiveness of the system.	• Similar to requirements trends; see the section Requirements Trends Leading Indicator for a detailed discussion. • Model-based tools may accelerate the pace of verification so historical data trend data may not be as useful.

Leading Indicators Most Likely to Be Partially Implemented with Use of a Model-Based Toolset

The second subset of leading indicators, as shown in Table 3, are those that are most likely to be partially implemented with the use of the program's model-based toolset. For example, technical performance risk information might be associated with the system model, but there may be other programmatic risk information that is tracked elsewhere. The extent to which the five leading indicators in this table are able to be generated from a model is dependent on what types of models the program uses, and how model-based toolsets are customized and extended.

Table 3. Leading Indicators Most Likely to Be Partially Implemented With Use of Model-Based Toolset (Roedler et al., 2010)

Leading Indicator	Insight Provided	Model-Based Implications
Risk Exposure Trends	Effectiveness of risk management process in managing/mitigating technical, cost, & schedule risks. An effective risk handing process will lower risk exposure trends.	• Model-based tool sets provide opportunity to have risk associated with or directly included within models.
Risk Treatment Trends	Effectiveness of the SE organization in implementing risk mitigation activities. If SE is not retiring risk in a timely manner, additional resources can be allocated before additional problems are created.	• Model-based tool sets provide opportunity to have risk associated with or directly included within models.
Technical Measurement Trends	Progress towards meeting the Measures of Effectiveness (MOEs)/Performance (MOPs)/Key Performance Parameters (KPPs) and Technical Performance Measures (TPMs). Lack of timely closure is an indicator of performance deficiencies in the product design and/or project team's performance.	• Model-based approaches, methods, and tools will enhance technical performance measurement. • Ability to project planned value and predict variances may be improved, so tolerance bands may vary from traditional engineering.
Defect/Error Trends	Progress towards the creation of a product or the delivery of a service that meets the quality expectations of its recipient. Understanding the proportion of defects being found and opportunities for finding defects at each stage of the development process of a product or the execution of a service.	• With model-based approach errors and defects may be found earlier in time; software can automate finding and fixing some defects. • Necessitates defining an alternative to "defects per page." • Historical defect discovery profiles from traditional engineering will likely not be suitable; defects models and discovery profiles will need to be developed as experience grows
Work Product Approval Trends	Adequacy of internal processes for the work being performed and also the adequacy of the document review process, both internal and external to the organization. High reject count would suggest poor quality work or a poor document review process each of which could have adverse cost, schedule, and customer satisfaction impact.	• Models may become tracked work products in model-centric programs; criteria would need to be developed. • Models may influence the approval rate of system work products.

Leading Indicators Less Likely to Be Implemented with Use of Model-Based Toolset

The third subset of leading indicators, as shown in Table 4, are those that are less likely to be implemented with the use of a program's model-based toolset. Presently, these leading indicators would likely be tracked in a separate technical management tool or tracking system. Model toolset experts view it as possible to extend model-based toolsets to include any programmatic and process models in a model-centric environment. So, while at present there are likely to be few programs that have implemented this, the likelihood will increase over time as model-based environments evolve.

Table 4. Leading Indicators Less Likely to Be Implemented with Use of Model-Based Toolset (Roedler et al., 2010)

Leading Indicator	Insight Provided	Model-Based Implications
Technology Maturity Trends	Risk associated with incorporation of new technology or failure to refresh dated technology. Adoption of immature technology could introduce significant risk during development while failure to refresh dates technology could have operational effectiveness/ customer satisfaction impact.	• Increased use of models may enhance ability to assess potential impacts.
Review Action Closure Trends	Responsiveness of the organization in closing post-review actions. Adverse trends could forecast potential technical, cost, and schedule baseline issues.	• Model-centric programs are likely to have more continuous action item review than traditional. • Technical-related action items may be directly linked to models.
Systems Engineering Staffing & Skills Trends	Quantity and quality of SE personnel assigned, the skill and seniority mix, and the time phasing of their application throughout the project life cycle.	• Model-based approaches, methods, and tools require additional staffing and skills, possibly at different points in program. • Insufficient model-based staffing/skills have impact on cost, schedule, and quality.
Process Compliance Trends	Quality and consistency of the project defined SE process as documented in SEP/SEMP. Poor/inconsistent SE processes and/or failure to adhere to SEP/SEMP, increase project risk.	• Model-based programs will be using newer and/or developing processes integrated with toolsets. • Compliance deviations and comments recorded within the model enable automated compliance measurement. • Process compliance measurement needs to accommodate modifications to process given learning on program and/or other programs.
Facility and Equipment Availability Trends	Availability of non-personnel resources (infrastructure, capital assets, etc.) needed throughout the project life cycle.	• See the section Facilities and Equipment Trend Leading Indicators for a detailed discussion.
System Affordability Trends	Progress toward a system that is affordable for the stakeholders. Understanding the balance between performance, cost, and schedule and the associated confidence or risk.	• Assessing affordability under the digital engineering paradigm is likely to require different approach. • Lacking historical data, model-based programs need to develop, approach, and adjust measurement of this.
Architecture Trends	Maturity of an organization with regards to implementation and deployment of an architecture process that is based on an accepted set of industry standards and guidelines.	• Model-based approaches/tools will have influence on assessing maturity. • Programs should tailor base measures as needed to reflect advantages of model-based approaches/tools.
Schedule and Cost Pressure	Impact of schedule and cost challenges on carrying out a project.	• Minimal historical data available for the digital engineering situation, and setting notional values for thresholds may be challenging.

Composability of Leading Indictor Measurement Data

Leading indicators are most useful when applied for predictive purpose to facilitate programmatic decisions and/or corrective actions. Requirements Trend indicators, for instance, are used to evaluate trends in the growth, change, completeness, and correctness of the definition of system requirements. Traditionally, this indicator provides insight into the rate of maturity of system definition against the plan. It also characterizes stability and completeness of system requirements which could potentially impact design, production, operational utility, or support. In traditional document-based engineering practice, requirements are central objects that can be used for assessing maturity of system definition. In model-based engineering, however, there are many other constructs and digital artifacts. With modeling languages (e.g., SysML, LML) there are requirements diagrams, use case diagrams, activity diagrams, state machine diagrams, parametric diagrams, and others. With the advantages of model-based approaches, a leading indicator used to assess progress of system definition that uses only requirements would be a limited indicator. In this case, one would want to consider progress of systems definition to include composition of measurement information from system diagrams of all relevant types.

Composability concerns selection of elements that can logically and reasonably be assembled. In context of this research, the focus is on composability of base measures extracted from a digital system model or digital process model used to produce a leading indicator. An initial step is to consider the existing 18 leading indicators. Future research is needed to explore new leading indicators (e.g., model volatility) that are made tractable through model-based toolsets. Automation and augmented intelligence offer opportunities to explore the future of leading indicators for digital engineering program decision-making.

Illustrative Case Discussion

An illustrative case has been used in the research to explore how digital engineering is expected to modify and/or enable the leading indicators most likely to be implemented with direct use of model-based toolsets. These five leading indicators (see Table 2) all relate to aspects of requirements management. In the current state of practice, requirements are typically collected and stored in a specialized requirements database, often using software (e.g., DOORS® or other similar packages suited to needs of the project/enterprise). These types of packages are generally interoperable with and/or loosely coupled to other systems engineering model-based toolsets.

It is the assumption of this research team that the specific details of this will vary based on the chosen model-based tools used. For the purposes of this illustrative case, Innoslate® was used to conduct a number of small scale exercises. Innoslate® is an integrated MBSE software package that implements the open source LML ontology, which is compact but comprehensive (Dam, 2019, p. 5). LML provides an organized and structured terminology for systems engineering, enterprise-defined extensions, and includes features that support the earliest concept stage throughout the life cycle to disposal (Dam, 2019, pp. 6, 10). Innoslate enforces the important principal of *concordance,* which facilitates single source of truth by requiring that a given piece of information in the systems engineering knowledge base will have the same meaning when viewed through different language or visualization lenses.

The *Action* entity in the LML ontology is the primary building block for functional models. Similarly, the *Asset* entity is the primary building block for physical models. Every entity has a "type" property, which allows many variants of Actions and Assets to be represented. For example, Actions may be described as Activity, Capability, Event, Function, Process, or Task. Assets may be described as Component, Entity, Service, Sub-system, or System as needed in a particular modeling context.

The other key basic concept in functional models is *Input/Output*, which represents the flow of information in or out of an Action, including Item, Trigger, Information, Data, and Energy. The corresponding basic concept in a physical model is the *Conduit*, which might be implemented as a Data Bus, Interface, or Pipe.

Relationships are used to make connections among entities. Decomposition is denoted with the *decomposed by/decomposes* relationships. A functional model Action can be allocated to a physical model Asset using the "*performed by/performs*" relationship. A functional model Input/Output entity may be allocated to a physical model Conduit via the "*transferred by/transfers*" relationship where the functional flow thereby becomes constrained by the properties of the physical device implementing the Conduit. Standard entity attributes are defined in the LML ontology along with standard relationships and the entity types they connect (LML Steering Committee, 2015).

Tracking the trends needed for the leading indicators (Table 2) requires taking snapshots of metrics values at intervals over time as the program proceeds. If integrated into an LML model, this program management data would be stored as objects in the database to facilitate integrative analysis with other program data.

Requirements Trends and Interface Trends

The metrics required for *Requirements Trends* and *Interface Trends* can be composed by counting explicit and implicit requirements identified in the Innoslate database. Explicit requirements are found in Requirements entities that contain natural language statements, which are (1) sourced from documents loaded into the system; (2) entered directly into the database by engineers; or (3) computed from other data and stored in the database.

Implicit requirements are derived from the functional and physical models developed by engineers during requirements analysis. Functional Requirements may be defined by Action entities and the flows, relationships, and properties that describe them. Innoslate also has a tool that converts Actions in a functional model into implied Assets and Conduits in a physical model.

Interface Requirements can be inferred from Conduit entities that connect Assets in the physical model. The technical characteristics of the endpoint Assets and the Conduit combine to specify the interface requirements. Performance Requirements often come from data related to Asset entities and connections. External interfaces would be represented by connecting a Conduit to an Asset that is outside the system boundary.

As requirements analysis progresses, the model and the requirements will grow deeper and broader. In traditional practice, the requirements are frozen in text and isolated from the models that engineers use for analysis. Whether explicit or implicit, a requirement in Innoslate is linked by relationships to other elements of the model, giving greater context to understanding the meaning of a requirement. For example, by running simulations on executable models, the engineer can identify whether a set of requirements has face validity or meets expectations. Spider charts and hierarchy charts can be used to visualize the structure of the model and the requirements.

As systems understanding develops, some information will be less refined than other information. For example, the value for a parameter in a requirement may be unknown (TBD) or estimated (TBR). LML Decision entities can be attached to the model to represent the both the uncertainty and the process for finding the missing information as well as defining assumptions. When the TBD/TBR is resolved, the updated Decision entities provide a record of how the value was obtained.

Requirements Validation Trends and Requirements Verification Trends

Systems engineering practice involves beginning requirements validation and verification early in the project as requirements are found and entered into the database. At the early stage, Innoslate and some other toolsets offer a natural language tool for checking the quality of requirements statements against six of the eight standard criteria (clear, complete, consistent, design, traceable, verifiable but not correct and feasible). Another tool applies heuristics to evaluate models and requirements in more depth. A roll-up of these quality metrics could provide leaders with early insight on how well the requirements are progressing and whether problems are being left to later in the life cycle where they will be more difficult to resolve. Innoslate also includes a Test Center where test plans and scenarios can be built for early or later use and VCRM Reports generated. The leading indicators for requirements verification and requirements validation could be improved by adding a measure for progress on developing test plans to complement the metric for successful completion of validation and verification testing. Product validation and verification also needs to be considered holistically as well as individually by requirement. The model can be used with simulation tools to predict the behavior of the whole system or subsystems.

Visualization and Interactivity

More complex leading indicators are likely in the digital engineering context, resulting from increased information, synthesis, and composability of measurement data. Accordingly, decision-makers will face challenges in comprehending the information, including a need to understand the underlying assumptions and uncertainties in the constituent data elements. Investigating the approach to display such leading indicators is an important area of inquiry. Measurement dashboards have been used extensively for decades, typically providing static display of information. Visual analytics and interactive technologies provide the opportunity to create dynamic dashboards that would enable a decision-maker to be able to interact with the data. This provides more transparency to underlying data, as well as enabling the development of understanding and trust in the information.

Visual Analytics

Visual analytics is fundamentally about collaboration between a human and a computer using visualization, data analytics, and human-in-the-loop interaction. More than just visualization tools, visual analytics aims to take advantage of a human's ability to discover patterns and drive inquiry to make sense of data. Thomas (2007) defined visual analytics as "the science of analytical reasoning facilitated by interactive visual interfaces" that "provides the last 12 inches between the masses of information and the human mind to make decisions." As engineering becomes increasingly model-based, the available information to draw on to generate measures of effectiveness is vast and complex. It is foreseeable that decision-makers could be presented with large amounts of data that would be cognitively challenging to comprehend and find patterns that could be used to judge the effectiveness of the engineering on an ongoing program. For this reason, the knowledge and recent advancements in visual analytics may offer significant support in processing and displaying measurement data.

Vitiello and Kalawsky (2012) state the "guiding process in visual analytics is a synergy between interactive visualization and automated analysis of the data." They discuss an approach that integrates a visual analytic-based workflow to the notion of sensemaking. The authors describe using visual analytics to support systems thinking to make sense of complex systems interactions and interrelationships, enabling rapid modeling of the systems of interest for systems engineering design and analysis processes. The visual analytic-based sensemaking framework they describe aims toward providing the means to rapidly gain valuable insights into the data.

Interactive Dashboards

Systems engineers, managers, and government sponsors all rely on creative work products of systems engineering and all need to glean an appropriate level of understanding of the work as it progresses. The mean time for a warfighting system to move from well-defined concept to initial operating capability can be substantial, regularly averaging 6–7 years (Dwyer, 2020). Leading indicators can help stakeholders see how a project or program is progressing throughout the life cycle and whether it is on target to deliver what is needed when it is needed at an affordable cost.

The complexity of understanding the status and trajectory of a program is high and larger than any one person can hold in one's head. Systems engineering methods, languages, and models are intended to leverage visualizations, structure, and compuational representations to make the task manageable for all the humans who must be involved. Model-based systems engineering incorporates all of those features. Sindiy et al. (2013) demonstrates how clean visual representations can help in making MBSE models accessible. Dam (2019) argues that in addition to visualizations, modeling language and ontology matters, since a representation that is inherently fragmented and lacks a well-structured ontology will be less cognitively accessible to users. Dashboards are often created as views into program data that has been extracted and loaded into a data warehouse. Dam (2020) proposes that stakeholders should be given controlled direct access to MBSE models to improve the speed and depth of understanding in system reviews. He also argues that prime contractors and subcontractors can achieve better coordination by using MBSE models as a vehicle for communication about the system that is being created, program progress, and how organizations with different roles and incentives will fit together to deliver the capability needed to meet customer objectives.

Selby (2009) argues that interactive dashboards facilitate effective management. Leading indicator project data can be presented in a compact form with tools for organizing data, drilling into the underlying data, and connecting data to analytic tools and models. Orlowski (2017) and Orlowski et al. (2015) extend Selby and propose a framework for guiding leading indicator development and usage. Recent work by Thiruvathukal et al. (2018) shows the potential for using open source software repositories in the development of software metrics dashboards. Nadj (2020) addresses how interactive dashboards help managers in gaining and maintaining situational awareness to understand the context of metrics.

Discussion and Future Directions

Potential impact of adapted and extended leading indicators is twofold: (1) to continue to provide visibility into the future state through use of leading indicators in model-centric programs; and (2) to enhance insights provided by the leading indicators related to digital models and artifacts that enrich the systems development practice. The current set of leading indicators is predicated on use of document-driven processes and major milestone reviews. The transformation to digital-model based engineering will result in the use of digital artifacts, with more frequent review of the expected system performance (Parrot & Weiland, 2017). One of the expected outcomes of digital engineering is to move away from milestone design reviews to more continuous reviews given access to the maturing system model. Leading indicators can be very supportive of this goal (Orlowski, 2017; Orlowski et al., 2015). An open question is how trend information for models and digital artifacts (e.g., SysML diagrams) could be used in a similar manner to predict when the model is in a state where a review activity is most useful.

Interim Research Findings

Interim outcomes of the research include knowledge for augmenting measurement specifications of existing systems engineering leading indicators to describe model-based implications, and illustrating what is possible for enhanced measurement through direct use of model-based toolsets. Another interim outcome is initial investigation of how visual analytics and interactive dashboards may be used to provide leading indicator information that provides a deeper understanding for program leaders. This research has identified considerations and additional interpretation guidance to augment existing systems engineering leading indicator measurement specifications. The results are aimed at assisting systems engineering organizations that have been using leading indicators as they transition to model-based systems engineering.

This research has also included initial investigation for how model-based toolsets enable the collection and composability of base measures to generate leading indicators. An initial investigation of interactive dashboards suggests that program leaders will be able to make improved and accelerated decisions using leading indicators if these are integrated with model-based environments to provide on-demand trend information. Implications identified in this research, including potential new leading indicators, can inform ongoing efforts in the systems community to define new or revised metrics for digital engineering programs and enterprises.

Related Research and Initiatives

Interim and final research outcomes are shared with several working groups as input to future research on digital engineering metrics and revision of the leading indicators guide. The DoD Systems Engineering Research Center (SERC) performed a research activity in 2020 that investigated digital engineering transformation metrics (McDermott et al., 2020). Metrics categories were derived as a general categorization, including adoption, user experience, velocity/agility, quality, and knowledge transfer. Literature review and a survey were performed, resulting in a set of top metrics categories related to the benefit of digital engineering. Although the focus of the leading indicators and digital engineering metrics are somewhat different, their relationship is worth considering as potential additional leading indicators are developed.

The outcomes of the SERC investigation, other ongoing measurement related efforts, and foundational work by PSM are employed in a broader Aerospace & Defense Digital Engineering Metrics Initiative. This initiative brings together a diverse team including industry associations, government agencies, FFRDCs, UARCs, academia, and industry (including AIA, NDIA, INCOSE, OUSD R&E, SEI, SERC, the Aerospace Corporation, PSM, MIT, and several companies). The effort aims to define an industry consensus measurement framework and candidate performance indicators for digital engineering, and to align measures with business information needs for project execution and organizational performance improvement.

Limitations and Future Research

The research largely draws from the defense systems engineering community and literature from that sector. Future research can investigate additional sectors, as well as related disciplines. Expert knowledge was gathered though available workshops and from prior leading indicator project participants in the early phases. The limitations imposed by the COVID-19 pandemic, especially on workshops and conference events other than virtual, resulted in reduced opportunities for access to the community of interest. Planned group discussions were replaced with individual interviews and discussions, which resulted in reduced iteration and feedback opportunities.

This research has included some experimentation with extraction of metric data based on a single systems engineering toolset (selection of toolset was based on ease of use and availability to research team). Future research is needed to investigate extraction and composition of measurement information across the available model-based toolsets. Additionally variation in implementing digital engineering practice need to be examined in regard to this objective. For example, some of the existing leading indicators depend on disciplined management processes for approval of key program artifacts (e.g., requirements, change orders, interfaces, and test plans). While these processes are not part of the system being developed, they can be modeled and/or tracked through model-based toolsets. This would enable measuring aspects of process compliance. Dam (2019) gives examples of how software could be used in support of measuring management processes.

Model-centric programs have the opportunity to leverage leading-edge technologies in the collection, composition and display of measurement data, as well as enable better decisions to be made throughout the program life span. Two aspects for future investigation are techniques emerging from visual analytics and data science. Model-based acquisition programs will be faced with dealing with four cited challenges of big data: *volume*: the magnitude of digital engineering information; *variety*: existence of digitized assets (e.g., drawings, etc.) that are not in themselves models; *velocity*: rapid information flow (e.g., operational digital twins sending information back to the digital system model); and *veracity*: uncertainty inherent in model data (e.g., artificial data from simulations, incomplete data, subjectivity in models).

Future research is needed to further elicit ideas from the systems community on program level indicators and enterprise-level indicators. Desirable research is to conduct industry case studies to learn from digital engineering early adopters concerning what metrics and leading indicators they have implemented, as well as novel approaches that have been developed. This includes extraction and composition of leading indicators, the implementation of measurement dashboards, and the specific practices used in making decisions with measurement information.

References

Bone, M. A., Blackburn, M. R., Rhodes, D. H., Cohen, D. N., & Guerrero, J. A. (2019). Transforming systems engineering through digital engineering. *The Journal of Defense Modeling and Simulation*, *16*(4), 339–355.

Dam, S. (2019). *Real MBSE: Model-based systems engineering (MBSE) using LML and Innoslate*.

Dam, S. (2020). *How to perform a model-based review* (White paper). https://www.innoslate.com/resource/model-based-review-whitepaper/

DoD. (2018, June). *Digital engineering strategy.* Office of the Deputy Assistant Secretary of Defense for Systems Engineering.

Dwyer, M. (2020). *Understanding acquisition speed for the Defense Department's costliest and most complex programs* (CSIS-AM-20-159). CSIS. https://dair.nps.edu/handle/123456789/4288

Elm, J., & Goldenson, D. (2013). Quantifying the effectiveness of systems engineering. *IEEE Systems Conference (SysCon)*, 6–13.

Elm, J., Goldenson, D., El Eman, K., Donatelli, N., & Neisa, A. (2008). *A survey of systems engineering effectiveness.* Carnegie Mellon University and National Defense Industrial Association.

Gerst, K. J., & Rhodes, D. H. (2010). Strengthening systems engineering leading indicators for human systems integration considerations: Insight from the Practitioner Community. *Proceedings of Conference on Systems Engineering Research.*

Gilbert, D., Yearworth, M., & Oliver, L. (2014). Systems approach to the development and application of technical metrics to systems engineering projects. *Procedia Computer Science, 28*, 71–80.

LML Steering Committee. (2015). *Lifecycle modeling language (LML) specification*. https://lifecyclemodeling.org/wp-content/uploads/2021/01/LML_Specification_1_1.pdf

McDermott, T. A., Hutchinson, N., Clifford, M., Van Aken, E., Slado, A., & Henderson, K. (2020). *Benchmarking the benefits and current maturity of model-based systems engineering across the enterprise* (Technical report SERC-2020-SR-001). Systems Engineering Research Center (SERC).

Micoun, P., Paper, P., Fabre, L., Razafimahefa, T., & Becquet, R. (2018*). Property model methodology: A landing gear operational use case*. INCOSE International Symposium (hal-01829910).

Montgomery, P., & Carlson, R. (2010). *Systems engineering applied leading indicators: Enabling assessment of acquisition technical performance.* Naval Postgraduate School, Graduate School of Business and Public Policy.

Orlowski, C., Blessner, P., Blackburn, T., & Olson, B. A. (2015). A framework for implementing systems engineering leading indicators for technical reviews and audits. *Procedia Computer Science: Complex Adaptive Systems Conference*, *61*, 293–300.

Orlowski, C. T. (2017). *A framework for implementing systems engineering measures at technical reviews and audits*. https://pqdtopen.proquest.com/doc/1878241332.html?FMT=ABS

Nadj, M., Maedche, A., & Schieder, C. (2020). The effect of interactive analytical dashboard features on situation awareness and task performance. *Decision Support Systems*, *135*, 113322. https://doi.org/10.1016/j.dss.2020.113322

Parrot, E., & Weiland, K. (2017). Using model-based systems engineering to provide artifacts for NASA project life-cycle and technical reviews. *AIAA Space and Astronautics Forum and Exposition*, p. 5299.

PSM. (2020). *Practical software & systems measurement*. http://www.psmsc.com

Rhodes, D. H. (2020). Investigation of leading indicators for systems engineering effectiveness in model-centric programs. *17th Annual Acquistion Research Symposium.* https://dair.nps.edu/bitstream/123456789/4209/1/SYM-AM-20-060.pdf

Rhodes, D. H., Valerdi, R., Gerst, K. J., & Ross, A. M. (2009). Extending leading indicators for human systems integration effectiveness. *7th Conference on Systems Engineering Research.*

Rhodes, D. H., Valerdi, R., & Roedler, G. J. (2009). Systems engineering leading indicators for assessing program and technical effectiveness. *Systems Engineering*, *12*(1), 21–35.

Roedler, G. J., & Rhodes, D. H. (2007). *Systems engineering leading indicators guide: Version 1.0.* Massachusetts Institute of Technology.

Roedler, G. J., Rhodes, D. H., Schimmoler, H., & Jones, C. (2010). *Systems engineering leading indicators guide: Version 2.* Massachusetts Institute of Technology.

Selby, R. (2009). Analytics-driven dashboards enable leading indicators for requirements and designs of large-scale systems. *IEEE Software*, *26,* 41–49. https://doi.org/10.1109/MS.2009.4

Shirley, E. (2016). A*pplication of system engineering leading indicators to scrum agile projects* (Master's thesis). Air Force Institute of Technology.

Sindiy, O., Litomisky, K., Davidoff, S., & Dekens, F. (2013). Introduction to information visualization (infovis) techniques for model-based systems engineering. *Procedia Computer Science*, *16*, 49–58. https://doi.org/10.1016/j.procs.2013.01.006

Systems Engineering Research Center. (2020). http://sercuarc.org

Thiruvathukal, G. K., Hayward, N. J., & Läufer, K. (2018). *Metrics dashboard: A hosted platform for software quality metrics*. https://arxiv.org/pdf/1804.02053.pdf

Thomas, J. (2007). *Visual analytics: An agenda in response to DHS mission needs*. National Visualization and Analytics Center.

Vitiello, P., & Kalawsky, R. S. (2012). Visual analytics: A sensemaking framework for systems thinking in systems engineering. *Proceedings IEEE International Systems Conference*.

Zheng, L., Baron, C., Esteban, P., Xue, R., Zhang, Q., & Yang, S. (2017). Considering the systems engineering leading indicators to improve project performance measurement. *IFAC-PapersOnLine*, *50*(1), 13970–13975.

Zheng, L., Baron, C., Esteban, P., Xue, R., Zhang, Q., & Yang, S. (2019). Using leading indicators to improve project performance measurement. *Journal of Systems Science and Systems Engineering*, *28*(5), 529–554.

Defining a Model-Based Systems Engineering Approach for Technical Reviews

Warren Vaneman—has 33 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 Navy Reserve Captain (Ret.). 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. Vaneman also has a Joint Professional Military Education Phase 1 Certificate from the Naval War College. He is a Certified Systems Engineering Professional (CSEP) from The International Council on System Engineering (INCOSE). [wvaneman@nps.edu]

Ron Carlson—served 26 years in naval aviation as a pilot, 7 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 10 years ago. He has a master's of philosophy in systems engineering 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 a bachelor's degree in nuclear engineering from the University of Michigan. [rrcarlso@nps.edu]

Abstract

Program technical reviews are discrete points in time, within a system's life cycle, where the system is evaluated against a set of specific accomplishments, known as "entrance criteria." These criteria are used to track the technical progress, schedule, and program risks. Technical reviews serve as gates used to demonstrate that the program is on track, and should be allowed to proceed. Current technical reviews are based around evaluations of static, contractually obligated documents that are used to demonstrate successful completion of the entrance criteria. These documents represent "snapshots" of the systems and do not represent a total view of the system. As a result, systems are often viewed by the entrance criteria individually, which fail to account for the system from a holistic perspective. Department of Defense (DoD) organizations are migrating to Model-Based Systems Engineering (MBSE) environments, with a vision of modernizing, developing, delivering, operating, and sustaining systems. Model-based reviews allow for complexity to be managed more efficiently because data, not "systems engineering products," is the commodity used to evaluate the entrance criteria. The data-driven MBSE technical reviews will provide greater insights and details across a program's life cycle. This paper discusses the applicability of current technical reviews criteria to MBSE.

Key words: Technical Review, Model-Based Systems Engineering

Background

Model-based processes are one of the most widely-discussed issues within the Department of Defense (DoD) today. For example, Model-Based Systems Engineering (MBSE) is a quarterly discussion at the Navy's Systems Engineering Stakeholders Group (SESG) and has been a tenant of the National Defense Industrial Association Systems Engineering and Mission Conference for the past several years. The DoD Digital Engineering Strategy (2018) provides a vision on how the DoD will modernize, develop, deliver, operate, and sustain systems. This strategy is important because advances in technology have led to larger and more complex systems. This implies a need for a clear, concise way to express the system design (clear, logically consistent semantics), and a need to represent systems differently to account for emergent behavior within the system due to the increased complexity.

When developed properly, models can provide a precise virtual representation of the functional, physical, parametric, and program entities of the systems. Increased emphasis is on the model itself, specifically the objects and relationships it contains, rather than the diagram, to encourage better model development, usage, and decision-making.

This paper evaluates the suitability of current systems engineering models for MBSE technical reviews. The section “Systems Engineering Technical Reviews” provides an overview of these reviews and serves as a point of departure of our evaluation. The section “SETRS in an MBSE Environment” discusses Systems Engineering Technical Reviews (SETR) in an MBSE environment. The section “Applicability of Current Technical Review Criteria to MBSE Technical Reviews” discusses applicability of current technical review criteria to MBSE technical reviews.

Systems Engineering Technical Reviews

The System Acquisition Life Cycle Model identifies five primary phases that take the system from concept development and material solution analysis through operations and support (AcqNotes, 2019). Systems Engineering Technical Reviews (SETR) are discrete points in time, within a system’s life cycle, during which the program is assessed against a set of program specific accomplishments (entrance criteria). The SETRs serve as gates that when evaluated successfully, demonstrate that the program is on track to achieve its final program goals, and should be allowed to proceed to the next acquisition phase. Figure 1 shows the technical reviews superimposed on the Systems Acquisition Life Cycle Model (derived from DAU, 2018). The acquisition phases, with their associated technical reviews, are described in Table 1 (derived from Manning, 2019).

The technical reviews that were considered most likely to benefit from MBSE are conducted during the Materiel Solution Analysis (MSA), Technology Maturation and Risk Reduction (TMRR), and the Engineering and Manufacturing Development (EMD) Phases—the phases that lead to Milestone C. Beaufait (2018) studied the applicability of MBSE to programs post-Milestone C, and concluded that MBSE can benefit programs post-Milestone C; however, introducing MBSE that far into the life cycle of the program will face challenges related to cost, schedule, and a lack of understanding of MBSE. At this stage of the program, the implementation of MBSE has an additional cost that is likely not planned in the budget, and skeptical program managers are reluctant to make that investment in exchange for the promised benefits of MBSE (Beaufait, 2018).

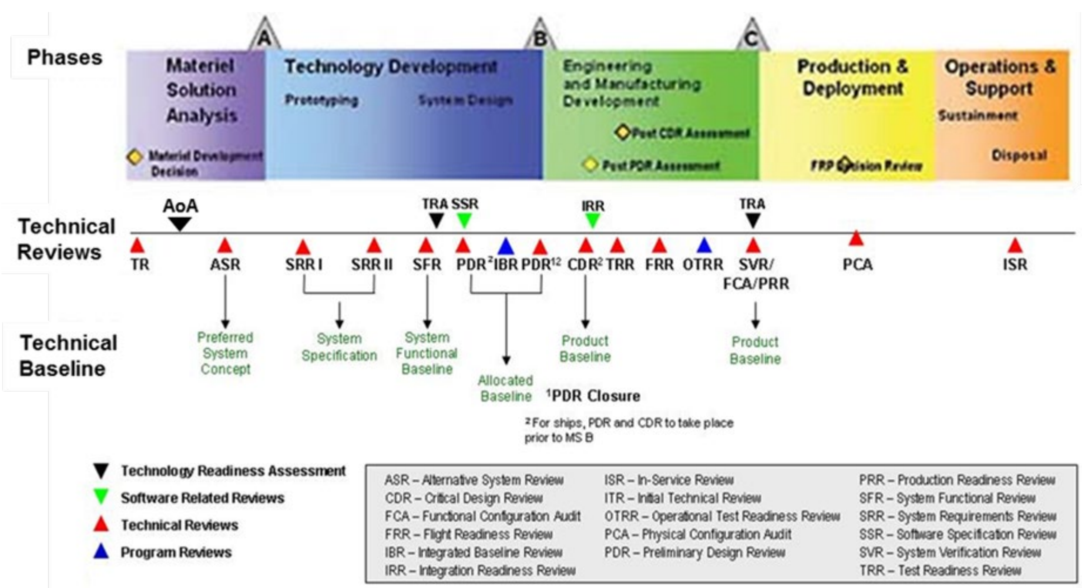


Figure 1. System Acquisition Life Cycle Model (DAU, 2018)

The technical reviews that this research considered for applicability to an MBSE environment are (AcqNotes, 2019):

- Initial Technical Review (ITR)—A multi-disciplined review to support a program's initial Program Objective Memorandum (POM) within the Materiel Solutions Analysis phase (MSA).
- Alternative System Review (ASR)—A review that assesses the preliminary materiel solutions that have been developed during MSA.
- System Requirements Review (SRR)—A review to ensure that system requirements have been completely and properly identified, and that a mutual understanding between the government and contractor exists during the Technology Maturation and Risk Reduction (TMRR) phase.
- System Functional Review (SFR)—A review to ensure that the system's functional baseline is established and can satisfy the requirements of the Initial Capabilities Document (ICD) or draft Capability Development Document (CDD) within the currently allocated budget and schedule, during TMRR.
- Preliminary Design Review (PDR)—A review that establishes the allocated baseline of a system to ensure a system is operationally effective. A PDR is conducted before the start of detailed design work and is the first opportunity for the government to closely observe the contractor's hardware and software design. This review is conducted during TMRR.

Current SETRs are based around lengthy reviews of static, contractually obligated "artifacts" that are used to demonstrate successful completion of the entrance criteria. Participants typically "freeze" these "artifacts" many days prior to the SETR in order to provide baselines from which to synchronize various products used during the review. This baselining and eventual loss of concordance[1] between "artifacts" are the primary drawbacks when conducting reviews using "artifact-based" methods.

[1] Concordance is the ability to represent a single entity, such that data in one view, or level of abstraction, matches the data in another view, or level of abstraction, when talking about the exact same thing. This allows for complexity to be managed more efficiently because each entity is ideally represented in the model only once, essentially creating a virtual representation of the system in the model. Systems engineering views are generated from the data (Vaneman, 2018).

Table 1. Summary of the DoD System Acquisition Life Cycle Phases (Vaneman & Carlson, 2019)

Life Cycle Phase	Description of the Life Cycle	Technical Reviews within Life Cycle
Materiel Solution Analysis (MSA)	MSA assesses potential solutions for a needed capability in an Initial Capabilities Document (ICD). The MSA phase is critical to program success and achieving materiel readiness because it's the first opportunity to influence systems supportability and affordability by balancing technology opportunities with operational and sustainment requirements.	• Initial Technical Review (ITR) • Analysis of Alternatives (AoA) • Alternative System Review (ASR) ♦ **Milestone A**
Technology Maturation and Risk Reduction (TMRR)	The purpose of TMRR is to reduce technology risk, engineering integration, and life cycle cost risk and to determine the appropriate set of technologies to be integrated into a full system. The TMRR phase conducts competitive prototyping of system elements, refines requirements, and develops the functional and allocated baselines of the end-item system configuration.	• System Requirement Review (SRR) • System Functional Review (SFR) • Preliminary Design Review (PDR) ♦ **Milestone B**
Engineering and Manufacturing Development (EMD)	A system is developed and designed during EMD before going into production. The phase starts after a successful Milestone B, the formal start of any program. The goal of this phase is to complete the development of a system or increment of capability, complete full system integration, develop affordable and executable manufacturing processes, complete system fabrication, and test and evaluate the system before proceeding into the production and deployment (PD) Phase.	• Critical Design Review (CDR) • Test Readiness Review (TRR) ♦ **Milestone C**
Production and Development (PD)	A system that satisfies an operational capability is produced and deployed to an end user during PD. The phase has two major efforts: (1) Low-Rate Initial Production (LRIP) and (2) Full-Rate Production and Deployment (FRP&D). The phase begins after a successful Milestone C review.	• Full Rate Production (FRP) • Initial Operational Capability (IOC) ♦ **Full Operational Capability (FOC)**
Operation and Support (OS)	During OS, a system that satisfies an operational capability is produced and deployed to an end user. The phase has two major efforts: (1) Low-Rate Initial Production (LRIP) and (2) Full-Rate Production and Deployment (FRP&D). The phase begins after a successful Milestone C review	• Sustainment ♦ **Disposal**

SETRs in an MBSE Environment

Current SETR reviews require various DoD Architecture Framework (DoDAF) views, and other systems engineering artifacts, to serve as evidence for various aspects of the system's progress and status. These views are often "document-based," and thus are developed statically without an underlying model structure. In an MBSE environment, the system is represented virtually; therefore, the data and relationships, not the views, are the "atomic" level of detail.

In an MBSE environment, the model is a virtual representation of the system and becomes the focus of a SETR. Using the model as the source for decision-making throughout the system acquisition life cycle is a significant departure since programs often generate unique artifacts for the sole purpose of the reviews. Each system element should be represented only once in the model just as it is in the real-world system. The data that comprises the model is iteratively developed and maintained throughout the system life cycle.

A significant difference between traditional document-based technical reviews and model-based technical reviews is model structure. Structure defines the relationships between the system entities, establishes concordance within the model, and allows for the emergence of system behaviors and performance characterizations. Structure provides decision-makers with insights that have been heretofore unavailable. This includes emerging system behavior, and the assurance that a common system baseline is used to report on various aspects of the systems. A discussion of model-development is beyond the scope of this paper, but a lengthy description can be found in Vaneman et al. (2020).

While an MBSE environment contains a virtual representation of the system, current SETR criteria rely on model-based data, which is depicted by views within a presentation framework, similar to a document-based review. While a virtual representation of the systems will exist, the acquisition community currently lacks the experience of viewing the data in this format. Thus, the SETR criteria still requires the information to be viewed in the standard document-based systems engineering format. This is acceptable, because the virtual system can represent data in any desired presentation framework (e.g., DoDAF), but there is additional information available to the reviewer in the model itself as described in this paper.

Table 2 shows the applicability of MBSE views to the system acquisition life cycle. The relationships in the matrix were made by correlating the generic criteria for each review, or content of the major documents, to the data in each system engineering view. The existing review criteria is designed to be addressed by document-based processes. Given that the MBSE environment creates a virtual system, the SETR criteria need to be revised to account not only for the current, but also the additional insights that can be gleaned through a model-based approach.

Table 2. Applicability of Systems Engineering Views within the Systems Acquisition Life Cycle (Vaneman & Carlson, 2019)

	Materiel Solution Analysis			Technology Develoment				Engineering and Manufacturing Development			Documents			
Systems Engineering Views	**Analysis of Alternatives (AoA)**	**Alternative Systems Review (ASR)**	**Milestone A**	**System Requirements Review (SRR)**	**System Functional Baseline (SFB)**	**Preliminary Design Review (PDR)**	**Milestone B**	**Critical Design Review (CDR)**	**Test Readiness Review**	**Milestone C**	**Initial Capabilities Document**	**Capability Development Document (CDD)**	**System Requirements Specifications**	**Test Report**
CV-2	x	x		x							x	x		
CV-3	x	x		x							x	x		
CV-6		x		x							x	x		
OV-1	x	x		x							x	x		
OV-2	x	x		x								x	x	
OV-4		x		x							x	x		
OV-5b	x	x		x	x	x						x	x	
OV-5b/6c	x	x		x	x	x					x	x	x	
OV-6c	x	x		x	x	x					x	x	x	
PV-2					x	x		x						
SV-1	x	x		x	x	x		x	x			x	x	x
SV-2					x	x		x	x				x	x
SV-5b				x	x	x		x	x			x	x	x
SV-7	x	x		x	x	x		x	x			x		x
Cost Esttimate	x			x				x						
Risk Matrix	x	x			x	x		x						
Similation Results	x			x		x		x	x				x	x
Test and Verification Matrix						x		x	x					x
Test Results								x	x					x
Work Breakdown Structure					x	x		x						

As an example of how data created in an MBSE environment can yield new insights into the status of the system, consider the Alternative Systems Review (ASR). The ASR assesses the preliminary technology solutions that have been developed during the Materiel Solution Analysis (MSA) Phase. The SETR ensures that one or more proposed materiel solution(s) have the best potential to be cost effective, affordable, operationally effective and suitable, and can be developed to provide a timely solution at an acceptable level of risk to satisfy the capabilities listed in an Initial Capabilities Document (ICD; Manning, 2019).

The systems engineering process typically has to progress to the point where the following information is available for the ASR (TTCP, 2014):

- Description of how the users will conduct operations, and how they expect to use the new system in this context of major mission areas and scenarios
- Statement of need, and capabilities, in terms oriented to the system users, the stakeholders, and independent of specific technology solutions
- The required system characteristics and context of use of services and operational concepts are specified
- Major stakeholder capabilities are identified and documented, but detailed system requirements analysis has yet to be completed
- The constraints on a system solution are defined
- Results of an AoA with a recommended preferred solution
- Initial plans for systems engineering (e.g. Overview and Summary information (AV-1), Systems Engineering Plan (SEP), Systems Engineering Management Plan [SEMP]) providing the notion of "how" this system can be realized, including the level of process and process maturity needed to generate a system of the required complexity
- Initial definition of the environment and the characteristics of the threat
- Initial test and evaluation strategy, including test cases derived from user operational vignettes, concept of operations, and capability description
- An understanding of where the greatest risks and challenges may reside

An analysis of the ASR generic entrance criteria[2] (DAU, 2019) against traditional and MBSE reviews is shown in Table 3 (Vaneman & Carlson, 2019). First the criteria are reviewed in the context of traditional reviews. Many of the criteria were assessed to be partially satisfied. These results do not suggest that ASRs have not been performed properly in the past. Rather, given the absence of concordance in document-based reviews, the criteria requiring different types of data using different artifacts is extremely difficult to achieve efficiently and effectively. All of the criteria were assessed to be satisfied in an MBSE environment because of the concordance. The MBSE views needed to address the criteria are also shown in Table 3.

[2] Entrance criteria are used to track the technical progress, schedule, and program risks.

Table 3. ASR Criteria and Related Views (Vaneman & Carlson, 2019)

Criteria	Satisfied by Traditional Review?	Satisfied by MBSE?	Views
Is the initial CONOPS updated to reflect current user position about capability gap(s), supported missions, interfacing/enabling systems in the operational architecture?	Partial	Yes	CV-2, CV-6, OV-1, OV-6c, OV-5b/6c
Are the required related solutions and supporting references (ICD and CDDs) identified?	Partial	Yes	CV-2, CV-3, CV-6, OV-4, OV-5b, OV-5b/6c
Are the thresholds and objectives initially stated as broad measures of effectiveness and suitability (e.g., KPPs)?	Yes	Yes	CV-2, OV-5b, OV-5b/6c, SV-7
Is there a clear understanding of the system requirements consistent with the ICD?	Yes	Yes	CV-2, CV-3, CV-6, OV-4
Are high-level description of the preferred materiel solution(s) available and sufficiently detailed and understood to enable further technical analysis in preparation for Milestone A?	Partial	Yes	OV-2, OV-5b, SV-1
Are interfaces and external dependencies adequately defined for this stage in life cycle?	Partial	Yes	OV-2, SV-1
Are system requirements sufficiently understood to enable functional definition?	Partial	Yes	OV-5b, OV-5b/6c
Is a comprehensive rationale available for the preferred materiel solution(s), based on the AoA?	Partial	Yes	CV-2, CV-3, CV-6, OV-2, OV-4, OV-5b, OV-5b/6c.
Can the proposed material solution(s) satisfy the user needs?	Partial	Yes	CV-2, CV-3, CV-6, OV-2, OV-5b, OV-5b/6c.
Have cost estimates been developed and were the cost comparisons across alternatives balanced and validated?	Partial	Yes	OV-2, OV-5b, SV-1
Have key assumptions and constraints associated with preferred materiel solution(s) been identified?	Partial	Yes	OV-2, OV-5b, SV-1

Applicability of Current Technical Review Criteria to MBSE Technical Reviews

An initial assumption for this research was that the approximately 85 systems engineering model visualizations that currently exist could be used to address all SETR questions for review through the TMRR phase. However, this research does recognize that some questions may have to be adjusted from binary (yes or no) questions (e.g., “Does the project have a Risk Management Guide?”) to questions that provide more concrete details to allow for better program and system analysis.

Our research found that MBSE, as it currently exists, can be used to satisfy the criteria found throughout the MSA phase and during most of the TMRR phase. However, we found that current MBSE environments do not adequately address the criteria for a PDR. Review criteria for PDRs were evaluated from the Defense Acquisition University (DAU), the Navy's Strategic Systems Program (SSP), and the Naval Air Systems Command (NAVAIR). The criteria from NAVAIR was eventually selected to be reviewed because it was found to be the most comprehensive.

During this research, 846 PDR questions were evaluated for applicability to be addressed by current MBSE. Of these 846 questions, only 80 questions could be addressed directly by current MBSE models. To make the problem manageable, the 864 questions were categorized into 56 PDR criteria categories. Of these 56 categories, only 11 categories were adequately satisfied by MBSE, 13 categories were partially satisfied by MBSE, and 32 categories were not adequately satisfied by MBSE. Tables 4a and 4b show the 56 PDR criteria categories and the assessed MBSE ability to satisfy those criteria.

These disappointing results do not mean that employing MBSE methods to PDRs should be abandoned. To achieve better PDR results, it is clear that new visualization techniques must be developed to fully realize the benefits of an MBSE environment. Developing new visualizations also makes sense because many of the approximately 85 current visualizations have been used by the systems engineering community for decades. While many of these models have been the cornerstone of technical reviews, a true MBSE environment, where the model is a virtual representation of the system, will glean additional insights that have heretofore been unrealized.

In addition to the PDR evaluation categories not being represented in MBSE visualizations, there is another issue. Over time, the scope of the PDR questions increased to the point where many senior leaders agree that questions were added without an appropriate audit of suitability. For PDRs to be more effective in their current form, and in an MBSE environment, a detailed evaluation of the review criteria needs to be explored, and questions need to be modified, to truly use MBSE to assess the program and system at PDR.

Table 4a. PDR Criteria Categories and the MBSE Ability to Satisfy Them

PDR Criteria Category	MBSE Ability to Satisfy Criteria
Schedule Planning	↑
Program Critical Path	→
Cost/Schedule/Performance/Key Performance Parameters (KPP)	↑
Latest Cost Estimate	→
Production Costs Estimates	↓
Operating and Support (O&S) Costs Estimate	→
Earned Value Management (EVM)	→
Work Breakdown Structure (WBS) review	↑
Software Metrics	→
Program Management	↑
Configuration Management (CM)	↑
Systems Engineering Processes	↑
Acquisition Logistics Support Management and Staffing	↓
Automated Information Technology (AIT)	↓
Risk Management (RM) Processes	↑
Logistics Budgeting and Funding	↓
Test Processes (TEMP, T&E Strategy, etc.)	→
Production Processes (ISO 9000, etc.)	↓
Software	→
Producibility	↓
Human System Safety	↓
Aeromechanics	↓
Structures	↑
Materials	↓
Mass Properties	↓
Human Systems Integration Engineering	↓
Environmental Regulations	↓
Safety and Health	↓
System Safety	↓
Hazardous Material Management	↓
Pollution Prevention Program	↓
Maintenance Planning	→
Key	
Adequately satisfies criteria in category	↑
Partially satisfies criteria in category	→
Does not satisfy criteria in category	↓

Table 4b. PDR Criteria Categories and the MBSE Ability to Satisfy Them

PDR Criteria Category	MBSE Ability to Satisfy Criteria
Testability and Diagnostics	→
Manpower, Personnel and Training (MP&T)	↓
Training Outline and Curricula Design	↓
Training Material	↓
Training Devices / Simulators	↓
Supply Support	↓
Organic Support	↓
Supply Chain Management / PBL Management	↓
Warranty Management	↓
Support Equipment	↓
Technical Data	↑
Product/Technical Data Package and Publications	↓
Computer Resources	↓
Facilities	↓
Packaging, Handling, Storage, and Transportation	↓
Design Interface	↑
Manufacturing Planning	↓
Parts and Materials Selection	↓
Commodity Management	↓
Root Cause Corrective Action	→
Obsolescence	↓
Platform Diagnostics Integration	→
Life Cycle Logistics	→
Performance Requirements	↑
Key	
Adequately satisfies criteria in category	↑
Partially satisfies criteria in category	→
Does not satisfy criteria in category	↓

Conclusions

Formalized planning for modeling and decision-making across the life cycle must include a new approach for SETRs. This not only includes the content of the reviews, but how the models will be assessed against the criteria (Dam, 2018). We found that current processes for assessing documents are not adequate in an MBSE environment. For example, many questions are binary in nature, and do not provide any insight into the "health" of a program. Consider, for example, a question that takes this form: "Does the program have a risk management plan?" The answer is "yes" or "no" and does not provide any insights into the quality of the plan content or the program "health."

The DoD Digital Engineering Strategy (2018) states that there is a strong need to ensure that decision-makers understand the different model types and what information can be gleaned from them. The results of our analysis of how MBSE will satisfy a PDR were unexpected because we believed that current MBSE visualizations would address a wider range of the PDR content. While our research found only 9.5% of PDR questions to be adequately addressed by current MBSE methods, we do not recommend abandoning the use of MBSE for PDR assessments. Instead, it is clear from this research that new visualizations must be developed to adequately address the needs, and provide greater insight with faster comprehension for the details across the life cycle.

As DoD organizations migrate to an MBSE environment, efficiencies will be gained by transitioning from the traditional paper-based reviews to model-based reviews. Model-based reviews allow for complexity to be managed more efficiently because data, in lieu of "systems engineering products," is the commodity that will be used to evaluate the entrance criteria. The MBSE milestone reviews will provide greater insight with faster comprehension for the details across a program's life cycle. This will not only provide efficiencies for the review, but will improve the program's cost and schedule efficiency.

References

Beaufait, J. (2018). *Decision tools to implement MBSE Post Milestone C* [Unpublished thesis].

Dam, S. (2014). *DoD architecture framework 2.02*. SPEC Innovations.

Deputy Acquisition University. (2019). *Defense acquisition guidebook.* https://www.dau.mil/tools/dag

Deputy Secretary of Defense for Systems Engineering. (2018). *Department of Defense digital engineering strategy.*

Manning, B. (2019). *AcqNotes: Defense acquisition made easy.* http://acqnotes.com

The Technical Cooperation Program. (2014). *Recommended practices: System of systems considerations in the engineering of systems* (Technical report TR-JSA/TP4-1-2014).

Vaneman, W. K. (2018). *Evolving model-based systems engineering ontologies and structures*. In Proceedings of the 28th Annual INCOSE International Symposium, Washington, D.C: International Council on Systems Engineering.

Vaneman, W. K., Carlson, R. R., & Wolfgeher, C. (2019). *Defining a model-based systems engineering approach for milestone technical reviews.* Acquisition Research Program Sponsored Report Series. In *Proceedings of the 16th Annual Acquisition Research Symposium, Monterey, CA*.

Vaneman, W. K., Carlson, R. R., & Wolfgeher, C. (2020). *Defining a model-based systems engineering approach for milestone technical reviews* (NPS-SE20-016). Acquistion Research Program Sponsored Report Series.

PANEL 11. THE ART AND SCIENCE OF PROGRAM MANAGEMENT

Wednesday, May 12, 2021	
10:45 a.m. – 12:00 p.m.	**Chair: Rear Admiral Douglas Small, USN,** Commander, Naval Information Warfare Systems Command ***Behavioral Acquisition*** Robert Mortlock, Naval Postgraduate School Nicholas Dew, Naval Postgraduate School ***Extending Program Management Theory Through Knowledge Flow Turbulence Modeling Methods*** Raymond Jones, Naval Postgraduate School ***"Is the Department of Defense a High-Risk Anomaly?" Theory to Practice*** Douglas Brook, Duke University Danelle Gamble, Duke University ***Fire Scout: The Navy, Northrop Grumman, and Acquisition in Adversity: A Case Study*** John Kamp, Kaena Point Consulting LL Amirhossein Etemadi, The George Washington University

Rear Admiral Douglas Small, USN—is a native of Birchwood, Wisconsin. He is a 1988 graduate of Marquette University in Milwaukee, Wisconsin, where he received a Bachelor of Science in Physics. He also holds a doctorate degree in physics from the Naval Postgraduate School in Monterey, California.

Small's operational tours included mechanical division officer and main propulsion assistant on USS Camden (AOE 2), in which he deployed in support of Operation Desert Storm, as well as combat systems officer (plankowner) on USS Iwo Jima (LHD 7). During Operation Iraqi Freedom he served as the first technical director of Joint Crew Composite Squadron (JCCS) 1, assembled by the Navy to assist with defeating then-prevalent radio-controlled improvised explosive devices (IED).

An engineering duty officer since 1997, Small had a number of tours in acquisition, starting at the Naval Surface Warfare Center, Dahlgren Division and then the Missile Defense Agency. Moving to the Program Executive Office for Integrated Warfare Systems (PEO IWS), Small was the electronic warfare assistant and then the major program manager for above water sensors (PEO IWS 2), overseeing the development of improvements for the Navy in electronic warfare and radar. Following this tour he served as the executive assistant to the Assistant Secretary of the Navy for Research, Development and Acquisition.

Small's first flag assignment, in 2016, he was the program executive officer for Integrated Warfare Systems, where he led the team responsible for development, delivery, and sustainment of all surface navy combat and weapon systems. He assumed command of the Naval Information Warfare Systems Command in August 2020, leading a global workforce of 11,000 civilian and military professionals who design, develop, install, and support Navy's networking, communications, information, and cyber capabilities and systems.

Small's personal awards include the Distinguished Service Medal, Legion of Merit, and Bronze Star; unit awards include the Army Meritorious Unit Commendation and various campaign and service awards.

Behavioral Biases within Defense Acquisition

Dr. Robert Mortlock, COL, USA (Ret.), is a professor of the practice for defense acquisition and program management in the Graduate School of Defense Management at the Naval Postgraduate School in Monterey, CA. He holds a PhD in chemical engineering from the University of California, Berkeley, an MBA from Webster University, an MS in national resource strategy from ICAF, and a BS in chemical engineering from Lehigh University. [rfmortlo@nps.edu]

Dr. Nick Dew is a Professor of Strategic Management at the Naval Postgraduate School in Monterey, CA. His teaching and research focuses on entrepreneurship, strategy and innovation in defense and homeland security. He has a Ph.D. and MBA from the University of Virginia and experience working in the international energy industry. [ndew@nps.edu]

Abstract

This paper contributes to the process of building knowledge about what we term as *behavioral acquisition*, which explores defense acquisition from a behavioral standpoint, including the impact of psychology, organizational behavior, and politics. Behavioral acquisition studies the decisions acquisition professionals make in Department of Defense (DoD) acquisition programs. The paper focuses on one aspect of these decision processes in the defense acquisition environment: behavioral biases. In three defense acquisition programs studied, we find strong evidence that planning fallacy, difficulty in making trade-offs, over-optimism, and recency bias affected the management and decision-making within these programs. This research helps us better understand and predict how acquisition professionals and senior leaders think and make decisions about program strategy, managing resources, and leading people. A key element in this perspective is that important insights into these decisions derive from models in which agents are not fully rational. Behavioral acquisition is analogous to behavioral finance, which has successfully applied social science theories—especially from psychology—to improve the accuracy of predictions about the behavior of actors across the entire financial landscape.

Keywords: behavioral acquisition, systemic biases, program management, decision making, defense acquisitions, culture, leadership, hierarchies

Introduction

Program managers (PM) are at the center of the defense acquisition process, yet there are substantial gaps in our knowledge about how PMs actually make decisions about the programs they manage on behalf of the defense community. Given the size of the defense acquisition portfolio in the United States, better knowledge of the intricacies of decision-making by PMs might be highly valuable for informing improvements in defense acquisition processes in the future, including for PM training and education. This paper contributes to the process of building knowledge about what we term as *behavioral acquisition*, which explores defense acquisition from a behavioral standpoint, including the impact of psychology, organizational behavior, and organizational politics.

Our paper focuses on one particular aspect of these decision processes in the defense acquisition environment: behavioral biases. These biases can be categorized into cognitive and emotional biases, but their common root is in the ways in which human brains use their limited capacities to process information. The results are decision-making capabilities that are, at times, stunning in their elegance and effectiveness in real world environments (Gigerenzer et al.,1999; Klein, 2009) and, at other times, shockingly flawed and error prone (Kahneman et al., 1982). That both outcomes are possible is one of the geniuses of the human brain. We provide a detailed treatment of four well-known behavioral biases and their scope to occur in acquisition programs. We examine the scope issue by doing a deep-dive into three significant acquisition programs using a case study-based approach to determine whether there is prima facie evidence that behavioral biases play a role in decision-making in acquisition programs. The programs are

- Enhanced Combat Helmet (ECH) program: Joint rapid acquisition effort executed in response to urgent warfighter needs leveraging new technologies.
- Joint Common Missile (JCM) program: Joint major defense acquisition program executed as a development effort with a deliberate acquisition approach with approved requirements.
- Ground Combat Vehicle (GCV) program: Service-specific major defense acquisition program executed as a development effort with a deliberate acquisition approach with approved requirements.

A key observation we make is that there is a lack of research studying the effects of behavioral biases on decision-making in the defense acquisition environment. Kiesling and Chong (2020) studied decision biases within acquisition programs by tracing the presence of keywords indicating specific biases from Government Accountability Office's (GAO) summary documents of the acquisition programs. However, the research did not study primary source program data (Kiesling & Chong, 2020). Therefore, at this point, it is sensible to motivate research on this topic with two basic questions: (1) How do behavioral biases affect decision-making in acquisition programs? And (2) to what extent do behavioral biases affect acquisition outcomes? The three acquisition programs studied were all wide-open to bias creeping into them in the forms of the planning fallacy, difficulty in making trade-offs, over-optimism, and recency bias. What the empirical cases illustrate best is that acquisition programs are environments where there is abundant opportunity for behavioral biases to play a significant role in decision-making; they are a perfect setting where one would expect to see these biases occurring. Furthermore, in some instances, the data is more suggestive. Not only was the opportunity in place, but there is at least some evidence that these biases were playing a role in program decision-making in ways that probably affected program outcomes. The data here is circumstantial but, when pieced together, the fact-pattern is suggestive of this conclusion. We cannot say anything more definitive than this based on the case data we have available, but certainly the patterns we see are consistent with biases playing a role.

It is worth pointing out that in our study we focused on the setting in which real acquisition managers do their work rather than the acquisition manager role in the abstract. This means we look at the acquisition challenge of juggling performance, schedule, and budget from a pragmatic perspective rather than as some kind of abstract rational optimization game. Recent acquisition reform directives and statutes require data-driven analysis and decisions, which put an emphasis on rational optimization. But whatever technologies of analysis are the fad or fashion of the day (and in the defense acquisition profession there have been many, over the years) decision-making inevitably still consists of boundedly rationally individuals operating inside a collective entity (a program) trying to make better decisions that deliver improved organizational outcomes (Levinthal, 2018). Hence, despite calls for more rationality, the organizational and political dimensions of decision-making still matter very much, and we shall see that these dimensions interact with behavioral biases in particular ways.

Background and Literature Review

Defense Acquisition Overview

Defense acquisition professionals facilitate the development, testing, procurement, and fielding of capability to warfighters. The program manager (PM) is at the center of defense acquisition, whose purpose is to deliver warfighter capability. The PM is responsible for cost, schedule, and performance (commonly referred to as the "triple constraint") of assigned projects—usually combat systems in the Department of Defense (DoD). The PM has a hierarchal chain of command (or authority) through the DoD in the executive branch. PMs report directly to a program executive officer, who reports to the service acquisition executive (an assistant secretary for that service—either Army, Navy or Air Force), and who reports to the defense acquisition executive (the Under Secretary of Defense for Acquisition and Sustainment). Depending on the program's visibility, importance and/or funding levels, a program's milestone decision authority (MDA) is assigned to the appropriate level of the chain of command.

Programs within defense acquisition require resources (primarily funding) and contracts (for execution of work) with industry. Congress provides the resources for the defense programs through the annual enactment of the defense authorization and appropriations acts, which become law and statutory requirements. The PM, through warranted contracting officers, enters contracts with private companies within the defense industry. As a backdrop to this complex acquisition environment for PMs, three decision support templates exist to guide programs: one for the generation of requirements, a second for the management of program milestones and knowledge points known as the Adaptive Acquisition Framework (often referred to as little "a"), and a third for the allocation of resources. Each of these decision support systems is fundamentally driven by different and often contradictory factors. The requirement generation system is capability needs-driven based on an evolving threat—requiring a responsive acquisition system. The resource allocation system is calendar-driven, with Congress writing an appropriations bill and the president signing the bill every fiscal year—providing control of funding to Congress and transparency to the public and media for taxpayer money. The Adaptive Acquisition Framework is event-driven by milestones—based on commercial industry best practices of knowledge points and off–ramps supported by the design, development, and testing of the systems as technology matures and integration and manufacturing challenges occur. The combination of the PM triple constraint, chain of authority, acquisition environment, and decision support templates provides a framework to view U.S. defense acquisition, referred to as the Defense Acquisition Institution (or big "A") and depicted in Figure 1.

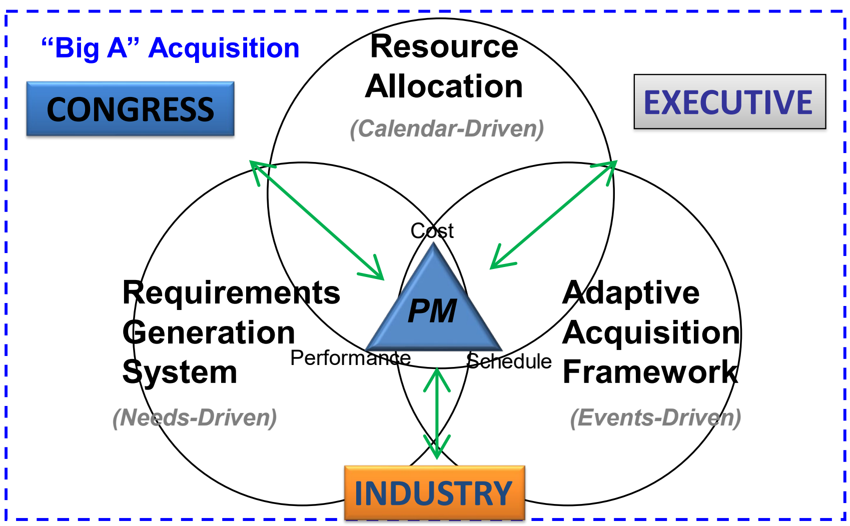


Figure 1. Defense Acquisition Institution or Big "A"

Due to the inherent complexity of the development, procurement, and fielding of sophisticated weapons systems that are required to operate reliably in challenging military environments, acquisition programs often fail to deliver required performance capabilities within cost and schedule constraints. Root causes of acquisition program failures (schedule slips, cost over-runs, or capability under-achievement) can be generally grouped into the following: ill-defined requirements, immature technology, integration challenges, poor cost estimating, unstable budgets, poor schedule planning, and schedule pressure from annual appropriation limitations. But an underappreciated reason for acquisition program failures and understudied part of big "A" is the "people part" of defense acquisition, which may have the largest effect on improving acquisition outcomes. Behavioral acquisition studies how acquisition professionals think, manage, and lead acquisition programs. Behavioral acquisition includes a study of organizations and hierarchies, and the intersection of individual behavior, leadership, culture and decision-making. In this study, we narrowly focus on how behavior biases affect acquisition decision-making at the institutional, organizational, and individuals levels. We recognize that decisions at the institutional level (DoD level) are often using a political conceptual model where decisions are a result of bargaining games. And decisions at the organizational level (Army level or PEO level) are based on the appropriateness of the actions fitting the organizations' cultural norms. Whereas at the individual level (program level), decisions use the rational conceptual model where decisions are based on logic and reasoning by assigning pros and cons (or risk and rewards) and deciding the best chance of success. We recognize that there may be important differences in how biases affect leader decision-making in organizations versus in institutions like the DoD. Figure 2 presents the overall model showing the connection of hierarchical, leadership, cultural, management, and behavior factors on decision-making and program outcomes. The model was adapted from the work of Shore (2000) who studied the effect of systemic biases within projects.

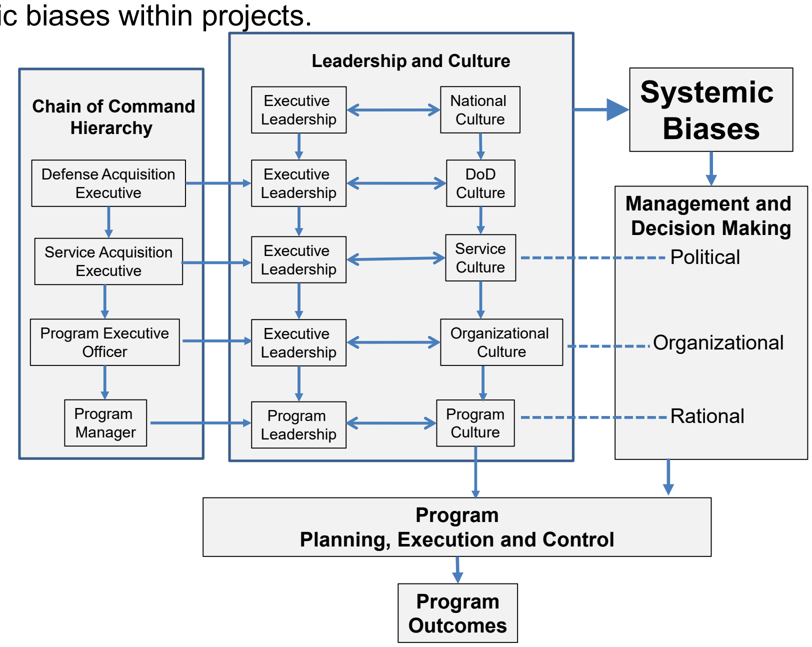


Figure 2. Connection of Hierarchical, Leadership, Cultural, Management, and Behavioral Factors on Decision-Making and Program Outcomes

Behavioral Biases Relevant to Defense Acquisition

Research centering on the acknowledgment and study of bounded rationality has long recognized that people process information in ways that may lead them to make biased judgments (Simon, 1955). Cognitive biases are a two-edged sword: On the one hand they have a positive function in helping people to make fast decisions using limited cognitive resources; on the other hand, cognitive biases also lead people to make errors in decision-making that deviate—often in important ways—from prescriptive (rational) decision-making. It is worth underlining that the baseline for making comparisons is important here: the identification of biases is based on comparisons to idealized "small world" rationality whereas in reality all decision-making is behavioral (Levinthal, 2011). Furthermore, research has also shown that at least some biases are driven by information presentation alone and may be mitigated by presenting information in ways that leverage the brain's information processing skills, rather than penalizing them (Gigerenzer, 1991). Nonetheless, a basic premise of research into biases is that as the volume and complexity of information increases, people are forced into using simplifying tactics that ration the limited cognitive resources they have available (Camerer, 2011). Hence. they adopt heuristics that ease the cognitive strain they experience. And because these heuristics involve rationing how information is processed (in a wide variety of ways) they develop systematic patterns of bias in decision-making, as compared to an ideally rational baseline (Kahneman et al., 1982).

Research on heuristics and biases is at this point a massive undertaking that has identified literally dozens of biases that people are predisposed towards (Kahneman, 2011). The objective of this paper is not to review them all but instead to focus on a handful of biases that illustrate the value of studying defense acquisition through an explicitly behavioral lens. Therefore, we focus on explaining four biases that have been widely studied in the literature under various guises: the *planning fallacy*, *over-optimism*, *recency bias*, and *difficulty making trade-offs*. These biases have different roots. Over-optimism and the recency bias are straightforwardly cognitive in nature, and their effects manifest in particular ways in acquisition programs. In contrast, the planning fallacy and difficulty making trade-offs are the result of how human cognitive factors interact with specific group and organizational processes. With all four biases, we are ultimately interested in how bounded rationality intersects with complex real world settings in ways that deviate substantially from what we might expect based on prescriptive rationality. We focus on these biases because the three defense acquisition programs we studied are wide open to being affected by all four of them.

Planning Fallacy

The planning fallacy addresses the unrealistic optimism about program management that numerous studies of program outcomes have documented across many decades in the defense, public and private sectors (Pressman & Wildavsky, 1984). Economic theory points to multi-level principal–agent issues as a key source of the gap between plans and outcomes. This leads economists to suggest that programs typically under-deliver because program managers have a vested interest in embellishing program projections to get programs approved through specific stage gates (Flyvbjerg et al., 2009). Within defense acquisitions, the Government Accountability Office (GAO) reported that PMs are incentivized to develop acquisition strategies focused on program approval at the milestone review but not acquisition strategies that could later be executed and deliver capabilities (GAO, 2015). There are also behavioral explanations for excessive optimism about programs, two of which we explain here: the planning fallacy, which occurs as the result of management practices, and dispositional optimism, which we discuss in the next section.

The main claim of the planning fallacy is that independently of other factors, planning processes themselves bias manager beliefs and lead them to make program forecasts that are too optimistic (Cassar, 2010). Of course, carefully planning projects is essential for good management and a legal requirement that establishes the acquisition program baseline of cost, schedule, and performance metrics. It is, therefore, problematic (and paradoxical) that planning induces a behavioral bias that actually undermines the intended outcome of planning. However, the explanation of this behavioral bias is quite simple. Planning processes lead managers to build an "inside view" of a project with detailed designs for the implementation of the project (Kahneman & Lovallo, 1993). These deep, well-explained designs enhance managers' perceptions of control over the project or program. Thus, they become more confident in the success of their plans (Heath & Gonzalez, 1995; Cassar, 2010). However, planning processes inevitably understate unpredictable events that will disrupt and delay the plan. Plans are always subject to the dreaded potential of "unknown unknowns" to intervene in what is otherwise carefully manicured expectations. Furthermore, the compound probability of even small individual disruptions can seriously undermine a program plan (Pressman & Wildavsky, 1984). Hence the fallacy of planning: it actually leads to control expectations and optimism that are unwarranted illusions when the context of programs is fully considered.

It is important to realize that the planning fallacy has deep roots in what are perceived to otherwise be good management practices, as well as in cultural characteristics that have their origins outside of project management. Control perceptions are central to these roots: managers are groomed to believe that their own actions substantially determine the results they get. Our accountability systems depend on these beliefs as well as their enforcement through bureaucratic oversight hierarchies. Program planning efforts are based on—and tend to reinforce—these idealized perceptions of control, resulting in managers typically perceiving they have more control over processes and outcomes than they have in reality. These control illusions may also lead them to believe they will avoid problems in a project or be able to easily overcome problems. What seem like good management practices may just compound the planning fallacy. For example, intuitively it seems like a good practice to focus intently on the specifics of a particular program, yet this tends to reinforce the "inside view" problem, which increases bias. Using incentives also seems like a good idea, yet research indicates that incentives also tend to encourage people to focus more intently on their plans, which increases bias (Buehler et al., 1997). In reality, the planning fallacy creates biased expectations that mask a control gap that will exact a price over the course of most programs.

Optimism Bias

In this section we discuss dispositional optimism, which is the tendency of individuals to see the world through "rose-colored glasses" or, more formally, their "tendency to expect positive outcomes even when such expectations are not rationally justified" (Hmieleski & Baron, 2009). The extant evidence suggests that in general individuals are over-optimistic, in the sense that their expectations for the future are more favorable than they will eventually experience (Cassar, 2010). Healthcare studies suggests that a degree of optimism bias is natural and healthy, since it tends to be correlated with psychological health and overall well-being. In other studies of the general population (Maltby et al., 2008), optimism has been found to be related to perceptions of luck (e.g., in global self-assessments luck beliefs are correlated strongly with optimism). Optimists *expect* good things to happen to them; they believe that chance events will break in their favor. Furthermore, when chance events happen, optimistic people tend to perceive them in a positive light. For example, optimists interpret an event such as narrowly avoiding an accident as lucky, whereas pessimists view the same event as unlucky (Hales & Johnson, 2018).

We don't know of any research that specifically examines the dispositional optimism of defense acquisition managers. However, we know that higher than average levels of optimism are present among individuals working in other domains that could broadly be construed as project or program management. For example, entrepreneurs are involved in start-up projects and studies have shown that on average entrepreneurs are distinctly over-optimistic, with some studies finding them to be off-the-scales on optimism (Hmieleski & Baron, 2009). Scholars have related these results to a willingness to initiate action, observing that individuals high in dispositional optimism are more willing to forge ahead in the face of daunting obstacles. This suggests a distinct selection bias in which optimism bias leads individuals to enter into activities for which they have little evidence to base beliefs about their eventual success (Meza & Southey, 1996). More difficult programs are more likely to attract more optimistic managers because these managers are more likely to believe everything will work out favorably for them, independently of having a plan to achieve success (Scheier et al., 2001). This suggests that some degree of optimism bias may be a prerequisite to becoming a program manager, with more difficult programs attracting the more optimism biased.

However, optimism is a double-edged sword. Research suggests that a curvilinear relationship between optimism and performance, with a distinct sweet spot. Individuals that are low in optimism believe that disaster awaits them in whatever they do. This makes them less likely to select into activities with uncertain outcomes in the first place and, if they are selected, it leads them to focus on negative information and have low motivation to complete tasks. On the flipside, over-optimistic individuals tend to focus only on positive information (they pick out good news stories), see positives in ambiguous situations (always look on the bright side), make suboptimal decisions such as setting unrealistic goals (the now infamous "stretch" goals), are less likely to learn from failure (i.e., update their beliefs), are more likely to persist with failing courses of action for longer periods (thus wasting resources) and be more at risk of escalation of commitment (another infamous problem in projects). Fundamentally these propensities tie optimism bias to less effective program performance. However, the extant evidence does also suggest that the same bias contributes positively to resiliency (Hmieleski & Baron, 2009).

Recency Bias

If you have ever had a project or presentation go badly right before your annual review is due, you already intuitively understand the recency effect: the widely recognized bias where recent data is given disproportionate emphasis in judgments (Beach & Connolly, 2005). Recency effects tend to occur when individuals process large amounts of information over time, which leads them to use heuristics to make judgments. This heuristic processing aids in sorting through the information but also introduces biases with regards to which information individuals pay the most attention. Because of this systematic variation in attention, with only a recent subset of all the available information getting the most attention, individuals make suboptimal decisions.

There are a number of explanations for why the recency effect occurs. It is more difficult to remember information that is older because of memory decay. In order to access or reconstruct information stored in memory, people rely of categorization processes. If information necessary for deciding has been stored in faulty or irrelevant categories, it may affect an individual's ability to recall it. Over long time spans, these issues become more pronounced compared to the accessibility of recent information, which suggests that this bias worsens if recent information is processed alongside much older information. It is worth noting that social processes that legitimate newer as compared to older data as more relevant and worthy of consideration may add to recency effects deriving from individual bias. In these cases, the bias may derive from social, political, and organizational factors, as well as personal ones.

While we do not know of any study that has examined the recency bias in defense acquisition, there are relevant studies that investigate this bias in the management literature. Recency bias has been most studied by accounting scholars interested in the effects that information presentation has on accounting judgments. Researchers widely use Hogarth and Einhorn's (1992) belief-adjustment model in these studies. The model proposes that people revise their beliefs using an anchoring and adjustment process. Current opinion is the anchor and new data potentially causes adjustments. In essence, the model suggests that when evaluating mixed information, individuals average the current piece of data and anchor. This results in more weight being placed on the latest information received (e.g., a recency effect). Accounting scholars find that many accounting judgments do seem to exhibit this effect, unless there is some mitigating factor. One review found that 21 out of 25 studies found some evidence of recency bias (Kahle, 2005). Studies have searched for factors that mitigate recency bias.

Some research has suggested that experience may mitigate recency bias but results of studies are inconclusive on this factor. Other studies (e.g., Kennedy, 1993) have found some support for accountability as a mitigating factor (e.g., recency effects are reduced when subjects are required to justify your decisions to others by explaining your reasoning in writing). These mitigating factors should be considered in understanding the complete effects of recency bias in the defense acquisition environment.

Trade-Offs Bias

Central to program management are trade-offs between program cost, schedule, and performance. Normative decision theory has proposed various methodologies for making such trade-offs by confronting them systematically and holistically, typically through some version of cost–benefit analysis. But as we already observed, the reality is that all decision-making is behavioral (Levinthal, 2011), and therefore models premised on idealized rationality quickly bump-up against the realities of bounded cognition in organizational settings, which is the setting in which defense acquisition processes actually have to work. Under these circumstances, it is germane to ask how acquisition managers actually make decisions. A basic understanding among behavioral researchers is that individuals avoid making trade-offs wherever possible (Slovic, 1975) because conflicting options are difficult to evaluate (Irwin & Davis, 1995). The lack of internal capability for judging trade-offs in the ways prescribed (e.g., by cost-benefit analysis) is a key reason that these analyses are typically externalized in the form of spreadsheets, diagrams, and tables (Clark, 2008). These "decision tools" often involve formulating trade-off criteria in numerical values, in order to make options easy to compare. This assists in making trade-offs that are otherwise too perplexing for individuals to make based on their cognitive capacities. However, behavioral research also notes that the assumptions made in financial models are also subject to biases (Lautliev & Menter, 2014).

Behavioral research suggests that individuals make trade-offs according to a different model: reason-based choice (Shafir et al., 1993). In this approach, individuals make conflicting choices easier to evaluate by constructing reasons for choices. If necessary, they make a list of reasons for choices (a relatively simple mental task) rather than trying to trade off costs and benefits across options (a much more complex mental task, typically significantly beyond human short-term memory limitations). By constructing reasons, individuals turn difficult-to-reconcile characteristics of options into a problem that is more readily comprehensible. Instead of making trade-offs, they use the heuristic approach of framing and re-framing an issue until they find a dominant option, which avoids the cognitive effort that real trade-offs require. In some cases, dominance may be based on a prominent attribute of the options. Hsee (1996) discusses how individuals find it easier to evaluate a good reason for purchasing an asset than whether they paid the right price for it, which is much harder to discern. If a good reason is not currently available for a choice, individuals may simply delay choosing or they may add other options that help clarify a dominant option. The essence of reason-based choice is that the human mind naturally prefers to find a dominant reason for a choice rather than delving into the complexities of cost-benefit analysis.

Furthermore, it is likely that reason-based choice may be even more important when groups are making decisions than when individuals make them. This is important in the acquisition world, since programs inherently involve numerous stakeholders and significant organizational oversight that means many decisions have to be justified to groups. Therefore, while individuals choose using reasons because of cognitive limitations, groups may also prefer reasons because of social dynamics (Barber et al., 2003). Accountability and group conflict are two explanations why reason-based choice may be affected by social dynamics. Work on accountability suggests that reasons become more important when individuals have to justify their choices to groups whose respect and approval are important to maintain (Lerner & Tetlock, 1999). No one wants to look foolish in front of a group, and good reasons that are simple to explain are a way of making sure they don't. Other work has suggested that reasons based on prominent attributes help avoid group conflict because prominent attributes may be less controversial (Irwin & Davis, 1995). For example, safety is a prominent and (mostly) non-controversial attribute in defense acquisition programs; therefore justifying choices based on safety (where possible) is likely to cause less group conflict. It is worth highlighting that because reasons depend on social legitimacy (i.e., what is accepted and approved of by particular groups) navigating organizational landscapes effectively requires having a detailed understanding of the nuances about which reasons are broadly acceptable to the community (and avoiding those that are likely to cause controversy).

Acquisition Program Case Studies

ECH Program

Combat helmets have evolved over time to offer improved performance because of technology advances and manufacturing capability improvements (see Figure 3). Soldiers wore the M1 helmet, nicknamed the "steel pot" because it was made from pressed manganese steel, from the 1940s through the late 1970s. The M1 helmet was heavy and uncomfortable, and it provided little blunt trauma protection. The maturation of ballistic fabrics based on para-aramid polymer technology enabled the Army to replace the M1 with the Personnel Armor System for Ground Troops (PASGT) helmet in the mid-1980s. These helmets were in the 3–4 pound range (lighter than the M1) and provided increased ballistic protection. The shell of the helmet consisted of layers of ballistic aramid fabric, the most famous of which is DuPont's Kevlar®—resulting in the "Kevlar" or "K-pot" nicknames. The ballistic aramid technology allowed helmets to provide not only fragmentation protection from explosions but also small caliber handgun protection at a reasonable weight. The Modular Integrated Communication Helmet replaced the PASGT helmet on a limited basis. By the mid-2000s, the Advanced Combat Helmet (ACH) was the Army's primary helmet. The ACH provided Soldiers important performance improvements like increased ballistic protection, reduced weight, and better blunt impact protection.

In the late 1990s and early 2000s, the Army research centers teamed with commercial industry to mature the next generation of ballistics materials, resulting in the development of high molecular weight polyethylene (HMWPE) ballistics fibers that could be weaved into fabrics with application to combat helmets. The basis of future combat helmets, both the enhanced combat helmet (ECH) and its replacement, the Soldier Protection System future combat helmet, rested in HMWPE technology.

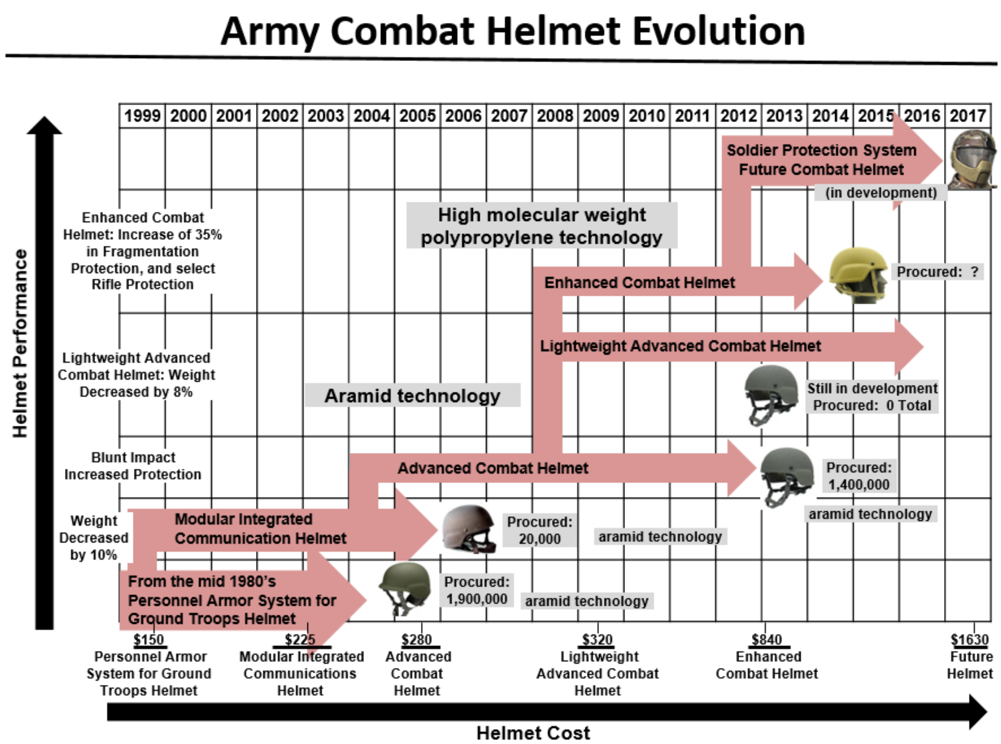


Figure 3. Evolution of Combat Helmets (Mortlock, 2018)

The ECH effort was driven by the urgent need for a new helmet to address protection for Soldiers and Marines against rifle threats in combat and reduce the combat weight. The guidance and priorities from the warfighting community and senior leaders were maximum protection and weight reduction. HMWPE technology allowed the Army and Marine Corps to consider increasing force protection by providing better ballistic protection or decreasing weight of the helmets. The basic options considered for the ECH requirements included: (1) maintain the protection levels of the current helmets with a reduced weight of up to 20%, or (2) increase the protection levels but maintain (or increase) the weight of the helmet. The helmet requirements had to balance acceptable minimum risk versus maximum safety for protective equipment, and weight reduction (combat load) versus protection (ballistic and blunt force)—not be an easy compromise for the program stakeholders. The ECH, however, would not be able to address both protection against the rifle threat and reduction in the helmet weight (*Trouble Making Trade-offs bias*).

To address the schedule aspect of the program, the Army and Marine Corps considered the options of pursuing a formal program of record through the deliberate acquisition process versus pursuing a rapid acquisition process supported by a directed or urgent requirement (*Planning Fallacy*). Establishing a formal ECH program involved a 4-year time period of contracting, development and testing. Year 1 allowed for the refinement, analysis, and approval of formal requirement documents and the development of testing protocols. Years 2 and 3 involved the development and testing of helmet prototypes resulting from competitively awarded contracts (cost-plus type contracts) to be awarded to industry companies. Year 4 allowed for the Army and Marine Corps to award procurement contracts to the successful companies for the manufacture and production of helmets. The alternative to a program of record was to use the rapid acquisition process. In rapid acquisition, the Services wrote a directed requirement (within a month) for the ECH and awarded competitive contracts (fixed-price contracts for certain quantities with production options) to industry within 6 months. Another 6 months would be required to test the helmets. So, ECHs could be on Soldiers' and Marines' heads in just over a year. With the rapid acquisition option, the new helmet's requirements would not be underpinned by analysis, and the test protocols had to rely on the protocols for current helmets because there would be no time to develop test protocols specifically for the ECH (*Planning Fallacy*). This was particularly important for the ECH, which would rely on novel HMWPE thermoplastic polymers that might perform much differently than the current para-aramid based helmets. For example, ECHs had the potential to lose their rigidity after being shot once and potentially offered less protection from multiple shots. Also, the ECH may deform excessively, leading to head trauma and skull fractures. There were legitimate testing and safety concerns that would have to be addressed during the development, testing, and manufacture of this new helmet.

The ECH program began in early 2009 (as shown in Figure 4) as rapid acquisition effort (*Planning Fallacy and Over-optimism*). The Army and Marine Corps approved urgent requirements based on the need for increased protection against enemy rifle threats and set broad requirements to include an increase in fragmentation and pistol protection, and rifle threat protection—all at the same weight of the current combat helmets. The acquisition strategy was a single step development in which competition was encouraged among industry manufacturers. The original request for proposal asked for each vendor to deliver test data validating their claim that their design met the new ECH requirements for rifle protection. Four vendors submitted proposals; however, only one vendor's design was acceptable. At the end of 2009, this vendor received a firm fixed price (FFP) contract to produce ECHs to undergo government validation testing with contract options for production deliveries after successful first article tests. In late 2010, after successful developmental testing, the Army and Marine Corps approved the program milestone to enter into low-rate initial production with the selected vendor. The decision permitted the production of a small number of helmets to undergo testing in order to validate that the contractor could successfully produce the helmets to performance requirements. The use of FFP type contracts for the development of the ECH was heavily influenced by Better Buying Power (BBP) 1.0, which encouraged a greater use in acquisition efforts (*Recency Bias*). Generally, however, for programs with only top-level warfighter operational requirements and with new technology, the use of cost plus–type contracts provides industry greater flexibility to innovate, allows the refinement of requirements based on knowledge learned in the development, and increases the chance of successfully producing a manufacturable, high quality product. In late 2011, the ECH passed the second round of first article testing after testing protocols were adjusted appropriately. To meet an aggressive production schedule for the Army and Marine Corps, the vendor submitted an engineering change proposal for a second and third production line. It took all of 2012 for the vendor to successfully pass the third round of first article testing.

The testing results demonstrated that the ECH met its requirements and offered Soldiers and Marines the potential for greater protection compared to current helmets. After passing testing and 4 years since program initiation, in the summer of 2013, the ECH successfully received a full rate production decision. The setting of requirements and testing protocols in the absence of quantitative analysis led to prolonged schedule, especially important with limited funding, intense scrutiny on program cost/schedule overruns, and pressures to field new capabilities to the warfighters quickly (*Difficulty Making Trade-offs and Planning Fallacy*).

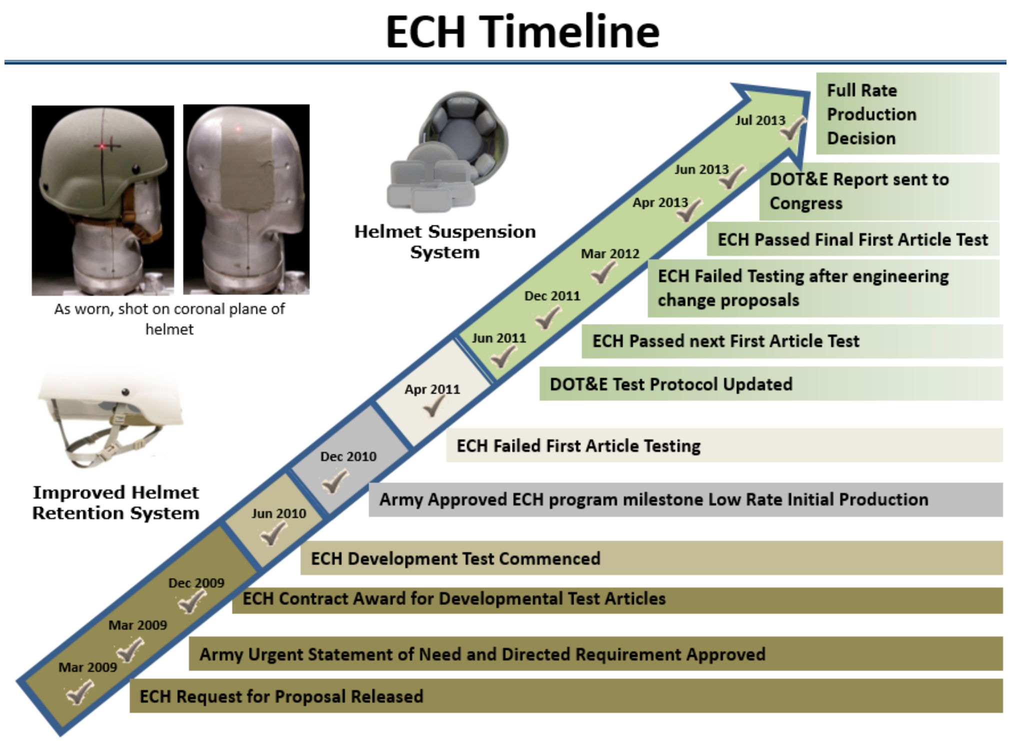


Figure 4. ECH Program Timeline (Mortlock, 2018)

The rapid acquisition program required that a directed requirement be written without a complete analysis of performance requirements. The new technology leveraged in the ECH could provide a 20% lighter (about a ½ pound) helmet while maintaining current protection, or could provide limited rifle protection at current helmet weights, or could strive for a substantial increase in rifle protection but with added weight. Pressure existed to lighten the fighting loads of warfighters in combat with benefits being improved speed and mobility, less fatigue and more endurance, and fewer long-term injuries. At the same time, there was the push for more capability (in this case, rifle protection), which meant increased weights. The Army and Marine Corps struck the easiest most expedient balance of increasing protection while constraining the weight to that of current helmets (*Difficulty Making Trade-offs and Planning Fallacy*)—essentially pushing the “easy button” attempting to “go fast” in a schedule-driven effort.

The testing protocols were important because they were placed in the helmet specification in the signed contracts for helmet deliveries and told industry the government requirements to be met. This was a typical “chicken or egg” scenario. Helmets with the new technology had never been manufactured. The testing protocols can only be established after making helmets and fully characterizing those helmets through design limit testing. However, a full-scale research and development effort required time and money. The testing protocol for the current helmets was refined after more than a decade of development, testing, and manufacture. The new technology in helmets behaved differently than the previous technology after being shot with a bullet. Thus, the required testing protocols differed. With nothing else to go on initially, the ECH testing protocols were set the same as current helmet test protocols (*Difficulty Making Trade-offs and Planning Fallacy*). The subsequent schedule slips in the ECH program were partially a result of a refinement of the test protocol learned over time in the development effort.

Programs that involve the application of a new technology inherently include high levels of integration, manufacturability, producibility, and quality risk. These programs should guard against being primarily schedule-driven (*Over-optimism*). Time is required to optimize the requirements and testing protocols while also encouraging the widest possible participation in the program by interested and innovative helmet manufacturers. In this case, an effort that originally planned to field helmets within a year took 4 years to reach a production decision—not so rapid. The industrial base suffered as the program relied on a sole-source vendor without the benefits of competition to keep costs and schedule in check. A program that is knowledge–driven from a research and development effort that includes many competitors from the industrial base may have proven more beneficial and had a similar execution timeline. The Services must be realistic about the risks associated with development efforts involving new technology and must avoid being primarily schedule–driven rather than knowledge–driven for acquisition decision-making.

The ECH program was initiated as a 1-year rapid acquisition effort that resulted in a 4-year effort to field a helmet (better than the current helmets) but with less than the theoretical possible optimal level of performance at a cost three times that of current helmets because of limited competition and low manufacturing levels of a capability and capacity (industrial base concerns). The Service's and the PM's decision-making in this program was greatly affected by *Trouble Making Trade-offs, Planning Fallacy, Over-optimism,* and *Recency Bias*.

JCM Program

The Joint Common Missile (JCM) program was initiated in the late 1990s. It was a Joint (Army, Navy, Marine Corps) effort to replace Hellfire, Maverick, and aviation-launched, tube-launched, optically-tracked, wire-guided (TOW) missiles fired from both rotary wing (AH-64 Apaches, AH-1 Cobras, and MH-60 Seahawks) and fixed wing (F/A18 D/F Super Hornets) aircraft. The JCM program had a successful milestone (MS) B in early 2004 with an approved capabilities development document (CDD) and subsequently awarded an Engineering and Manufacturing Development (EMD) contract for a planned 4-year EMD phase. The approved JCM acquisition strategy had a planned single-step development approach to meet all required capabilities (*Difficulty Making Trade-offs, Planning Fallacy, Over-optimism Bias*).

The Army and Navy planned the JCM program for a decade prior to the 2004 Milestone B or official designation of the program of record and start of the EMD phase. The science and technology (S&T) communities matured the underlying missile technologies through S&T objectives and a technology maturation and risk reduction phase. A high-level government work breakdown structure described the missile design and used to assess a medium risk assessment for the JCM development effort as well as technology readiness level (TRL) determinations of 6 for the critical technology elements (CTE) of the missile in support of the Milestone B decision. During this same time, the requirements generation system completed both a capabilities-based assessment (CBA) and analysis of alternatives (AoA). The CBA and AoA supported the approval of the JCM capability development document (CDD), which contained key performance parameters (KPP), initial operational capability (IOC) dates, acquisition objective (AO), and an average unit procurement cost (AUPC). Simultaneous with the technology maturation and requirements solidification, the resourcing plan for a JCM program was being worked and funding was planned and programmed. The JCM business case analysis supported the JCM program office estimate, the Army and Navy program objective memorandum (POM) submissions, and the JCM program was deemed affordable by the Army and Navy. However, the independent cost estimate (ICE) by Cost Analysis Improvement Group (CAIG) that supported the JCM Milestone B decision determined that the effort would take considerably longer than planned (from 74–144 months in EMD rather than the planned 48 months) and cost considerably more than planned (an AUPC of $153,000 rather than the planned AUPC of $108,000) (*Planning Fallacy, Over-optimism Bias*).

Despite the ICE conclusions, the JCM acquisition strategy recommended by the Army and Navy, supported by the warfighters, and approved by the defense acquisition executive (DAE) in the spring of 2004 after a successful Milestone B was a single-step development effort that planned to meet all the KPPs (*Difficulty Making Trade-offs, Planning Fallacy, Over-optimism Bias*). Late in 2004 (approximately 6 months after program approval), the JCM program was canceled as a program of record, and the effort was eventually renamed as the Joint Air to Ground Missile (JAGM) program. In 2015, the JAGM program applied the key lesson learned from the failed JCM effort—adoption of an incremental development approach. The JAGM program was approved as a program of record and successfully awarded an EMD contract after a Milestone B approval in 2015 (11 years after the JCM Milestone B approval). The capabilities to be delivered under the JAGM program were greatly reduced from the capabilities desired in the JCM program. Figure 5 displays the differences between the JCM and JAGM programs. The documented lessons learned emphasized the avoidance of extensive unprioritized requirements, multiple threshold platforms, and the fixed-wing F18 platform. The Army and Navy lessons applied to the JAGM effort emphasized an incremental development of the warfighter's highest priorities, reduced the threshold platforms, and leveraged the existing Hellfire missile warhead and motor to reduce risk, cost, and schedule (*Difficulty Making Trade-offs, Planning Fallacy, Over-optimism Bias*).

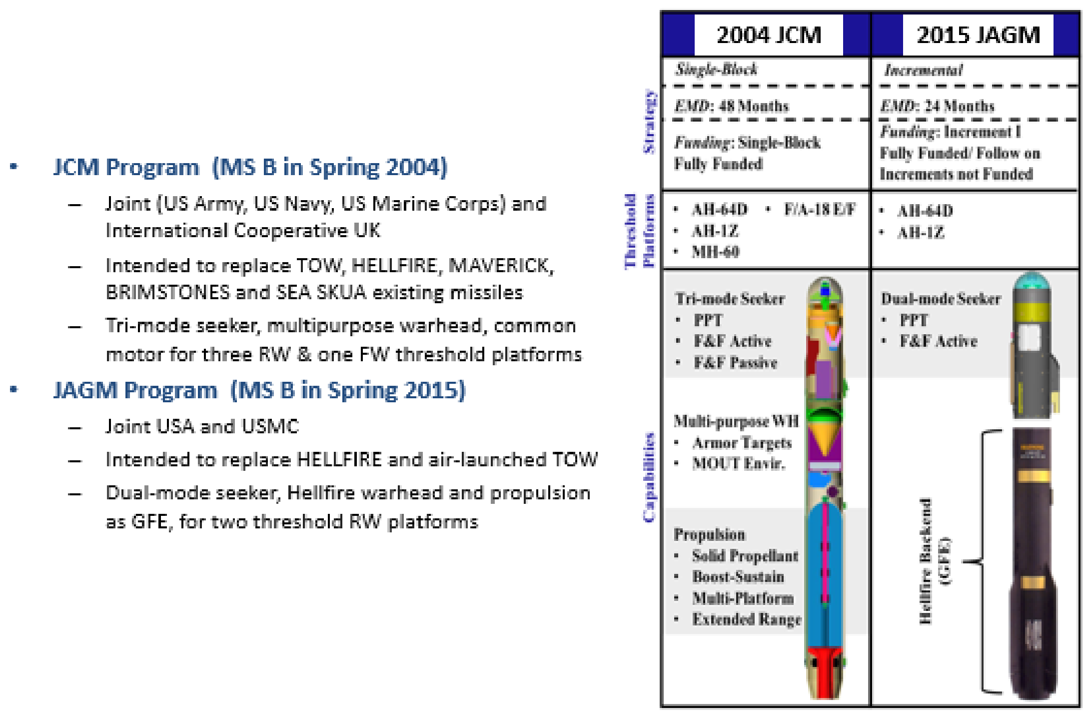


Figure 5. JCM/JAGM Acquisition Strategy Comparison (Mortlock, 2020)
[note the following acronyms: RW = rotary wing, FW = fixed wing, USA = US Army, USMC = US Marine Corps, GFE = government furnished equipment, PPT = precision point targeting, F&F = fire & forget, MOUT = military operation in urban terrain]

The plight of the original JCM program approval, subsequent cancellation, and then transition to the JAGM program offers an example supporting the GAO (2015) conclusion that the defense acquisition system often provides incentives for Services and PMs to promote *successful* acquisition strategies (defined as approved and leading to successful milestones) rather than *sound* acquisition strategies (defined as executed within cost, schedule, and performance constraints, and leading to fielding capability). *Difficulty in Making Trade-offs* bias makes it difficult for the Services and PMs to develop acquisition strategies to optimally balance near-term program milestone approval and long-term program executability in terms of maintaining cost, schedule, and performance baselines and delivering capability.

The Services and the JCM PM basically had two choices to reduce programmatic risk when formulating the JCM acquisition strategy—plan more time and money for the effort as defined by the warfighter or reduce scope (achieve less performance requirements) for the time and money planned. Allocating more money or additional schedule was not considered because JCM had a TMRR phase deemed successful and planned EMD phase aligned with funding in the Service POMs. A reduction in scope by recommending reducing performance capabilities was not considered because that risked losing the support of the warfighting communities who strongly supported achieving the full required capabilities (*Planning fallacy*, *Difficulty in Making Trade-offs*).

The JCM requirements were well established and supported by years of analysis with a set capability need date. The technologies needed to turn those requirements into capabilities for the warfighter had matured to the point that they were deemed mature (TRL 6) and ready for integration and development work. Additionally, the funding to support the JCM program of record for a development and engineering work and procurement of missiles was aligned to the required need date (IOC). The PM triple constant of cost, schedule, and performance was all synchronized and set within the planned acquisition program baseline (APB). However, for the JCM program, a single-step acquisition strategy to deliver all required capabilities was eventually canceled and the warfighter received no capability. Had an incremental development approach like the subsequent JAGM acquisition strategy been adopted initially, the warfighter could have received improved capability more than a decade sooner.

In the case of the JCM program, the cost and schedule constraints and the ICE determinations indicated the need to consider an incremental development approach and delay some capability to later increments (*Planning Fallacy*). The JCM program was canceled, and it took more than 10 years for the new JAGM program to pass an Milestone B—this time with an incremental approach that leveraged existing government furnished equipment (GFE) components and non-development item (NDI) technologies. Meanwhile, during this “lost decade,” the warfighter got none of the desired capabilities required. The defense acquisition system incentivized the Services and PMs to get an approved milestone—but with a program that was soon canceled, failed to meet performance requirements, and delivered no capability. The JCM program serves as an example of a program in which the cost, schedule, and performance constraints were unrealistically established. The Services and the PM decision-making were affected by the *Planning Fallacy, Difficulty in Making Trade-offs, and Over-optimism Bias*.

Army Infantry Combat Vehicle Programs

The Bradley Fighting Vehicles (BFV) remains the backbone of the Army mechanized infantry warfighter formations. Developed in 1970s, the Bradley has been upgraded several times to offer Soldiers enhanced capabilities; however, since the early 2000s, the Army has been trying to replace the BFV due to size, weight, and power constraints that severally restrict any potential upgrade options. Figure 6 shows the evolution of the BFV over time.

Figure 6. BFV Over Time (Roth & Hames, 2019)

One attempt at a BFV replacement was the Infantry Carrier Vehicle (ICV) as part of a family of eight manned ground vehicles (MGV) within the planned Future Combat Systems (FCS) Brigade Combat Team construct for the Army Future Force. The FCS program entered the acquisition framework as an official program of record at Milestone B to begin engineering and manufacturing development (EMD) efforts in 2003 with a planned Milestone C (low rate initial production) in 2010, but the program was canceled in 2009. Figure 7 describes the planned ICV as presented at the preliminary design review in early 2009. Over the past decade, defense acquisition experts have referenced the FCS program as an example of everything wrong with defense acquisition—a canceled program that wasted billions of dollars and delivered no capability to warfighters (Pernin et al., 2012). The FCS program was an ambitious effort that attempted to integrate technologies using a system-of-systems approach to transform the way Army brigades would fight. Additionally, the program started as a Defense Advanced Research Projects Agency (DARPA) effort contracted through other transaction authority (OTA) with Boeing and its industry teammate, Science Applications International Corporation (SAIC). The OTA incentivized Boeing to get the Army to an approved Milestone B, to start the formal program of record and system development and demonstration phase as quickly as possible. The Army planned, programmed, and budgeted funding—at risk if the program was not executed on schedule (schedule-driven). Boeing and the Army achieved that Milestone B in 2003. The OTA also enabled Boeing to become the lead system integrator (LSI) for the FCS program of record. Despite warnings from the GAO of immature technologies and lack of adequate funding, the Army marched forward until 2009, when the Secretary of Defense canceled the FCS program because of affordability concerns, immature technologies, and requirements not reflecting the current threats faced by soldiers in Iraq and Afghanistan (GAO, 2004). The FCS program was an example of a rush to failure and resulted in no capability delivered to the warfighter (*Planning Fallacy*). The program was also hampered by *Difficulty in Making Trade-offs* and *Over-optimism* (too many requirements, too many immature technologies), as well as *Recency Bias* (use of DARPA, OTA, system of systems approach and LSI concept) to use the latest acquisition reform initiatives despite other more suitable, less risky acquisition approaches.

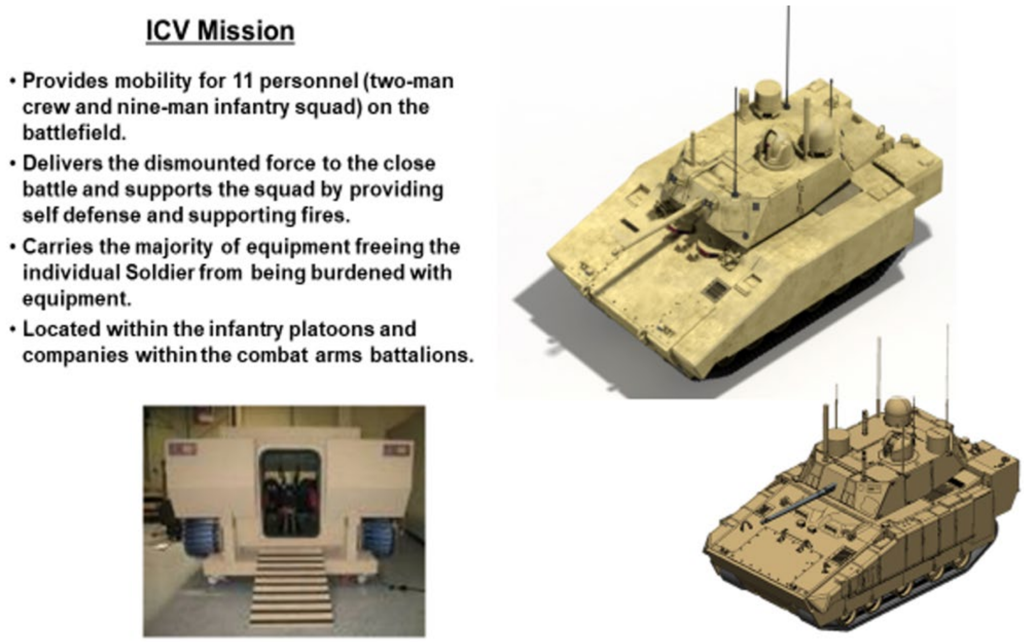


Figure 7. FCV ICV in 2009 (Adapted from PdM ICV, personal communication, 2009)

After cancellation of the FCS program in 2009, the Army embarked on the Ground Combat Vehicle (GCV) program (see Figure 8) to replace the BFV. At the time, Fort Benning served as the home of the infantry, and Fort Knox served as the home of armor. All resources that had been supporting the oversight and management of the development of a family of eight FCS manned ground vehicles were now applied to the development of the GCV. The Army designated Fort Knox as the lead in the defining the requirements for the GCV. The GCV program pushed for a materiel development decision and Milestone A in 2010 to begin awarding technical maturation and risk reduction (TMRR) contracts to industry. The same two industry partners that were teamed together in the FCS EMD phase for manned ground vehicles now competed against each other in a TMRR phase for the GCV. The Army began the GCV program and awarded FFP type research and development (R&D) contracts to BAE Systems and General Dynamics for designs and prototypes. The new vehicle's requirements called for a heavy reliance on mature commercial technologies. Better Buying Power (BBP) initiatives strongly encouraged and incentivized resulted in awarding FFP type R&D contracts (*Recency Bias*) despite the lack of appropriateness based on the level of system integration complexity and risk.

With the Armor School in the lead for defining the GCV requirements for a BFV replacement vehicle for the mechanized infantry, the requirements for GCV resulted in mixture of requirements from the BFV, the FCS ICV, the recently fielded mine resistant ambush protected (MRAP) vehicles, and the M1A2 Abrams tank. Based on the GCV requirements, the program office, the interested industry competitors, and engineers at the research, development, and engineering center (RDEC) at the Tank and Automotive Command (TACOM), determined that the GCV would weigh between 50 and 70 tons—nearing the weight of the 72-ton M1A2 Abrams tank and almost twice as heavy as the 30-ton BFV or planned 30-ton FCS ICV.

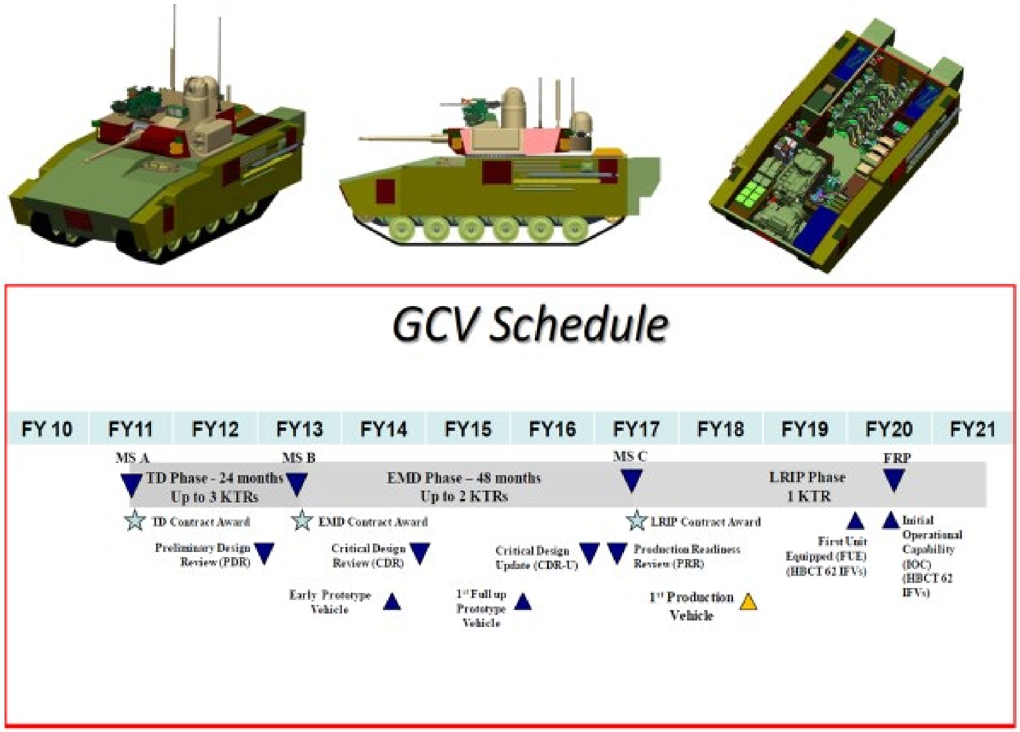


Figure 8. GCV as depicted in 2010 (Adapted from PdM GCV, personal communication, 2010)

The GCV had force protection, survivability, and lethality requirements for a mechanized infantry vehicle written by armor warfighters or tankers. In subsequent reviews with the Headquarters Department of the Army (HQDA) staff (including the Vice Chief of Staff of the Army), the potential weight of the GCV and excessive requirements were highlighted; however, the Army pushed ahead and awarded two TMRR contracts based on schedule pressure for protecting the planned and programmed resources of the old FCS MGV program (*Planning Fallacy, Over-optimism and Difficulty in Making Trade-offs*). Four years later, the Army canceled the GCV program because the vehicle was going to be too big and heavy and had excessive requirements. The GCV effort was not focused on the mechanized infantryman—it was focused on other Army priorities.

In recent years, after several failed attempts of initiating the Next General Combat Vehicle because of aggressive requirements and lack interest by industry, the Army is trying again—this time calling the BFV replacement the Optionally Manned Fighting Vehicle (OMFV). The OMFV program is susceptible to same behavior acquisition biases (*Planning Fallacy, Over-optimism, Difficulty in Making Trade-offs, and Recency Bias*) as contributed to the failures of the predecessor BFV replacement acquisition efforts. How can the design and development of a mechanized infantry vehicle be optimized for troop transport and protection, lethality, and remote autonomous operations simultaneously? The answer is unfortunately, it can't—this will require difficult requirement trade-offs to avoid the *planning fallacy* and *over-optimism* bias. A vehicle that is optimized for lethal autonomous operations would be an inefficient combat vehicle to protect the crew and protect the troops being transported. It appears that *Recency Bias* has also played a significant role in the OMFV program. Is the Army more interested on riding the autonomous vehicle hype wave? Or does the Army have other priorities like proving the value of the high-profile, newly established Army Futures Command or Next Generation

Combat Vehicle (NGCV) Cross Functional Team (CFT)? The acquisition strategy for the OMFV program appears to be heavily influenced by *Recency Bias*. The acquisition strategy leverages the newly formed middle tier acquisition (MTA) pathway to avoid forming an acquisition program of record to enter the EMD phase after a successful Milestone B. The OMFV program will use MTA authorities to rapidly prototype vehicles for experimentation and demonstration and then establish a formal acquisition program of record at Milestone C to enter low-rate initial production. This strategy is the exact opposite strategy that the GAO has recommended for more three decades for major defense acquisition programs—knowledge-based acquisition strategies. Defense acquisition programs have routinely rushed to production decisions without well-defined requirements, complete detailed design drawings, fully mature technologies, and mature manufacturing processes, and without demonstrating production representative systems in a relevant operationally environment. The OMFV program is attempting to do in an MTA rapid prototyping effort what a major defense acquisition program achieves in a formal EMD effort—a classic "schedule-driven" rush to failure.

Summary, Conclusions and Recommendations

Summary

We studied three defense acquisition programs and found strong evidence that systemic behavioral biases affected the management and decision-making within these programs. Table 1 summaries the research results. The outcomes of the ECH, JCM, and Army Infantry combat vehicles programs were affected by *planning fallacy, difficulty in making trade-offs, over-optimism, and recency bias* (except JCM).

Table 1. Summary of Biases Present in Defense Acquisition Programs

	Behavioral Biases			
Programs	**Planning Fallacy**	**Difficulty in Making Tradeoffs**	**Over-Optimism**	**Recency Bias**
ECH Program	√	√	√	√
JCM Program	√	√	√	
Army Infantry Vehicles	√	√	√	√

Conclusions

The presence of the effect of behavioral biases within the management and decision-making of acquisition programs comes as no surprise. However, the extent and frequency of the *planning fallacy, difficulty in making trade-offs, over-optimism, and recency bias* within the three acquisition programs studied sharpens the point on which biases are most relevant within defense acquisitions. For the past three decades, acquisition management has been highlighted on the GAO's high-risk list for excessive waste and mismanagement. Notable programs have failed to deliver capability, and failed to meet performance, cost, and schedule management targets. The root causes of program failure vary from ill-defined requirements, immature technologies, integration challenges, poor cost and schedule estimating, and the acceptance of too much development risk. Underappreciated and understudied is the effect that the *planning fallacy, difficulty in making trade-offs, over-optimism, and recency bias* have in contributing to root causes of acquisition program failures. The better we understand the effect of these systemic behavioral systemic biases, the better we can mitigate the risks of program failures.

The complexity of the defense acquisitions makes study of systemic biases interesting. The culture and leadership at different levels of the DoD from the institutional level to the organizational level to the program level affect the impact of the biases. In the DoD's hierarchical chain of command, the PMs are responsible for the program's cost, schedule, and performance (triple constraint). PMs make decisions for the proper management and execution of the program within the triple constraint parameters. However, the PMs do not establish the performance requirements, cost, or schedule objectives of the acquisition program baseline—the Services do. Additionally, PMs report to a milestone decision authority (MDA) who approves the program and determines overall program strategic direction. The systemic biases at the various levels of the chain of command manifest differently in the decision-making models used at different levels. *Behavioral acquisition* explores defense acquisition from a behavioral standpoint, including the impact of psychology, organizational behavior, and organizational politics on how culture, leadership, and decision-making affect the management and execution of program, as well as program outcomes.

Recommendations

For further research, we recommend rigorous study of more acquisition programs that can clearly show the distinctions we suspect are in play. We recognize that in the real world these distinctions are hard to show, even though we all know what's going on. Surveys of PMs and MDAs are an excellent way to get primary data to understand *behavioral acquisition* more fully. Figure 9 to highlights a model designed to study how the chain of command, culture, and decision-making moderate the effect of behavior biases in acquisition managers.

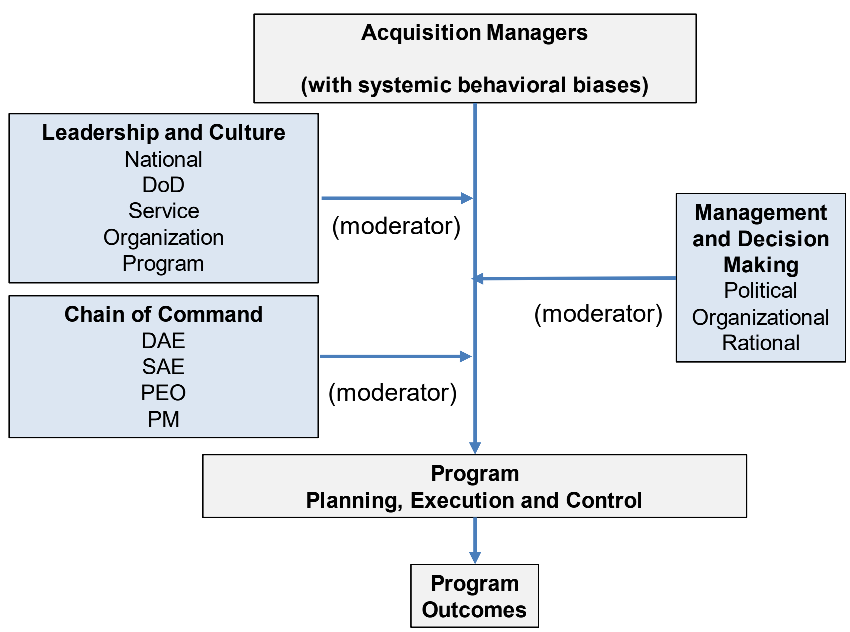


Figure 9. Moderator effects on acquisition manager behavior that affect program outcomes

References

Barber, B. M., Heath, C., & Odean, T. (2003). Good reasons sell: Reason-based choice among group and individual investors in the stock market. *Management Science*, *49*(12), 1636–1652.

Beach, L. R., & Connolly, T. (2005). *The psychology of decision making: People in organizations*. Sage Publications.

Buehler, R., Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Cassar, G. (2010). Are individuals entering self-employment overly optimistic? An empirical test of plans and projections on nascent entrepreneur expectations. *Strategic Management Journal*, *31*(8), 822–840.

Clark, A. (2008). *Supersizing the mind*. Oxford University Press.

Flyvbjerg, B., Garbuio, M., & Lovallo, D. (2009). Delusion and deception in large infrastructure projects: Two models for explaining and preventing executive disaster. *California Management Review*, *51*(2), 170–194.

GAO. (April 1, 2004). *Defense acquisitions: The Army's Future Combat Systems' features, risks, and alternatives* (GAO-04-635T). *Statement of Paul L. Francis, Director, Acquisition and Sourcing Management. Testimony before the Subcommittee on Tactical Air and Land Forces, Committee on Armed Services, House of Representatives*. U.S. Government Printing Office.

GAO. (2015). *Joint action needed by DoD and Congress to improve outcomes* (GAO-16-187T). U.S. Government Printing Office. https://www.gao.gov/products/GAO-16-187T

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology*, *2*(1), 83–115.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart* (pp. 3–34). Oxford University Press.

Hales, S. D., & Johnson, J. A. (2018). Dispositional optimism and luck attributions: Implications for philosophical theories of luck. *Philosophical Psychology*, *31*(7), 1027–1045.

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: Evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes*, *61*(3), 305-326.

Hmieleski, K. M., & Baron, R. A. (2009). Entrepreneurs' optimism and new venture performance: A social cognitive perspective. *Academy of Management Journal*, *52*(3), 473-488.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluation of alternatives. *Organizational Behavior Human Decision Processes*, *67*, 247–257.

Irwin, J. R., & Davis, J. H. (1995). Choice/matching preference reversals in groups: Consensus processes and justification-based reasoning. *Organizational Behavior Human Decision Processes*, *64*, 325–339.

Kahle, J., Pinsker, R., & Pennington, R. (2005). Belief revision in accounting: A literature review of the belief-adjustment model. *Advances in Accounting Behavioral Research*.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982*). Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research*, *31*, 231–245.

Kiesling, J. K., & Chong, D. M. (2020). *Examining the role of decision biases in shaping acquisition decisions within defense acquisition programs* [Master's thesis, Naval Postgraduate School].

Klein, G. A. (2011). *Streetlights and shadows: Searching for the keys to adaptive decision making*. MIT Press.

Lautliev, G. N., & Menter, P. (2014). *Behavioural perspective on financial modelling choices in the U.S. power generation industry* [Master's thesis, Copenhagen Business School].

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255.

Levinthal, D. A. (2011). A behavioral approach to strategy—what's the alternative? *Strategic Management Journal*, *32*(13), 1517–1523.

Maltby, J., Day, L., Gill, P., Colley, A., & Wood, A. M. (2008). Beliefs around luck: Confirming the empirical conceptualization of beliefs around luck and the development of the Darke and Freedman beliefs around luck scale. *Personality and Individual Differences*, *45*(7), 655–660.

Meza, D. D., & Southey, C. (1996). The borrower's curse: Optimism, finance and entrepreneurship. *The Economic Journal*, *106*(435), 375–386.

Mortlock, R. F. (2020). Studying acquisition strategy formulation of incremental development approaches. *Defense Acquisition Research Journal*, *27*(3), 264–311.

Mortlock, R. F. (2018, February–March). Protecting American soldiers: Development, testing and fielding of the enhanced combat helmet. *Project Management Journal*, *49*(1), 96–109.

Pernin, C. G., Axelband, E., Drezner, J. A., Dille, B. B., Gordon IV, J., Held, B. J., McMahon, S., Perry, W. L., Rizzi, C., Shah, A. R., Wilson P. A., & Sollinger, J. M. (2012). *Lessons from the Army's Future Combat Systems program*. RAND Corporation.

Pressman, J. L., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland; Or, why it's amazing that federal programs work at all, this being a saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation* (Vol. 708). University of California Press.

Product Manager Ground Combat Vehicle, Program Executive Office Integration. (2010, May). *Ground combat vehicle overview briefing* [PowerPoint briefing prepared for Army Leadership].

Product Manager Infantry Carrier Vehicle, Program Manager Future Combat Systems. (2009, March). *Infantry carrier vehicle overview briefing* [PowerPoint briefing prepared for Program Review].

Program Manager Soldier Protection and Individual Equipment, Program Executive Office Soldier. (2013, May 21). *Soldier protection system milestone B* [PowerPoint briefing prepared for Army Leadership].

Roth, M. C., & Hames, J. M. (2019, April–June). Then & now: Next generation on track. *Army AL&T*. https://asc.army.mil/web/news-alt-amj19-then-now/

Scheier, M. F., Carver, C. S., & Bridges, M. W. (2001). *Optimism, pessimism, and psychological well-being*.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, *61*(2), 195.

Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, *49*, 11–36.

Shore, B. (2008, December). Systemic biases and culture in project failures. *Project Management Journal*, *39*(4), 5–16.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118.

Slovic, P. (1975). Choice between equally valued alternatives. *Journal of Experimental Psychology: Human Perception Performance*, *1*, 280–287.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Advanced Earned Value Management: A Proposal for Extending Program Management Theory Through Value Centric Turbulence Flow Methods

Raymond Jones—Naval Postgraduate School [rdjone1@nps.edu]

Abstract

Current methods of measuring acquisition project performance are based upon generations of old cost models that compare budgets with expenditures. These methods provide little in the way of forecasting, leaving the project manager to react to past performance rather than making decisions of future potential. The principal means of measuring project performance in large complex projects is earned value management (EVM), which measures actual work performed to budgeted work performed. While this method provides valuable insight into how well a project is doing relative to a predetermined plan, new methods of forecasting when projects begin to become volatile are needed. Knowing when a project will enter a phase of volatility will provide greater transparency and flexibility to decision-makers before critical events have occurred. This paper describes the state of current project management theory and how it has driven the project management community to a point of stagnation. Additionally, this paper will outline a new and innovative approach to forecasting when a project will become volatile long before current EVM and risk management tools can today. This research approach proposed in the paper intends to extend current theory and practice in project management by examining several information systems programs and providing new, more effective decision-making tools for future project managers. The methods defined in this research will be known as advanced earned value management (AEVM).

Introduction

When managers try to develop complex products with many interdependent subsystems, the high informational processing load can overwhelm the organization. Errors, poor decisions, and bad communication can result in additional levels of effort later in the development process leading to inefficiency or failure. Additionally, decision support systems currently being used in defense acquisition are focused on analysis of historical cost data rather than the effect the volatile environment has on future performance. This problem is particularly acute in information-centric acquisition programs in which the boundaries of hardware and software are less obvious and where the socio-technical boundaries are critical to successful performance of the system being developed.

Defense acquisition programs remain vulnerable to underperformance and excessive cost growth during times of increasingly constrained budgets. History suggests that the "normal" condition of complex programs is one of failure in that they typically exceed their preplanned cost, schedule, and performance targets. This phenomenon is especially true in large complex programs such as information systems and network acquisition programs. Since 1975, an annual array of acquisition reform studies, beginning with the Packard Commission, have had virtually no impact on the ever-increasing trend of cost growth and substandard program performance. The increasing need for complex artificial intelligence and networked technologies capability within the DoD portends even more challenges for the defense acquisition process. Current management theory reinforces classical approaches to project management in which development programs are viewed from a preplanned production process that can be controlled through a robust systems management approach. It is critical for the acquisition community to examine new analytical and decision-making approaches to managing the acquisition of these systems. It is also important to explore the theory of program management and assess if change is necessary to the existing paradigm of managing risk within the iron triangle paradox of cost, schedule, and performance.

Project planning is an output-oriented process that attempts to "decide" future events though rigorous advanced planning and process controls. Projects tend to be planned and executed using classical methods, in which a project follows a standard system model that requires an input, transformation, and output flow model intended to align resources with objectives (see Figure 1).

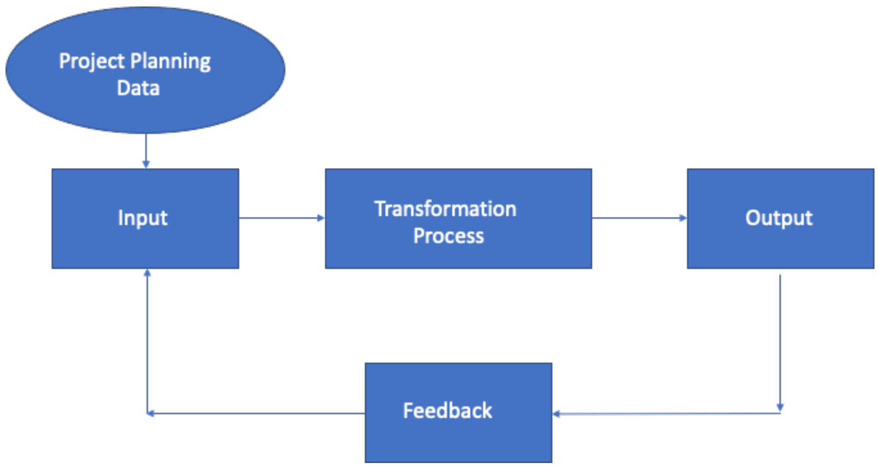


Figure 1. Standard System Feedback Process

The process involves problem-solving focusing on "what" and "how" the project will be accomplished. All activity in the project is devoted to the accomplishment of the objective through planned future time periods. Project performance is typically measured using techniques such as earned value management (EVM), which derives perceived "value" from planned work relative to actual work performed. While these processes provide historical perspectives of a program, they do not provide objective predictive insight by which a program management team can make better decisions that lead to higher probabilities of success of a development acquisition program.

Current program management methods are grounded in classical and system management theory, which is based on the belief that workers have physical and economic needs rather than social needs or job satisfaction and that projects are a system that is represented by a collection of parts brought together to accomplish some end goal or objective. Classical management theory advocates specialization of labor, centralized leadership and decision-making, and profit maximization. In the case of defense program management, program teams are focused on budget and cost optimization rather than profit maximization, resulting in an approach that assumes a development program can be executed in accordance with a centrally derived and executed cost and schedule baseline. Unfortunately, this view of project management prevents project teams from adequately anticipating future events that could lead to increasing project volatility. Project managers are left to deductive reasoning based upon regressive data and experience, significantly increasing the probability that complex programs will exhibit unacceptable levels of variability with regard to the planned program performance baseline.

Other scientific fields such as aerodynamics have analogous problems in that prediction and forecasting is necessary to improve system design and performance. Using aerodynamics as an analogous benchmark with which scientists have developed methods by which the turbulent and volatile fluid environment can be more accurately predicted, there may be an opportunity to extend this methodology into program management theory to better understand the turbulent and volatile project environment in which program's flow during the development phase of a project life cycle. If it is possible to more accurately forecast the chaotic environment in which a program must interact, than perhaps there is an opportunity to fundamentally advance the underlying theory upon which project management is based and develop new methods to better forecast project performance.

Statement of the Problem

The absence of a strong and inquisitive theoretical base in program management is perpetuating a management strategy that is grounded on historical cost accounting and traditional manufacturing models rather than methods that offer improved forecasting and decision-making approaches. Hence, *the problem is that current cost accounting and system thinking methods, based upon classical management and system theory, do not provide adequate forecasting opportunity for complex acquisition programs during the execution and control phase of the program life cycle.* Other disciplines such as fluid and thermodynamics have similar challenges in that they attempt to predict physical outcomes from complex and dynamic environments. This research will show how a complex business environment can be treated as a knowledge flow system that moves from calm laminar states to volatile turbulent states, much the same as other Newtonian fluid environments. The following research questions are proposed to address the problem.

1. Do current project management and control methods lack the ability to adequately forecast project future performance during the developmental phase of the project life cycle?
2. Does the project environment display knowledge flow characteristics similar to other Newtonian fluids?
3. Is it possible to forecast project management performance using turbulent flow theory such as the Reynolds-averaged Navier Stokes (RANS)?

Objective

The purpose of this research is to advance program management theory and practice by applying knowledge flow and turbulence forecasting methods to traditional business decision support systems. While traditional decision support systems such as earned value management (EVM) provide an accurate accounting view of project performance, turbulent flow modeling will allow for improved forecasting, leading to an advanced EVM (AEVM) approach to project management. Turbulent flow methods have shown themselves to provide a more predictive construct to other disciplines such as weather forecasting, aerodynamics, and hydrodynamics. This approach is becoming more relevant to other fields outside of natural physics and is revealing itself as a potential framework by which human-centered organizational outcomes may be predicted. This research will also be grounded in an information science domain in that we will be using structured and unstructured project data in a new way that allows improved project forecasting. Additionally, within the information sciences paradigm, business intelligence and analytics is gaining an increasing role with regard to how data is transformed into actionable insights that inform an organization's strategic and tactical business decisions. Business analytics such as project cost data, lexical link analysis of key project reports, and turbulent flow methods may provide predictive project forecasting information, increasing the business intelligence environment of the project team and subsequently improve the project decision support strategy.

Why Is This Research Important?

This research is important because current methods of measuring project performance such as earned value management are not sufficient to adequately provide forecasting insight into project volatility, limiting decision-makers' ability to make informed decisions in a risk-based project environment. Current theory and practice are not well suited to managing the complex nature of programs intended to develop artificial intelligence and adaptive reality capability. As the DoD undertakes ever-increasing complex programs such as Project Overmatch and Project Convergence, the need to have more accurate forecasting methods is vital to ensuring success and preventing repeating the mistakes of the past.

Relying on current methods that are based upon a dearth of theoretical explanation is not an acceptable strategy if the United States intends to maintain its technological lead into the 21st century. This research will provide the project team a new tool by which they can more accurately predict the impact of planning decisions before the project exceeds its performance baseline metrics such as cost and schedule, resulting in a chaotic and volatile program. Current models, such as EVM and the risk management framework (RMF) are regressive in nature and do not provide adequate forecasting insight into project management performance. A project becomes unstable and volatile when it begins to vary off its predetermined performance baseline. While there will always be some level of variability relative to the program performance baseline, when project variability exceeds its ability to maintain project baseline convergence, the decision-making environment becomes chaotic, resulting in an increasingly turbulent project with increasing volatility. For the purpose of this research, increasing project volatility is considered to be a dynamic state in which project risk realization becomes oscillatory dynamic with regards to mitigation strategies. In other words, the risks begin to outpace the project's ability to mitigate them toward an acceptable project conclusion.

If project teams could accurately diagnose potential areas of failure through improved forecasting methods, proactive measures could then be employed to prevent program failure. Engineers use the Reynolds number in fluid mechanics as a metric that predicts whether the flow of a fluid will be smooth and stable (laminar flow) versus unstable and chaotic (turbulent flow). Knowledge flow in an organization acts in the same way. Fyall (2002) argues that sufficient organizational capacity may allow for good information flow through a team while an overwhelmed group will suffer from turbulent information flow and risk total failure. An organizational Reynolds number, for example, could determine if the information flow through an organization was going to be efficient and reliable versus volatile and chaotic. This would help project teams to better understand the project risk profile and make choices to mitigate risk realization. Reynolds number is a value that helps to predict viscous and inertia effects in Newtonian fluids in order to characterize the stability of the fluid environment under varying conditions.

Literature Summary

Organizations across many business sectors and geographic borders are steadily embracing project management as a way to control spending and improve project results. Executives emphasize that adhering to project management methods and strategies reduces risks, cuts costs, and improves success rates. As a profession, project management has gained increasing interest over the past several years for both its failures and its relevance (Morris, 2000). Often referred to as a means for dealing with change (Cleland & Ireland, 2007), project management has become a topic of real interest from both the practitioners' and researchers' perspective (Ika, 2009). Recent trends within the DoD suggest that project management is critical to effectively delivering key weapons capability to the warfighter. For example, the Defense Acquisition University (DAU) is currently attempting to fundamentally change the pedagogy by which project management training is delivered to the defense acquisition workforce. The DoD has identified six critical functional areas, one of which is project management, that are necessary as the DoD moves into the next century.

Additionally, Packendorff (1995) argues that project management is emerging as a true scientific discipline with its own academic journals, conferences, and language that are characteristic of claiming scientific status. Yet, despite the level of interest and significant amount of investment in the profession of project management, its effectiveness remains in question. Significant project delays, cost overruns, and underperformance seem to be the rule rather than the exception (Ika, 2009). This initial search of the literature will establish the current *state of project management theory*, provide an introduction *to the tools by which project managers* attempt to understand and control their programs, and offer a glimpse into how *turbulent flow model* can help move project management into a more productive management paradigm and why this is critical to improving project forecasting.

The State of Project Management Theory

There seems to be a growing realization that the fragmented theoretical foundation in the field of project management is contributing to the stagnant performance being observed in actual project performance (Cicmil, 2006). The prevailing practices that seem to embrace project management practices and research are grounded in traditional economic and manufacturing methods, which view cost as the dependent variable for project success. The foundation of project management theory goes back to Frederick Taylor, Max Weber, and Henry Fayol. Taylor's *The Principles of Scientific Management* introduced the concept that management can be viewed as a science in which deliberate scientific processes replaced the rule-of-thumb method and that work can be divided between workers and managers based upon their respective levels of training and responsibility (Taylor, 1911).

Additionally, Weber (1930) argued that organizations should be hierarchically based with decisions and processes that are based upon a prescribed set of rules. Weber argued that relationships within an organization should be impersonal and focused on the scope of their respective authorities. These classical views are critical to understanding current project management in that today's organizations fundamentally follow these principles of division of labor and hierarchical decision-making, which structures a project into discrete packages of work that are assigned to individuals in a project office. These work packages are subsequently allocated resources in the form of time and money and sequenced into a project schedule known as the performance measurement baseline (Brotherton, 2008). The effect of this hierarchical and linear management approach suggests that if work can be allocated and assigned to competent workers, then projects can be planned to the smallest level of effort, thus minimizing uncertainty and potential volatility as the project moves through its life cycle. While this is a generally accepted planning framework, this classical method provides little insight into uncertainty due to less tangible factors, such as individual behavior and changing environmental conditions.

Classical theories of management needed to be extended to allow for a more human-centric influence that accounted for human behavior. Behavioral theorists such as Follet, Maslow, and McGregor extended the more classical theories by arguing that employees must be considered active participants in the decision-making process, thus introducing a new dimension of ambiguity into management decision outcomes (Prietto, 2015). McGregor in particular theorized that people work for inner satisfaction and are not necessarily driven by materialistic rewards (McGregor, 1960). His Theory X and Theory Y introduced the concept that a manager's assumptions about human nature influences how that manager manages the workforce (Prietto, 2015). From a program management perspective, the notion that individuals within a project environment do not comply with the "economic man" principle and that outputs can be influenced by the bias of both the manager and the worker, opens up the potential for arguing that the knowledge flow within a project organization cannot be precisely predicted based upon discrete planning processes as suggested in the more classical management theories and methods. This leaves us with the dilemma of trying to manage a project based off of precise pre-project planning information that could be significantly influenced by less tangible factors such as human behavior and project environmental changes.

Karl Ludwig von Bertalanffy introduced the idea that a system could be thought of from a more biological perspective which is self-correcting and is made up of processes that transform inputs to outputs (Prietto, 2015). From a management perspective, this seems to extend the underlying classical theory in that it provides the necessary rigor by which to explain the socio/managerial relationship between a hierarchical and human-centered process. His work sought to demonstrate the pervasiveness of open systems in nature and the potential of reducing human thought into a systems concept (Lopreato, 1970). The idea of viewing organizations as systems was not new. The trend, however, was to view management processes as closed systems with predictable outcomes. Pareto's 1897 law, for example, had demonstrated a system-like behavior in society by showing how 80% of the wealth belonged to 20% of the population, regardless of the total amount of available capital (Newman, 2006). A systems approach to project management suggests that the boundaries are closed and clearly delineated, allowing for well-defined interdependencies (Lorreato, 1970). By suggesting that organizations and acquisition projects behave like systems, one must accept the premise that the structure and components of an organization are interdependent and therefore affect one another.

If an organization, however, were more like a biological open system, as Bertalanffy suggested (Prietto, 2015), then one is left with the challenge of capturing random interdependent relationships of highly complex networks within an organization. By entertaining the possibility that organizations behave like biological open systems in which people and processes impact the preplanned nature of a project, we are able to accept the potential of neo-classical approaches that provide better real forecasting of project behavior and outcomes.

Path analysis is used to describe the directed dependencies among a set of variables and used to try to understand organizational behavior with regard to independent and depend variables. By viewing an organization as a system, path analysis could be considered a viable technique by which causal inference can be explored. With regard to organizational analysis, however, path analysis is often plagued by intervening variables that significantly influence the organizational model's highest path coefficients. This would suggest that systems modeling of project management organizations need additional tools and methods by which to more accurately capture the uncertainty of external intervening variables that impact the knowledge flow through an organization.

Classical and systems theory provide a relevant framework by which to begin understanding how a project management organization makes decisions and how knowledge flows through the organization, but may be limited in its ability to forecast future activities in that the complex nature of the project organization appears to behave more like an open system that is influenced by a complex network of intervening pathways and variables. The more we understand all the pathways of knowledge flow and how they affect the productivity of an acquisition project, the more relevant other analogous engineering and physics models could help us to improve project performance forecasting. Knowing how to improve the flow of knowledge requires a framework that considers all the pathways within a project (Cohen & Prusak, 2001). Pathways for knowledge flow are similar to networks in organizations. Within a project management structure, knowledge flow can be correlated with the work breakdown structure and how efficiently the preplanned work breakdown structure (WBS) is being executed relative to the *planned performance measurement baseline* (PMB). The use of network ties stimulates innovation as an interactive process and represents the main source of competitive advantage for an organization (Swan et al., 1999). Viewing WBS productivity as a knowledge flow network provides us with the opportunity to explore other models of forecasting within the engineering domain. These analogies could allow us to apply theories such as chaos, thermodynamics, and fluid dynamics to problems that have heretofore been characterized as social constructions measured through the lens of economics and accounting.

Current Project Management Measurement and Control

As information-centric projects become more complex, project management must become more effective and dynamic in order to keep up with the ambiguity and speed of technology. Sociotechnical systems and artificial intelligence will require new ways of understanding how future needs will be transformed into real and virtual capability. Cost-based manufacturing methods will not provide the necessary agility and insight into how capabilities are evolving to meet customer needs and how resources are being optimized to achieve maximum return on investment (ROI) within predefined budget constraints.

Project management methods that are used to control programs are based upon a cost management that provides little insight into project value or volatility. Changes in a project's requirements, stakeholders, schedule, and budget can cause changes in the broader environment just as desperate business processes impact project deliverables. Changes in key project metrics such as cost and schedule that vary off the original project baseline create a project environment that has the potential to be uncontrollable and results in increasing cost and schedule delays (Pitagorsky, 2018). Variance off the PMB is reflected in terms of cost and is considered to be increasing in volatility the larger the variance. Degrees of project volatility follow a continuum from rapid and unpredictable to infrequent and predictable. Changes in project certainty can be the result of many factors such as people, requirements, policies, and budget changes. While monitoring cost is an important metric, it does little in terms of improving a project manager’s ability to forecast future performance, particularly in terms of future volatility.

There are several methods by which project teams attempt to manage and reduce uncertainty in a project. The literature, however, supports the argument that there are principally only five tools that are considered most relevant for managing complex acquisition programs (Housel & Mun, 2020). These include techniques such as PERT, Lean Six Sigma, balanced score card, knowledge value-added (KVA), earned value management (EVM), and integrated risk management (IRM). The only methods that seem to suggest some level of forecasting is IRM and EVM.

Project managers typically follow an ongoing risk management process that helps them identify, understand, and respond to threats and opportunities (O’Conner, 2020). IRM is a forward-looking, risk-based decision support system incorporating various methods such as Monte Carlo Risk Simulation, Parametric Forecast Models, Portfolio Optimization, Strategic Flexibility, and Economic Business Case Modeling. Economic business cases using standard financial cash flows and cost estimates, as well as non-economic variables such as expected military value and strategic value (Housel & Mun, 2020). These metrics are often forecasted using system dynamic models based upon preplanned hierarchical structures, assumptions, and probabilistic models that make assumptions based upon the uniqueness of the project under consideration. They generally do not consider the dynamic nature of mediating variables over time such as human performance, requirements variability at both the operational and project level, and impacts of the bureaucratic and political environment.

Risk analysis, real options analysis, and portfolio optimization techniques are enablers for estimating ROI and estimating the risk-value of various strategic options. These techniques are favored in portfolio management approaches to project management and attempt to provide greater insight into future performance in order to optimize project investment. Real options theory informs the risk management process in that it provides insight into how to make decisions regarding investments in tangible assets when the future is uncertain (Hayes, 2020). Current statutory guidance such as the Clinger–Cohen Act of 1996 mandates the use of portfolio management for all federal agencies. The GAO (1997) requires that IT investments apply ROI measures, and DoD Directive 8115.01, issued October 2005, mandates the use of performance metrics based on outputs, with ROI analysis required for all current and planned IT investments. The DoD (2017) Risk Management Guidance Defense Acquisition Guidebook requires that alternatives to the traditional cost estimation need to be considered since cost models do not adequately address costs associated risks across complex projects. This admission by the DoD and GAO reinforces the weakness of current risk-based methods with regard to understanding project uncertainty, requiring new methods of forecasting. Unfortunately, current risk-based methods, including real options methods, are still based upon underlying cost analysis techniques, in that risk management processes require mitigation steps that show cost potential of each step that is proposed to reduce the potential of the underlying risk.

Risk is the probability that an adverse or opportunistic event will occur sometime in the future. Risks are characterized as both the probability and consequence that this event will happen and how much it will cost if the risk is realized as well as mitigated (DoD, 2014). To better understand the overall risk of a project, the project is reduced into packages of work, each having some level of risk associated with its execution. The risk in each package of work is represented in terms of the cost and time that is required to mitigate the risk. These packages of work are commonly referred to as cost accounts within a project WBS (DoD, 2005). Following classical management theory, the WBS represents a hierarchical structure that is subsequently aligned within an organization to the various work units and cost account managers. The cost and schedule elements for each task are modeled and risk identified within the WBS to determine the total cost and schedule risk of a certain program.

The WBS represents a product-oriented grouping of project work elements shown in graphical display to organize and subdivide the total work scope of a project. A WBS is the cornerstone of effective project planning, execution, controlling, statusing, and reporting. All the work contained within the WBS is identified, estimated, scheduled, and budgeted from product start to first article delivery. The WBS is the structure and code that integrates and relates all project work with regard to scope, schedule, and cost. Therefore, the WBS contains the project’s scope baseline necessary to achieve the technical objectives of the work described. The WBS is used as a management tool throughout the life cycle of a project to identify, assign, and track its total work scope. The WBS allows the project team to manage a project within the framework of EVM, the principal management tool used to assess work performed relative to work planned (DoD, 2019). Within the project management profession, EVM is perceived to provide the project team with some level of predictivity and is the principal decision support system for project management. As we will see, however, forecasting within the EVM framework is based upon recursive cost data, and actually provides little insight into future project performance or more importantly, the intangible variables that influence a program’s stability and performance.

Earned Value Management

Current program management practices suffer from deficiencies in its theoretical base, creating self-inflicted problems in program execution in addition to hampering the effective professionalization of program management. The predominant metric by which acquisition projects are measured based on the belief that complex programs can be precisely planned and executed using a classically informed linear cumulative labor curve. The budget at complete (BAC) for a project is used as the objective by which resources are acquired and the project plan is developed and subsequently put-on contract to a competent developer. The closer the actual cost gets to the ubiquitous BAC determines the overall success of the program. As projects deviate from BAC, a new cost number to tracked and the variance off the BAC determines how well the project is performing. Decisions are made based upon current expenditures and future estimates (estimate at complete [EAC]) based upon preplanned remaining work. Koskela and Howell suggest that this approach is flawed and is the result of overly “narrow" project management theory based upon implicit linearity, which is used but rarely acknowledged (Koskela & Howell, 2002).

Earned value management is a method that allows the project manager to measure the amount of work actually performed on a project relative to the amount of work planned (DoD, 2019). EVM provides a method that permits the project to be measured by progress achieved. The project manager is then able, using the progress measured, to forecast a project’s total cost and date of completion, based on trend analysis. The relative relation to work performed to work planned is known as “Earned Value (EV)” (Reichel, 2006). Value in this sense should not be construed as value from an economic perspective. Economically, value is assessed as the ratio between revenue and cost. In this context, earned value is simply a misleading analogy that suggests that the allocated work on schedule and to cost results in some sort of value to the project. Earned value provides very limited insight into future performance and yet project managers rely on this process as their principal decision support system.

EVM is used for providing a disciplined, structured, objective, and quantitative method to integrate performance, cost, and schedule objectives for tracking contract performance. The term *earned value* is also defined as the “budgeted cost of worked performed” or BCWP. BCWP allows the project manager to calculate performance indices for cost and schedule, which provides information on how well the project is performing relative to its original plans. These indices, when applied to future work, are used by project managers to forecast how the project will do in the future, assuming the cost and schedule indices do not fluctuate. The basis of EVM is an accurate plan with completion rates and budgeted costs (Reichel, 2006). This plan begins with the program work breakdown structure, in which the estimated cost and schedule is allocated to the performance measurement baseline (PMB). Once the PMB is approved, this becomes the project Integrated Master Schedule (IMS) from which the project is measured. EVM would be most effective to monitor a stable process that has minimal complexity or uncertainty. It is not an effective forecasting method for complex programs such as hardware/software artificial intelligence or networking projects. Commonly used methods such as risk and EVM do not reveal causal chains responsible for management actions, or the socio-political complexities and interconnectedness within the project environment (Williams, 2018).

The assumption is that if project volatility is adequately managed, then costs can be controlled and therefore preplanned budgeting methods continue to remain a valid technique for embarking on projects regardless of their complexity. In other words, if we espouse the belief that a project can be decomposed into its smallest unit of value, then we can apply a cost and risk variable to these units and manage them accordingly. According to the PMBOK® Guide (the Project Management Body of Knowledge), project control is a “project management function that involves comparing actual performance with planned performance and taking appropriate corrective action (or directing others to take this action) that will yield the desired outcome in the project when significant differences exist.” The belief is that management and control keep projects on a predetermined track. The PMBOK does not, however, provide an explanation or insight into reliable forecasting; rather, it focuses on measuring planned performance relative to actual performance, accurate reporting of cost and schedule data, and updating cost information based upon trending information.

The tools and methods used by project teams are fundamentally grounded in classical and system theory that emphasizes planning and cost management rather than data driven forecasting of projects that are informed by complex and chaotic environments. With this understanding, the nature of the research problem becomes even more relevant. Our problem stated that, *current cost accounting and system thinking methods, based upon classical management and system theory, do not provide adequate forecasting of complex acquisition programs during the execution and control phase of the program life cycle.* This problem, exacerbated by the relative lack of project management theory that focuses on project forecasting and future performance, creates a wonderful opportunity to extend theory as well as potential opportunities for more advanced processes that will help project management grow as a profession. Up to this point, we have examined the theoretical foundation on which project management is based. This foundation in classical and systems management theory is reflected in the fundamental tools used to plan, resource, and execute acquisition projects designed to provide unique products to meet a customer's needs. We have also introduced the concept of network knowledge flow from which more robust theoretical positions can be claimed and methods introduced. In the next section of this literature review, we will extend this line of reasoning to explore possible physics-based applications that might allow use to develop theory and methods that will provide enhanced forecasting in complex and highly volatile project environments. Relevant readings and thought into how theories from other fields such as fluid dynamics and knowledge value might be used to provide improved forecasting models in project management based upon turbulent flow analogies will be introduced. We shall also examine the necessary relationship the field of information science has with project forecasting and turbulent flow theory.

Forecasting Project Management Performance

Project managers need new practices that incorporate data analytics and information from the complex and chaotic environments with which to make decisions. These practices should also be forward looking with intent to capture future events from a more deliberate and information-informed strategy, rather than extrapolation based predominately upon historical data. In her article "Making Fast Strategic Decisions In High-Velocity Environments" Kathleen Eisenhardt revealed that fast decision-makers use more, not less, information than slow decision-makers (Eisenhardt, 1989). Additionally, the greater number of alternatives considered simultaneously, the greater the speed of the strategic decision. Her research showed that executives immersed themselves in real-time information on their environment and the firm operations. The result of this, according to Eisenhardt, was a deep personal knowledge of the enterprise allowing for rapid decision-making. Consequently, the greater the speed of the strategic decision process, the greater the performance in high-velocity environments (Eisenhardt, 1989). Developing methods that incorporate structured and unstructured data into a more predictive decision-making process provides similar information immersion that Eisenhardt identified in her qualitative research on how key business leaders made effective decisions. The knowledge-based decision-making required these executives to consider the effects of the volatile environment and not just cost and revenue data. In effect, the successful leaders were able to quickly integrate and forecast based upon structured and unstructured data. New methods by which more information provides better insight into project integrity through better volatility management is necessary in order to improve project outcomes.

While the business community recognizes that project management is critical to providing a competitive advantage, the lack of program management theory seems to reveal a lack of theoretical thought in the profession of program management. A sound theoretical basis is critical to moving a profession forward, and an absence of theoretical explanation and curiosity has the potential of rendering a profession stagnant and task oriented. Defense department leadership and non-project management communities within the DoD still regard project management as a branch of operations management, and therefore believe that it is simply a series of management tasks that are structured within a system engineering construct where projects are instruments used to achieve higher-level organizational goals and objectives (Kwak & Anbari, 2009). Within industry, however, there is still a belief that project management helps create a strategic value chain that gives companies an edge on their competitors, particularly in high-risk sectors and markets. Being able to deliver projects on time and within budget often determines whether a company will get the next job or whether its new product hits the market. Without new theory, project management will reach a natural limit with regard to its effectiveness.

Complex products with many interdependent subsystems and high information processing load requirements can cause organizational failure. While there are many decision support systems that are designed to interpret historical project data, there is currently no way for managers to forecast when the demands placed upon a project have exceeded the risk tolerance level of the project team, resulting in projects moving from a somewhat stable state to a more volatile and turbulent state. Engineers use the Reynolds number in fluid mechanics as a metric that predicts whether the flow of a fluid will be smooth and stable versus turbulent and chaotic. Reynolds number is a dimensionless parameter widely used in fluid dynamics and is defined as the ratio of inertia force to the viscous force of a fluid. The concept of Reynolds number was introduced by George Stokes in 1851 in the development of the Navier-Stokes equation intended to solve the complex nature of turbulent flow (Anderson, 2017).

Fyall (2002) argues that it is possible to establish an organizational Reynolds number that estimates when a project approaches a turbulent region, alerting the project team to apply proactive measures before project turbulent behavior occurs. Using project workflow networks, subtask interdependency can be seen as exhibiting fundamental fluid properties exhibiting probabilistic behavior. Routine tasks, for example, may exhibit a higher probability of success, whereas more innovative and complex tasks could have a lower degree of success, suggesting that the knowledge flow between tasks may vary based upon their relative level of complexity. Establishing a metric similar to Reynolds number could provide insight into the nature of information flow in an acquisition project, allowing a project team to more reliably forecast if the information flow through an organization resulted in a stable predictable environment or one approaching turbulence as manifested in increasing project volatility.

Determining the tripping point at which a developmental program moves from a smooth predictable state to a chaotic and turbulent state is conceptually similar to predicting the turbulent flow characteristics across an airfoil in flight. By constraining the boundary conditions and identifying and measuring the multitude of dynamic variables within the project frame of reference, one can begin to define the nature of how these variables transition from a laminar (smooth) controlled state, to a turbulent state**.** The Reynolds number is the ratio of inertial forces to viscous forces and is a convenient parameter for predicting if a flow condition will be laminar or turbulent. It can be interpreted that when the viscous forces are dominant, they are sufficient enough to keep all the fluid particles in line, or in laminar flow (Batchelor, 1967, pp. 211–215). The ratio between viscous and inertia properties of a fluid establishes the relative volatility of the fluid. A low velocity fluid tends to be more stable than a high velocity fluid, indicating that there are increasing eddies being created as a result of specific environmental variables. As the fluid increases in velocity, failure to dampen or control these eddies causes the fluid to move from a laminar to a turbulent state (Anderson, 2017). Analogously, when a program is moving in a controlled fashion with little unpredictable external or internal variance from the original estimates, one could say a program is in a laminar state. However, as the program activity begins to increase in both speed and numbers of interdependent and unanticipated inputs, the program begins to move toward a turbulent and volatile state. However, stable business environments, which control for extraneous factors, could provide stability for acquisition projects, leading to less project volatility and a higher probability of success.

If we consider information flow in an organization to be analogous to fluid flowing through a pipe and if the capacity of the organization is large enough to manage and control the information flowing through it, then project stability is more likely. However, when an organization becomes overwhelmed and demand exceeds capacity, information flow throughout the organization becomes less controlled and more ambiguous, resulting in program volatility and greater risk of failure. Continued research is necessary to determine whether or not information flow through an organization changes from laminar to turbulent at a predictable point. Additionally, if an organizational Reynolds number exists, which variables and characteristics would define the capacity of the organization and the nature of the information being processed? If we are able to identify an analogous Reynolds number for an acquisition project, perhaps the project organization and strategy can be designed to produce the greatest amount of work in the most efficient way possible without risking chaotic information flow. Determining an organizational Reynolds number would provide enhanced understanding of the current EVM methods, allowing us to extend both project management practices and theory.

Method and Scope

This paper argues for a quantitative study that will use project data from EVM and selected acquisition reports (SAR) and correlate these data with Reynolds Averaged Navier Stokes (RANS) turbulent flow variables in order to better predict acquisition program performance. RANS is a mathematical technique that decomposes flow variables of the Navier-Stokes equations into their mean and fluctuating components. The Reynolds number is a dimensionless number used to categorize fluid systems in which viscosity controls the flow pattern of the fluid.

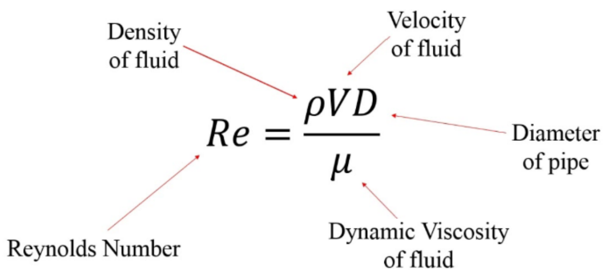
EVM is a project management methodology that integrates schedule, costs, and scope to measure project performance. Based on planned and actual values, EVM provides the project team an understanding of how well the project is performing but does not provide any real predictive power. The project is expected to perform according to the future plan if the project is within a reasonable variance of the project performance baseline.

Data Sources

Data will be obtained from the DoD Defense Acquisition Visibility Environment (DAVE). This data repository provides a consolidated source for programmatic data, including earned value management and selected acquisition report data. The DAVE database is a registered site that requires CAC entry. It can be accessed at: https://dave.acq.osd.mil/. Specific data that will be required for this research will be EVM, schedule, and budget data for acquisition category I (ACAT I) information technology programs. ACAT I programs are defined to have a research and development budget over $250 million. SAR data will provide the qualitative descriptions for these programs and will allow the researcher to develop a category that can be subsequently correlated to knowledge flow variables.

Variables

The Reynolds number helps predict flow patterns in different fluid flow situations. At low Reynolds numbers, flows tend to be dominated by laminar flow, while at high Reynolds numbers, flows tend to be turbulent. Reynolds number is a measure of volatility. For the purpose of this research the independent variables are velocity, distance, density, viscosity, and force. In addition to defining a Reynolds number for the project, the research will identify the project drag coefficient (Cd) by which Re will be compared and plotted, creating a Project "Moody" plot that reveals the predicted laminar and turbulent regions for the project.



Equation 1. Reynolds Number

$$C_D = \frac{2F_D}{\rho A V^2}$$

Equation 2. Drag Coefficient

The dependent variables for this research are the Re and Cd, which provide the foundation for project turbulent forecasting via the project "Moody" plots. Figure 2 represents a notional view of the project "Moody" plot that would clearly identify when a project begins to enter the critical zone between a smooth, well-managed project and a turbulent volatile project that could become unmanageable.

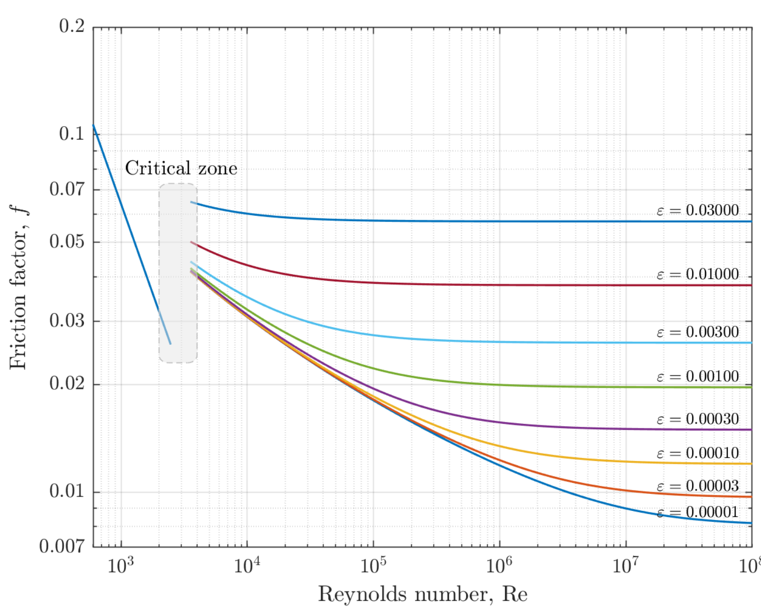


Figure 2. Sample Moody Plot

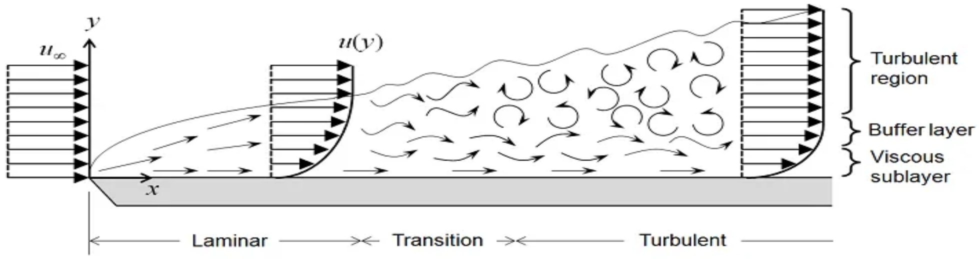


Figure 3. Flow Transition from Laminar to Turbulent Volatility Begins in Transition Region

Knowledge Flow

A necessary component of EVM is the work breakdown structure. Breaking work into smaller tasks is used to make the work more manageable and provides insight into the complexity of each task and its associated risks. The Project Management Institute (PMI) PMBOK defines the WBS as a "deliverable oriented hierarchical decomposition of the work to be executed by the project team." If work packages can be viewed as units of knowledge, then as work packages are completed, the flow of knowledge can be mapped across the project as the various tasks are completed. For the purpose of this research, a task within a WBS work package is considered to be the smallest unit of measurable knowledge, or value, within a project, allowing us to measure knowledge flow over time. Figure 4 represents a sample WBS with individual task units representing units of knowledge value.

Work Breakdown Structure

L 1	L 2	L3	L4	L5	L6	L7	L8	WBS #	Description
								1	Airborne and Maritime/Fixed Station (AMF)
1	1							1.1	AMF Joint Tactical Radio System PMP
1	1	1						1.1.1	Subsystem 1 (JTR)
1	1	1	1					1.1.1.1	Development Stations
1	1	1	2					1.1.1.2	JTR-M Unique
1	1	1	2	5				1.1.1.2.5	JTR M Subsystem Systems Engineering / Program Management
1	1	1	2	5	1			1.1.1.2.5.1	JTR-M Program Management
1	1	1	2	5	2			1.1.1.2.5.2	JTR-M Systems Engineering
1	1	1	2	6				1.1.1.2.6	HW1100 INFOSEC/Processor, Red, Dual
1	1	1	2	6	1			1.1.1.2.6.1	Pre-EDM HW1100 INFOSEC/Processor, Red, Dual
1	1	1	2	6	2			1.1.1.2.6.2	EDM HW1100 INFOSEC/Processor, Red, Dual

Figure 4: Sample Work Breakdown Structure, AMF JTRS

As work packages are completed, the knowledge flow of the project increases either positively or negatively depending on whether or not the tasks were completed consistent with the planned performance baseline. A project completed on time and within its budget profile is considered to reflect optimal knowledge flow with no viscous effects, while projects that deviate from their baseline are considered to have achieved less than optimal knowledge flow.

Using basic EVM metrics from existing information technology programs, these metrics will be compared to actual work progress as represented by the workflow plan in the project WBS. For example, the cost performance and schedule performance indices represent how well a project is accomplishing previously planned work. As work begins to fall behind, the CPI/SPI reflect the variance and the project's cost and schedule begins to deviate from the original plan. CPI and SPI calculations are represented by

Schedule Performance Index = Earned Value/Planned Value
Cost performance Index = Earned Value/Actual Cost

Where:

Earned Value = Budgeted Cost of Work Performed
Planned Value = Budgeted Cost of Work Scheduled

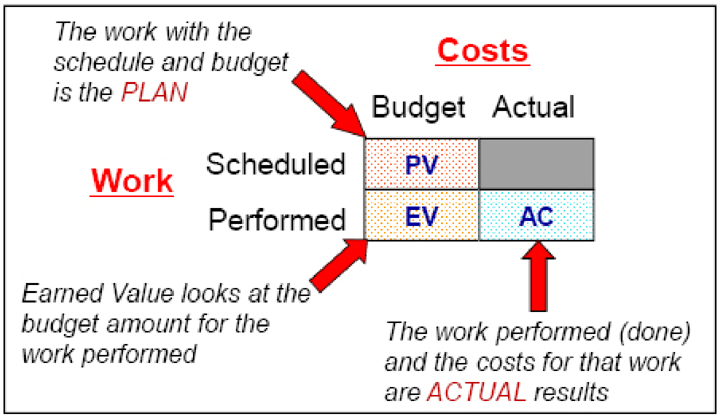


Figure 5: EVM Metrics (PMI.org)

The results of this comparison will represent a knowledge value (KV) variable by which knowledge flow can begin to be visualized. Since each task within the WBS represents a unit of knowledge, knowledge flow can be accurately determined throughout the project life cycle within the context of the accepted performance measurement baseline. Consequently, knowledge velocity (KVv) is the speed of actual knowledge value (KVa) relative to planned knowledge value (KVp). Additionally, KVv represents the property of the "fluid" environment within the developmental phase of the acquisition program.

Lexical Link Analysis

While characterizing the knowledge flow provides interesting insight and initial variable characterization, it is not sufficient to understanding future performance. Characterizing knowledge flow simply provides a visual into a moment in time with little understanding underlying nature of the knowledge flow state. In order to gain increased understanding into the nature of the knowledge flow, a method known as lexical link analysis (LLA) will be used to interpret the nature of the fluid environment at discrete points in time. This will begin to provide the researcher with a more coherent understanding of "why" the KVv is behaving in a particular manner. Lexical Link Analysis (LLA) is a form of text mining in which word meanings represented in lexical terms are treated as if they are in a community of a word network. LLA can provide automated awareness for analyzing text data and reveal previously unknown, data-driven themed connections (Zhao, 2016).

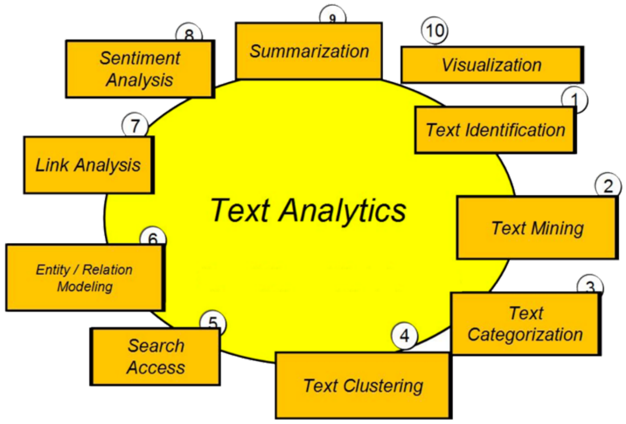


Figure 6: Graphical Representation of Lexical Link Analysis

Data for LLA will be obtained from the selected acquisition reports (SAR) from the projects under consideration for this research. A SAR is a comprehensive summary of a Major Defense Acquisition Program (MDAP) ACAT I program that is required for periodic submission to Congress by the Secretary of Defense. It's mandated by Title 10 USC § 2432 "Selected Acquisition Reports" (Acqnotes.com)

The SAR will provide the researcher with a rich understanding of the nature of the project's KVv by identifying key events through LLA that influence the nature of the knowledge value fluid. LLA can be mapped similar to scale free networks in which critical hubs that influence the network are easily identified.

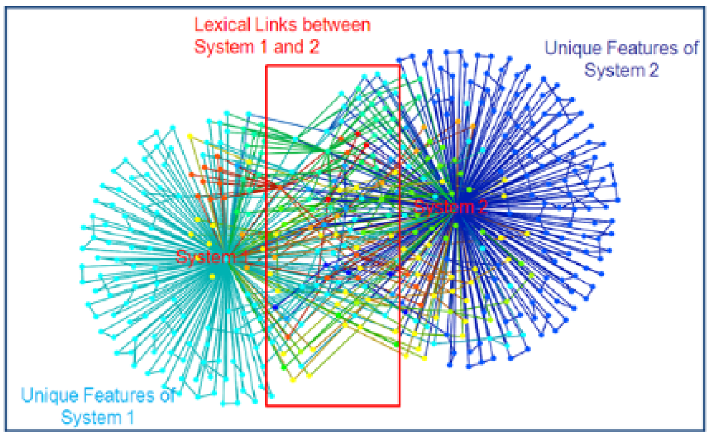


Figure 7: Nodal Relationship Using Lexical Link Analysis

Using the Pareto principle, which specifies that 80% of consequences come from 20% of the causes, the dominant hubs identified at each phase of the life cycle in conjunction with the KVv, key variables can be identified and chosen as candidates for inclusion into a project turbulence prediction model known as the Reynolds-averaged Navier Stokes (RANS) equation.

Finally, an ANOVA and a sensitivity analysis will be conducted to analyze how the different values of the independent variables affect the dependent variables at specific times in the project life cycle and will compare the variation among the means for the various programs studied. This analysis will improve our confidence in how accurately the turbulent flow advance EVM method measures what is intended to be measured.

Research Scope

The research scope will be limited to information technology programs that are in the developmental stage of their project life cycle. In the DoD acquisition framework, this is defined as being a program in the Post Milestone B phase known as the Engineering Manufacturing and Development phase. The EMD phase is the start of an official program. The purpose of this phase is the development of a capability. This phase starts after a Milestone B review and consists of two efforts, Integrated System Design (ISD) and System Capability and Manufacturing Process Demonstration (SC&MP) (Acqnotes.com). Figure 8 represents this phase graphically.

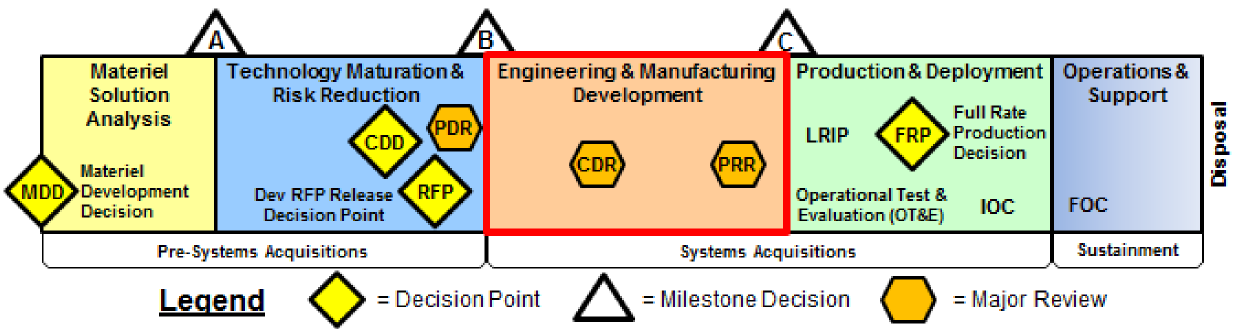


Figure 8: Defense Acquisition Life-Cycle Framework

The EMD phase is the phase in which EVM data is collected for a program in that the specific requirements are clearly delineated in a statement of work in the awarded contract. This phase requires that a performance measurement baseline be developed and EVM data collected for designated programs.

Potential Outcomes and Limitations

Two potential outcomes are the focus of this research. The first and perhaps most important is to extend current project management theory from a classical cost-based view of management into a more dynamic physics-based view that will more readily lend itself to improved forecasting and prediction. The second outcome is to provide a new model for project management forecasting and control that will improve decision-making for complex projects such as those typically found in information technology systems. Early analysis of this new fluid dynamics-based approach provides positive indications that the fundamental theory and method are sound. Notional data shows indications of increasing volatility significantly earlier than traditional EVM methods. Since this research methods uses traditional EM data and further refines the data with novel insight derived from LLA and fluid dynamics theory and methods, the methods for this improved measurement approach will be defined as Advanced Earned Value Management (AEVM).

The results of this research will be the basis upon which a library of Reynolds number "Moody" plots can be generated that will provide the initial planning tools by which complex projects can be architected. Reynolds number "Moody" plots reflect the fully developed flow of a fluid within the boundary conditions established for the project. By having these plots, projects can more accurately be structured and monitored for conditions that might create instability and turbulence or volatility resulting in projects that are difficult to manage.

Potential limitations for this project are data quality. The data for this research is available in multiple databases across the DoD, and obtaining access requires coordinating with multiple organizations willing to release program-specific data. Additionally, defining the independent variables will be based upon subject matter expertise and concurrence, leaving open the possibility for continued debate from external experts. This debate will subside as more data is used to validate the turbulent flow models and the library of Moody plots begins to grow over time.

Follow-on research can focus on moving from a more advanced form of project measurement developed in this research effort to application and policy development. As project volatility is better understood through this research, project managers will be able to more accurately management their projects with increased transparency in budgetary and technical decision-making. This increased transparency could provide the basis of fundamental reform in how budgets are determined and the flexibilities a project manager can be afforded with regard to financial and technical management.

References

Anderson, J. (2017). Fundamentals of aerodynamics (6th ed.). McGraw Hill

Batchelor, G. K. (1967*). An introduction to fluid dynamics.* Cambridge University Press.

Bertalanffy, L. (1968). *General system theory: Foundations, development, applications.* Penguin University Books.

Brotherton, S. A., Fried, R. T., & Norman, E. S. (2008). *Applying the work breakdown structure to the project management lifecycle*. Paper presented at PMI® Global Congress 2008—North America, Denver, CO. Project Management Institute.

Cicmil, S. (2006). Understanding project management proactive through interpretive and critical research perspectives. *Project Management Journal.*

Cleland, D., & Ireland, L. (2002). *Project management: Strategic design and implementation.*

Cleland, D., & Ireland, L. (2007). *Project manager's handbook*. McGraw Hill.

Cohen, D., & Prusak, L. (2001). *In good company: How social capital makes organizations work.* Harvard Business School Press.

DoD. (2017). *Risk management guidance defense acquisition guidebook*.

DoD. (1998). *Work breakdown structures for defense materiel items*.

DoD. (2019). *Earned value management implementation guide*.

Eisenhardt, K. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal*, *32*(3), 543–576.

Fayol, H. (1930). *Administration general and industrial administration*. Pitman & Sons.

Fyall, M. (2002). *When project information flow becomes turbulent: Toward an Organizational Reynolds Number* (CIFE Technical report no. 138). Stanford University.

GAO. (1997). *Assessing risks and returns: A guide for evaluating federal agencies' IT investment decision-making*.

Hayes, A. (2020). *Real option business essentials*. Investopedia. https://www.investopedia.com/terms/r/realoption.asp

Housel, T., & Mun, J. (2020). Acquiring artificial intelligence systems: Development challenges, implementation risks, and cost/benefits opportunities. *Naval Engineers Journal*, *132*(2).

Ika, L. (2009). Project success as a topic in project management journals. *Project Management Journal*. https://doi.org/10.1002/pmj.20137

Judea, P. (2018). *The book of why*. Basic Books.

Kauffman, S. (1993). *The origins of order: Self-organization and election in Evolution.* Oxford University Press.

Kwak, Y., & Anbari, F. (2009). Analyzing project management research: Perspectives from top management journals. *International Journal of Project Management*, *27*(5).

Koskela, L., & Howell, G. (2002, June). *The underlying theory of project management is obsolete*. Proceedings of PMI Research Conference, Seattle, WA. pp. 292–302.

Levitt, R. E. (1996). The virtual design team: A computational simulation model of project organizations. *Computational and Mathematical Organization Theory*, *2*(3), 171–196.

Marouf, L., & Doreian, P. (2010). Understanding information and knowledge flows as network processes in an oil company. *Journal of Information Management*.

McGregor, D. (1960). *The human side of enterprise*: Review by Joseph Lopreato. *American Sociological Review, 35*(3), 543–545.

Morris, P. W. G. (2000). *Researching the unanswered questions of project management.* Project Management Research at the Turn of the Millennium: Proceedings of PMI Research Conference, Paris, France. 87–101.

Newman, M., Barabási, A.-L., & Watts, D. J. (Eds.). (2006). *The structure and dynamics of networks.* Princeton University Press.

Packendorff, J. (1995). Inquiring into the temporary organization: New directions for project management research. *Scandinavian Journal of Management.*

Pitagorsky. (2018). *Managing volatility: How to manage change in projects.* Project Insight. https://projectinsight.com/blogs/whats-new/managing-volatility

Prieto, B. (2015). Project management and management of large complex projects. *PM World Journal*, *4*(6).

O'Connor, S. (2020). *The project risk management process: Five tips for success*. https://www.northeastern.edu/

Rozinat, A., & Wynn, M. (2009). Workflow simulation for operational decision support. *Data & Knowledge Engineering.*

Swan, J., Newell, S., Scarbrough, H., & Hislop, D. (1999). Knowledge management and innovation: Networks and networking. *Journal of Knowledge Management*, *3*(4), 262–275.

Taylor, F. W. (1911). *The principles of scientific management*. Harper & Brothers.

Reichel, C. W. (2006). *Earned value management systems (EVMS): "You too can do earned value management."* Paper presented at PMI® Global Congress 2006 North America, Seattle, WA. Project Management Institute.

Warburton, R. D. H., & Cioffi, D. F. (2014). *Project management theory: Deriving a project's cost and schedule for its network structure*. Paper presented at Project Management Institute Research and Education Conference, Phoenix, AZ. Project Management Institute.

Weber, M. (1930). *The Protestant ethic and the spirit of capitalism and other writings*. Penguin.

Williams, T. (2018). *The nature of risk in complex projects.* Program Management Institute.

Winslow, T. (1911). *The principles of scientific management.* Harper & Brothers.

Ylva, S., & Stehn, L. (n.d.). *Managing transformation, flow and value generation: A solid timber frame housing case*. Luleå University of Technology Structures.

Zhao, Y. (2016). Leveraging Lexical Link Analysis (LLA) to discover new knowledge. *Military Cyber Affairs*. https://scholarcommons.usf.edu/mca/vol2/iss1/3/

Is the Department of Defense a High-Risk Anomaly: Theory to Practice

Douglas A. Brook, Ph.D.—visiting professor in the Sanford School of Public Policy at Duke University. Professor Brook is also Emeritus Professor of Public Policy at the Naval Postgraduate School. [doug.brook@duke.edu]

Danelle R. Gamble, MPP—is a graduate of the Sanford School of Public Policy at Duke University. She is also a Major in the United States Army. [danelle.gamble@duke.edu]

Abstract

In phase one of this research, (Gamble, 2020, Brook 2020), DoD areas on the GAO's High-Risk List (HRL) were compared to similar longstanding high-risk non-Defense programs to determine if the DoD is a high-risk anomaly. Three attributes characterizing risk emerged: (1) the more technical programs have greater risk; (2) defense and national security areas have greater financial risk; (3) larger programs have greater, more prolonged risk. The study concluded DoD is a high-risk anomaly as the agency, and every area within, has two of the three attributes; but is not an anomaly in one as this attributes are present in non-Defense areas as well.

From this analysis questions emerged regarding whether the DoD can ever get off the HRL and whether the DoD should prioritize getting off the HRL. Additionally, we perceived a theory of interaction between the DoD, the GAO and Congress that would influence the answers. To explore these questions, we examined the interaction theory and its practice, and interviewed over twenty current and former officials and staff experts in the GAO, the DoD and on Capitol Hill. The analysis amplifies the earlier findings and reveals mixed and inconclusive views on whether the DoD could ever get off or should even prioritize getting off the HRL in its management agenda.

Introduction

This research report is a continuation of research conducted for presentation at the 17th Annual Acquisition symposium exploring DoD management and specifically, the persistent presence of DoD areas on the Government Accountability Office's (GAO) High-Risk List (Gamble 2020). Efforts to improve DOD management are persistent. Brook (2020) reports that proposals and initiatives come from a variety of internal and external sources, including internal sources such as politically-appointed leaders including the Chief Management Officer and internal advisors such as the Defense Business Board. External sources addressing Defense management include special commissions, presidential management initiatives, defense-related think tanks, and Congressional mandates.

The GAO, however, is arguably the source of the largest volume of studies, audits, initiatives and recommendations for Defense management. For instance, between 2014 and 2017 the GAO made a total of 1,122 individual recommendation to the DOD, averaging 280 per year. Of those reports, the GAO's High-Risk List stands alone in that it is a biennial report which can be used to track the persistence and prominence of major management issues over time. It also can reveal historical changes in the GAO's approach.

These GAO reports also provide information to the Congress. They can review and recommend actions that the Congress have taken or could take to support DoD management reforms. The GAO reports are used for congressional oversight of DoD management through committee hearings and they sometimes pique the interests of individual members with constituency concerns over the topics addressed.

Review of Prior Research

In phase one of this research, we considered whether DoD's persistent presence on the GAO HRL classifies the Department as a risk anomaly. As of the 2017 High-Risk Update, DoD owned half of the high-risk areas that remained on the list since its inception in 1990. The Department's efforts since 1990 have resulted in removal of only two areas: DoD Supply Chain Management recently removed in the 2019 report after 29 years on the list; and DoD Personnel Security Clearance Program. But, as of the 2019 HRL update report, the DoD still owns half of the areas that have been on the list for more than two decades. The Department's persistent presence on the list brought to question whether the DoD is a risk anomaly. Utilizing comparison analysis of longstanding DoD high-risk areas (DoD Contract Management and DoD Weapons Acquisitions) with comparable longstanding non-DoD high-risk areas (DoE Contract Management and NASA Acquisition Management), we determined that DoD is and is not a high-risk anomaly. The analysis revealed three high-risk trends that assist in answering the question: (1) the more technical the program, the greater and more prolonged the risk; (2) association with defense and national security lends to greater financial risk; and (3) the larger the program and portfolio, the greater and more prolonged the risk (Gamble 2020). The DoD meets the anomaly qualification in that its high-risk areas are always related to defense and the DoD will always be a large organization managing large and costly portfolios. However, we found that "DoD is not a high-risk anomaly in that it is not the only agency that succumbs to these high-risk trends" (Gamble 2020). For instance, the DoD is not the only HRL-listed agency that deals with highly technical matters or matters related to defense that appear to increase their propensity for risk. So, by this analysis, the DoD *is* and *is not* a high-risk anomaly.

These findings led to addition questions. What is the theoretical framework that appears to undergird the HRL?[1] Does this apparent framework actually work in practice? Is it plausible to think that the DoD will ever get off the HRL? Should getting off the HRL be a management objective for the DoD's chief management officer? How do the DoD and GAO view this list and the roles they play?

Methodology

To explore these and other questions, we first undertook an extensive documentary search consisting of GAO reports and Congressional testimony. We then gathered the observations, opinions and viewpoints of subject matter experts (SMEs). We analyzed the data from these documents and interviews to draw a more complete picture of how the theoretical framework operates in practice and to understand the viewpoints of knowledgeable current and former officials and staff experts about the DoD's presence on the HRL.

We identified a small number of current and former officials and staff experts in the GAO, the DoD and on Capitol Hill. In each case we employed the *snowball* technique asking all the interviewees to identify others whom they recommended for interviews. In total, we conducted 18 interview sessions involving individuals with a total of a total of 27 organizational affiliations, interviewed either singly or in small groups from one office. The tally of interviewees by organization is shown below. Some interviewees had experience in more than one organization and are therefore counted in more than one category.

[1] This framework is the authors' analytical construct derived from this and prior research.

Table 1
Organizational Affiliations of Interviewees

GAO	13
DoD	8
Capitol Hill	2
Other Agencies	4

Due to restrictions required by COVID-19, all interviews were conducted remotely, most on Zoom and a few on conference call. All interviews were conducted on a not-for-attribution basis so as to encourage open and candid conversations. Thus, interviewees will be identified only as "SME #" with the date of the interview.[2] The interviews were recorded and transcribed. The transcripts were then coded and examined to identify common themes and particularly salient statements. Finally, the data was categorized and this categorization is reflected in the following discussion and analysis.

All the interviews were conducted as in-depth interviews, utilizing an interview guide from San Jose State University (SJSC 2020). The expressed purposes of the interviews were to generate ideas, test hypotheses from the phase one study, and gain insights based on expert opinion. A short list of questions was prepared for each interview and supplementary questions and neutral probes were employed to draw out the interviewees' responses. Each interview was closed with a catch-all question, such as: "is there anything else you would like to tell us?" To assure completeness and reduce interviewer bias, both researchers participated in the interviews.

Theory to Practice Construct

Although not mandated by Congress, GAO publishes the high-risk list "to keep the Congress up to date on areas needing attention" within the federal government (GAO 2000). Since 1990 the GAO's process for evaluating areas and making a high-risk designation has evolved. Ten years into the high-risk list program, the GAO developed rating criteria, as displayed in a five-point star, that shows each area's progress in leadership commitment, action plan, monitoring, demonstrated progress, and capacity. Multiple congressional testimonies and reports from GAO contend that GAO's framework for determining and subsequently addressing high-risk areas will result in removal from the high-risk list, provided each participant in the process takes the recommended actions. During phase two of this research, we sought to identify how GAO's theory plays out in practice with respect to the DoD. Interviews provided more detailed information on GAO's process and how each member of the process – GAO, Congress, OMB, and the agency (the DoD in this case) – experience it.

The GAO High-Risk List, in Theory

GAO publications and testimonies dating back to 1990 provide a robust account of what the HRL is, why it exists, and how the program is intended to function. An agency of the Congress, the GAO is the legislative branch's "investigative arm" and the HRL provides "marching orders" for congressional subcommittees on government oversight (Cummings 2019; Jordan 2019). As the investigative arm, the GAO developed a Performance and Accountability Series that focused on "major program and mission areas" within each federal agency that meets certain criteria (GAO 2000). This Series is managed by the GAO's Strategic Issues Team

[2] Interviews with each SME took place only once. The full and proper citation with the date of the interview is included in the first reference to each respective interviewee. For ease of reading, in subsequent references for each interviewee, the full citation is omitted and only the SME# is identified.

which further designates program or mission areas as high-risk when they meet quantitative and qualitative criteria as outlined in our report from phase one. A new addition to the HRL is never a surprise, whether added to the HRL as a result of ongoing work that rises to the designation of high-risk, such as the recent addition of Veteran's Affairs Acquisitions added in 2019, or added as a new area like The Year 2000 Problem.

The GAO's commitment to the high-risk program is based on the belief that it is the mechanism needed to address the inaction of federal agencies to correct deficiencies and problems within management (GAO 1990). However, during the course of research it became apparent that the success of the high-risk program is predicated on underlying assumptions and the interplay between participants. One of the underlying theoretical assumptions the program operates from is that each area on the list can successfully be removed with improvements to agency management of the area. The GAO maintains this assumption despite more than a third of the original 14 areas remaining on the HRL after three decades. To identify why success has been illusive for most DoD high-risk areas, we must first understand the theoretical framework that the GAO utilizes. To do this, we identified what we've classified as key principles and assumptions, and examined the expectations of interplay amongst participants.

Key Principles and Assumptions

During the course of research, we identified several key principles that the GAO applies to the high-risk program. The first principle we identified is that the HRL is the mechanism to bring about change within the federal government to reduce agency fraud, waste, abuse, and mismanagement in high-risk areas. The second principle we saw is that risk does not have to be eliminated in order to achieve success. Rather, the GAO views success as risk management that results in a reduction in vulnerability to fraud, waste, abuse, and mismanagement (personal communication December 17, 2019). Our third observed principle is that the GAO's framework relies on transparency and cooperation between participants. The GAO's work is only as good as the information it receives from agencies, the OMB, and the Congress. Lastly, the GAO high-risk program and framework demand a non-partisan approach aimed solely at reducing risk.

In addition to these key principles, the GAO also employs three key assumptions in the high-risk program framework. First, the GAO assumes it is possible for all areas on the list to successfully get off the list. Secondly, the GAO assumes that the federal government should maintain each of the areas on the HRL and work towards removal from the list as opposed to eliminating the area. Finally, the GAO's framework assumes that agencies want to get off the list and will make an effort to do so.

Participants Theoretical Interplay

In addition to the principles and assumptions, the GAO high-risk program framework relies on an interplay between participants. The high-risk program involves four key participants: the GAO, the Congress, the federal agency in charge of the high-risk area, and the OMB. GAO's framework assigns roles and expectations to each of these with the aim of achieving success. These roles and expectations are best understood when considering how GAO envisions the interactions amongst the participants.

The GAO and the Federal Agency

As the creators and managers of the HRL program, the GAO expects to play an active role throughout an agency's tenure on the list. Utilizing core teams for each area, the GAO's Strategic Issues Team engages with agency points of contact for high-risk areas both prior to and during tenure on the HRL. By the time an area is added to the HRL, the GAO team is expected to be fairly familiar with the area and agency. Likewise, GAO expects the agency to have a minimum of one agency point of contact or team to communicate with GAO on the high-risk issues (personal communication May 29, 2020). The GAO's framework design expects the

GAO and agency teams to meet periodically to review the status of the high-risk area and any actions taken by the agency. The GAO's framework does not prescribe the frequency of meetings between the GAO and the agencies, but the GAO's best practices encourage agencies to have regular engagements with GAO teams. The GAO's best practices for HRL progress and removal focus heavily on the agency. Agencies must:

- Know the criteria for removal and how GAO is assessing them
- Develop action plans with measures and agreed upon goals
- Engage with GAO as actions taken
- Demonstrate sustained senior leadership commitment and build congressional support
- Ensure support and capacity to execute plans
- Demonstrate sustained progress, capable of spanning Administrations
- Coordinate on shared and government-wide areas (GAO 2019a).

Additionally, The GAO's dedicated teams and willingness to assist attest to its commitment to the high-risk program. The GAO routinely identifies and publishes best practices for high-risk and program specific issues, and encourages inter-agency crosstalk to share lessons learned.

The biennial high-risk updates prepared for Congress are integral to the high-risk program framework. GAO has set the precedent to supply agencies with a draft of the updates, along with an opportunity to provide feedback, before publication. Agencies are expected to provide feedback to ensure an accurate portrayal of progress is captured. In conjunction with the biennial reports, the GAO provides additional support to the agencies through detailed reports on specific issues within a high-risk area. These reports supplement the high-risk update and provide a more detailed account of each area. The GAO's essential expectation for interaction with agencies is that the agencies are willing to cooperate and work toward change.

The GAO and the Congress

For over twenty years, the GAO's framework has included issues requiring congressional attention and recommended congressional actions on the high-risk updates (U.S. Congress 2005). In a March 2019 House Committee on Oversight and Reform hearing, Comptroller General Eugene Dodaro testified that "where we've seen progress, congressional hand has been at play" (Dodaro 2019). The GAO has no authority in and of itself to mandate agency action, but as an agency of the Congress, the GAO's framework relies on the Congress's willingness to oversee the management of high-risk areas and encourage or even mandate action. The GAO's expectations of the Congress also include cooperation during preparation for the biennial report publication by providing feedback on drafted recommendations for Congressional actions. With the release of the biennial reports, the Comptroller General testifies before congressional committees with a team of experts to provide robust assessment and feedback on the state of the HRL. The GAO expects that Congress will continue to discuss and act on the recommendations for Congressional action.

The Congress and the Federal Agency

The high-risk framework includes interactions between the Congress and the agencies via congressional testimony, correspondence, NDAA mandates, and relationship building. Best practices from the GAO include engagement on the side of the agency and the Congress to ensure effective change. The framework also requires Congress to conduct the agency oversight required to ensure mandates are adhered to by the agencies.

The Office of Management and Budget

The GAO has long-held the view that the Office of Management and Budget, in the Executive Office of the President, needs to take a more active role in the high-risk program by designating an individual to lead the fight against fraud, waste, abuse, and mismanagement (U.S. Congress 2005). The OMB is seen as a key role player as well by Congress, with Senator Voinovich highlighting in a 2005 hearing that the President's Management Agenda has the ability to "create a confluence of ideas and synergy" that can lead to significant reform (Voinovich 2005). However, the OMB's role in the high-risk process was only a recommendation until late 2016 with the passing of the Program Management Improvement Accountability Act (PMIAA). The PMIAA is an example of Congress fulfilling expectations and acting on GAO recommendations. Specifically, the PMIAA mandates the OMB's Deputy Director for Management to review HRL area portfolios and receive recommendations from a Program Management Policy Council focused on improving management and high-risk areas (GAO 2019b). The theoretical GAO framework relies on the active participation of the Deputy Director of Management to mandate and oversee the implementation of recommended changes to high-risk areas. Additionally, the principle of transparency applies to the OMB. Between high-risk update issuances, the GAO provides the OMB with a draft of the HRL report and opportunity to provide feedback on the assessment and recommendations (U.S. Congress 2019). In order for the high-risk program to be successful, the theoretical framework requires OMB's active participation, transparency, and cooperation with GAO.

In theory then, with a synergistic approach by the GAO, the Congress, the DoD (agency), and the OMB, one could expect each evaluation criterion in the DoD's high-risk areas to be eventually satisfied and the areas removed from the HRL. However, five DoD areas remain on the list after two decades with progress fluctuating over the course of the tenure. We attempt to answer why by assessing how the theory plays out in practice.

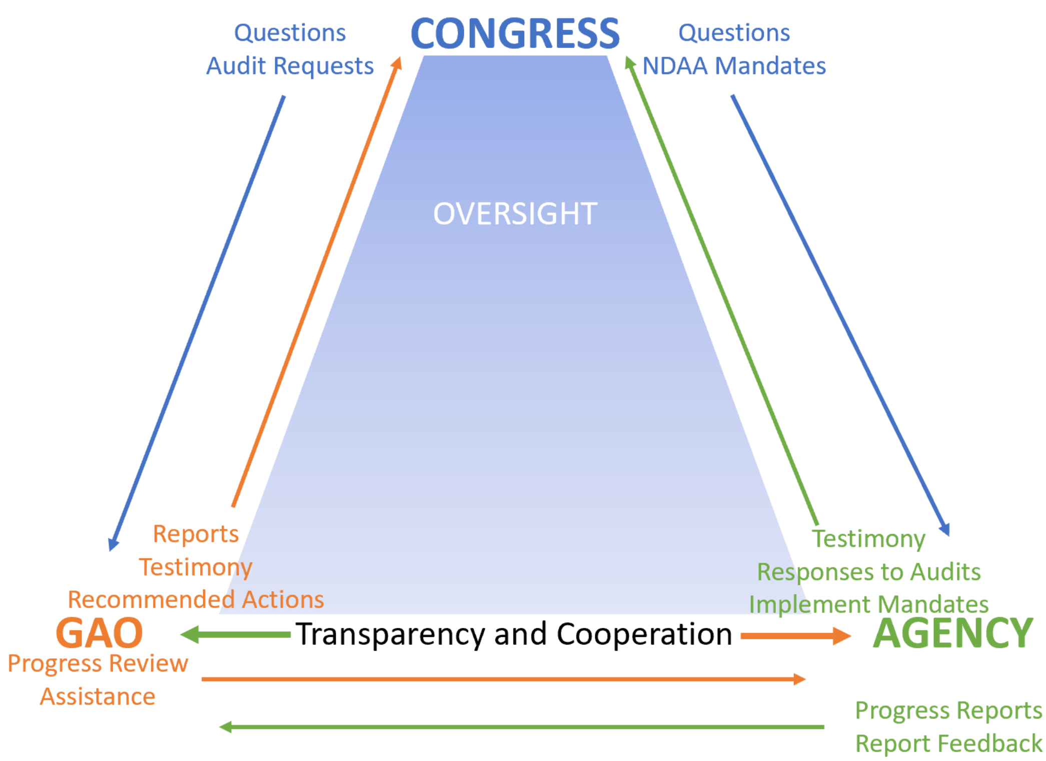


Figure 1

Observed High Risk List Theoretical Framework

The High-Risk List in Practice: Challenges to the Apparent Framework

From the GAO's perspective, the integrity of the framework is often compromised by congressional and agency inaction. GAO officials identified that the HRL is not of the necessary level of importance within the DoD to result in the required progress (personal communication October 28, 2019). Agencies that have seen great progress, such as the Department of Homeland Security, have done so by making the HRL an internal priority and increasing cooperation with the GAO (personal communication August 17, 2020). Across various high-risk areas, the GAO routinely sees as lack of implementation of best practices and recommendations. Agencies may list best practices in their action plan but fail to establish measures of accountability for their implementation.

The GAO, agencies, and the OMB all see leadership as a lynchpin to addressing risk. However, all three players acknowledge that leadership transitions, as often a result of political appointment, greatly hinder agency efforts (U.S. Congress 2005; U.S. Congress 2019). In addition to leadership longevity, much of the effort placed on the HRL by agencies and the OMB is driven by leadership priorities and personality. In 2019, the Comptroller General testified to Congress on the White House's unwillingness to cooperate with the GAO, greatly hindering GAO's assessment of cybersecurity and denying the GAO information regarding the Administration's actions to address issues of concern (U.S. Congress 2019).

From the agency perspective, the framework is unreliable because the GAO fails to meet expectations when it comes to expertise in agency operations. While the GAO attempts to understand the DoD's management practices and processes, a lack of understanding can lead to recommendations that are not feasible for execution.

The GAO's identification of issues requiring congressional attention is necessary, but the framework requires congressional action. While the HRL is an effective tool for Congress and a helpful reference, individual program reports (often mandated by Congress) often garner more attention and subsequent action (personal communication May 27, 2020). Overall, however, Congressional interest in oversight of agency reform is often limited to just a few members and occasional subcommittee hearings.

After considering challenges of executing the theoretical framework, we took the next step to explore the findings through interviews with subject matter experts. Our purpose was to test and validate the analysis and conclusions that were derived at this point in our research. We sought to determine how the theoretical construct of the HRL plays out in actual practice.

Theory to Practice – The Viewpoints of Experts

Data gathered from interviews, presented below, was categorized and analyzed based first on questions that arose from the phase one study and secondly based on topics which arose in the course of the interviews. The interview data fall into five identifiable categories of research questions: how is the theory of interactions between agencies, GAO, and Congress put into practice concerning the DoD; could the DoD ever get off the HRL; should the DoD prioritize getting off the HRL and, if so, what are proven strategies and best practices for making progress on HRL areas; and ultimately, is the DoD an anomaly?

Interactions Between Agencies, GAO and Congress

The interactions between the three major actors concerned with the HRL, the agencies, the GAO and Congress, reflect the roles each plays in public management. Concerning specifically the DoD's presence on the HRL, the DoD is responsible for the day-to-day and long-term management of the department; the GAO conducts management audits and provides evaluations and advice on management improvement; and the Congress provides oversight, receives reports, and can mandate management actions. Both the bilateral relationship between the GAO and the DoD, and the trilateral interaction between the GAO, the DoD, and Congress, were explored in interviews with current and former officials associated with each participant.

Bilateral Interaction between the GAO and the DoD

The foundational theory behind the HRL suggests that the three actors – the DoD, The GAO and the Congress play specific and complementary roles in a collective effort to reduce the high-cost risk of fraud, waste, abuse or mismanagement. Our interviews indicated that both bilateral and trilateral relationships exist in this framework. The most frequently observed relationship is that between the GAO and the DoD. We explored the nature of that relationship with SMEs from the two organizations. SME 2 (personal communication, July 1, 2020) said that in the early periods of the HRL the relationship between the GAO and the DoD was mostly adversarial, but other interviewees said the relationship today has more interaction and collaboration. For instance, SME 6 (personal communication, May 25, 2020), a GAO source, observed that DoD and GAO teams touch base several times a year to discuss high risk issues with senior DoD personnel and that GAO meets nearly weekly with mid-level personnel. SME 8 (personal communication, October 28, 2019) cited quarterly senior level meetings between the DoD and GAO and saw there was "a lot of good interaction throughout the year." SME 12 (personal communication, July 3, 2020), a DoD source, described reaching out to establish relationships with senior staff and instituting consistent ongoing dialogue.

Notwithstanding this current collaborative environment between the DoD and GAO, there are issues. One is the extent to which DoD believes the GAO understands its business operations. SME 2**,** a GAO official, said "We know DoD's business very well from an outsider's perspective. We will never say that we know the things from the inside." SME 9 (personal communication, June 15, 2020) a DoD official, claims that the GAO's findings often tend to corroborate what DoD knows internally but that sometimes GAO recommends actions that DoD probably cannot do; "the organization of the department doesn't always allow certain solutions for complex problems." The DoD does attempt to explain its operations to the GAO. SME 12 says the DoD does make an effort to explain the way they do business to the GAO. But, SME 8 is not quite so patient: "When it is just an academic exercise and people don't have the experience, DoD is not going to listen to them."

Bilateral Interaction between the GAO and Congress and between the DoD and Congress

A second set of bilateral relationships exist between the GAO and the Congress independent of the DoD and between the DoD and the Congress independent of the GAO.

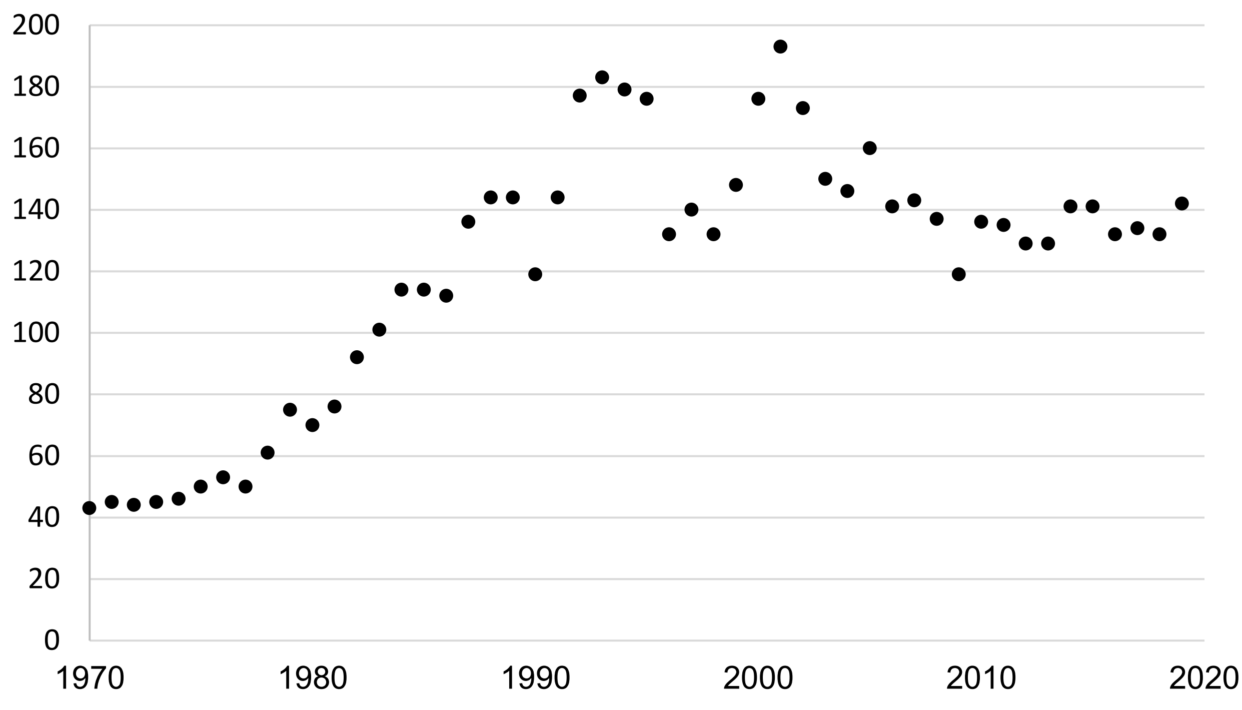
The GAO was created in the Budget and Accounting Act of 1921 as the General Accounting Office to be the principal financial analytical agency of the Congress. Over time its work expanded to include monitoring executive branch agencies' programs and spending. In 2004 its name was changed to the Government Accountability Office. The GAO is an agency of the legislative branch and reports to and takes direction from the Congress. (GAO 2020).

As an agency of the Congress, the GAO has a relationship with the Congress and key congressional committees well beyond the topic of the DoD and the HRL. GAO conducts other audits of the DoD, either self-initiated or requested by Congress. The GAO issues individual reports on these audits and its officials testify frequently before congressional committees. Between 2014 and 2017 the GAO made a total of 1122 individual recommendation to the DOD, averaging 280 per year. For this period, it identified 68 priority recommendations in the areas of acquisition and contract management (25 priority recommendations), readiness (14), financial management (11), health care (7), cyber security (5), headquarters management (3), support infrastructure (2), and information technology (1) (GAO, 2018, p. 25). 72% of the GAO's recommendations in this four-year period addressed management issues in categories both similar to and different from those found in the HRL. Any of these reports could become the topic of congressional oversight hearings or a source for legislation. Thus, GAO has much greater potential for interaction with the Congress on DoD-related topics other than the HRL. The biennial HRL, however, is always a subject of congressional hearings with the GAO upon release of the report.

The DoD was formulated with the creation of the National Military Establishment in 1947 and eventually as the Department of Defense in an amendment to the 1949 National Security Act (DoD, 2020). The organization and functions of the Department are codified in law in USC Title X. Like the GAO, the DoD has multiple interactions with the Congress outside the scope of the HRL. Many of the interactions deal with national security issues, such as the annual posture statement testimonies by DoD and the military departments, but many of them also address management topics. Congress directly addresses DoD management issues when it reviews funding and program authorizations through the annual Defense Appropriations Act and the National Defense Authorization Act, both of which can include mandates and requirements concerning Defense management. In fact, as shown in charts 1 and 2 below, Congress has been steadily increasing its organizational and management mandates through the general provisions of the annual appropriations act in the overall increase in the size of the annual authorization act.

Chart 1

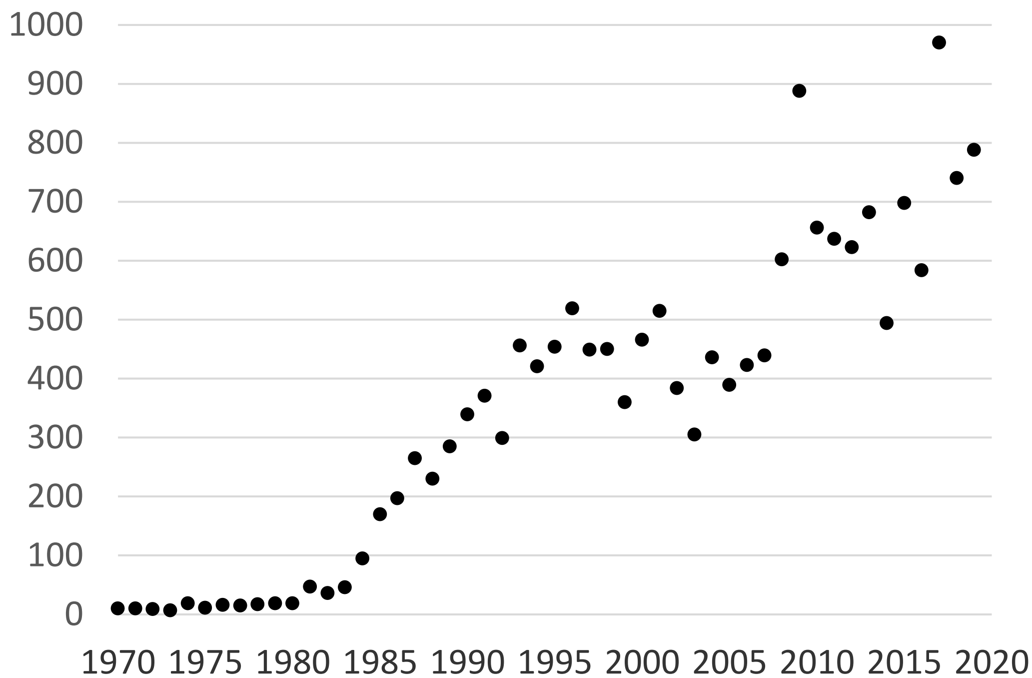
Number of General Provisions in the Defense Appropriations Acts, 1970-2019



Source: Candreva (2017), p.47.

Chart 2

Number of Pages in the NDAA



Source: Candreva (2017), p.49.

Trilateral Relationship Between the Congress, the GAO and the DoD

How then does Congress exercise oversight of the DoD and the GAO within the context of the HRL? SME 1 (personal communication, June 10, 2020) explained that the HRL assessment is not "sponsored" (i.e., directed) by Congress. Rather it is an initiative of the GAO (as contrasted with individual studies that are directed by or sponsored in Congress). Nevertheless, the entire biennial HRL update is always the subject of Congressional hearings. These hearings generally take place before the House Committee on Oversight and Reform and the Senate Committee on Homeland Security and Governmental Affairs. In addition, HRL areas specific to the DoD can be the subject of hearings before the respective House and Senate armed services committees. SME 4 (personal communication, November 12, 2019) said there were times when the Senate Armed Service Committee would hold a hearing on the HRL, but it was more likely for the committee to hold a hearing on a set of specific defense management issues arising from individual GAO reports where the HRL would come out in discussion.

What then is congressional interest in the HRL? SME 11 (personal communication, May 27, 2020) explained that some HRL-related areas have been included in the NDAA. Sponsors of the legislation will cite the HRL when they put something into the NDAA. In hearings, when GAO wants to make a recommendation, it would have some members of Congress reach out and ask about that recommendation and what they could put in the NDAA to fix it. Sometimes Members will have a political agenda that coincides with what is on the HRL and they will use the list to assert their agenda is credible because GAO is producing these reports. SME 9 identified that Capitol Hill interactions on the HRL depend on the broader congressional agenda, explaining that there is no separate callout in saying that "this is on the HRL," but sometimes there is a concern that aligns with the HRL. SME 6 thinks the HRL also directly influences the NDAA pointing out the 2019 HRL report acknowledges the involvement of Congress, saying "Congress also continues to take important actions. For example, Congress has enacted a number of laws since our last report in February 2017 that are helping to make progress on high-risk issues." (GAO, 2019b, p. 1) SME 11 noted that the HRL is mostly used as a "helper tool," not as a standalone product that the Committees look at. Instead, it appears the committees are more interested and responsive to individual GAO reports rather than the broader HRL.

Congressional interest, sometimes encouraged by GAO, can be used to motivate DoD to address areas on the HRL. SME 6 thinks the committee hearings have great impact if the Administration and Congress both support a recommendation. SME 10 agrees, saying when Congress holds hearings, they are well-attended and, in some cases, GAO's presentation of the HRL to Congress gives reason for DoD to act on its own internally-generated management reforms. Thus, both the GAO and the DoD can use Congress to leverage action on topics in the HRL but according to SME 4, the DoD doesn't come to the Senate to ask for legislation to help it in dealing with a high-risk program just because it's on the HRL.

Role of the Office of Management and Budget

The GAO framework envisions a significant role for the Office of Management and Budget. We explored this issue with SMEs knowledgeable about the OMB's involvement with the HRL. SME 7 said there were times when OMB tried to broker agreements on some management issues in sessions with the GAO and the DoD. SME 9 said it depended on who the "risk owners" were. The DoD owns the risks specific to it and OMB is more concerned with government-wide risks. SME 17 (personal communication, July 20, 2020), an OMB official, agreed that OMB's interests were in management themes crosscutting across agencies but said there was "no systematic connection between HRL and the OMB-managed President's Management Agenda." Nor has OMB analyzed what would be required to get off the HRL. In fact, the President's Management Agenda addresses only some of the GAO's high-risk issues.

In summary, it seems clear in the interviews that there is a good deal of HRL-related interaction between staffs of the GAO and the DoD, including periodic meetings at very senior levels. If the relationship between them was once adversarial it appears to be more collaborative today, though not without some friction points. Collaboration can vary by high-risk area and in some cases the collaboration amounted to periodic pro forma meetings, after which both the DoD and GAO returned to doing things their own way. There also appears to be significant bilateral interaction between the GAO and Congress, as one might expect. As an agency of Congress, the GAO reports regularly to the relevant committees about Defense management issues including the HRL. But the greater attention seems to be paid to individual GAO studies and reports rather than to the every-two-years HRL. The GAO can also leverage its position with Congress to motivate the DoD to address certain issues including those on the HRL. There seems to be less bilateral direct interaction between the DoD and Congress regarding the HRL outside of committee hearings and periodic meetings with congressional staff. Finally, there is scant evidence from the SMEs to suggest much productive trilateral collaborative interaction on HRL areas between the three entities.

Getting Off the HRL

A second theme that arises deals with the issue of getting off the HRL. This theme considers questions about (1) whether the DoD could ever really get off the list, and (2) whether the DoD's management agenda should prioritize getting off the list, and, if so, what successful practices could the DoD employ to make progress toward removal from the HRL.

Could the DoD Ever Get Off the HRL?

The long-tenure of some of the DoD's areas on the HRL raise questions about whether the DoD could ever get completely off the list. Some of the interviewees were skeptical, others were more optimistic. Virtually all of the GAO officials we interviewed expressed the opinion that the DoD could get its areas off the HRL. Some in the DoD, principally those who had been involved with a successful effort to get an item off the list were also somewhat optimistic. But most DoD officials we interviewed saw the DoD's HRL areas to be inherently risky and costly, and thus think the DoD will always be on the list. GAO officials, however, tended to acknowledge that risk exists in the DoD's areas and cannot be eliminated completely, but with proper management reform the DoD can come off the list. SME 4 acknowledges that good Defense management is important but that the HRL is largely not a list of solvable problems for the DoD. SME 7 agrees, saying everyone wants the DoD to be more efficient but the complexity of everything that DoD manages is going to bring risk.

On the other hand, SME 3 and SME 2, both of them GAO officials, are more positive. They point out that the DoD has recently had two high-risk areas removed from the list - personnel security clearances and supply chain management - and other complex HRL areas like terrorism and information sharing have come off the list. High-risk areas don't have to eliminate risk to get off the HRL, they need only to move from "high risk" to "risk". SME 9 a DoD official, somewhat concurs, observing that the Department's management agenda now is looking more closely at HRL areas in addition to individual GAO reports. So, there is a perceived difference in perspective here. While the GAO views getting off the list as managing the risk so that it is lower, the DoD perspective seems to some to mean eliminating the risk.

Should the DoD Prioritize Getting Off the HRL?

If there is at least disagreement over whether DoD could ever get off the list, the obvious next question is whether the DoD should prioritize getting off the list. This question involves both the probability of success and the incentives for trying. There is some disagreement among the SMEs interviewed about the extent to which the DoD should prioritize getting off the HRL in its management agenda. SME 5 makes the case for trying, saying that being on the HRL is fair but that getting off the list is not really central to the DoD management agenda. Both SME 10 a GAO official and SME 12, a DoD official answered simply "yes" to a question about whether DoD should be trying to get off the list, SME 12 adding emphatically, "I want to get off the list."

SME 8 believes the DoD just isn't going to prioritize the HRL. Instead, they are going to do things aligned with executing the national security strategy, prioritizing the mission over the HRL. Added to the negative argument now is the disestablishment of the Office of the Chief Management Officer in the FY 2021 NDAA. Some SMEs had pointed to the CMO as the logical responsible party to take ownership of the high-risk areas. Furthermore, neither the incentives nor disincentives clearly align with getting off the list. There seems to be little penalty for being on the list, except for some discomfort in occasional congressional hearings on the subject of the HRL. SME 16, a DoD official said "Congressional pressure has been the biggest" incentive to get DoD off the HRL. Conversely, SME 15, a GAO interviewee said "It should be in their (DoD's) interest because they could get more money." There is little indication however that Defense appropriations decisions are influenced by the HRL; SME 8 argued that the DoD did not receive any more money for recently getting areas off the HRL. Even though SME 15 said it is "embarrassing" to be on the HRL, there are few rewards for getting off the list aside from the intrinsic reward felt by individual managers whose areas make progress or get off the list. SME 8 said the only incentive was "personal satisfaction and that is it."

Successful Practices Toward Removal from the HRL

If DoD were to prioritize getting areas off the HRL are there successful practices that it might follow? Reflecting on the early days of the HRL, SME 2 said that at the time DoD contract management and weapons systems came onto the HRL, a lot of the decisions about what to put on the list were judgement calls, not very objective or methodical. In some cases, such as weapons systems acquisition, size and visibility were the major factors in getting on the list. SME-10 (personal communication, December 17, 2019), speaking from a GAO perspective, admitted that when the HRL began there were also no clear criteria for getting on or off the list. Now, however, the GAO's use of the five-pointed star rubric for measuring progress provide some credit for DoD efforts. SME 5 (personal communication, June 8, 2020)**,** a DoD official, doesn't disagree that the DoD belongs on the high-risk list, but just wants the GAO to "give us credit for where progress is being made."

And, in fact, the DoD has been successful recently in getting two areas off the HRL. SME 13 (personal communication, July 30, 2020) explained the DoD's process for getting supply chain management off the HRL. It essentially involved attacking each element of GAO's five-star evaluation system separately or sequentially. Initially the focus was on leadership, getting the two-star logisticians to establish a plan. Next came a comprehensive action plan and the dedication of resources. Metrics were established and periodic monitoring took place and policies were changed as needed. DoD was transparent throughout the process with the GAO in attendance for the DoD's internal meetings.

Goal setting and accountability are cited by SME 1 (personal communication, June 10, 2020) pointing out that the Department of Homeland Security has prioritized getting off the HRL within three years. They hold their executives accountable toward addressing the GAO's recommendations. SME 16 (personal communication, August 17, 2020) said Congress played a role in motivating DHS by requiring DHS to cooperate more with GAO. DHS thus realized over time that the HRL was going to be the cause of a lot of congressional pressure so the department's leadership there got pretty serious about it. With DHS's approach being mentioned multiple times as a model, the question arises: are the DoD and DHS analogous when it comes to the HRL? SME 3 answered "yes" but was quick to point out the difference in scale. For instance, both agencies have business systems modernization on the HRL but DHS's total IT budget was $9 billion while DoD's spending for business systems alone would exceed that amount. SME 16 however, thought DHS is just as complicated if not more than DoD. SME 16 went on to say that the argument that DoD is unique is convenient, getting off the list is just a matter of how much the leadership in each department wants to take it seriously.

Is the DoD making any progress toward getting more high-risk areas off the list? SME 6 currently sees a strong commitment in DoD to address high risk issues. There is more leadership commitment scored in the latest report and they are trying to create robust action plans. That is what is needed now. SME 17 advises that if DoD wants to prioritize getting off the list it should work to get specific language in support of those efforts into the NDAA.

In summary, the SMEs who participated in the interviews reflect mixed views of the questions about whether the DoD could get off the HRL and if it should even prioritize doing so. Strong arguments suggest that the DoD's size, complexity and the nature of its programs will always result on exceeding the GAO's threshold for high risk. Other equally powerful arguments, that an agency consuming half of the nation's discretionary budget has a responsibility to strive for better management and that success is possible as seen with the removal of two DoD areas from the HRL, prove that it should and can be done. In either case, it's clear that there are steps that the DoD could take to make progress on HRL areas, get credit for progress, and work toward getting at least some of them off the list.

Again, the question of getting off the HRL and whether the DoD should prioritize it seems highly influenced by the perspectives of DoD and GAO officials. The GAO perspective is that the DoD doesn't need to eliminate risk, just manage it to a lower, more acceptable level. But the apparent DoD perspective is that its programs are inherently risky because they are large in scale, high cost, and usually involve development of new technologies and new products. The perspective of those SMEs who were involved with areas that got off the list, saw it as possible and worth striving for. Perspective seems to be a material consideration in this question.

Is DoD an Anomaly?

To return to our original research question, is the DoD a high-risk anomaly? We end up in the same position as in the first phase of this research (Gamble 2020) as indicated in the views of three SMEs: SME 6 says yes, DoD is huge in scope and that makes it different; SME-12 (personal communication, June 30, 2020) disagrees, at least when compared to some other large complex agencies like DHS; SME 17 cautions that the DoD's scale may be a factor in GAO's capacity for analyzing defense management even though it devotes more of its capacity to national security than to any other area.

Conclusion and Observations

The answer to the question of whether the DoD is a high-risk anomaly remains ambiguous. The DoD is certainly characterized by the three main attributes of high-risk. But, so are some other agencies on the HRL. On the other hand, some DoD areas have managed to get off the HRL. When asked about the DoD's presence on the HRL, GAO, DoD and Congressional subject matter experts disagreed, with most GAO experts and a few DoD experts believing that it would be possible for the DoD to get off the list. More DoD experts thought the DoD could never get completely off the list owing to its size, complexity and high-dollar programs.

Our analysis concludes the theoretical framework of the HRL, with its expected roles and responsibilities for the GAO, the DoD, the OMB and the Congress is not fully functional in practice. It is true that there is cooperation and collaboration between the GAO and the DoD at both the functional and senior leadership levels that can lead to management improvement even if not achieving complete removal from the HRL. But the roles of the OMB and the Congress seem less active. The OMB's response to the PMIAA by citing the HRL in its circular A-11 doesn't appear to actually require the agencies or Deputy Director to do anything to further the goal of getting off the HRL other than to make a report. Similarly, the interest of Congress in the HRL seem only episodic and person- or constituency-driven. Interviews across the agencies suggest that the individual GAO audit reports receive more attention and prompt more corrective action than does the HRL. In the absence of tangible incentives there seems to be little in the way of motivation for the DoD to prioritize getting off the HRL. Instead, as more than one DoD SME argued, if the Department just works to improve its management policies and practices, making progress on the HRL and perhaps even getting off the list, would be a pleasant by-product.

Finally, it is clear that the HRL is an invention of the GAO. The HRL was not mandated by Congress nor does it have a particular congressional owner, either a committee or individual legislators. The GAO also lacks enforcement authority beyond persuasion and the public scrutiny that comes from biennial congressional hearings.

Ultimately, though the DoD is and is not a high-risk anomaly, whether or not it prioritizes making progress and getting off the HRL is a question that characterizes all of the DoD's management initiatives. The DoD will undertake to respond to the HRL when doing so directly serves the mission of the Department.

Works Cited

Brook, Douglas A. (2020). Persistent Perspectives on Defense Management Reform. In *Proceedings*, U.S. Naval Postgraduate School, 17th Annual Acquisition Research Symposium. Monterey, CA. May 13-14, 2020.

Candreva, Philip J. (2017). *National Defense Budgeting and Financial Management*. Charlotte, NC, Information Age Publishing, 2017.

Cummings, Elijah (MD). (2019) Quote from U.S. Congress: Hearing before the Committee on Oversight and Reform. House of Representatives. "GAO's 2019 High Risk Report." (Date: 3/6/2019). Retrieved from https://docs.house.gov/meetings/GO/GO00/20190306/109019/HHRG-116-GO00-Transcript-20190306.pdf, October 10, 2020.

Dodaro, Eugene. (2019) Quote from U.S. Congress: Hearing before the Committee on Oversight and Reform. House of Representatives. "GAO's 2019 High Risk Report." (Date: 3/6/2019). Retrieved from https://docs.house.gov/meetings/GO/GO00/20190306/109019/HHRG-116-GO00-Transcript-20190306.pdf, October 10, 2020.

Gamble, Danelle (2020). Is the Department of Defense a High-Risk Anomaly? An Analysis of the Government Accountability Office's High-Risk List's Persistent Residents. In *Proceedings*, U.S. Naval Postgraduate School, 17th Annual Acquisition Research Symposium. Monterey, CA. May 13-14, 2020.

Jordan, James (OH). (2019) Quote from U.S. Congress: Hearing before the Committee on Oversight and Reform. House of Representatives. "GAO's 2019 High Risk Report." (Date: 3/6/2019). Retrieved from https://docs.house.gov/meetings/GO/GO00/20190306/109019/HHRG-116-GO00-Transcript-20190306.pdf, October 10, 2020.

U.S. Congress (2005). Hearing before the Oversight of Government Management, the Federal Workforce, and the District of Columbia Subcommittee. "Programs in Peril: An Overview of the GAO High-Risk List." (Date: 2/17/2005). Retrieved from https://www.govinfo.gov/content/pkg/CHRG-109shrg20174/html/CHRG-109shrg20174.htm, October 10, 2020.

U.S. Congress (2019). Hearing before the Committee on Oversight and Reform. House of Representatives. "GAO's 2019 High Risk Report." (Date: 3/6/2019). Retrieved from https://docs.house.gov/meetings/GO/GO00/20190306/109019/HHRG-116-GO00-Transcript-20190306.pdf, October 10, 2020.

U.S. Department of Defense (DoD 2020). About the Department of Defense – History. Retrieved from https://archive.defense.gov/about/#history, December 3, 2020.

U.S. GAO (GAO 1990). High Risk: Letter to Congressional Committees Identifying GAO's Original High-Risk Areas. Retrieved from https://www.gao.gov/assets/660/659029.pdf, November 7, 2019.

U.S. GAO (GAO 2000). Determining Performance and Accountability Challenges and High Risks. Retrieved from https://www.gao.gov/assets/210/200448.pdf, June 15, 2020.

U.S. GAO (GAO 2020). History at a Glance. Retrieved from https://www.gao.gov/about/what-gao-is/history/, December 3, 2020.

U.S. GAO (GAO 2019a). Slide presentation. Making Progress on High-Risk Issues.

U.S. GAO (GAO 2019b) HIGH-RISK SERIES: Substantial Efforts Needed to Achieve Greater Progress on High-Risk Areas. Retrieved from https://www.gao.gov/assets/700/697245.pdf , January 8, 2021.

U.S. GAO. 2018. Status of GAO Recommendations Made to the Department of Defense (Fiscal Years 2014-2017). Washington, DC: GAO. Retrieved from https://www.gao.gov/assets/700/696753.pdf , January 3, 2021.

San Jose State University (SJSU 2020). In depth Interview Guide. Retrieved from https://www.sjsu.edu/faculty/weinstein.agrawal/urbp298_phI_handout_Interviews.pdf.

Voinovich, George (OH). (2005) Quote from U.S. Congress: Hearing before the Oversight of Government Management, the Federal Workforce, and the District of Columbia Subcommittee. “Programs in Peril: An Overview of the GAO High-Risk List.” (Date: 2/17/2005). Retrieved from https://www.govinfo.gov/content/pkg/CHRG-109shrg20174/html/CHRG-109shrg20174.htm, October 10, 2020.

Acknowledgements

The authors appreciate the substantial contribution in data organization and analysis provided by undergraduate research assistant, Kyle Melatti; and the institutional support provided by Michele Grow, senior grants and contracts administrator in the Duke Campus Grants Management Team.

Fire Scout: The Navy, Northrop Grumman, and Acquisition in Adversity: A Case Study

Amir Etemadi—received a Doctor of Philosophy degree in electrical engineering from the University of Toronto, Toronto, Canada, a Master of Science degree in electrical engineering from Sharif University of Technology, Tehran, Iran, and a Bachelor of Science degree in electrical engineering from the University of Tehran, Tehran, Iran. He joined the faculty at the George Washington University in 2013 where he is currently an Assistant Professor in the School of Engineering and Applied Science in the Department of Electrical and Computer Engineering and the Department of Engineering Management and Systems Engineering. He teaches master- and doctorate-level courses in electrical engineering and engineering management.

Etemadi's research interests include distributed generation, power system dynamics and control, quantitative risk analysis, and engineering management. He is the Principal Investigator on several current research projects, including a National Science Foundation project on geomagnetic disturbance impacts on power system operation, and an Acquisition Research Program grant on defense acquisition strategies. In addition, he is also a faculty graduate research advisor in the School of Engineering and Applied Sciences, and has supervised more than 50 masters and doctoral candidates for degrees in both electrical engineering and engineering management. Etemadi is a member of the Institute of Electrical and Electronics Engineers (IEEE). [etemadi@gwu.edu]

John Kamp—received a Doctor of Engineering in engineering management from the George Washington University in 2019, a Master of Engineering in nuclear engineering from Iowa State University, and a Bachelor of Arts in mathematics and French from the University of Nebraska–Lincoln. Kamp joined the George Washington University research staff in 2019 and is supporting Etemadi's acquisition strategy research project. He is a retired naval submarine officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

The Department of Defense is experimenting with how to deliver new capabilities in 2 to 5 years. Program offices recognize that they are dependent upon their contractors for successful delivery. The MQ-8 Fire Scout started in 1999 and achieved initial operational capability of the MQ-8C in 2019, after 20 years and effectively three program restarts. After each restart, the contractor developed, manufactured, and delivered a functional product deployed by the Navy within *5 years* of contract award.

Conventional wisdom says that senior leadership support and customer urgency are critical to fast delivery. This study shows how a program office and prime contractor were able to deliver a new capability despite changes in procurement objectives, evolving technologies, and requirements.

Results Statements

This research developed a case study from publicly available sources. This case study highlights how intangible assets such as prime contractor employee intellectual capital, goodwill, and a sustained corporate interest in strategically positioning for a future market sector are complementary to a program's acquisition strategy and essential to program execution.

Research Limitations/Implications

This research used publicly available data from budget submissions, program-related reporting, contractor annual reports, and contemporaneous press releases. The findings are specific to the Fire Scout program, and suggest that factors previously considered in industrial base arguments are relevant to rapid product delivery.

Disclaimer

This material is based upon work supported by the Acquisition Research Program under Grant No. HQ00341910004. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Department of Defense nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Introduction

The Department of Defense (DoD) buys products and services collectively described as a *capability*. Acquisition *strategies* are business plans developed by program offices and approved by senior leadership, containing a statement of need for the capability, estimated cost and schedule, and the contracting and support plans (General Services Administration, 2019, pt. 7). The rate of change of both technology and adversary capabilities is pushing the DoD and defense contractors to speed capability development and delivery. MDAP capability requirements such as maximum speed, endurance, and payload capacity change over time. Programs proceed in phases from program start through program decision and assessment gates[1] to initial operational capability (IOC).

This research was part of a 2019 Acquisition Research Program grant study. We developed a case study from the 1999 Vertical Take-off and Landing Unmanned Air Vehicle (VTUAV) development through the delivery of the MQ-8 Fire Scout in 2019. The development and delivery occurred within the context of a defense-unique market defined by the contractor and a government stakeholder. Major policy changes enacted during the Fire Scout program include the Weapons System Acquisition Reform Act of 2009 and the 2016 Section 804 Middle Tier of Acquisition[2] and had little effect on Fire Scout.

Contemporaneous news articles and press releases provide the context for Fire Scout development. Programmatic information is from contract award data from FPDS.gov and publicly released Selected Acquisition Reports.

The research examined *contractor acquisitions and teaming on program outcomes.* These affect not only market competition by changing the numbers of buyers and sellers, but also represent long-term contractor strategies faced with substitute goods, regulation, and peripatetic demand[3].

Background

Northrop Grumman History

Northrop Grumman has a long history of aviation innovation. The company designed the first flying wing bomber in the 1940s, produced lightweight fighters such as the F-5 and target drones in the 1960s, and the B-2 stealth bomber in the 1980s. Northrop acquired Vought in 1992 and Grumman Aerospace in 1994, bringing additional product lines such as the F-14 and E-2 into the company and consolidating aerospace market position[4][5].

---

[1] Examples include Engineering and Manufacturing Development, Critical Design Review, and Milestone C approval for production and deployment. Most but not all MDAPs have these gates as part of their acquisition strategy.

[2] For example, the 2016 National Defense Authorization Section 804 changes requires capability delivery within 5 years of program start to use these authorities.

[3] Porter (2008) discussed these as market forces.

[4] See https://www.northropgrumman.com/who-we-are/northrop-grumman-heritage/.

[5] For example, Vought was a B-2 subcontractor, with Northrop as the prime contractor. See: https://www.af.mil/About-Us/Fact-Sheets/Display/Article/104482/b-2-spirit/

Today's Northrop Grumman corporation resulted from a series of strategic acquisitions by the former Northrop Corporation beginning in 1992, when the company acquired a 49% interest in the Vought Aircraft Company, a designer and builder of commercial and military aerostructures. In 1999, Northrop Grumman acquired Ryan Aeronautical from Allegheny Teledyne for $140 million. Ryan was a small company of about 300 employees, but was one of the national leaders in Unmanned Air Vehicle (UAV) development, designing the U.S. Air Force Global Hawk and the DARPA Unmanned Combat Air Vehicle (Muradian, 1999).

With the Ryan acquisition, Northrop Grumman had product lines and revenue from operational Navy and Air Force drone target and reconnaissance systems, becoming the prime contractor (through Ryan) on the Global Hawk, the Miniature Air Launched Decoy and the Unmanned Combat Air Vehicle, and now had organic capability to design and produce unmanned air systems, and active development and delivery contracts (Kresa, 2001).

Northrop Grumman also acquired advanced electronics and radar expertise and radar contracts with the purchase of Westinghouse Electric Corporation's defense electronic systems group, and information systems expertise by acquiring Logicon, a leading defense information technology company. By 2007, Northrop Grumman had established themselves as a key supplier of UAV and defense electronic systems. Table 1 provides a summary of Northrop acquisitions.

Table 1. Northrop UAV-Related Acquisition Actions. (Northrop.com).

Year	Company	Action	Notes
1992	Vought Aircraft	Acquisition	Aerostructures manufacture
1994	Grumman Aerospace	Acquisition	Aircraft and Apollo Lander expertise, F-14 support
1996	Westinghouse Electric Corporation	Acquisition	Defense Electronic Systems Group, aircraft radar systems
1997	Logicon	Acquisition	Information technology and battle management systems
1998	Inter-National Research Institute	Acquisition	Command and Control, data fusion expertise
1998	California Microwave	Acquisition	Airborne ISR, mission planning
1999	Teledyne Ryan Aeronautical	Acquisition	UAV expertise
2000	Vought Division	Sale	Divest—metal structures production
2001	Aerojet General	Acquisition	Smart Weapons expertise
2007	Scaled Composites	Acquisition	Specialty composites and flight test expertise

These were part of and in response to the larger consolidation of the defense aerospace industry.[6] These acquisitions allowed Northrop Grumman to acquire not only production contracts, but also the tacit knowledge necessary for creating autonomous unmanned air vehicles (UAVs). Northrop Grumman acquired the remainder of Vought Aircraft in August 1994, and sold the metal structures expertise to Carlyle Group,[7] allowing Northrop Grumman to concentrate on composite structures.

[6] In 1999, Northrop Grumman noted the intense competition from this consolidation in its annual report to the Securities and Exchange Commission (SEC; Kresa, 2000).
[7] See https://www.thestreet.com/investing/stocks/northrop-grumman-to-sell-aerostructures-unit-to-carlyle-group-957901.

Merger and Acquisition Literature Overview

There is extensive literature on corporate mergers and acquisitions. Bertoncelj and Kavcic identify three types of corporate relationships, namely ad-hoc teaming, contractual, and ownership, and qualitatively characterize these relationships in terms of relative trust, protection, control, and learning (Bertoncelj & Kavcic, 2011). In any merger, be it two households or two companies, a fundamental issue is what to do with the combined assets. Anand identifies two approaches—redistribution or consolidation (Anand, 2004, p. 387).

Hensel (2016) argued that defense industry consolidations are a response to cost pressures on contractor workforce and facilities. Zullo and Liu (2017) described major contractor responses to budget reductions as expansion-merger, expansion-diversity, consolidation—focus on core or consolidation-specialization. Smaller suppliers, or subprime contractors, tended to develop strategies that exploited unique or proprietary capabilities or advantages (Zullo & Liu, 2017, p. 363). Jackson (2007) adds another strategy, that he calls strategic market positioning, where acquisition decisions are enhance an existing competitive advantage. He argues that this strategy drove the Northrop Grumman decision to buy Ryan Aerospace in 1999 (Jackson, 2007).

Brock (2009) considered the effects of consolidation on contractor economic power, arguing that concentration prevents economic efficiency gains passing to the buyer, suppresses innovation, restricts buyer choice, and protects sellers from market penalties for poor performance. His point is the market acts to limit the power of a single buyer or seller (Brock, 2009, p. 397).

A company has tangible and intangible assets. Tangible assets are physical property owned by a company, and used to produce a product or service. Intangible assets are non-physical assets (such as a patent) that create value for the company. Allen (2010) describes three classes of intangible assets—intellectual capital, intellectual assets, and intellectual property. In her model, intellectual capital represents the employee's inherent knowledge and skill; internal processes, methods, are formulations are intellectual assets; and intellectual properties are those intangible assets with legal protections for right of use (Allen, 2010).

Company annual financial reports reflect intangible assets as goodwill (brand recognition and loyalty) or purchased intangible assets (patents, trademarks, and such). Ievdokymov et al. (2020) showed that intangible asset value[8] is related to company market value and this relation is stronger for larger companies (Ievdokymov et al., 2020, p. 169). Bollen et al. (2005) used surveys to analyze the relationship between intellectual capital and company[9] performance. They followed Bonti and used three elements—human, structural, and relationship capital and showed statistically significant but indirect relationships between each intellectual capital element and aspects of company performance. In particular, they found intellectual property, *but not the supporting intellectual capital* significantly related to company performance (Bollen et al., 2005, p. 1180).

The DoD has a renewed emphasis on intellectual property (IP).[10] Defense acquisition programs are required to have a life cycle IP plan, codified in an new policy requiring intellectual property planning throughout a program life cycle for both for acquisition and sustainment, balancing DoD technology needs with fair intellectual property owner treatment (Lord, 2019).

[8] The conclusions are specific for Germany, France, and the United Kingdom.
[9] Conclusions are specific to the pharmaceutical industry.
[10] DODI 5010.44 defines IP as "Information, products, or services that are protected by law as intangible property, including data (e.g., technical data and computer software), technical know-how, inventions, creative works of expression, trade names."

A common belief is that vendors use intellectual property to extract economic rents and impose switching costs that cause the government not to compete work, called vendor "lock-in." Berardi and Cameron (2019) considered under what conditions software architectures and intellectual property rights favor vendor lock-in. They found that open software architectures are sufficient to prevent intellectual property or rights-based lock-in (Berardi & Cameron, 2019, pp. 69–72)

The Navy and Unmanned Helicopters

The DoD uses UAVs today for intelligence, surveillance and reconnaissance (ISR) missions and strike operations (USD AT&L, 2012). Unmanned air vehicles in the military were initially used as targets and as weapons and were derived from multiple enabling technologies, including propulsion, navigation, and controls (Erhard, 2000).

The Navy experimented in the 1950s with using shipborne helicopters for airborne anti-submarine warfare, but shifted to remotely piloted systems fielding the first maritime UAV weapons platform, called DASH,[11] in 1962. DASH operated from Navy ships until 1971 and were commonly used for naval gunfire spotting, surveillance, and torpedo or depth charge delivery (Gyrodyne Historical Foundation, 2020). The DASH control system suffered reliability and operability problems; controllability was so bad that few personnel could operate the system (Erhard, 2000, Chapter 7).

In 1998, the Navy posted a solicitation to develop an unmanned vertical takeoff and landing vehicle (VTUAV). A 1998 Navy press release named several teams intending to bid on the solicitation; one was led by Schweizer Aircraft Corporation (Lopez, 1998). At the same time, Northrop Grumman was acquiring Ryan Aerospace, a small 300-person company with more than 20 years' history in the development of unmanned aircraft. Northrop Grumman and Ryan Aerospace decided to submit a joint bid with Schweizer Aircraft on the VTUAV (Norris, 2003). Northrop Grumman and Ryan defined unmanned flight control as the critical development for program success. Northrop Grumman bought a Schweizer helicopter *in advance of contract awards* and worked with Ryan to develop a flight control system demonstrator (Norris, 2003).

The Navy canceled the program in 2002 and restarted the program in July 2003. Northrop restructured their efforts, moving prototype demonstrations to Patuxent River, MD, and redesigning the prototype for a more powerful engine (Norris, 2006). In July 2005, Northrop Grumman conducted flight and live fire testing at Yuma Proving Grounds (Paynter, 2005), and demonstrated autonomous landings on USS *Nashville* (LPD-13) at-sea in January 2006 (Staff Writers, 2006).

Northrop Grumman sought new customers and new roles for the Fire Scout, including the Army, Coast Guard, and foreign governments. The Army included Fire Scout in the Future Combat System, and procured eight systems before canceling the Future Combat System program.[12] Other prospects were interested, but generated no additional sales.

The program office and Northrop Grumman developed a plan to get early production systems to the fleet, which would accelerate identification and correction of design and employment issues. The Navy canceled procurement of five Littoral Combat Ships between 2006–2008 (O'Rourke, 2019), resulting in Fire Scout procurement deferrals (Smith, 2013). The Fire Scout program reached Milestone C in May 2007, allowing low-rate initial production. On December 10, 2008, Fire Scout embarked on USS *McInerney* (FFG-8) and deployed in 2009

---

[11] Drone Anti-Submarine Helicopter, manufactured by Gyrodyne Company of America.

[12] The Navy converted these systems to fit Navy specifications.

(Northrop Grumman, 2009). During the deployment Fire Scouts were used in intercepting drug smugglers (Evans, 2014).

In 2009, Defense Secretary Gates directed cancellation of the Future Combat System (Robinson, 2009). In 2011, Fire Scouts deployed with Helicopter Anti-Submarine Squadron Light (HSL) 42 Detachment 2 aboard USS *Halyburton* (FFG-40; HSL-42, 2011), and provided forward observation and targeting over Libya (Axe, 2020).

According to the 2012 Selected Acquisition Report (SAR) the inventory objective was 168 production systems by 2032 (Dodge, 2013). However, the same SAR shows a two year gap in procurement after ordering 23 systems, reportedly to align with Littoral Combat Ship (LCS) delays (Dodaro, 2013). The Fire Scout program saw user demand but procurement quantities decline.

Some Key Corporate Actions

Northrop Grumman had some corporate expertise in unmanned systems with their target drone product lines. They did not have extensive vertical flight experience, so they teamed with Schweizer Aircraft on the Vertical Takeoff UAV (VTUAV) solicitation. Northrop's acquisition of Ryan Aeronautics immediately them design and production expertise with the Global Hawk and Unmanned Combat Air Vehicle UAVs, but also experience with vertical takeoff and landing platforms. Combined with prior acquisitions of airborne sensor, command and control, and mission planning companies, Northrop had the experience within the company for adapting a commercial helicopter into a VTUAV system.

Northrop Grumman (Ryan) focused on the historical problem with remotely piloted systems—flight control (Erhard, 2000). They eliminated significant technical challenges by scaling to a proven commercial platform—the Schweizer 330—with known performance characteristics. Schweizer had already modified the helicopter and redesigned the transmission to accommodate a 420-horsepower turboshaft engine, and developed a 4-blade hub, allowing for future growth in lift capacity (Norris, 2006).

The program office supported Northrop Grumman's efforts and helped create new capabilities. The Navy made significant design choices, including adaptable interfaces and payloads, ground systems and datalinks (NAVAIR, 2000), a common launch and recovery system, and a ground control system that could operate from land or ships and feed existing data networks (Smith, 2013). The Fire Scout was re-designated as a multi-mission platform, and in July 2005, a Fire Scout fired rockets during flight testing at Yuma Proving Grounds (Paynter, 2005). The final variant had increased payload and sensor capacity (Soderberg, 2019).

Figure 1 shows the acquisition and teaming relationships Northrop Grumman continued with commercial helicopter manufacturers after acquiring Ryan, enabling an organizational balance of expertise between the platform and the unmanned system.

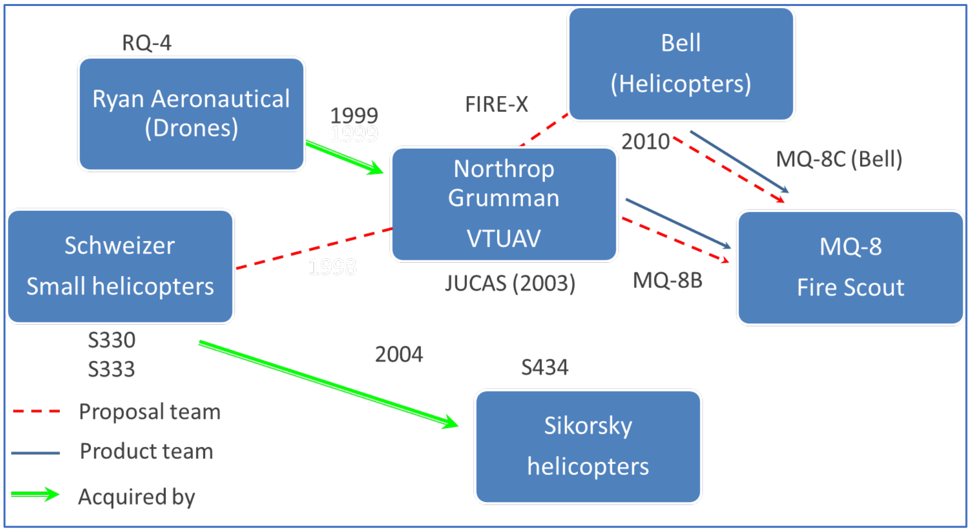


Figure 1. Fire Scout Teaming Arrangements

Findings

Fire Scout Budgets and Procurement Objectives

The Fire Scout budget profiles from 2000 to 2020 show the effects on research and development from program cancellation in 2002 and 2012 procurement cuts.[13] The two procurement budget phases in Figure 2 are related to conversion of Army-configured systems in 2008 and MQ-8C procurements in 2016.

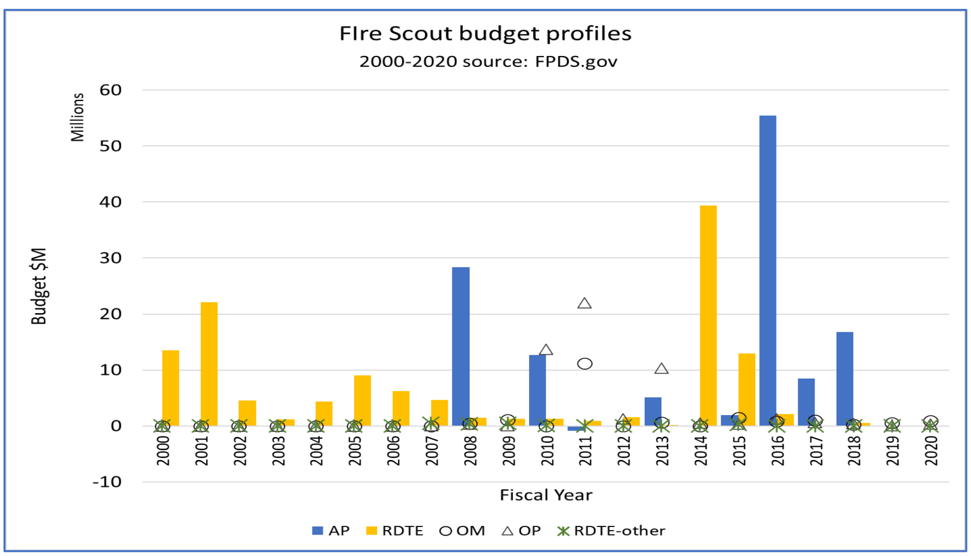


Figure 2. Fire Scout Budget Profiles. (FPDS.gov).

Contract work continued at about this same rate through 2020, in part due to the company continuing to develop new capabilities ahead of demand.[14] Notwithstanding the program office and contractor efforts, the program delivery was limited by funding, the realized

[13] Due to Future Combat Ship program reductions, which was to use the Fire Scout in mission packages.
[14] For example, the ability to launch weapons, the upgraded turboshaft engine system, and the adaptation of a new airframe to create the larger and more capable MQ-8C.

annual production capacity of about six systems per year, and the dependency on other acquisition programs.

Two events affected production quantities. The first was the Army's 2010 decision to cancel MQ-8 procurement plans with the Future Combat System termination. The second was the 2011 restructuring of the Littoral Combat Ship program and resulting delay of Fire Scout production. The reduction of both ship and mission module inventory objectives reduced required buys (Smith, 2013). The loss of Littoral Combat Ship demand contributed to a Nunn–McCurdy Breach (Smith, 2013). By 2019 the Navy inventory objective was reduced to 68 platforms (9 development and 59 production) as shown in Figure 3.

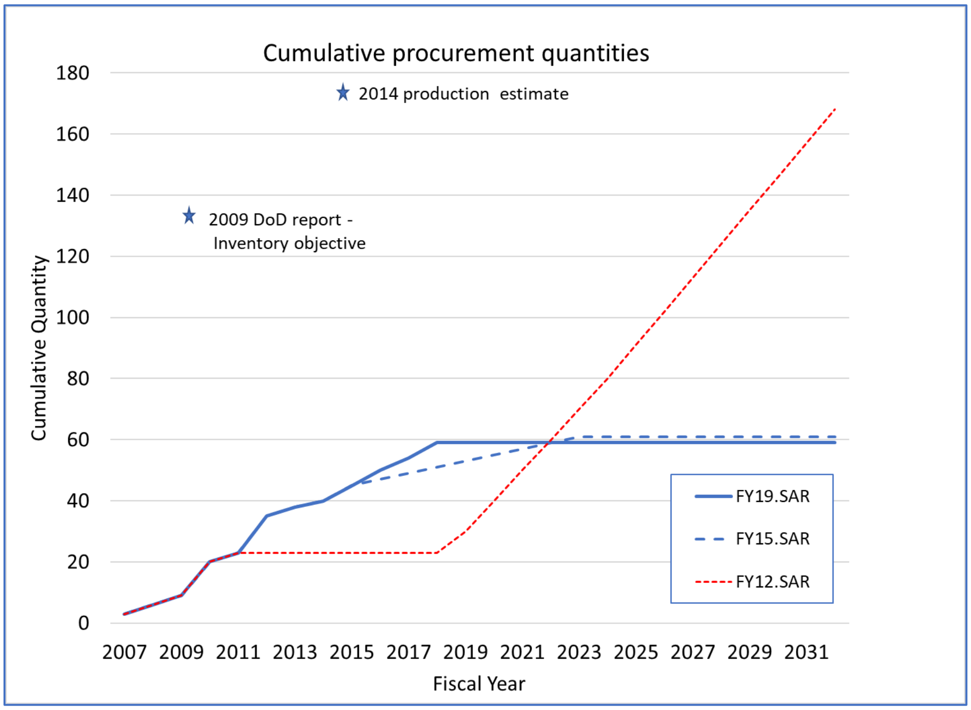


Figure 3. Cumulative Procurement Quantities. (MQ-8 SARs).

The final SAR showed required production quantities dropped from 168 in 2013 to 61 in 2015 (Soderberg, 2019). The red dashed line (2012 SAR data) shows the short-term result of the Future Combat System termination and Littoral Combat Ship procurement delay. The blue lines (2015 and 2019 SAR data) reflect the reduced procurement quantities (blue lines) to 61 plus 9 developmental units (Dodge, 2015) resulting from the decision to truncate Littoral Combat Ship procurements.

Northrop Grumman Impacts

Northrop Grumman is a diversified technology company, with multiple government and commercial customers. Contract award data between 1977 and 2020 shows obligations to Northrop Grumman in 200 different Product and Service Codes (PSCs). The Navy had the most overall contract activity with Northrop Grumman, in part due to Northrop Grumman shipyard ownership.

Figure 4 shows Northrop Grumman government-reported obligations for specific aviation PSCs by service.

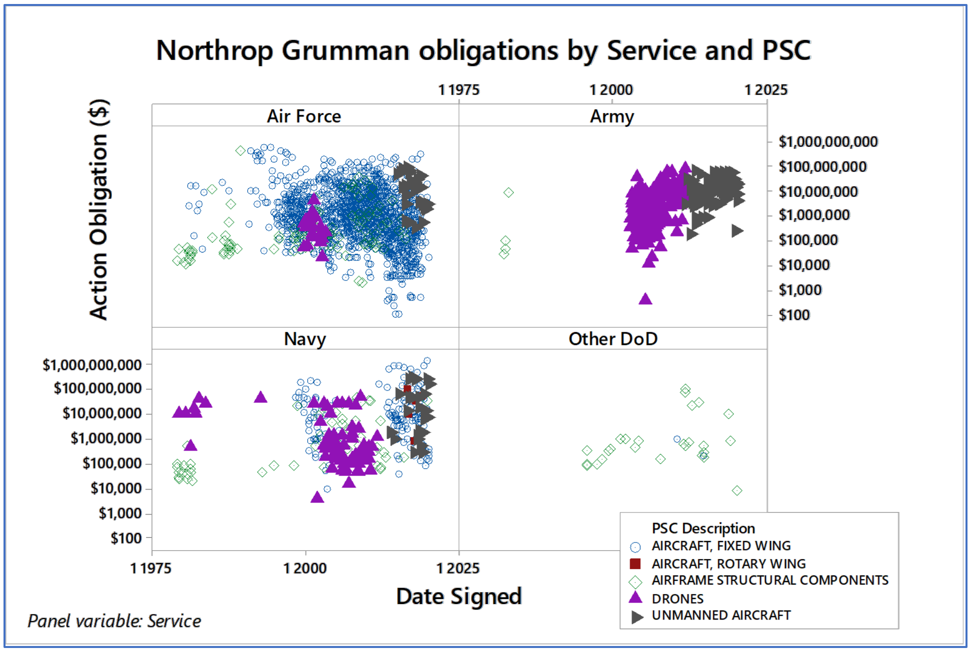


Figure 4. Northrop Grumman Obligations by Date Signed for Aviation PSCs. (FPDS.gov).

The figure highlights Northrop Grumman's base aircraft and airframe work activity, and shows the significant and increasing work in these areas in all services. While Northrop Grumman provided drone and unmanned systems and support to all services, the Navy had a long-term relationship covering multiple platforms, including Fire Scout early and later Triton (a Global Hawk variant). The drone and unmanned aircraft work (purple and grey triangles in Figure 2) became important after 2000.

The results of Northrop Grumman's strategy to establish a market position in unmanned air systems are shown by contract award data. Northrop Grumman earned twice the revenue on Fire Scout related contract work as other performers. The Ryan acquisition cost was recovered in 5 years and was a steady revenue source for the next 15 years as shown in Figure 5.

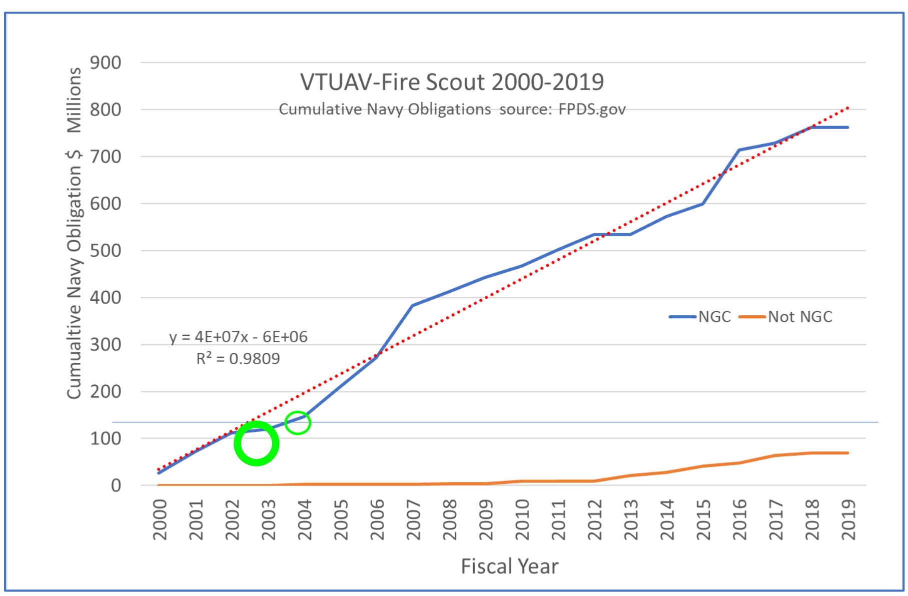


Figure 5. Fire Scout Obligations 2000–2020. (FPDS.gov).

In 1998 Northrop Grumman had one flagship aircraft acquisition program—the E-2 Hawkeye. By 2018, unmanned air systems had become major acquisition programs with annual budgets exceeding $1 billion per year. After years of strategic acquisitions and sustained activity, Northrop Grumman was one of the major unmanned air systems (the Global Hawk/Triton) manufacturers for the Department of Defense as shown in Table 2.

Table 2. Selected Major Weapons System Summary

System	Name	1998 ($M)	2018 ($M)	NOC
AH-64	Longbow Apache (C/D/E/reman)	609.2	1,441.9	Sub
E-2	Hawkeye (C//D)	374.8	1,116.4	*Prime*
F-18	Hornet (E/F) //Super Hornet	3274.6	1,253.1	Sub
B-2	Spirit	307.6	0	*Prime*
E-8	JSTARS	850.3	0	*Prime*
F-35	Joint Strike Fighter//Lightning II	909.1	10,837.9	Sub
V-22	Osprey	985.1	961.8	Sub
F-15	Eagle	274.8	963.1	Sub
F-22	Raptor	2,406.5	915.5	Sub
C-130	Hercules	0	886.1	Sub
P-8	Poseidon	0	1,609.4	Sub
MQ-1	Predator UAS	0	174.4	*
RQ-4	Global Hawk UAS	0	1,282.3	*Prime*
MQ-9	Reaper UAS	0	1,009.8	*
UAV	Smaller UAVs	0	129.7	*

Northrop Grumman was the prime contractor of four systems in Table 2—the E-2, B-2, E-8, and RQ-4. Only acquired programs were receiving production funding in 2018. Northrop Grumman's long-term acquisition strategy helped them remain a significant military aircraft producer.

Discussion

The Department of Defense is experimenting with how to deliver new capabilities in 2 to 5 years. Conventional wisdom says that senior leadership support and customer urgency are critical to fast delivery. This case study shows how contractor decisions and actions satisfying long-term strategic interests of the company affected program office success.

The program started in 1999 as the VTAUV and achieved initial operational capability as the MQ-8C in 2019, after 20 years of change and several programmatic restarts and changes. The Navy and Northrop Grumman were able to sustain program progress despite adversities such as program defunding in 2002, loss of a customer in 2008, and early operational deployments.

Northrop Grumman's long-term strategy to acquire a market presence in unmanned air systems aligned with the program office objectives. Their strategy ensured that the people with the experience and understanding to address the critical technical needs behind emerging government requirements. They anticipated government requirements evolution. Northrop Grumman and the program office kept finding new user bases and made critical adjustments to satisfy operational requirements. They transformed assets built for the Army Future Combat System into Navy assets, and deployed them for operational use. The sustained emphasis on operational use and future markets created confidence in system capabilities and identified new modifications meeting emerging needs.

The MQ-8 Fire Scout remains an unusual program. Navy program office and contractor tenacity and complementary objectives mattered in final program outcomes. This study shows how a program office and prime contractor were able to mature and deliver a new capability despite changes in procurement objectives, evolving technologies, and requirements. The Program Office benefitted from Northrop's experience and willingness to assume risk. Early field deployment and operational is a high-risk, potentially high-reward strategy. The Navy and Northrop Grumman had to respond when systems failed or were lost in combat (Axe, 2020). This required extraordinary dedication, but resulted in an extraordinary record of development and delivery—in the face of adversity.

References

Allen, K. (2010). *Entrepreneurship for scientists and engineers* (1st ed.). Prentice Hall.

Anand, J. (2004). Redeployment of corporate resources: A study of acquisition strategies in the US defense industries, 1978–1996. *Managerial and Decision Economics*, *25*(6–7), 383–400. https://doi.org/10.1002/mde.1197

Axe, D. (2020, March 3). This was the U.S. Navy's very first modern combat drone carrier. *The National Interest*. https://nationalinterest.org/blog/buzz/was-us-navy%E2%80%99s-very-first-modern-combat-drone-carrier-139267

Berardi, C., & Cameron, B. (2019). Intellectual property and architecture: New research on how to avoid lock-in. *Defense AR Journal*, *26*(1), 44–79. https://doi.org/10.22594/dau.18-803.26.01

Bertoncelj, A., & Kavcic, K. (2011). Hierarchy orientation in equity alliances: Core capabilities perspective. *Kybernetes*, *40*(5/6), 685–696. https://doi.org/10.1108/03684921111142241

Bollen, L., Vergauwen, P., & Schnieders, S. (2005). Linking intellectual capital and intellectual property to company performance. *Management Decision*, *43*(9), 1161–1185. https://doi.org/10.1108/00251740510626254

Brock, J. W. (2009). Merger policy and the Antitrust Modernization Commission: Economic power and the miasma of efficiency. *Antitrust Bulletin*, *54*(2), 337. https://doi.org/10.1177/0003603X0905400206

Dodaro, G. L. (2013). *Defense acquisitions: Assessments of selected weapon programs* (GAO-13-294SP).

Dodge, J. (2013). *Selected acquisition report MQ-8 vertical takeoff and landing tactical unmanned aerial vehicle Fire Scout (VTUAV)* (Selected Acquisition Report DD-A&T(Q&A)823–253). Department of the Navy. https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Selected_Acquisition_Reports/FY_2014_SARS/15-F-0540_MQ-8_Fire_Scout_SAR_Dec_2014.PDF

Dodge, J. (2015). *MQ-8 Fire Scout unmanned aircraft system (MQ-8 Fire Scout) as of FY 2016 president's budget* (Selected Acquisition Report No. 15-F-0540_MQ-8_Fire_Scout_SAR_Dec_2014). Naval Air Systems Command (NAVAIR). https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Selected_Acquisition_Reports/FY_2014_SARS/15-F-0540_MQ-8_Fire_Scout_SAR_Dec_2014.PDF

Erhard, T. P. (2000). *Unmanned aerial vehicles in the United States armed services: A comparative study of weapon system innovation*. The Johns Hopkins University.

Evans, M. (2014, February 7). *McInerney (FFG-8)*. Naval History and Heritage Center. https://www.history.navy.mil/research/histories/ship-histories/danfs/m/mcinerney--ffg-8-.html

General Services Administration. (2019). *Federal acquisition regulations*. http://farsite.hill.af.mil/reghtml/regs/far2afmcfars/fardfars/far/Far1toc.htm#TopOfPage

Gyrodyne Historical Foundation. (2020, June 20). DASH history—The DASH weapon system. *DASH History*. http://gyrodynehelicopters.com/dash_history.htm

Hensel, N. (2016). The defense industry: Tradeoffs between fiscal constraints and national security challenges. *Business Economics*, *51*(2), 111–122. https://doi.org/10.1057/be.2016.16

HSL-42. (2011, February 2). HSL-42 Det "Motley Two" poised to make history with Fire Scout UAV. *Florida Times-Union*. https://www.jacksonville.com/article/20110202/NEWS/801258041

Ievdokymov, V., Ostapchuk, T., Lehenchuk, S., Grytsyshen, D., & Marchuk, G. (2020). Analysis of the impact of intangible assets on the companies' market value. *Natsional'nyi Hirnychyi Universytet. Naukovyi Visnyk*, *3*, 164–170. https://doi.org/10.33271/nvngu/2020-3/164

Jackson, S. E. (2007). Creating value through acquisitions. *Journal of Business Strategy*, *28*(6), 40–41. https://doi.org/10.1108/02756660710835923

Kresa, K. (2000). *Annual report pursuant to section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended December 31, 1998* (10-K No. 1–3229; p. 149). Northrop Grumman Corporation.

Kresa, K. (2001). *Annual report pursuant to section 13 or 15(d) of the Securities Exchange Act of 1934 for the fiscal year ended December 31, 2000* (10-K No. 1–3229; p. 73). Northrop Grumman Corporation.

Lopez, R. (1998, December 22). Turning peanuts into gold. *FlightGlobal*. https://www.flightglobal.com/turning-peanuts-into-gold/24523.article

Lord, E. M. (2019). *Intellectual property (IP) acquisition and licensing*. Department of Defense. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/501044p.PDF?ver=2019-10-16-144448-070

Muradian, V. (1999, May 28). Northrop-Grumman to buy UAV-Maker Ryan for $140 million. *Defense Daily; Potomac*, *202*(42), 1.

NAVAIR. (2000, March 1). *Navy announces winner of VTUAV competition.* Navair.Mil. https://navairu.navair.navy.mil/

Norris, G. (2003, January 27). *Fire Scout takes flight*. FlightGlobal. https://www.flightglobal.com/fire-scout-takes-flight/46796.article

Norris, G. (2006, December 13*). Born survivor—An in-depth look at the Northrop Grumman MQ-8B Fire Scout vertical take-off and landing UAV*. FlightGlobal. https://www.flightglobal.com/born-survivor-an-in-depth-look-at-the-northrop-grumman-mq-8b-fire-scout-vertical-take-off-and-landing-uav-/71076.article

Northrop Grumman. (2009, May 5). *MQ-8B completes 2nd test period on USS McInerney (FFG-8)*. https://www.helis.com/database/news/mq-8b_ffg-8/

O'Rourke, R. (2019). *Navy Littoral Combat Ship (LCS) program: Background and issues for Congress* (No. RL33741; p. 25). Congressional Research Service. https://crsreports.congress.gov/product/pdf/RL/RL33741/238

Paynter, T. (2005, July 25). *Photo release—Northrop Grumman's Fire Scout UAV successfully fires test rockets* [Archive]. Internet Archive Wayback Machine. http://web.archive.org/web/20051124090308/http://www.irconnect.com/noc/press/pages/news_releases.mhtml?d=82646

Robinson, B. (2009, June 24). *FCS cancellation confirmed, Army modernization changes course*. https://fcw.com/Articles/2009/06/24/Army-future-after-FCS.aspx

Smith, P. (2013). *MQ-8 vertical takeoff and landing tactical unmanned aerial vehicle Fire Scout (VTUAV) As of FY 2015 president's budget* (Selected Acquisition Report No. 14-F-0402_DOC_76). Naval Air Systems Command (NAVAIR). https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Selected_Acquisition_Reports/FY_2013_SARS/14-F-0402_DOC_76_VTUAVDecember2013SAR.PDF

Soderberg, E. (2019). *MQ-8 Fire Scout unmanned aircraft system (MQ-8 Fire Scout) as of FY 2021 president's budget* [Selected Acquisition Report]. Naval Air Systems Command (NAVAIR). https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Selected_Acquisition_Reports/FY_2019_SARS/20-F-0568_DOC_60_MQ-8_Fire_Scout_SAR_Dec_2019_Full.pdf

Staff Writers. (2006, January 25). Autonomous Fire Scout UAV lands on ship. *Space Daily*. https://www.spacedaily.com/reports/Autonomous_Fire_Scout_UAV_Lands_On_Ship.html

USD AT&L. (2012). *Department of Defense report to Congress on future unmanned aircraft systems training, operations, and sustainability*. Department of Defense.

Zullo, R., & Liu, Y. (2017). Contending with defense industry reallocations: A literature review of relevant factors. *Economic Development Quarterly*, *31*(4), 360–372. https://doi.org/10.1177/0891242417728793

PANEL 12. FOLLOWING THE MONEY IN DEFENSE ACQUISITION

<table>
<tr><th colspan="2">Wednesday, May 12, 2021</th></tr>
<tr><td>10:45 a.m. – 12:00 p.m.</td><td>Chair: Major General Cameron Holt, USAF, Deputy Assistant Secretary for Contracting, SAF/AQ

Microeconomics, Competition and Major Defense Acquisition Program Cost
Kenneth McElroy, The George Washington University
John Kamp, The George Washington University

Do Accelerated Payments for DoD Contractors Help Small Businesses?
Justin Marion, Naval Postgraduate School
Jesse Cunha, Naval Postgraduate School

Assessing the Reliability of the Future Years Defense Program and Building a Forecast
Gregory Sanders, Center for Strategic and International Studies

Evaluating the Impact of Federal Improvement and Audit Readiness (FIAR) Compliance
William Lucyshyn, University of Maryland
Dylan Hunt, University of Maryland</td></tr>
</table>

Major General Cameron Holt, USAF—is the Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Technology and Logistics, Washington, D.C. He is responsible for all aspects of contracting relating to the acquisition of weapon systems, logistics, and operational support for the Air Force and provides contingency contracting support to the geographic combatant commanders. He leads a highly skilled staff of mission-focused business leaders supporting warfighters through $825 billion of Space, Global Power/Reach and Information Dominance programs. He also oversees the training, organizing and equipping of a workforce of some 8,000 contracting professionals who execute programs worth more than $65 billion annually.

Prior to this assignment, General Holt served as the Commander, Air Force Installation Contracting Agency, Office of the Assistant Secretary of the Air Force for Acquisition, Wright-Patterson Air Force Base, Ohio. He led an over 700 personnel agency with a total contract portfolio of $55 billion. In this capacity, he directed enterprise-wide installation strategic sourcing efforts for the Air Force and oversaw $9.1 billion in annual obligations for mission and installation requirements.

General Holt received his commission through the ROTC 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.

Microeconomics, Competition, and Major Defense Acquisition Program Cost

Kenneth L. McElroy Jr.—is a Staff Systems Engineer and Systems Engineering, Integration and Test (SEIT) Lead for the Northrop Grumman Corporation. He is a retired naval officer with 20 years of service as a Surface Warfare Officer and over 15 years of industry experience. He holds a Bachelor of Science degree in economics and general engineering from the U.S. Naval Academy, Master of Science degrees in physics from the Naval Postgraduate School and in systems engineering from The Johns Hopkins University, and a Doctor of Engineering in engineering management from the George Washington University. His research interests include engineering management and systems engineering. [kenneth.mcelroy@ngc.com]

John Kamp—received a Doctor of Engineering in engineering management from the George Washington University, a Master of Engineering in nuclear engineering from Iowa State University, and a Bachelor of Arts in mathematics and French from the University of Nebraska–Lincoln. Dr. Kamp joined the George Washington University research staff in 2019 and is supporting Etemadi's acquisition strategy research project. He is a retired naval submarine officer with extensive experience in research and development and program management. His research interests include engineering management, maritime systems, and acquisition system research. Dr. Kamp is a Fellow in the Royal Institution of Naval Architects and a member of several professional associations. [jckamp2018@gwu.edu]

Abstract

The Major Defense Acquisition Program (MDAP) market is a monopsony facing oligopoly. In the last four decades, the Department of Defense has placed a great deal of emphasis in its acquisition reform efforts on the power of competition to help control cost overruns and cost growth. In this research, quantitative analyses were used to determine the effect of two reform measures—competitive prototyping and competitive contracting—on cost overruns and cost growth during the Engineering and Manufacturing Development (EMD) phase of the acquisition life cycle. We performed a case study of 63 hardware MDAP contracts from all services. The findings show that while competitive prototyping and competitive contracting lead to greater competition, as the defense acquisition community believes, they fail to control cost overruns and cost growth, just as microeconomic theory predicts.

Introduction

Concerns over the increasing cost of U.S. defense programs are not new. In fact, according to Cancian (2010), these concerns date back as far as the earliest days of our republic when Congress, in its oversight role, began questioning the rising costs of the first naval ships, which it authorized in 1794 (Cancian, 2010). Since then, there have been numerous attempts at defense acquisition reform with the goal of controlling cost overruns and cost growth. Two of the most recent and significant acquisition reform initiatives are Congress's Weapon Systems Acquisition Reform Act (WSARA) of 2009 and a series of U.S. Department of Defense (DoD) initiatives championed by Under Secretary of Defense for Acquisition, Technology, and Logistics (AT&L) Frank Kendall, called Better Buying Power (BBP). Both the WSARA and BBP place a great deal of emphasis on the power of competition in controlling defense acquisition costs and in particular, the cost of Major Defense Acquisition Programs (MDAPs).

Despite decades of effort and frequent attempts to control cost overruns and cost growth in MDAPs, these latest initiatives and their predecessors have failed to achieve their objectives regarding cost (Ritschel et al., 2019). Defense acquisition reforms based on competition fail to control cost overruns and cost growth in hardware MDAPs. The question is why. The straightforward answer is microeconomics.

Extensive research exists studying defense acquisition program costs—both overruns and growth. Most studies focused on program management techniques and acquisition policies from the point of view of the government customer with scant attention to defense acquisition reforms from the perspective of industry and the market (i.e., microeconomics). This research uses quantitative analysis to provide insight into whether reform initiatives based on competition contribute to the reduction of cost overruns and cost growth in MDAPs.

Background

To meet the performance requirements, engineers must work within specific design constraints such as system weight, size, and shape, but ultimately systems are constrained by fundamental laws of nature such as the laws of motion, gravitation, and thermodynamics. Similarly, there are many constraints imposed on cost, but ultimately cost is subject to the laws of economics such as the law of supply and demand.

Summarizing from Edwin Mansfield's (1982) undergraduate-level textbook, *Microeconomics Theory and Applications*, economics is in two broad branches—microeconomics and macroeconomic. Microeconomics deals with how individual consumers, firms, and resource owners behave, while macroeconomics is concerned with the behavior of economic aggregates such as inflation, gross national product, and level of employment. In microeconomics, firms operate in markets that are a collection of buyers and sellers for a particular good or service, and the behavior of each market type can be modeled by a demand side and a supply side—the market structure.

According to microeconomics, there are four types of market structures: perfect competition, monopoly, monopolistic competition, and oligopoly. How the price (cost) of a good or service is determined is different depending on the market structure. In perfect competition, the equilibrium price—the price at which there is no tendency to change—is determined where the quantity versus price schedule of buyers (i.e., the demand curve) crosses the quantity versus price schedule of sellers (i.e., the supply curve). In an oligopoly, the equilibrium price occurs at the profit maximizing quantity where the marginal cost equals marginal revenue. Firms in an oligopoly adjust their outputs to gain a share of the profit maximizing quantity (Mansfield, 1982).

A key component of many cost control strategies involves the use of measures to increase competition based on the idea that competition leads to lower price (cost) or can control cost overruns or cost growth. Competitive contracting has been required by law since 2000, and in 2009, WSARA included a requirement to use competitive prototyping to promote competition and control cost despite microeconomics suggesting otherwise.

Our belief was that this confidence in the power of competition is based on a misapplication of microeconomic theory. It assumes that the perfect competition model, where there are many buyers and many sellers, is representative of the MDAP market.

> The U.S. economic system is built on the concept of free enterprise regulated by competition. ... The defense industry does not fit that model. Many defense acquisition problems are rooted in the mistaken belief that the defense industry and the government–industry relationship in defense acquisition fit naturally into the free enterprise model. (Fox, 2011)

However, the MDAP market has a monopsony–oligopoly structure, so the perfect competition model does not apply. The supply side of the MDAP market is an oligopoly, and according to microeconomic theory, firms in an oligopoly market do not compete on price. This is because, as Fudenberg and Tirole (2013) explained, if one firm cuts price to gain market share, this tends to lead to a price war where others in the market react by cutting their price.

The result hurts all players, since “the long-run costs of the price war outweigh any short-run gain” (Fundenberg & Tirole, 2013).

Competition in the defense market is so often cited as an important tool in fighting cost overruns and cost growth that it appears to have become a matter of faith that few bother to challenge. Typical examples include the Government Accountability Office’s claim that “competition is the cornerstone of a sound acquisition process and a critical tool for achieving the best return on investment for taxpayers” (GAO, 2015). The enormous confidence that government places in competition comes despite what O’Neil (2011) points out when he notes that in a pair of foundational studies of defense acquisition from 50 years prior, Merton J. Peck and Frederic M. Scherer of the Harvard Business School revealed significant issues that are still largely unaddressed by intervening management efforts (Peck & Scherer, 1962; Scherer, 1964). In particular, Peck and Scherer (1962) argued at length that price competition, which is widely favored as a mechanism for controlling costs, is almost certain to be largely ineffective in major defense system acquisition and is actually much more likely to be counterproductive (O’Neil, 2011). We contend that, while measures to stimulate competition may result in an increase in the number of bids received for MDAP Engineering and Manufacturing Development (EMD) contracts, that increase in competition will not result in cost control.

To establish which is correct—Congress and the DoD’s faith in competition or our belief in microeconomic theory—we sought the answers to the following questions:

1. Does competitive prototyping lead to more competition (an increase in the number of bids) in the MDAP market?
2. Does competitive contracting lead to more competition (an increase in the number of bids) in the MDAP market?
3. Does more competition lead to lower cost growth or overruns in the MDAP market?

If market competition does not control cost overruns in MDAP hardware acquisitions, then the defense acquisition community must devise and adopt strategies that do not need competition to reduce cost overruns and cost growth.

Methodology

This research is a case study of 36 MDAP hardware programs of various types and from all the services. We performed a quantitative analysis to determine the answers to our questions by testing the following research hypotheses:

1. Research Hypothesis 1 (RH1): In the MDAP market, competitive prototyping leads to more competition.
2. Research Hypothesis 2 (RH2): In the MDAP market, competitive contracting leads to more competition.
3. Research Hypothesis 3 (RH3): In the MDAP market, competition does not lead to lower cost growth or lower cost overruns.

Research Logic Flow

As a framework for our analysis, we devised a research logic flow to describe the potential outcomes (see Figure 1).

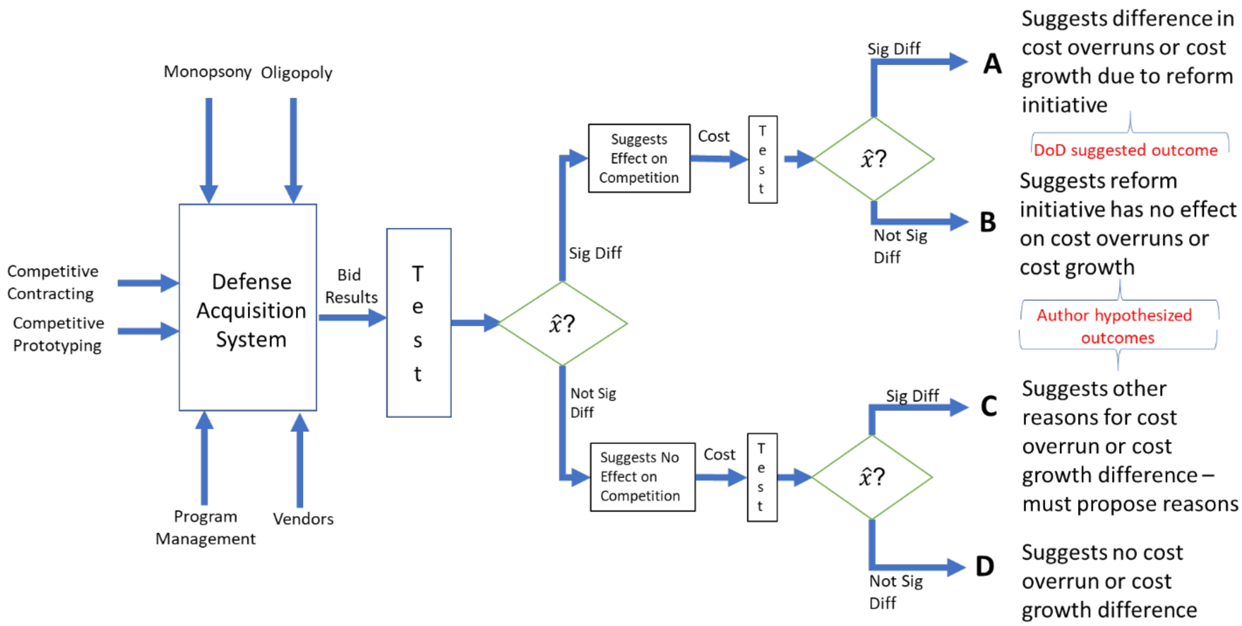


Figure 1. Research Logic Flow

Table 1 shows the potential outcomes.

Table 1. Potential Outcomes Logic Table

Result / Measure	Increase Competition?	Reduce Cost Overruns/Cost Growth?	Outcome
Competitive Prototyping	Yes	Yes	A
	Yes	No	B
	No	Yes	C
	No	No	D
Competitive Contracting	Yes	Yes	A
	Yes	No	B
	No	Yes	C
	No	No	D

Data

The data for our analysis came from six sources covering the years 2003 through 2019:

1. U.S. Government Accountability Office (GAO) *Defense Acquisitions: Assessments of Selected Major Weapon Programs* reports,
2. DoD Selected Acquisition Reports (SARs),
3. Federal Procurement Data System (FPDS),
4. Corporate 10-K reports filed with the U.S. Securities and Exchange Commission (SEC),
5. Kamp (2019), and
6. Fast (2016)

The first step in our analysis was to confirm that the MDAP market is, in fact, an oligopoly. We accomplished this as described by both Hayes (2020) and Kenton (2020) by calculating the standard CR_4 concentration ratio, which is the ratio formed by taking the sales of the top four firms and dividing by the total industry sales, which gives us the strength of the oligopoly power in the market. Concentration ratios range from 0.00 to 1.00 where 0.00 indicates perfect competition and 1.00 indicates a perfect monopoly. Table 2 lists typical rules of thumb that characterize the level of concentration.

Table 2. Mapping of Market Concentration to Market Structure and Concentration Indicators (Hayes, 2020; Kenton, 2020)

Level of Concentration	**Market Structure**	**CR_4**
Low	Perfect Competition to Oligopoly	0.00–0.40
Medium	Oligopoly	0.40–0.80
High	Oligopoly to Monopoly	0.80–1.00

The higher the market concentration, the less competitive the market. Table 3 shows the authors' calculated CR_4 MDAP market concentrations for the years 2003 to 2018, which shows market concentrations typical of an oligopoly market as described by (Hayes, 2020; Kenton, 2020).

Table 3. MDAP CR_4 Market Concentration for Years 2003–2018

Year 20XX	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18
Market Concentration	.83	.90	.94	.87	.86	.83	.87	.82	.86	.82	.81	.80	.80	.72	.77	.78

Once we confirmed our initial premise—the MDAP market is an oligopoly—we then proceeded with our data analysis.

Next, we performed correlation analysis to verify our hypothesized relationships between our variables. The relationships of interest are between the measures claimed to increase competition—competitive prototyping and competitive contracting—and the change in competition and between the change in competition and cost overruns and cost growth.

Additionally, we looked to see if there was a relationship between competitive prototyping and competitive contracting and between cost overruns and cost growth directly (i.e., due to some cause other than a change in competition). For the purposes of this research, we define competitive prototyping like Fast (2016)—as prototyping where two or more contractors develop prototypes prior to Milestone B (MS B), which are tested or demonstrated to the government to verify that they meet requirements. In addition, we define competitive contracting, in accordance with U.S. law, as a contracting strategy that relies on full and open competition. According to the Competition in Contracting Act (CICA, 2000), a procurement is considered as competed under full and open competition if *all responsible sources* are permitted to submit sealed bids or competitive proposals. Further, we define cost overrun as described by Cancian (2010)—as costs that exceed the estimate for a contract, in our case the EMD phase contract. EMD begins with a MS B decision to continue with development and ends with the Milestone C (MS C) decision to proceed into the production phase. Also, we define cost growth for all cost growth variables as an increase in cost from the estimate at program start to MS C. Finally, we use the definition of competition as "the attempt by two or more companies or other organizations to secure the business of a customer" (Farlex, n.d.). In this research, we measure the degree of competition by the number of bids received for the MDAP's EMD phase contract. Therefore, the greater the number of bids received for the EMD contract, the greater the degree of competition.

The cost overrun variables are Cost Overrun (Cost Overrun $M) and Percent Cost Overrun (% Cost Overrun). Cost growth variables are Percent Change in Unit Cost (UC.M.PCT), Program Acquisition Unit Cost (PAUC), Total Procurement Cost (TPC), and Percent Change in Total Procurement Cost (TPC.Pct). We calculated Cost Overrun $M by taking the difference between the baseline estimate at MS B and the current estimate at MS C in Base Year millions of dollars (in BY $M) from the SARs. Base Year is the prescribed DoD reference for measuring cost change because it removes the effect of inflation (Defense, 2020). % Cost Overrun was calculated by dividing Cost Overrun $M by the baseline estimate at MS B. UC.M.PCT came from the Kamp data set and is the percent change in the unit price since program start as reported in the GAO reports at MS C. It includes research and development (R&D) and procurement costs. A broader variable, PAUC, is the Program Acquisition Cost divided by the Program Acquisition Quantity as reported in the SAR reports at MS C. It includes all costs involved in the acquisition, not just R&D and Procurement. TPC was derived by adding the R&D funding (RD.M) and Procurement funding (P.M) since the program start to MS C. RD.M and P.M were included in the Kamp data set and came from the GAO reports. TPC.Pct was taken from the GAO report for MS C and is the percent change in total program cost from program start without regard to changes in quantity.

For our correlation analysis, we used the Spearman rho correlation because it is useful for both linear and nonlinear relationships. There are no universally accepted ranges for weak, moderate, and strong correlation coefficients. For our purposes, we were simply looking for an initial way to focus our analysis. Therefore, we used the following rules of thumb in our analysis: Strong correlations are those relationships with a Spearman correlation coefficient greater than or equal to 0.667; moderate correlations are indicated by Spearman rho values greater than 0.333 but less than 0.667; and weak correlations are those with coefficients that are less than or equal to 0.333. In addition, we considered significance levels for $\alpha = 0.05$ and 0.01. We found that there is a strong relationship between our competition variables pair combinations of competitive prototyping and competitive contracting and Num_bids that is statistically significant. This indicates that the use of competitive prototyping and competitive contracting is related to an increase in the number of bids. However, there appears to be no relationship between number of bids and our cost variables: Cost Overrun $M, Pct Cost Overrun, RD.M, RD.M.Pct, UC.M.Pct, and PAUC Chg from SAR. These results are favorable to our hypothesized proposition that while competitive prototyping and competitive contracting may increase the number of bids, the increase in competition does not lead to a reduction in cost overruns or cost growth. Armed with this information, we proceeded with our statistical analysis.

Statistical Analysis

We selected the Mood's Median Test as our analysis method because a visual examination of the histograms of our response data and probability plots from the Kolmogorov–Smirnov tests revealed that our data are not normally distributed. This indicates that a nonparametric method is called for, and the Mood's Median Test is a particularly good choice because our visual inspection also indicated the presence of outliers in our response data, and Mood's is insensitive to outliers. To test our research hypotheses, we evaluated 20 predictor–response variable pairs summarized in Table 4.

Table 4. Summary of Test Pairs

Predictor Variables	**Response Variable**
Competitive prototyping (CP)	Number of bids
Competitive contracting (CC)	Number of bids
CP	Cost Overrun $M
	% Cost Overrun
	RD.M
	RD.M.PCT
	UC.M.PCT
	PAUC
CC	Cost Overrun $M
	% Cost Overrun
	RD.M
	RD.M.PCT
	UC.M.PCT
	PAUC
Number of bids	Cost Overrun $M
	% Cost Overrun
	RD.M
	RD.M.PCT
	UC.M.PCT
	PAUC

Results

From our hypothesis testing, we found that there is evidence to support the claim that competitive prototyping leads to greater competition in the MDAP market, as seen by the statistically significant higher number of bids received on EMD contracts that included competitive prototyping versus those that did not. Similarly, the evidence supports the hypothesis that, when used, competitive contracting also increases the number of bids received and thus the degree of competition on MDAPs. However, we found no evidence to support the claim that an increase in competition (i.e., an increase in the number of bids received for an EMD contract) led to a decrease in cost overruns or cost growth. Furthermore, we found that there is no evidence to support any suspicion that competitive prototyping and competitive contracting themselves affect cost overruns or cost growth in MDAPs with one exception. Competitive contracting does appear to lead to a reduction in PAUC. There is enough evidence to reject the null hypothesis that the median PAUC for MDAPs that used competitive contracting and those that did not are equal. As a result, we can conclude that competitive contracting does lead to a reduction in PAUC. Since we also found that competitive contracting leads to an increase in competitions, one might conclude that this leads to Research Logic Flow Outcome A from Figure 1. However, we believe this would be a mistake. PAUC is the total of all development, procurement, and military construction cost divided by the number of units procured, and since we found no correlation between an increase in competition and procurement cost (P.M), nor did we see evidence that competition reduced development cost (RD.M, RD.M.PCT, UC.D.PCT), we must conclude that this reduction in PAUC is due to the portion from military construction. This seems reasonable because the construction industry is characterized by many small buyers and sellers more closely described by perfect competition where competition does influence price. A summary of these results is provided in Table 5.

Table 5. Summary of Hypothesis Test Results

Research Hypothesis	Explanatory Variable	Response Variable	H_0	*p* Value	Conclusion
RH1	CP (0) CP (1)	# of bids	Medians equal	0.000	CP leads to greater competition
RH2	CC (0) CC (1)	# of bids	Medians equal	0.004	CC leads to greater competition
RH3	# of bids	Cost overruns	Medians equal	0.166	More competition does not lead to lower cost overruns
RH3	# of bids	% change cost overruns	Medians equal	0.360	More competition does not lead to lower % cost overruns
RH3	# of bids	R&D cost growth	Medians equal	0.480	More competition does not lead to lower R&D cost growth
RH3	# of bids	% change R&D cost growth	Medians equal	0.145	More competition does not lead to lower % change in R&D cost growth
RH3	# of bids	% change in unit cost growth	Medians equal	0.802	More competition does not lead to lower % change in unit cost growth
RH3	# of bids	PAUC	Medians equal	0.298	More competition does not lead to lower PAUC
Similar results for CP and CC versus cost variable except CC versus PAUC					
	CC (0) CC (1)	PAUC	Medians equal	0.006	CC leads to lower PAUC
Research Hypothesis 1 (RH1): In the MDAP market, competitive prototyping leads to more competition. Research Hypothesis 2 (RH2): In the MDAP market, competitive contracting leads to more competition. Research Hypothesis 3 (RH3): In the MDAP market, competition does not lead to lower cost growth or lower cost overruns.					

Conclusions and Recommendations

The subject of cost overruns and cost growth in weapon systems programs is clearly not new; nor is defense acquisition reform. These topics reach back to the earliest days of our republic. A major theme of the modern defense acquisition reforms is that competition is good and will control cost. This may be true for items purchased by the DoD that are in a perfect competition market, such as copy paper and other consumables, where there are many buyers and sellers in both the civilian and defense sectors. However, attempting to apply the price minimizing characteristics of perfect competition to the oligopoly structure of the MDAP market—which behaves much differently—is inappropriate. The literature discusses some virtues of competition in the MDAP market, such as helping to maintain the defense industrial base and improving innovation, but cost control is not among them.

Our research shows that for the 63 hardware MDAPs we investigated in this case study, our assertion regarding the response to competition is correct. We demonstrated that the MDAP market is an oligopoly and that, while it may be possible to increase competition, the resulting competition will not lead to lower cost overruns or cost growth as Congress and the DoD believe.

Moreover, we answer all three of our research questions. Competitive prototyping does lead to more competition (an increase in the number of bids) in the MDAP market. Additionally, we show that competitive contracting also leads to more competition (an increase in the number of bids) in the MDAP market. Finally, we demonstrate that more competition does not lead to lower cost growth or cost overruns in the MDAP market. Table 6 and Figure 2 show these results in terms of our Research Logic Flow. We believed that either Outcome B or C from our Research Logic Flow would be proven to be correct. Our results show that Outcome B is the correct logic flow path as we proposed.

Table 6. Logic Table Outcome

Result Measure	Increase Competition?	Reduce Cost Overruns/Cost Growth?	Outcome
Competitive Prototyping	Yes	Yes	A
	Yes	No	B
	No	Yes	C
	No	No	D
Competitive Contracting	Yes	Yes	A
	Yes	No	B
	No	Yes	C
	No	No	D

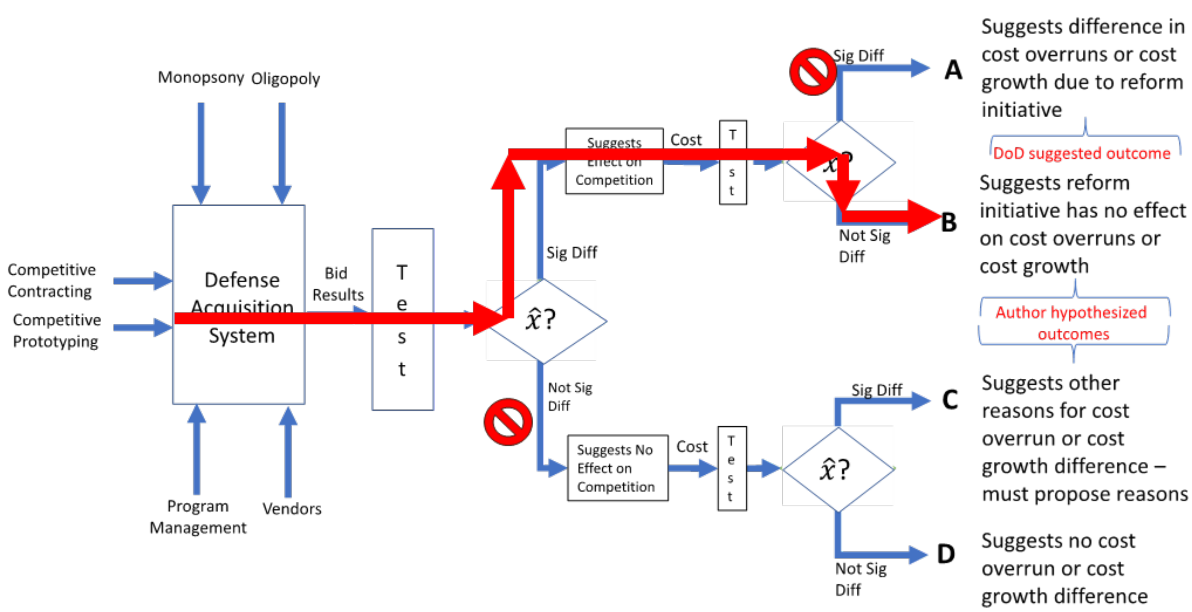


Figure 2. Research Logic Flow Outcome

Lastly, our analysis shows that, in addition to competition not affecting cost outcome for the hardware MDAPs we studied, competitive prototyping and competitive contracting themselves do not affect cost outcomes.

Unfortunately, the DoD's confidence in the power of competition to control cost in MDAPs appears to be based on the perfect competition model rather than the appropriate oligopoly model. This is a misapplication of microeconomic theory. We have shown that for our case study programs, competition does not control cost under oligopoly as microeconomic theory predicts. Therefore, we suggest that the DoD and Congress must look elsewhere for solutions to the problem of cost overruns and cost growth in MDAPS.

References

Cancian, M. F. (2010). Cost growth: Perception and reality. *Defense Acquisition Review Journal*, 389–405.

Competition in Contracting Act, 41 U.S.C. § 253 (2000).

Farlex. (n.d.). Competition. In *The free dictionary*. Retrieved March 24, 2021, from https://financial-dictionary.thefreedictionary.com/competition

Fast, W. R. (2016). An assessment of early competitive prototyping for major defense acquisition programs. *Proceedings of the 13th Annual Acquisition Research Symposium*, 47–66.

Fox, J. R. (2011). *Defense acquisition reform, 1960–2009: An elusive goal*. U.S. Army Center of Military History.

Fundenberg, D., & Tirole, J. (2013). *Dynamic models of oligopoly*. Routledge.

Hayes, A. (2020, February 11). *Financial ratios*. Investopedia. https://www.investopedia.com/terms/h/hhi.asp

Kamp, J. C. (2019, January 10). *Integrating immature systems and program schedule growth* [Doctoral praxis, The George Washington University]. ProQuest.

Kenton, W. (2020, September 6). *Economics*. Investopedia. https://www.investopedia.com/terms/c/concentrationratio.asp

Mansfield, E. (1982). *Microeconomics theory and applications*. W. W. Norton & Company Inc.

O'Neil, W. D. (2011). Cost growth in major defense acquisition: Is there a problem? Is there a solution? *Defense Acquisition Research Journal*, 277–295.

Ritschel, J. D., Lucas, B., White, E., & Mrla, D. (2019). The impact of WSARA on the cost of Air Force weapon systems. *Journal of Public Procurement*, 2–14.

Weapons Systems Acquisition Reform Act of 2009, Pub. L. No. 111-23 (2009). https://www.congress.gov/111/plaws/publ23/PLAW-111publ23.pdf

Do Accelerated Payments for DoD Contractors Help Small Businesses?

Jesse Cunha—Naval Postgraduate School [jcunha@nps.edu]

Justin Marion—Naval Postgraduate School [justin.marion@nps.edu]

Abstract

In 2011, the Department of Defense (DoD) set a goal of paying small business contractors within 15 days of invoice receipt rather than the standard 30 days. In 2012, other federal agencies also set a goal of accelerated payments to small businesses, and all agencies later expanded this goal to include all contractors regardless of size. We study whether small businesses benefited from these accelerated payment goals. Using a difference-in-difference design, we find that small business participation in government contracts rose following the setting of accelerated payments goals. Importantly, contracts for perishable foods and construction services, which were unaffected by the new accelerated payments policies, did not see an increase in small business participation. We also find that the benefits of accelerated payments are concentrated among small businesses with a backlog of ongoing projects. This is consistent with the hypothesis that accelerated payments alleviate the liquidity constraints that may be particularly acute for small businesses.

Introduction

The federal government has long been interested in supporting small businesses. An often-used policy lever is the intentional purchasing of goods and services from smaller, as opposed to larger, businesses, commonly by setting aside prime contracts for small businesses or by requiring that prime contractors utilize small business subcontractors. However, simply directing more purchases toward small business may not be sufficient on its own if other challenges that are faced by these firms are not addressed. Credit constraints are one such challenge. Access to credit could be costly, and small businesses often do not have sufficient cash reserves to smooth short-term liquidity requirements. Cash constraints may, therefore, limit the amount of business that can be undertaken by small contractors, and this will limit the effectiveness of programs meant to increase small business purchases.

One specific difficulty that contributes to the cash crunch faced by small businesses is the mismatch between when firms incur costs and when they receive payments for the goods and services provided. One way to improve the cash flow of small businesses is to reduce the amount of time between when goods and services are provided and when payment from the government for those goods and services is received. Accelerated payments have recently received substantial interest, with several federal policy actions calling for the acceleration of payments to small business contractors. In this report, our objective is to evaluate whether small businesses performed better when payments were accelerated, which we measured by the degree of participation by small businesses as prime contractors in federal procurement.

In 2011, the Department of Defense (DoD) announced a policy of accelerated payments to small business contractors. Under this policy, payment to small businesses occurs within 15 days of the receipt of invoice, rather than the standard 30-day payment period. Over the months following the initial announcement, the DoD phased in the policy into its major payments systems beginning in June 2011 with the DoD's largest payment system, the Mechanization of Contract Administration Services (MOCAS) system. The Office of Management and Budget (OMB) in September 2011 outlined a goal to pay invoices within 15 days for all federal procurement. The policy of accelerated payments was later expanded to include all prime contractors regardless of size. Neither the DoD, nor the broader federal policy, altered the penalties the government faced for late payment; these penalties continued to only be incurred after the standard 30-day payment period. While it is clear that the DoD implemented the accelerated payments policy as intended, we are uncertain about the extent to which the other federal agencies complied with the OMB's directive to apply accelerated payments for all federal procurement.

To evaluate whether accelerated payments achieve the goal of positively impacting small businesses, we utilize data from the Federal Procurement Data System (FPDS), which contains the universe of nonclassified federal contract actions. These data do not contain information regarding direct measures of firm success such as employment, investment, or profits. Instead, we examine whether participation by small business prime contractors increased when accelerated payments were in place. While this is an indirect measure, as we discuss in more detail, this is an appropriate measure of the benefits to small businesses from accelerated payments. One reason is that encouraging small business participation is, in itself, often a goal of procurement policy. Moreover, economic theory suggests that by examining participation we can provide indirect evidence about whether small business performance is aided by accelerated payments. This is for two reasons. First, small businesses are more willing to enter the competition for contracts if they benefit from accelerated payments. Second, conditional on entry, they will be willing and able to bid more competitively to win the contract.

Our empirical approach utilizes the change in accelerated payments policy over time, as well as the differential impact that accelerated payments have on affected relative to unaffected products. Not all products were affected, since the payment period for construction contracts and for some food products was already less than 15 days. Contracts for these products were, therefore, unaffected by accelerated payments. The empirical approach is a difference-in-difference design, where we consider the change in the number of contracts awarded to the average small business after the implementation of accelerated payments to the similar change for the average large business. Furthermore, this differential change for small versus large firms can be compared between affected and unaffected products, since contracts for construction services and many types of food products already had payment periods of less than 15 days. The appeal of this approach is that it can control for a myriad of factors that affect small businesses at a given time, so long as those economy-wide shocks similarly affect firms that provide affected and unaffected products. It can also account for changes over time that might affect all firms, both large and small.

Early payment receipt can help small businesses in several ways. First, accelerated payments increase the present value of contract payments, which results in a small benefit to contractors. Second, and likely more importantly, accelerated payments help the cash flow of a business. Prime contractors have obligations to their subcontractors and other suppliers and must manage the cash flow necessary to make on-time payments. Either suitable cash reserves, or lines of credit, must be maintained to meet short-term obligations. Firms may face borrowing constraints that limit their ability to meet short-term needs, necessitating holding cash reserves. This constraint may be particularly acute for small businesses, which have less access to both external and internal capital, which could be a barrier to taking on new work on top of ongoing projects. Accelerating payments could, therefore, assist small businesses in participating in government procurement and in supporting their growth. Encouraging small business participation and growth may, in fact, be cost effective for the government, as current small businesses become future, possibly more efficient, competitors.

For two similar reasons, helping small businesses through accelerated payments could be costly. First, the present value of contract payments is higher under accelerated payments. Given the favorable terms of federal borrowing, this is a lesser issue for government agencies than it is for the seller, though in the aggregate the impact is nontrivial. To get a sense of the scale of this cost, consider the thought exercise of advancing the payment on all DoD contracts by 2 weeks.[1] DoD contracts in 2017 were worth $320 billion, and advancing the payment on these contracts by 2 weeks increases the present value of this expenditure by around $300 million.[2] Second, the procuring agency must itself also manage cash flow, and the DoD has raised concern that the expansion of accelerated payments has created cash constraints for some defense activities. Such constraints may pose a greater challenge for agencies than the increase in the present value of contract payments.

This report proceeds as follows. In the Literature Review section, we review the evidence regarding small businesses and government programs to assist them. We then describe the accelerated payments policies and the time line of their adoption in the section titled Accelerated Payments Policy. The next section contains a description of the data used in this study. Then, we provide a model that motivates the empirical specifications. Next, we specify our empirical approach and present the results. Finally, we provide our conclusion.

Literature Review

It is often argued that small businesses are important because of their role as job creators and innovators (Birch, 1979) and because entrepreneurship is a potential pathway out of poverty (Glazer & Moynihan, 1970). Based on these arguments, lawmakers often enact public policies meant to help small businesses, commonly through public procurement programs that steer contracts or by providing terms that are particularly favorable to small, rather than large, businesses. The efficacy of these policies rests (1) on their ability to address the factors that contribute to the success of small businesses and (2) the cost that the policies impose on the government. Our research fits within a robust literature in economics, which examines the factors that influence the success of small businesses and entrepreneurship as well as the subset of this literature that examines the costs and benefits of small business policies implemented via public procurement. Such policies may target small businesses generally, or they may take the form of affirmative action programs specifically intended to benefit minority- and women-owned enterprises.

Several factors may influence entrepreneurial activity and the formation, expansion, and survival of small businesses. Liquidity is a key constraint faced by small businesses, as it determines the access to capital needed to start a business and to meet short-run cash flow needs. It is also, therefore, closely related to accelerated payments. Fairlie and Krashinsky (2012) documented that easing liquidity constraints significantly boosts entrepreneurship. Holtz-Eakin et al. (1994) and Holtz-Eakin and Rosen (2005) also found

[1] We do not attempt a cost–benefit analysis in this report. A proper measure of the cost of the program would require data on exactly how much more quickly payments were made. These data exist but are not publicly available. A proper measure of benefits would require dynamic estimates of how firm survival and efficiency may positively impact the operation of the procurement market. As we discuss in the conclusion, this is outside of the scope of our research.

[2] This is likely an overstatement of the costs of the program that we consider for two reasons. First, the accelerated payments policies we consider here are not expected to affect contracts for food or construction, which indicates that less than $320 billion in DoD contracts were impacted. Second, compliance may not have been complete, signaling that some payments that should have been accelerated were, in fact, not accelerated.

similar effects of liquidity on entrepreneurs.

Most firms that qualify for minority- and women-owned business subsidies are small businesses, and so the constraints faced by these firms will be highly relevant for small business policy. Fairlie and Robb (2007) studied minority-owned enterprises and examined the factors that contribute to the disparities in business outcomes experienced by these firms compared to white-owned firms. Related to the role of liquidity in small business success, they found that these disparities are influenced by access to startup capital. Black-owned businesses may face lending discrimination, as found by Blanchflower et al. (2003) and Cavalluzzo and Wolken (2005) and may also have less family wealth (Bates, 1997).

The costs and benefits of small business subsidies and affirmative action implemented through public procurement have been examined by several authors. Chatterji et al. (2014) examined the effects of affirmative action in city contracting on minority entrepreneurship, and Marion (2011) studied how affirmative action affects government purchases from minority- and women-owned enterprises. Nakabayashi (2013) examined the set-aside of government contracts for small enterprises in Japan, finding that approximately 40% of Small and Medium Enterprises (SMEs) would exit the procurement market if set-asides were removed. This would lead to a lack of competition that would ultimately increase government procurement costs. Marion (2007), Krasnokutskaya and Seim (2011), and Hubbard and Paarsch (2009) studied bid preferences for small businesses in government contracting. Denes (1997) assessed the cost of small business set-asides in federal contracting. Finally, Marion (2009) and Rosa (2020) examined the effect of affirmative action via subcontracting goals on government procurement cost.

Accelerated Payments Policy

Prompt Payment Act

The Prompt Payment Act (PPA), originally enacted in 1982 and subsequently revised in 1988, set several standards with respect to the speed at which federal contractors are paid for work performed or orders fulfilled (FAR 52.232-27, 2021). This act addressed the typically slow rate of payment by many federal agencies by setting a maximum time between receipt of invoice and payment, while including a series of exceptions that would provide even faster payments for particular types of goods. In this section, we describe these regulations, as well as the more recent efforts to further accelerate payments to the benefit of small business enterprises.

The PPA specified that for most government procurement, the due date of contractor payment is the latter of the “30th day after the designated billing office receives a proper invoice from the contractor” or “the 30th day after government acceptance of supplies delivered or services performed” (Prompt Payment Act [PPA], 1988). Should the payment occur after the 30-day window, the contractor is automatically due interest penalties from the federal agency, and the calculation of those penalties is also codified into law. Furthermore, the PPA also governs payments by the contractors to its subcontractors. The contractor has 7 days upon receipt of payment from the government to pay its subcontractors or suppliers, provided that the subcontractor has provided satisfactory performance. The subcontractor in turn has 7 days to pay its own lower-tier subcontractors or suppliers.

For some types of goods and services, payments must be made in a much shorter time frame than the standard 30-day window, and these exceptions play a key role in identifying the impact of accelerated payments in our empirical work. Broadly speaking, the exceptions apply to all construction contracts and many types of food products, particularly perishable foods. For construction contracts, the payment deadline is 14 days after the payment request is made. For food products, suppliers of poultry, eggs, and frozen fish must be paid within 7 days. Suppliers of perishable agricultural commodities must be paid within 10 days. Finally, dairy products and those foods made from edible fats or oils must be paid within 10 days.[3] For products with a payment period of less than 15 days, the new accelerated payment goals should have, at most, a small impact on payment speed.

Accelerated Payments Policies

In early 2011, the DoD announced that it was setting new guidelines for the speed of payments to its contractors. This policy is codified into Subpart 232.9 of the Defense Federal Acquisition Regulations System (DFARS). Initially, the policy applied only to invoice payments to small business contractors. Soon after the adoption of accelerated payments by the DoD, guidelines from the OMB extended this policy to federal contracting more broadly, and accelerated payments were later further expanded to apply to all contracts regardless of the size of the contractor. Here, we describe the time line of the policy announcement and implementation.

Defense Acquisition Regulations System (DARS) Memorandum 2011-O0007 was issued on April 27, 2011, announcing that the DoD would commence implementation of DFARS 232.9303 accelerated payments to small business contractors (Assad, 2011). The new policy described a plan to pay contractors within 15 days of submitting a proper invoice. The implementation would occur in phases, with the initial phase focusing on the modification of the DoD's largest payment system, the MOCAS system. On June 28, 2011, the DoD announced in DARS Memorandum 2011-O00013 that "the Defense Finance and Accounting Service has completed modifications of MOCAS to provide for these accelerated payments to small business" (Ginman, 2011). Payment systems that account for a smaller share of DoD contracts followed. The list of payments systems and the dates they were implemented is listed in Table 1.

Table 1. Dates of DoD Payment System Implementation

Date of announcement	Payment system(s)
June 2011	MOCAS
October 2011	Standard Automated Voucher Examination System (SAVES)
January 2012	Automated Voucher Examining and Disbursement System (AVEDS) the Fuels Automated System (FAS)
May 2012	Corps of Engineers Financial Management System (CEFMS)
August 2012	Transportation Financial Management System (TFMS)
April 2013	Enterprise Business System (EBS) Financial Accounting and Budget System (FABS)
February 2014	OnePay, Integrated Accounts Payable Systems (IAPS) Computerized Accounts Payable System-Windows (CAPSW)
April 2014	Defense Agencies Initiative System

The relevant DFARS memos are 2011-O00013, 2012-O0001, 2012-O0002, 2012-O0006, 2012-O0011, 2013-O0008, 2014-O0006, and 2014-O0015 respectively.

[3] This rather broad category includes "liquid milk, cheese, certain processed cheese products, butter, yogurt, ice cream, mayonnaise, salad dressings, and other similar products" (FAR 52.232-25, 2021).

In September 2011, after the initial DoD announcement, the federal government announced a directive for all agencies to make accelerated payments to small business prime contractors within 15 days of invoice (Lew, 2011). Then, in July 2012, the federal government extended the accelerated payments policy to include all federal prime contractors—not just small businesses—including those vendors selling to the DoD.[4] While it is not clear when each particular agency implemented this in practice, the DoD announced that its payment systems had incorporated this change in August 2012 (Ginman, 2012). In February 2013, the DoD canceled the accelerated payments for large prime contractors, and the policy reverted to only accelerating payments for small businesses, though it was required to reinstate accelerated payments for large prime contractors in July 2014 (Ginman, 2013). In summary, accelerated payments was the stated policy of the DoD from April 2011 onward for small businesses, and from July 2012 until February 2013 and from August 2014 onwards for large businesses (see Table 2). Accelerated payments was the stated policy of federal contracting more broadly from September 2011 onward for small businesses and from July 2012 onward for large businesses.

Evidence presented by the DoD indicates that they complied with the policy: "In practice, the Defense Financial Accounting Service (DFAS) currently provides accelerated payments to nearly all DoD contractors, as permitted by law" (DoD, 2019). The average time to pay an invoice by MOCAS was under 15 days, as of early 2018 (McDuff, 2019). Unfortunately, we were not able to find direct evidence about whether other federal agencies complied with the policies. The issue of whether or not federal agencies strictly adhered to the accelerated payments policies is important and potentially relevant. These policies did not carry with them financial penalties for noncompliance. Neither the DoD in its actions previously described, nor the broader federal action, altered the rules regarding interest penalties for late payments by agencies. Some federal agencies may have, therefore, viewed accelerated payments as nonbinding and may not have implemented the policy as directed. Only for the DoD, with the pronouncements that MOCAS and other key payments systems had been altered, can we say with confidence that the accelerated payments policies were in fact implemented. To assess compliance more thoroughly, we would need to observe data on time-to-payment. Such data are likely contained within the Contract Performance Assessment Reporting System (CPARS), but these data are not publicly available, and we were not able to gain access in the time frame of the current project.

Table 2. Time Line of Accelerated Payments Policies

Date	Action	Referenced order
April 2011	DoD accelerated payments for SB primes	DARS 2011-O0007
June 2011	DoD implements acc. pay in MOCAS	DARS 2011-O00013
Sep. 2011	Acc. pay for federal SB primes	OMB M-11-32
July 2012	Acc. pay for all federal primes	OMB M-12-16
Aug. 2012	Implements M-12-16 for DoD primes	DARS 2012-O0014
Feb. 2013	Cancels acc. pay for DoD large primes	DARS 2012-O0014
Aug. 2014	Reinstates for DoD large primes	DARS 2014-O0019

Qualifying as a Small Business

Which firms qualify as a small business varies substantially across product types. The Small Business Administration (SBA) outlines the criteria that are used to determine qualification as a small business in federal contracting. The SBA defines threshold values

[4] This policy change also inserted a clause directing the prime contractor to accelerate payments to their small business subcontractors. It is unclear whether this additional clause was binding, since prime contractors were already required to pay any subcontractors within 7 days.

for employment and revenue that vary by industry classification of the firm. If either employment or revenue exceed these values, the firm does not qualify as a small business. The most common thresholds for manufacturing companies are 500 employees and average annual revenues of $7.5 million. A majority of firms in the FPDS at one time or another are listed as a small business, and small business status can, therefore, change over time as a firm expands or contracts. In our empirical work, rather than using a dichotomous indicator for small business status, we instead use the fraction of contracts in which the firm participated as a small business across all years of the data.

Data

Our data come from the universe of contracting actions for Fiscal Years 2010 to 2015, obtained from the Federal Procurement Data System–Next Generation (FPDS). All unclassified contracts above a mandated minimum value must be reported in this system. Contracts are identified by the contracting firm's Data Universal Number System (DUNS) number, a unique firm identifier that allows us to track firms across procurements and form firm-specific measures of contract backlog. As mentioned above, backlog is a measure of how many contracts a firm has underway at any point in time, and high backlog can limit firms' ability to take on new work. Other important characteristics of a contract that are reported in the FPDS are the number of offers, whether the firm is a small business, the agency requesting proposals, whether the contract was competitively awarded, the date the contract was signed and the date it was effective, and the expected completion date. There are also several variables that we use to determine the type of action, including any referenced indefinite delivery vehicle (IDV), whether the referenced action is a modification, and the reason for this modification.

These data also provide the product service code, describing the type of product provided and the North American Industry Classification System (NAICS) code for the associated industry. The good or service being purchased can then be categorized as food, construction, or some other type of good, which allows us to determine whether the accelerated payments policy was likely to be binding or if the product already had a payment time under 15 days. The focus of our empirical work is on new contract awards—not revisions, change orders, or exercised options of existing contracts. Our objective is to evaluate how accelerated payments affect small business willingness and ability to take on new work. This is best reflected in new contracts. For the summary statistics presented in this section, and for all the other empirical estimation that follows, the sample of contract actions are only new contract awards.

Types of Contracts

There are two broad categories of federal contracts, direct contract awards (DCA) and task order awards (TOA). A DCA is the simplest type of contract. It is awarded to a single vendor, with specifications set by the agency to which funding has been obligated. DCAs can be either a definitive contract, which is typically a contract agreeing to purchase a good or service that is noncommoditized (i.e., unique), or it can be a purchase order for a commoditized good or service. The initial agreement is recorded in FPDS, as well as any further modifications to that agreement.

A TOA is also with a single vendor but is awarded under a broader contract or an IDV with that vendor. The IDV is with a particular vendor, and then the vendor can be issued delivery orders or purchase agreements under the IDV. The IDV will typically have a ceiling on the amount of the order. The FPDS lists detailed information about the broader IDV with which the order is associated, as well as each order against that IDV.

The distinction between TOAs and DCAs is important to note, as these two types of actions need to be treated differently. For the empirical analysis, the most appropriate focus is on new DCAs and on orders against existing IDVs that are not simply fulfilling an already-agreed-upon delivery. The objective of this report is to document whether small businesses are able to take on more work and participate in more contracting when accelerated payments are in place, and DCAs clearly reflect this. However, it is ambiguous whether a new delivery under an existing IDV represents a willingness to take on new work or a vendor simply fulfilling prior obligations. On one hand, some IDVs may be under a broader federal supply schedule, and when a federal agency orders under this supply schedule, it will buy from vendors that are willing to sell. This type of order will, therefore, reflect the firm's contemporaneous interest in supplying the good. On the other hand, with other forms of the IDV, such as an indefinite delivery contract, the vendor may have agreed in advance to deliver the goods to an agency on demand up to a cap. New orders against this IDV, therefore, do not reflect the willingness of the contractor to participate in that moment but rather their willingness to participate when the IDV was entered into.

To evaluate whether an order against an IDV is appropriate for inclusion in the empirical analysis, we consider the extent of competition. Specifically, if an order was subject to competition, then it suggests that the order was effectively new, and participating in the competition reflects the firm's contemporaneous interest in taking on further work.

As was described previously, payment periods shorter than 15 days already applied for construction services and many types of food products. For procurement of these goods, accelerated payments should have no effect. The product and service code (PSC) of the product is listed in the data, which allows us to identify those products for which the treatment is expected to have no effect. The codes for food are 8905 (meat, poultry, and fish); 8910 (dairy, foods, and eggs); 8915 (fruits and vegetables); and 8945 (oils and fats). Similarly, accelerated payments policies were not binding for construction contracts, which already had payment time frames shorter than 15 days. The PSC can also be used to distinguish these contracts. The code Y1 (construction of structures and facilities) includes all federal construction work, from roads, bridges, and buildings to fuel supply facilities and heating and cooling plants.

In Table 3, we present the composition of federal IDV contract awards across the relevant product categories separately for DoD and non-DoD procurement. In Table 4, we present similar figures for DCAs. In the DoD, most new IDV contract actions are not food or construction and are, therefore, covered by accelerated payments. The average year sees 63,351 new IDV food awards and 2,904 construction awards compared with 534,151 contracts for other types of goods. The dollar value of new IDV contracts is also heavily tilted toward non–food and construction, though not to quite the same degree the number of contracts. For non-DoD contracts, we see that most IDV awards are neither food nor construction. The average year sees 449,104 new IDV awards for goods that would be affected by accelerated payments and only 2,439 between perishable food and construction. When considering DCAs, we again see that most new contract awards are affected by the accelerated payments policy. Within the DoD, only 193 DCAs were made in an average year for food products that would be unaffected by accelerated payments. Only 1,696 new DoD DCAs are for construction contracts in the average year. These compare to an average of 273,825 awards that are affected by accelerated payments.

Table 3. Average Annual IDV Actions by Product Type

	Non-DoD		DoD	
	Actions	Obligated amount (millions)	Actions	Obligated amount (millions)
Construction	1377	776	2904	3320
Food	1059	341	63351	2783
Other	449104	32630	534151	57461

Table 4. Average Annual DCAs by Product Type

	Non-DoD		DoD	
	Actions	Obligated amt (millions)	Actions	Obligated amt (millions)
Construction	2315	2692	1696	5607
Food	8732	1633	193	4.1
Other	248121	13739	273825	30982

Small business participation declines substantially as the size of the award grows. Table 5 shows the share of federal contracts awarded to small businesses separately for DCAs and IDVs. Across all action types, small businesses are awarded 47% of contracts. For most contracts (98.8%), the obligated amount is below $1 million in value. For contracts between $10 million and $20 million, 31% go to small businesses. Contracts between $20 million and $30 million are awarded to small businesses 21% of the time, and this drops to just 12% of contracts over $30 million. The pattern is highly pronounced for both DCAs and IDVs, though overall IDVs are less frequently awarded to small businesses.

Table 5. Small Business Share of Awards by Obligated Amount

	All	DCA	IDV
0 to 1mm	0.469	0.613	0.396
1 to 10mm	0.435	0.547	0.390
10 to 20mm	0.309	0.374	0.274
20 - 30mm	0.207	0.241	0.188
>30mm	0.116	0.097	0.132
All	0.470	0.612	0.396

Model of Contract Participation

In this section, we present a simple model of contract participation. This provides the theoretical basis for examining participation as an outcome. In a competitive procurement, winning a government contract reflects two factors, the likelihood of entry into the competition for the contract and the likelihood of winning the contract conditional on entry. If small businesses benefit from accelerated payments, then both of these dimensions will be affected.

A firm's decision to enter the competition for a contract depends on their expected profits. The expected profits are the likelihood that the submitted offer is successful multiplied by the profits the firm will earn if it is awarded the contract:

$$E[\Pi] = Pr(win) * (\Pi|win).$$

If it were costless to participate in the competition for contract, then the firm would submit an offer for all contracts where it stood a chance to earn positive expected profits. However, submitting an offer is not costless. In addition to the administrative cost of preparing a bid, the firm needs to invest time and resources in understanding the contract requirements and the cost of meeting those requirements. Let the cost of participating in the competition for a contract be K. Entry will occur if the expected profits of entry exceed these fixed costs:

$$E[\Pi] = Pr(win) * (\Pi|win) \geq K. \qquad (1)$$

Accelerated payments improve both the probability of winning and the profits of the firm conditional on winning. In a first-price auction, a commonly used mechanism to award government procurement contracts, the probability of winning is $Pr(b_i < b_{-i})$, where b_{-i} represents the lowest bid of the firm's competitors. A widely accepted result in the auction literature is that the firm's optimal bid is monotonically related to its cost (Krishna, 2009). In other words, as the firm's costs decline, the firm is able to submit a lower bid, and the lower the bid is, the higher the likelihood that $b<b_{-i}$. Under accelerated payments, the bidder can complete the project more inexpensively because the bidder's cost of capital is lower since they receive payments more quickly. Furthermore, the opportunity cost of taking on the project has gone down by expanding the firm's capacity constraint. Put differently, receiving a contract may reduce the firm's ability to take on other work. This opportunity cost will be incorporated into the firm's bids. So, we expect that if small businesses benefit from accelerated payments, then the probability of winning increases for those firms.

Similarly, the firm's profits conditional on winning could positively depend on accelerated payments as well. The profits upon winning a first-price auction are

$$\Pi|\text{win} = b - c$$

which is the firm's bid less its cost. For all the reasons just described, firm cost c declines with accelerated payments and, therefore, Π|win increases.

We also now discuss two added dimensions that we do not model directly. First is the role of project backlog. A common finding in the procurement literature is that firms have limited capacity, and firms with a backlog of incomplete projects have difficulty taking on new work; this can interact with programs that affect the volume of business.[5] We anticipate that firms with existing projects underway will find accelerated payments particularly valuable, as they can take on new work with additional confidence that cash flow will be sufficient to meet obligations to suppliers on projects that are already underway.

A second dimension that we do not model is the possibility that firms may substitute across contracts in response to incentives that vary across contracts at a point in time. Limited productive capacity to complete the work in the contract, or limited managerial capacity to prepare bids, require firms to be selective about the contracts for which they will

[5] For instance, Balat (2014) estimated how the effectiveness of the American Recovery and Reinvestment Act was affected by the sudden surge of projects when highway construction firms have upward sloping marginal cost curves. Jofre-Bonet and Pesendorfer (2003) estimated the response of forward-looking firms, where contractors anticipate the effect that winning an auction today will have on the likelihood of winning future contracts via increased backlog.

compete. If the accelerated payments policy makes one type of procurement contract more appealing, then this could induce a substitution across procurement auctions. While we do not model this phenomenon directly, we will discuss the empirical implications of this issue below.

Empirical Implications

This simple model illustrates how we can uncover the benefits of accelerated payments for small businesses in the federal procurement data. Using the FPDS, we cannot measure the profitability of the firm. Nor can we directly measure other outcomes of interest that relate to accelerated payments, such as the firm's cash flow. However, the model suggests that we can study the benefits of accelerated payments by examining contract awards. Consider again Equation 1. Accelerated payments can increase both terms on the left-hand side of the inequality. As both Pr(win) and $\Pi|$win increase, the likelihood that the firm participates in the bidding for an auction increases. As both participation in competition for contracts and the likelihood of winning those contracts increase, then firms will win more contracts. Looking at the number of awarded contracts is, therefore, a valid measure of whether the accelerated payments policy increases the profits of small businesses. In the empirical section, we connect this to the data in two ways. First, we examine the number of contract awards by month at the firm level. The hypothesis that we test is whether the average small business receives more government contracts in months where accelerated payments were in place. In light of the theory discussion above, we separately consider firms with and without a backlog of existing projects in order to test whether the benefits of accelerated payments depend on the backlog.

The second outcome we examine is the number of offers made by firms for contracts set aside for small businesses. As suggested by Equation 1, the desire to participate in an auction should increase under accelerated payments. In general, the data do not contain information about the bidders for a contract, only the number of offers submitted. Therefore, we are not able to observe the number of bids by small businesses. Many contracts are specifically set aside for small businesses. Others are set aside for disadvantaged business enterprises, which are a subset of the small businesses. Thus, we can observe the number of small businesses that enter the competition for set-aside contracts.

The question of substitution across auctions is highly relevant here. It is possible that accelerated payments increase auction participation of small businesses across the board, which would show up in the data as an increase in the number of offers for set-aside contracts. Alternatively, accelerated payments could allow small businesses to be more competitive for procurement contracts that are open to general competition. Small businesses might substitute toward general competition auctions, and set-aside participation could, in fact, decline. Even if the accelerated payments policy raises the desire of small businesses to participate in the competition for contracts, the predicted impact on the number of set-aside offers is ambiguous.

Empirical Approach

Using these policy changes described above, we can estimate the effect of accelerated payments on small businesses. We do so by examining the participation of small businesses in federal contracts and determining whether participation becomes more likely when accelerated payments are in place. We take a difference-in-difference (DD) approach, examining the difference in small business participation on contracts when accelerated payments are in effect compared to times when they are not, and we compare this difference to the same difference for large businesses. We also conduct this exercise separately for affected and unaffected products. We expect no effect for unaffected products, so performing this estimation is a placebo exercise. If other factors aside from accelerated payments were affecting small business participation, then these should be witnessed in participation in contracts for placebo products as well.

We aggregate the data to the firm-month level, so that the variable of interest is the number of contracts that a firm is awarded in a given month.[6] Doing so creates a data set that is too large to be practical—approximately 28 million observations. We therefore take a 25% random sample of the firms in the data, which leaves 7.2 million firm-month observations. Let y_{it} denote the number of contract awards received by firm *i* in time period *t*. Let Y^S_0 be the average level of participation of small businesses prior to the adoption of accelerated payments, and let Y^S_1 be the participation after. The change from the period without accelerated payments to after is given by $\Delta Y^S = Y^S_1 - Y^S_0$. The DD estimate of the effect of accelerated payments on small business participation is the difference in this change between small and large firms: $\gamma = \Delta Y^S - \Delta Y^L$. The linear regression specification that yields the DD estimates is as follows:

$$y_{it} = \beta_0 + \gamma A_t * S_i + \rho_i + \varphi_t + \varrho_{it} \qquad (2)$$

In this regression, the variable A_t is an indicator for being in a period of time where accelerated payments are in effect, and S_i is an indicator for whether firm i is a small business. Since small business status can change over time endogenously, when we implement this regression specification we instead include the fraction of contracts in which the firm participated as a small business. The parameters ρ_i and φ_t represent firm and time fixed effects, respectively. The coefficient of interest is γ, which is the DD effect of accelerated payments on small business contract awards.

To connect this regression equation to the intuitive description of the DD estimates above, the first difference for small firms is $\Delta Y^S = E[y_i|A = 1, S = 1] - E[y_i|A = 0, S = 1]$, with a similar definition for large firms. This difference removes the firm fixed effect, ρ_i, which accounts for all differences across firms that are fixed over time, including any time-constant effects of being a small business. The second difference is $E[\Delta Y^S] - E[\Delta Y^L]$. This difference removes any time-specific factors, φ_t, that affect all firms equally.

The main confounder in a DD specification is the presence of unobserved time-varying shocks that differentially affect the treated group. Put differently, if an unobserved variable increases small business participation in contracting (and does not impact large business participation), and this variable happened to increase in the accelerated payments period, then the DD results would be biased in favor of finding an effect. As an example, changes in credit availability may differentially affect small businesses, and if credit availability changed for some reason during the accelerated payments period, then the impact of accelerated payments would also include this effect of credit on participation. This concern is difficult to address directly, though it can be addressed indirectly using a placebo exercise. We estimate the parameter γ for food and construction contracts separately,

---

[6] We measure participation using the number of contracts rather than the dollar value of those contracts. We do so for two reasons. First, there is a close connection between this measure and the theoretical concepts discussed above. Second, the measures of contract value in the data may not reflect the expected value of the contract. Contractors form expectations regarding change orders, whether the options in the contract will be exercised, whether they intend to agree to extra work, and so on. Expectations about these variables may be influenced by the accelerated payments policy itself. For these reasons, measuring participation by the number of contracts has a clearer interpretation and is likely to more accurately measure the underlying theoretical concepts.

which should be affected little by accelerated payments, if at all. If there was an unobserved shock affecting small businesses in the accelerated payments period, it should show up in this estimate. If the estimated γ is instead small and statistically indistinguishable from zero for food and construction, then we can rule out this source of bias under the assumption that unobserved small business shocks that are correlated with accelerated payments impact all firm types equally.

In practice, there are two treatment periods that are relevant—the time where accelerated payments applied only to small businesses and the period where the policy applied to all firms. We, therefore, introduce two separate indicators for the accelerated payments treatment period rather than a single indicator, A_{it}. Our expectation is that small businesses are impacted more by accelerated payments and, therefore, should see a benefit relative to large businesses even when the policy applies to all firms. That said, the effects of these two periods on small business contract participation could be different. On one hand, expanding the application of accelerated payments to large businesses may reduce any advantage in bidding that small firms enjoyed when they alone received accelerated payments. On the other hand, the period of time where accelerated payments applied only to small businesses was fairly brief. If the policy took time to have an effect, then the impact may only be observed during the later treatment period. The reason for a delayed impact could be because some DoD payment systems were slow to be converted. Alternatively, as previously described, one mechanism for the effect of accelerated payments that we explore is the role of a firm's backlog. Faster payments may improve a firm's ability to take on multiple projects, so the effects of the policy could amplify over time.

Results

Small Business Contract Awards

In this section, we present our main set of results. We show how small business contract awards changed during the period of accelerated payments, providing separate estimates for DoD contracts versus awards by other federal agencies. We begin by providing estimates for all product types together, and we subsequently break the data apart by broad product category.

In Table 6, we present the results for contract awards made by the DoD. In the first column of this table, we present the results from regressing the logarithm of the number of contract awards made to a firm during a month on the interaction between the small business indicator and the two treatment windows. We find very little impact for the average small business of the accelerated payments program. The interaction between the small business indicator and the first treatment window was negative and very small in magnitude.[7] It is statistically significant, but it is important to note that this is because the coefficient is precisely estimated with over 7 million observations. With 95% confidence, we can rule out an effect size smaller than −0.009 log points, which is less than a one percentage point effect on contract participation by small businesses.

The main result masks a heterogeneous effect depending on firm backlog. In Columns 2 and 3, we split the sample between firms with and without unfinished contracts. For small businesses with a positive amount of backlog, there is a positive and statistically significant effect of accelerated payments. The first treatment window has a moderately small positive effect—0.026 log points—on the contract awards to small businesses. During

[7] In this section, we refer to the period where accelerated payments applied only to small businesses as the "first treatment window" and the period where accelerated payments applied to all firms as the "second treatment window."

the second treatment window, small businesses saw higher contract awards of 0.064 log points compared to when accelerated payments were not in place. For firms without backlog, the estimated impact was nearly zero. Compared to the period without accelerated payments, the difference in contract awards to small businesses was virtually zero—around 0.0005 log points.

In Table 7 we present a similar set of results for other federal agencies. The pattern of results is similar, though the effects are much smaller in magnitude than for DoD small business contractors. Accelerated payments are associated with a small negative effect on log contract awards to small businesses. Though statistically significant, the effect is very small and precisely estimated. The difference in small business participation from the time of no accelerated payments to the first treatment window is just −0.005 log points, and small businesses in the second treatment window receive only −0.0024 fewer log points of contracts. As was the case for DoD contractors, any positive effect of accelerated payments is observed for firms with a backlog of incomplete contracts. The first treatment window effect is 0.011, and the second treatment window effect is 0.026. These point estimates are statistically significant and meaningful, yet they are smaller in magnitude than for the DoD. The smaller effect of accelerated payments in non-DoD contracts is likely due to the clearer implementation of accelerated payments.

By Product Type

As previously discussed, accelerated payments policies should have little effect on construction contracts and many types of food contracts; these already are paid inside of 15 days, so a policy of accelerating payments will not be binding. This fact can be used as a placebo exercise to verify the results.

Table 6. Log Participation by Firm-Month, DoD Contracts

	(1) All firms	(2) Have backlog	(3) No backlog
SB*DoD Treament for SB	-0.0059***	0.026*	0.00054
	(0.0014)	(0.010)	(0.00037)
SB*DoD Treatment for all firms	-0.0028	0.064***	0.00049
	(0.0015)	(0.013)	(0.00039)
Observations	7195104	540386	6649598
R-Squared	0.45	0.66	0.10

The dependent variable is the log of one plus the number of contracts won by a firm in a month. The estimation sample is 25% random sample of the firms ever receiving contracts between 2008 and 2015. The variable SB is the percentage of times the firm participated as a small business from 2008 and 2015.
In parenthesis is the standard error clustered by firm.
*,**,*** denote significance at the 10%, 5%, and 1% level, respectively, and refers to inference using the asymptotic standard error.

In Table 8, we present the results of estimating our base specifications separately for contracts for affected versus unaffected products. In the first three columns, we present the results for unaffected food and construction contracts. For these contracts, the effect of accelerated payments is extremely small for the average small business. The point estimates of the coefficient on the interaction term between the small business indicator and the first and second treatment windows are 0.00025 and 0.00018. These coefficient estimates are small and precisely estimated. In Column 2, we present the estimates for firms with backlog. We see that there is an effect of accelerated payments on the participation of small businesses for food and construction contracts. The effect is not as large as the main effect found earlier, but it is noticeable. The estimated treatment effect of accelerated payments on the participation of small businesses on these contracts is 0.0064 during the first treatment window and 0.011 during the second treatment window. It is worth noting that even though these products did not directly benefit from accelerated payments, it is still possible that firms with backlog could benefit. Such firms would more quickly receive payment for the backlog contracts. In Column 3 of Table 8, we find that small businesses without backlog had virtually the same likelihood of contract participation during the accelerated payments period as when it was not in place.

In Columns 4 through 6 of Table 8, we present the results for contracts for non–food and construction products—those directly affected by the change in payments policy. We see that the effects of accelerated payments are strongest for these types of contracts, and in particular for participation by firms with backlog. Small businesses with backlog experienced an increase in participation by 0.022 log points during the first treatment window and 0.056 log points during the second treatment window. It is not surprising that these values are close to the main findings presented in Table 6, as most contracts are not for construction products or for the subset of food products that already had accelerated payments. The important point is that the effect for the affected products is approximately 5 times as large as the effect for unaffected products.

Table 7. Log Participation by Firm-Month, Non-DoD Contracts

	(1) All firms	(2) Have backlog	(3) No backlog
SB*Non-DoD Treament for SB	-0.0052***	0.011*	-0.0024***
	(0.00075)	(0.0049)	(0.00042)
SB*Non-DoD Treatment for all firms	-0.0024***	0.026***	-0.0010**
	(0.00072)	(0.0069)	(0.00031)
Observations	7195104	856910	6330678
R-Squared	0.28	0.49	0.081

The dependent variable is the log of one plus the number of contracts won by a firm in a month. The estimation sample is 25% random sample of the firms ever receiving contracts between 2008 and 2015. The variable SB is the percentage of times the firm participated as a small business from 2008 and 2015.

In parenthesis is the standard error clustered by firm.

*,**,*** denote significance at the 10%, 5%, and 1% level, respectively, and refers to inference using the asymptotic standard error.

Table 8. Log Participation by Firm-Month, DoD Contracts

	Food/Constr.			Non-Food/Constr.		
	(1) All firms	(2) Have backlog	(3) No backlog	(4) All firms	(5) Have backlog	(6) No backlog
SB*DoD Treament for SB	0.00025	0.0064**	0.000088	-0.0062***	0.022*	0.00045
	(0.00028)	(0.0024)	(0.000086)	(0.0013)	(0.010)	(0.00036)
SB*DoD Treatment for all firms	0.00018	0.011***	0.00011	-0.0031*	0.056***	0.00038
	(0.00047)	(0.0031)	(0.00011)	(0.0015)	(0.013)	(0.00037)
Observations	7195104	540386	6649598	7195104	540386	6649598
R-Squared	0.32	0.66	0.13	0.45	0.65	0.098

The dependent variable is the log of one plus the number of contracts won by a firm in a month. The estimation sample is 25% random sample of the firms ever receiving contracts between 2008 and 2015. The variable SB is the percentage of times the firm participated as a small business from 2008 and 2015.
In parenthesis is the standard error clustered by firm.
*,**,*** denote significance at the 10%, 5%, and 1% level, respectively, and refers to inference using the asymptotic standard error.

Restricting to Active Firms

One concern with the results just presented is that many firms are not active at a particular point in time. Inactive firms will not be affected by accelerated payments, and the overall effect of accelerated payments may be larger than what we estimated. In this section, we restrict the estimation sample to only active firms, which we define as firms who won at least one contract within the same year as the sample observation. In other words, if a firm did not win a contract in 2012, we do not include that firm's 2012 observations in the regression.

In Table 10, we present the base results for the DoD using this restricted sample, separately for food/construction and for other types of goods. The results mirror the estimates presented in Table 8 but are larger in magnitude. The average effect of accelerated payments is estimated to be small and insignificant for food and construction products for the average firm. For firms with backlog, there is a statistically significant increase in small business participation. Firms without backlog do not witness an increase in participation for food or construction contracts.

In the final three columns of Table 10, we present similar estimates for non–food or construction contracts. For the average active firm, the estimated effect of accelerated payments is small in magnitude. The active firms with backlog have a substantial increase in participation, particularly in the second transfer window. This estimate is larger in magnitude than the estimate for all firms. Active small businesses without backlog do not witness an increase in contract participation.

Offers on Set-Aside Contracts

In this section, we examine the number of offers by small businesses on set-aside contracts. If accelerated payments benefit small businesses, we expect that the desire to participate in auctions increases. In general, it is not possible to determine the number of bids submitted by small businesses by action. Only the total number of offers can be determined. With small business set-asides, all bids are presumably from small businesses. By examining the number of offers for set-asides, we can then determine whether the number of bids submitted by small businesses increases when accelerated payments are in place.

Table 9. Log Participation by Firm-Month, Non-DoD Contracts

		Food/Constr.			Non-Food/Constr.	
	(1) All firms	(2) Have backlog	(3) No backlog	(4) All firms	(5) Have backlog	(6) No backlog
SB*Non-DoD Treament for SB	0.0000087	-0.00045	-0.000023	-0.0052***	0.028***	-0.0027***
	(0.00014)	(0.00035)	(0.00015)	(0.00074)	(0.0053)	(0.00060)
SB*Non-DoD Treatment for all firms	0.00033**	0.00017	0.00029*	-0.0027***	0.048***	-0.0018**
	(0.00012)	(0.00031)	(0.00013)	(0.00071)	(0.0082)	(0.00056)
Observations	7195104	540386	6649598	7195104	540386	6649598
R-Squared	0.34	0.15	0.35	0.28	0.57	0.18

The dependent variable is the log of one plus the number of contracts won by a firm in a month. The estimation sample is 25% random sample of the firms ever receiving contracts between 2008 and 2015. The variable SB is the percentage of times the firm participated as a small business from 2008 and 2015.
In parenthesis is the standard error clustered by firm.
*,**,*** denote significance at the 10%, 5%, and 1% level, respectively, and refers to inference using the asymptotic standard error.

Table 10. Log Participation by Firm-Month, DoD Contracts, Active Firms

		Food/Constr.			Non-Food/Constr.	
	(1) All firms	(2) Have backlog	(3) No backlog	(4) All firms	(5) Have backlog	(6) No backlog
SB*DoD Treament for SB	0.00018	0.014**	0.000077	-0.0064***	0.025	0.00086
	(0.00034)	(0.0050)	(0.00012)	(0.0015)	(0.018)	(0.00053)
SB*DoD Treatment for all firms	0.00014	0.019**	0.00015	-0.0034	0.081**	0.00096
	(0.00062)	(0.0065)	(0.00016)	(0.0019)	(0.025)	(0.00054)
Observations	5552407	236316	5312924	5552407	236316	5312924
R-Squared	0.42	0.79	0.20	0.58	0.76	0.17

The dependent variable is the log of one plus the number of contracts won by a firm in a month. The estimation sample is 25% random sample of the firms ever receiving contracts between 2008 and 2015. The variable SB is the percentage of times the firm participated as a small business from 2008 and 2015. Active firms are those that participated in at least one other contract during that calendar year.
In parenthesis is the standard error clustered by firm.
*,**,*** denote significance at the 10%, 5%, and 1% level, respectively, and refers to inference using the asymptotic standard error.

One disadvantage of this approach is that it cannot account for substitution across set-aside contracts and those with open competition. In particular, accelerated payments may allow a small business to participate to a greater extent in open competition auctions. However, we expect that accelerated payments should lead to greater bidding participation in both set-aside and non–set-aside contracts rather than a substitution between the two.

Table 11 presents the results. We do not find evidence of an increase in the offers for set-aside contracts. This is true when considering all federal set-aside contracts or when estimating the effect specifically for DoD contracts. In fact, the opposite held true during the second transfer window. For DoD set-aside contracts, fewer offers were made during the second treatment window than during the time when accelerated payments were not in effect. To the extent that accelerated payments increase the desire of small businesses to compete for government contracts, this result suggests that small businesses are substituting toward the open competition auctions.

Table 11. Number of Offers for Contracts Set-Aside for Small Business

	All set-aside contracts				DoD set-aside contracts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SB treatment*Affected product	0.15	0.060	-0.028	-0.035	-0.14***	-0.14***	-0.038	-0.032
	(0.20)	(0.20)	(0.024)	(0.024)	(0.034)	(0.034)	(0.044)	(0.037)
LB treatment*Affected product			-0.0017	-0.0017			-0.16***	-0.17***
			(0.016)	(0.016)			(0.035)	(0.029)
Food	7.37***	7.32***	1.02***	1.01***	-0.45***	-0.45***	-0.49***	-0.48***
	(0.32)	(0.33)	(0.036)	(0.036)	(0.040)	(0.040)	(0.036)	(0.037)
Construction	0.42***	-0.063	0.065***	0.033	0.064**	0.064**	0.077***	0.051*
	(0.16)	(0.16)	(0.021)	(0.021)	(0.025)	(0.025)	(0.024)	(0.026)
Food*DoD	-9.91***	-9.40***	-1.35***	-1.32***				
	(0.36)	(0.36)	(0.043)	(0.045)				
Constr*DoD	1.87***	1.66***	0.085***	0.075**				
	(0.30)	(0.24)	(0.029)	(0.029)				
Log obligated amount		0.23***		0.015***				0.0096**
		(0.018)		(0.0029)				(0.0042)
IDV	1.21***	1.25***	0.22***	0.22***	0.23***	0.23***	0.23***	0.23***
	(0.090)	(0.074)	(0.016)	(0.014)	(0.026)	(0.026)	(0.026)	(0.024)
Observations	1496174	1484886	1491331	1480098	897286	897286	897286	892136
R-Squared	0.055	0.061	0.074	0.076	0.026	0.026	0.026	0.027

Standard errors corrected for clustering at the year*month level are in parentheses. *,**,*** denote significance at the 10%, 5%, and 1% level, respectively.
The unit of observation is a contract action that is set aside for small business, including any set aside where size was a criteria, such as set asides for Disadvantaged Business Enterprises and firms in Hubzones.
SB treatment is the period of time in which payments to small business contractors were supposed to be accelerated, with a similar definition for "LB treatment." Affected products are not construction contracts or most types of foods, both of which were already subject to a payment period of less than 15 days.
All specifications include year*month fixed effects.

Conclusion

In this report, we consider the impact of a federal procurement policy that accelerated payments to contractors. The policy was initiated by the DoD, first applying only to small business contractors. It was later adopted by all federal agencies and subsequently extended to all federal contractors regardless of size. Reducing the time between invoice and payment is desirable for contractors because of the lag between when costs are incurred and payments are received. Firms rely on internal and external sources of capital to fill this gap. This poses particular challenges for small businesses, which are likely to have lower cash reserves and less access to inexpensive credit. The impact of accelerated payments will likely be largest for this set of firms, allowing them to be more competitive for contracts and take on additional work.

Our findings indicate that small businesses participated in more contracts during the time when accelerated payments were in place. The estimated effect was stronger for DoD contracts, for which the adoption of accelerated payments was apparently more widespread. Our empirical design exploits the fact that accelerated payments did not affect all products equally; invoice payments for contracts for perishable foods and construction services were already accelerated, and the policy should not affect the payment of these goods. We find that the modest rise in small business participation after the introduction of accelerated payments was observed only in contracts not involving food or construction.

Our findings lend support to the contention that small businesses benefit from accelerated payments. More generally, our findings are also consistent with the hypothesis that liquidity constraints pose a greater challenge for small businesses, which suggests that policies such as set-asides that direct more contracts to small businesses may be more effective if coupled with policies that alleviate constraints faced by small businesses.

Further research is called for along two dimensions. First, our conclusions would be bolstered by evidence of how the payment behavior of agencies responded to the accelerated payments policies. We currently have only indirect evidence on this point. Second, evaluating the costs and benefits of accelerated payments would be a key input into policy discussions. Conducting such an analysis may require an understanding of the long-run effects of policies on small businesses. If the survival and growth of firms is enhanced by accelerated payments, then this improves the operation of federal procurement markets and should therefore be counted among the benefits of the policy. This is a nontrivial exercise that is outside the scope of this report.

References

Assad, S. D. (2011, April 27). *Class deviation: Requirement for accelerated payments to small businesses* [Memorandum]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/USA001787-11-DPAP.pdf

Balat, J. (2014). *Highway procurement and the stimulus package: Identification and estimation of dynamic auctions with unobserved heterogeneity*.

Bates, T. (1997). *Race, self-employment, and upward mobility: An illusive American dream*. Woodrow Wilson Center Press.

Birch, D. L. (1979). *The job generation process*. MIT Program on Neighborhood and Regional Change.

Blanchflower, D. G., Levine, P. B., & Zimmerman, D. J. (2003). Discrimination in the small-business credit market. *Review of Economics and Statistics*, *85*(4), 930–943.

Cavalluzzo, K., & Wolken, J. (2005). Small business loan turndowns, personal wealth, and discrimination. *The Journal of Business*, *78*(6), 2153–2178.

Chatterji, A. K., Chay, K. Y., & Fairlie, R. W. (2014). The impact of city contracting set-asides on black self-employment and employment. *Journal of Labor Economics*, *32*(3), 507–561.

Denes, T. A. (1997). Do small business set-asides increase the cost of government contracting? *Public Administration Review*, *57*(5), 441–444.

DoD. (2019). Defense Federal Acquisition Regulation Supplement: Prompt payments of small business contractors. *Federal Register, 84*(105), 25225–25227.

Fairlie, R. W., & Krashinsky, H. A. (2012). Liquidity constraints, household wealth, and entrepreneurship revisited. *Review of Income and Wealth*, *58*(2), 279–306.

Fairlie, R. W., & Robb, A. M. (2007). Why are black-owned businesses less successful than white-owned businesses? The role of families, inheritances, and business human capital. *Journal of Labor Economics*, *25*(2), 289–323.

FAR 52.232-25, Prompt Payment (2021). https://www.acquisition.gov/far/52.232-25

FAR 52.232-27, Prompt Payment for Construction Contracts (2021). https://www.acquisition.gov/far/52.232-27

Ginman, R. (2011, June 28). *Class deviation: Update to accelerated payments to small businesses* [Memorandum]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/USA003585-11-DPAP.pdf

Ginman, R. (2012, August 15). *Class deviation: Providing accelerated payment to small business subcontractors* [Memorandum]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/USA004783-12-DPAP.pdf

Ginman, R. (2013, February 21). *Rescission of class deviation: Providing accelerated payment to small business subcontractors* [Memorandum]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/USA001002-13-DPAP.pdf

Glazer, N., & Moynihan, D. P. (1970). *Beyond the melting pot: The Negroes, Puerto Ricans, Jews, Italians, and Irish of New York City*. MIT Press.

Holtz-Eakin, D., Joulfaian, D., & Rosen, H. S. (1994). Entrepreneurial decisions and liquidity constraints. *The Rand Journal of Economics*, *25*(2), 334.

Holtz-Eakin, D., & Rosen, H. S. (2005). Cash constraints and business start-ups: Deutschmarks versus dollars. *Contributions to Economic Analysis & Policy*, *4*(1), 1–26.

Hubbard, T. P., & Paarsch, H. J. (2009). Investigating bid preferences at low-price, sealed-bid auctions with endogenous participation. *International Journal of Industrial Organization*, *27*(1), 1–14.

Jofre-Bonet, M., & Pesendorfer, M. (2003). Estimation of a dynamic auction game. *Econometrica*, *71*(5), 1443–1489.

Krasnokutskaya, E., & Seim, K. (2011). Bid preference programs and participation in highway procurement auctions. *American Economic Review*, *101*(6), 2653–2686.

Krishna, V. (2009). *Auction theory*. Academic Press.

Lew, J. J. (2011, September 14). *Accelerating payments to small businesses for goods and services* [Memorandum]. https://www.crowell.com/files/2011-Executive-Office-of-the-President-new-policy.pdf

Marion, J. (2007). Are bid preferences benign? The effect of small business subsidies in highway procurement auctions. *Journal of Public Economics*, *91*(7), 1591–1624.

Marion, J. (2009). How costly is affirmative action? Government contracting and California's Proposition 209. *The Review of Economics and Statistics*, *91*(3), 503–522.

Marion, J. (2011). Affirmative action and the utilization of minority- and women-owned businesses in highway procurement. *Economic Inquiry*, *49*(3), 899–915.

McDuff, C. (2019, April 19). *Accounts payable MOCAS update.* Defense Finance and Accounting Service. https://www.dfas.mil/Portals/98/Documents/Contractors-Vendors/DIL_MOCAS-AP-Update.pdf?ver=2020-02-29-143109-110

Nakabayashi, J. (2013). Small business set-asides in procurement auctions: An empirical analysis. *Journal of Public Economics*, *100*, 28–44.

Prompt Payment Act, 31 U.S.C. § 3902 (1988). https://uscode.house.gov/view.xhtml?path=/prelim@title31/subtitle3/chapter39&edition=prelim

Rosa, B. (2020). Affirmative action subcontracting regulations in dynamic procurement auctions. *International Journal of Industrial Organization*, *72*. https://doi.org/10.1016/j.ijindorg.2020.102657

Assessing the Reliability of the Future Years Defense Program and Building a Forecast

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 in applied economics from the Johns Hopkins University and a BA in social studies from Harvard University. [ahunter@csis.org]

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, as well as a BS in computer science, from the University of Maryland. [gsanders@csis.org]

John Severini—is a graduate student at Columbia University's School of International and Public Affairs, where his research focuses on international security policy and quantitative methods. He currently works for the Arnold A. Saltzman Institute of War and Peace Studies on projects primarily dealing with maritime military issues. Severini holds a BA in political science and in linguistics from The Ohio State University. He can be reached at [john.severini@columbia.edu]

Contributing Authors: **Chitrakshi Bhardwaj, Gabriel Coll, Cuong Nguyen, Gerhard Ottehenning**

Abstract

Discerning, negotiating, and communicating priorities are necessary tasks for the U.S. defense acquisition system to effectively implement its portion of the National Defense Strategy. One of the Department of Defense's central tools for doing so is the Future Years Defense Plan (FYDP), a projection of the cost and composition of the force over the next 5 years. This project created a dataset and employs it to study FYDP reliability, focusing on two sources of uncertainty: differences in approach between military departments and differences in volatility between those line items in the base budget and those that include contingency operation spending.

Introduction

For the U.S. defense acquisition system to properly implement its portion of the National Defense Strategy, it must effectively discern, negotiate, and communicate its priorities. One of the Department of Defense's (DoD's) central tools for this process is the Future Years Defense Plan (FYDP), a projection of the cost and composition of the force over the next 5 years.

Annually updated and submitted as part of the president's budget submission projection, the FYDP provides important insights into the DoD's priorities and projections of the future, both internally and externally. Internally within the DoD, wherein the FYDP is constructed, the process forces the stakeholders involved to debate tradeoffs and outline their visions of the future. Externally, it lays out for Congress a vision of how U.S. national security strategies could be implemented in practice, which the legislature must then choose whether to fund or alter. It helps the U.S. defense industry understand where the DoD plans to invest and thereby allows companies within the industry to align themselves with current priorities. It helps scholars identify trends and do research on major capital-intensive projects, which can be used to inform future projects, both defense and nondefense. It helps U.S. citizens identify how the government plans to spend its taxpayers' dollars. However, the FYDP has a few major drawbacks for these stakeholders that undercut its ability to communicate priorities.

The first drawback is the inherent tension between the FYDP's role expressing the funding amount that the executive branch deems necessary to support the strategy and its role in creating a plan that can be implemented within the funding amount authorized and appropriated by Congress.

A second, related, shortfall is the absence of any measure of reliability or predictive intervals for the projections. Some parts of the DoD budget are easier to predict than others, but the point estimate provided by the FYDP does not differentiate between known quantities, like the purchase of uniforms, and cutting-edge technology, like the development of a next-generation alloy. That said, by design, Overseas Contingency Operations (OCO) budgets operate as a pressure valve for uncertainty by taking some of the most volatile spending out of the base budgets and FYDP and managing them through OCO methods instead. However, the intended functionality of the OCO accounts is muddled when predictable spending is moved to OCO accounts to avoid budget caps.

Third, the unclassified FYDP is released in a form that makes it straightforward to study topline spending or individual line items or programs but challenging to analyze anything in between. This is because the FYDP is released in dozens of PDFs through separate justification books, and not as a centralized database or even in summary documents. Collectively, these limitations present a higher barrier to entry to stakeholders and make it laborious for specialists and unappealing for anyone else to put investment plans in a meaningful context.

The FYDP is a system for planning rather than a forecasting tool, but there are nonetheless multiple benefits to understanding the relationship between its projections and actual spending. First, stakeholders can better employ the system and its results if its strengths and biases are more transparent. Second, this analysis can put common assumptions to the test, for example, the volatility of OCO spending or if the long-term Navy Shipbuilding planning process results in more accurate projections of future needs. Third, for the defense industry, the difference between projected and actual spending can be a key indicator of risk and aid in investment planning. Finally, defense spending must often respond to external changes and updated strategies; large gaps can indicate not just failure of prediction but also speed in adapting the larger defense enterprise to new priorities.

This project has created a dataset to ease research of the FYDP and uses that dataset to address the question, *How reliable are projections within the FYDP as an indicator for actual spending?* It tests the value of the unclassified FYDP for investment spending, RDT&E and Procurement, as a bottom-up indicator of DoD priorities by comparing cumulative projections through fiscal year (FY) 2019 spending from 1, 3, and 5 years in advance with the actual cumulative spending. In addition, the paper examines whether there are differences between military departments and between line items that do and do not include OCO spending.

Literature Review

There are a multitude of challenges in defense planning even within the base budget. The United States, despite its resources and robust analytical staff, faces more difficult challenges than those of many of its peer countries. First, the United States is a presidential system with projections prepared by the executive branch but funding authority resting with Congress. A projection process could be designed that does more to incorporate congressional opinions into the planning process, but the role of two co-equal branches of government means that some degree of uncertainty for both the topline funding and for individual projects of interest to the legislature is irreducible.

Second, the United States is the global leader in defense research and development, and as Light et al. (2017) find, "there is a considerable amount of cost and schedule growth risk facing all [Major Defense Acquisition Programs] at [Milestone] B" when the DoD commits to significant development spending (p. 44). Even a better estimation approach would be highly unlikely to eliminate uncertainty in defense research and development.

In analyzing the 2020 President's Budget (PB2020), Matthew Woodward and David Arthur (2019) draw on the FYDP and project specific reporting. Employing historical factors developed from studies by the RAND Corporation and the Institute for Defense Analysis, including those referenced previously, they find that "using the resulting cost estimates instead of DoD's cost estimates raises total projected acquisition costs by 3.5 percent over the FYDP period and by 6.1 percent over the 2026–2035 period" (p. 16).

The relationship between these project estimates and the larger FYDP projections is complex, with estimated cost influencing budget requests and a sense of total available funds influencing what the budget is able to fund and at what level. For example, the CBO estimates explore the cost implications of the President's Budget by keeping present plans constant. In practice, MDAPs and other budget lines can be descoped, slowed down, or canceled outright.

Topline FYDP projections should not be treated simply as a sum of composite budget lines but as a consequential form of estimating in their own right. As Todd Harrison and Seamus Daniels (2020) note,

> Previous inflection points in the defense budget, both up and down, have been influenced by wars, shifts in strategy, changes in the threat environment, and economic conditions. ... While there does not appear to be appetite to cut the defense budget in the remainder of FY 2020 or in FY 2021, as is evident by the inclusion of additional defense funding in stimulus bills, the political environment could shift markedly once an economic recovery is underway in FY 2022 or FY 2023. (pp. 57–58)

While the pandemic makes the present particularly challenging to forecast, the problem has never been an easy one for the DoD. Kevin Lewis (1994) finds that plans routinely expect small incremental changes, but in practice, changes, in topline spending or individual programs, are regularly more dramatic and often cyclical (pp. 110–113). Leland Jordan (2015) goes further and argues that historically most administrations project more funding than materializes, showing "systematic fiscal optimism" (p. 274). Jordan (2015) analyzed budgets from 1975 to 1995 and discovered that 70% of the projections exceeded the appropriated amounts (pp. 282–283).

Effective projection should introduce difficult choices in the present, before expensive commitments have been made, rather than in the future, where cost overruns or budget shortfalls may lead to the termination of programs experiencing difficulty or redirection of funds from programs that are presently successful. Jordan (2015) concludes that "those administrations having demonstrated the greatest bias in their real growth projections also most seriously handicapped program managers" (p. 288).

Enduring Budget in Overseas Contingency Operations

As mentioned in the introduction, OCO budgets acknowledge the existence of uncertainty due to external events such as wars or other cases where events beyond the control of the DoD require rapid responses. However, the distinction between the base budget and OCO has been undercut by budgetary maneuvering since the passage of the Budget Control Act (BCA) of 2011. In an attempt to reduce federal budget deficits, the BCA established spending limits on discretionary budget authority, applying to both defense and non-defense programs. These limits do not apply to the OCO budget. Subsequently, the executive and legislative branches have evaded the caps by transferring some predictable enduring spending out of the base budget and into OCO spending. This approach has been criticized by some defense experts and government officials from both parties. Katherine Blakeley and Lawrence Korb (2014) from the Center for American Progress voice their concerns that "financially, the free flow of war funding has decimated any pretense of fiscal discipline at the Pentagon. … Unclear budget guidance and poor financial management have allowed DoD to pay for substantial enduring costs with war funding rather than the base budget, further muddying the waters" (p. 28). Then Acting White House Chief of Staff Mick Mulvaney also criticized in strong terms the "use of OCO funding for base budget requirements" (McGarry & Epstein, 2019, p. 9).

The Congressional Budget Office (CBO) estimates that from 2006 to 2018, more than $50 billion in OCO funding per year (in 2019 dollars), on average, has gone toward the costs of enduring activities rather than the temporary costs of overseas operations. This is particularly transparent in the FY 2020 President's Budget. As Harrison and Daniels (2020) report, "the request shifted entire categories of funding, such as Army Ammunition Procurement, from the base budget into OCO" (pp. 4). With the coming expiration of the BCA caps in FY 2021, the FYDP released with the FY 2020 budget request then shifts all the money back to base spending after the expiration.

The insertion of enduring items within the OCO budget does not necessarily undermine the value of OCO in those cases where it is still used as intended. Andrew Hunter (2019) defends OCO by noting that in the last decade, a major portion of OCO funds was used to support the operations of Afghan security forces. As the actual size, operational employment, and equipage of these forces have changed rapidly, OCO funds were extremely useful to sustain the mission. Due to the fact that the levels are not planned out a full 5 years in advance, OCO allows the performance of missions that might be practically impossible otherwise. Moreover, the origin of OCO was an attempt to bring more oversight and transparency to wartime emergency supplemental bills. Senator McCain, objecting to the inclusion of non-emergency procurement in a supplemental bill,

> demanded that DoD submit its request for war funding along with the regular budget so that it might receive a similar level of congressional scrutiny as the base budget and so that it would be less easy to embellish with non-war-related funding. (Hunter, 2019)

The Strategic and Communication Role of the Future Years Defense Plan

The FYDP helps the executive branch make strategic choices and communicates them to internal and external audiences. Critically, the FYDP is released as part of the President's Budget and thus linked to the negotiation between branches of the U.S. government. The numbers are provided to justify the funding requests the executive branch is making to Congress. As Todd Harrison and Seamus Daniels (2020) put it,

> The FYDP is therefore best understood as a statement of policy rather than a prediction of where the budget is headed. It is an indication, with considerable detail, of the Defense Department's priorities and trade-offs among modernization, force structure, and readiness. (p. 11)

This emphasis on policy and strategy provides another set of criteria that can be used to evaluate the effectiveness of the FYDP. In particular, Thomas-Duerrel Young (2018) is critical of long term defense planning as practiced in the United States. While other authors have emphasized the benefits of budgetary stability making efficient choices, Young believes this fails to acknowledge the way adversaries can unpredictably shape choices. Instead, Young highlights two tasks that defense planners can achieve: "to produce costed priorities" and "creating an understanding of future financial projection of current obligations" (p. 366). For Young, the quality of cost estimates matters for costing priorities, but the value of the FYDP is not to provide reliable predictions. Instead "the utility of these financial projections should be judged by how much flexibility they can provide ministers and senior defense officials to change the way money is being spent to produce relevant defense outcomes" (Young, 2018, p. 370).

While strategic flexibility is inherently appealing, it does still face challenges noted by Kevin Lewis (1994) in his argument for defense planning humility. Lewis (2014) observes a range of cyclical factors in the defense budget and a shrinking portion of the budget going to combatant forces. He cautions, "we should expect inefficiencies, and substantially more negative effects on capability than might be expected from these inefficiencies, because of the increasing role of 'fixed-cost' overhead items" (p. 132). That said, the larger emphasis on the FYDP's role in achieving flexibility harks back to Alain C. Enthoven and K. Wayne Smith (2005), who reject the idea that long-range plans limit the president's ability to implement a new strategy and argue that "an organization's flexibility to move in a new direction is greatly reduced if it lacks a clear picture of the direction in which it has been heading" (p. 50).

Young's (2018) emphasis on the strategic flexibility provided by the FYDP presents a challenge; while there has been extensive research comparing projections to actual costs, measuring the agility of the acquisition enterprise is a less intuitive problem. Will Domke (1984) provided one possible answer by analyzing how the Defense Budget responded to presidential priorities going back to the Eisenhower administration by analyzing the winners and losers among DoD funding accounts. He found that the balance between military departments change most at the start of an administration and more popular president's have greater influence (p. 389).[1]

Picking up on Domke's (1984) approach, Travis Sharp (2019) argues that what the current debate on gaps between strategy and resources "generally [does] not provide, however, is any objective sense of whether DoD is doing relatively better or worse aligning resources with strategy" (p. 9). Sharp (2019) considers three diagnostic tests including a Winners test that looks at whether spending has shifted into areas identified by the National Defense Strategy as priorities.[2] He looks at the FYDP's Major Program Categories and investment areas, with the latter only available through the President's Budget. In both cases, Sharp (2019) finds that the 2020 President's Budget failed this test (p. 24).

Data and Methods

Data Sources and Structure

This project focuses on budget lines rather than major programs for both financial and policy reasons. Todd Harrison (2016) reports that as of the FY 2016 President's Budget, "these smaller programs account for an average of 57 percent of the total acquisition budget over the FYDP" (p. 24). Moreover, during the study period, the DoD is increasingly experimenting with alternate channels, including mid-tier acquisition and other transaction authority, responding to pressure from the Executive Branch and Congress to pursue speed and innovation.

To better understand this era of reduced reliance on the major weapons system pipeline, this report chooses to focus its attention on procurement line items and RDT&E program elements. This is not the most detailed level of analysis available; however, it has the advantage of being available from multiple sources. The first pair of these sources are the P-1s for Procurement and R-1s for RDT&E. These budget documents are provided as spreadsheets that

[1] Domke (1984) does not include FYDP data in his analysis.

[2] "Instead, a winner is best defined as one that receives the largest increase in its proportional share of DoD's total spending compared to what DoD's prior outyear plans forecasted" (Sharp, 2019, p. 45).

cover the entire DoD enterprise, going back to the 1998 President's Budget (DoD Comptroller, 2020). However, these detailed and convenient documents do not include FYDP's out years.

Instead, for detailed future year projections at the procurement line item (P-40) and RDT&E program element (R-2) level, it is necessary to turn to the Justification Books. In these documents, the military departments and agencies lay out their spending request and describe what is being bought as well as providing program management details. A major challenge for open-source researchers is that for investment spending alone, each President's Budget is accompanied by dozens of PDF files, splitting the information based on organization and funding account.

This project overcomes the limitations of the justification books and does so with the benefit of two external sources. First, a predecessor FYDP analysis led by Gabriel Coll bulk laid the foundation for this project by downloading many of the justification books and conducting an initial analysis. A parallel effort by the CSIS's Defense Budget Analysis (DBA) group greatly accelerated this effort with the discovery that the justification books, starting with the FY 2013 FYDP, have included XML encoded spreadsheets that do not require the intensive data cleaning effort necessary when scraping PDFs. This past and parallel work assisted in the creation of the dataset, and, in the DBA case, provided a valuable source for cross checking totals and budget line classifications.

To allow for cross-comparisons, the team has imported R-1 and P-1 budget requests from FY 2011 to FY 2021, using the most recent files and most recent columns within those files. Much of the effort of the dataset focused on the creation of unique identifiers, called CSIS budget line keys, that ease the process of tracking a budget line across different sources and over time. While many of the same columns are available in the R-1s and R-2s as well as in the P-1 and P-40s, there are discrepancies in the labels used, which are more challenging in early years and with procurement data in particular. For example, in some years and for some agencies, the line number, that is to say, the order in which it is presented in that PB, is the same as a line item. This causes problems as P-1s and P-40s include slightly different budget lines, leading to misalignment, and because order regularly changes from year to year. Moreover, line item standards change over time, with many budget lines changing from having a six-character line number to a 10-character one while still having the same broad topical focus. For procurement, for the analysis in this paper, all cost types pertaining to a single program are combined.[3]

A related challenge is that in a typical year, a few score PEs and LIs will cease to receive funding or be tracked for the first time. This is a natural outflow of changes in strategy, priorities, and technology. However, some of the time, a new PE or LI does mean a genuinely new project but may instead reflect a change in identifiers for an existing budget line of greater magnitude than just an increase in the length of the identifier used. This may be a matter of a change of agency, such as the move from the Defense Health Agency to the Defense Health Program, or a reclassification, such as in the lead-up to the creation of the Space Force.

The ties between the original sources and their unique identifiers are recorded within the dataset's repository for transparency and reproducibility reasons. The study team has taken the additional step of classifying budget lines based on our confidence that potential confounding factors have been removed.

---

[3] In the early years of the dataset, Advanced Procurement LIs sometimes had a different line item than the rest of a program. These divergent advanced procurement budget lines have been combined by the study team with the rest of their program under the same CSIS budget line key.

Figure 3 shows a summary of the R-1 and R-2 portions of the dataset by confidence level in the left and right columns, respectively. The top row shows the total actual spending associated with the projections of each PB. For the R-1, this includes only the actual spending for the budget year in question. The R-2s show substantially more spending because their projection window covers the President's Budget and four out years. For those cases where projections go beyond FY 2019, the spending amount projected is shown in gray.

Focusing on those projections for which we know the actual spending, there are five confidence levels of interest:

- No FYDP Expected: This covers budget lines, such as classified spending, that are typically not included in the unclassified FYDP as reported by the R-2s or P-40s.
- Unanticipated Budget Line: This covers budget lines that did not yet exist when the PB was published. They may be a genuinely new budget line, or perhaps a transfer whose predecessor was not identified by the study team.[4]
- Not Confident: This category covers budget lines that are present in one source but not in the other.[5]
- Semi-confident: This category covers budget lines that are present in both sources but that have one of a range of known discrepancies. This includes having a difference projection between the two sources of more than $2,000. In the case where there is a missing cost type, budget lines are reported as semi-confident if there is a net difference between the two projections of $2,000 or less but the spending occurs in different years depending on the source. Finally, if a budget line projected spending in a future year, but the budget line was not reported in the year in question, it is marked as semi-confident.
- Confident: This is the category that is the basis of most of the charts in this paper. To be classified as confident, both sources must report projections within $2,000 of one another. A budget line may end during the projection period and still be classified as confident, but only if the ending was anticipated by the President's budget. These strict criteria are intended to limit the sample to those budget lines that are genuinely starting and ending rather than having overlooked connections.

Turning again to Figure 1, the larger light blue blocks present in the R-2 column indicate that there is a substantial amount spent by PEs that the FYDP does not see coming years in advance. The lower row of Figure 1 shows the same budget lines, but using the metric of the count of lines rather than the amount of actual spending.

[4] Unanticipated Budget Line is smaller in more recent years because it is a lagging indicator and only available once actual spend figures are known.
[5] For procurement, this also covers cases in which there is a cost type under the budget line, for example, advanced procurement, that is not present in the other source and there is a net difference in projection of more than $2,000 between the two sources.

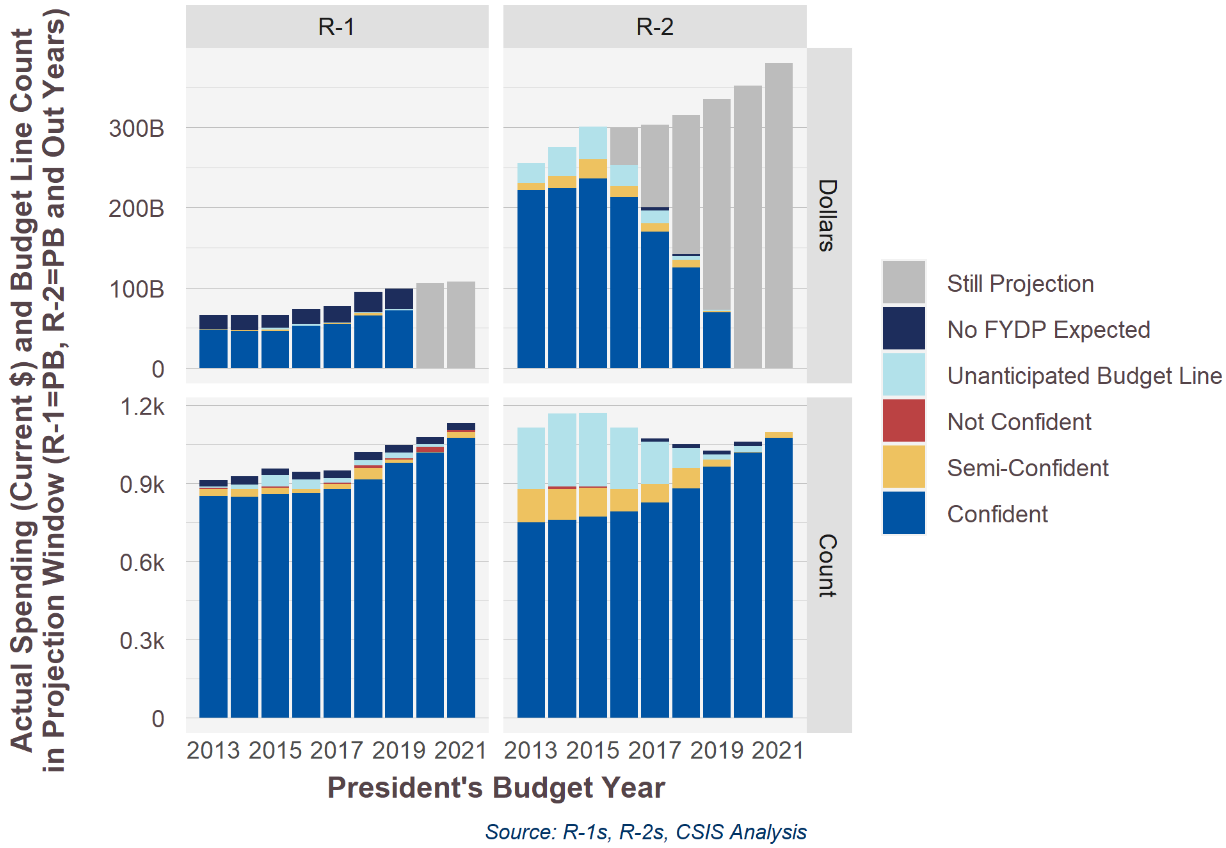


Figure 1. RDT&E Actual Spending and Budget Line Count by Level of Confidence

The Procurement dataset is summarized in Figure 2, and by the prevalence of red and yellow, shows that there are more disagreements between the P-1s and P-40s than between the R-1s and R-2s. There were multiple contributors to these problems. First, there were gaps in reporting on the P-40 side, though some of this the study team has already overcome by manually transcribing the PDF files that were missing XML files.[6] Second, cost type categories, particularly reductions to adjust for prior year past procurement, were sometimes missing from P-40 reporting and merit closer examination.

[6] The most problematic absence in dollar terms had been in the Navy Shipbuilding and Construction account. For some President's Budgets, key columns went unreported, but these were overcome by imputing the values using other available data. For both the R-2s and P-40s PB2014 and PB2015, the justification books did not include OCO spending with a note that those figures were to be released later. The study team imputed these values from the P-1s and R-1s. In PB2016, the total President's Budget spending column was missing from the P-40s, but the study team imputed that number by adding base and OCO spending together for each row. In PB2017, both base and OCO spending were amended after the R-2s and P-40s were published, and the study team again drew from the R-1s and P-1s to impute the amended values.

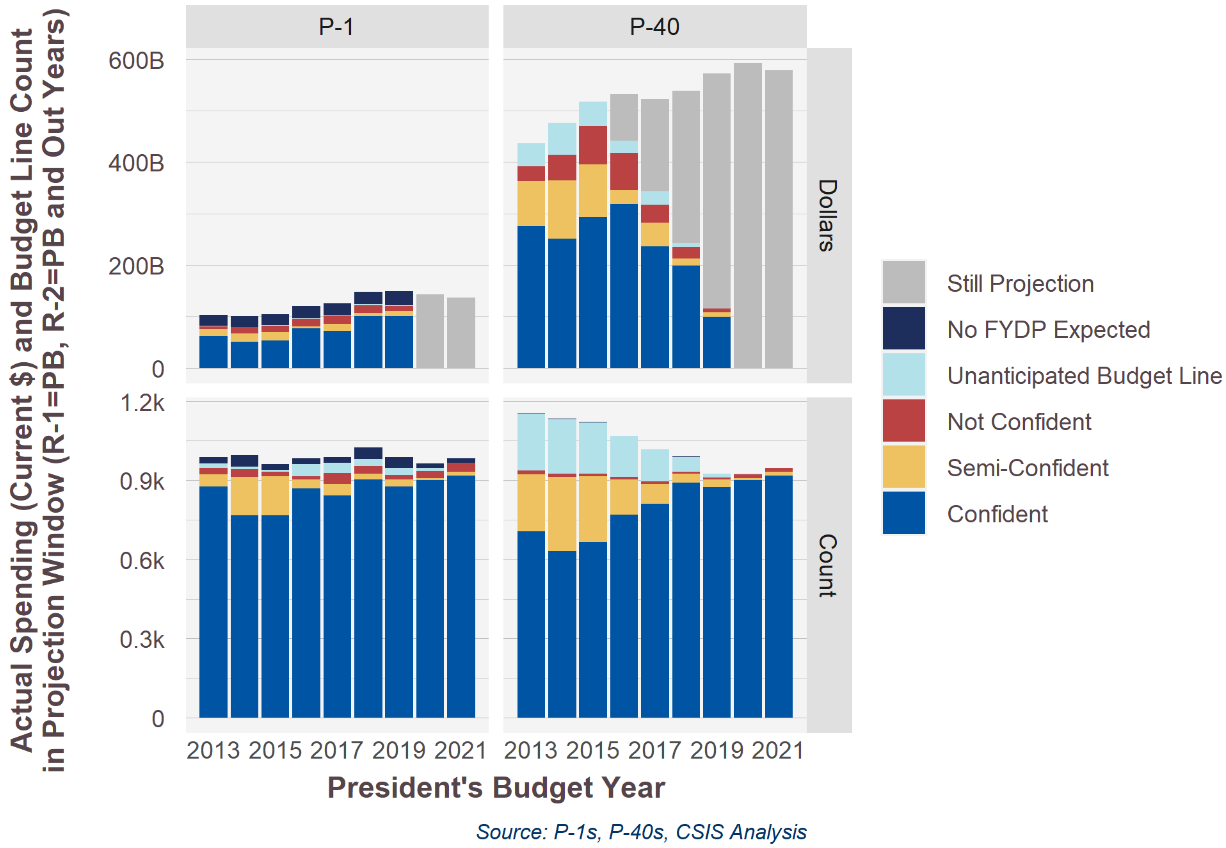


Figure 2. Procurement Actual Spending and Budget Line Count by Level of Confidence

All dollar amounts in this report represent current dollars. Although changes in spending timing (delays, shifts, etc.) occur throughout the time period, much of the FYDP estimations themselves have inflation concerns built into their reasoning. For statistical purposes, values across multiple-years have been aggregated into a cumulative expenditure; for example, analyzing the full 5-year window involves comparing the sum of 5 years of projection and 5 years of actual spending.

There are several tradeoffs to this methodology. The analysis is exchanging year-specific sensitivity for a more robust measurement better representative of the discrepancies being assessed over the time periods in question. This sacrifices the ability to directly compare the predictive strength of different FYDP years (i.e., testing the extent to which the second and third out years are more reliable than the fourth and fifth out years). Likewise, this approach makes it harder to account for any single year having an abnormal occurrence, such as sequestration. In return, the variables being compared serve as a much more accurate representation of what is being estimated in total. From an industrial standpoint, this serves as a more natural portrayal of how spending is being looked at overall.

Results

How Reliable Are FYDP Projections?

Figure 3 shows cumulative FY 2019 actual spending plotted against FYDP projections starting in 2015.[7] The Y-Axis is the Total Spend, the logged set of actual expenditures fully realized as outlays over given years. The X-Axis is the Projected Budget, the logged set of prior expenditure estimates for all budget lines expected to exist in those future actual spending years. The black diagonal lines represent the boundary where the projected budget is exactly equal to the total spend. Points to the upper left of the line have more spending than expected, and points to the lower right projected more spending than actually occurred. Points lying on the axes are the result of projected budget lines that ended up having no actual spend (x-axis) or budget lines appearing with actual spend that projected no spending or were unanticipated (y-axis). The graph's different columns correspond to different comparison periods: the PB2015–PB2019 graph includes 5 years of projections, while the PB2019–PB2019 graph includes only the first year with no out years. The upper row covers all budget lines, while the lower row includes only those budget lines where the study team is confident in the data quality.

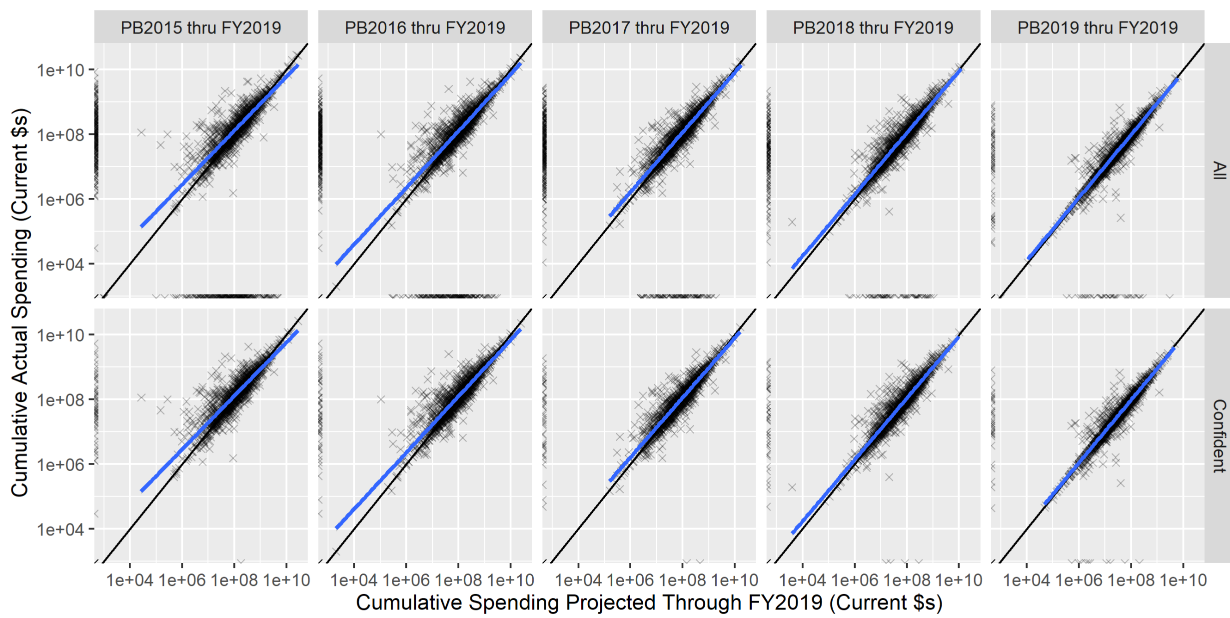


Figure 3. Scatter Plot showing FYDP projections versus Actual Total (Investment)

Each of the plots has a blue regression line of best fit that estimates the relationship between the projections and the actual spend. The closer the line is to the black diagonal line, the better the projections. When limiting the sample to those budget lines where the study team is more confident of the data's quality, as shown in the second row of Figure 3, the quality of the relationship strengthens.

Figure 4 compares the projected and total values shown in Figure 3 but uses a histogram to focus on the differences between the projected and total values. Only the 5-year, 3-year, and 1-year projections are shown in this and subsequent graphs to allow the display of more detail. The X-Axis denotes the relative difference between projected expenditure on any

---

[7] Due to the extreme range in spending across different items, all figures will employ logarithmically transformed axes for statistical purposes. Logging both sides results in more normalized sets, rather than most of the points falling into the lower left corner due to the great variation in size of budget lines.

single budget line and the actual spend on that same budget line in FY 2019. Relative difference is used to allow the scale to include cases where the projected or the actual spending figures are zero.[8] The Y-Axis denotes the count of budget lines with that level of difference.

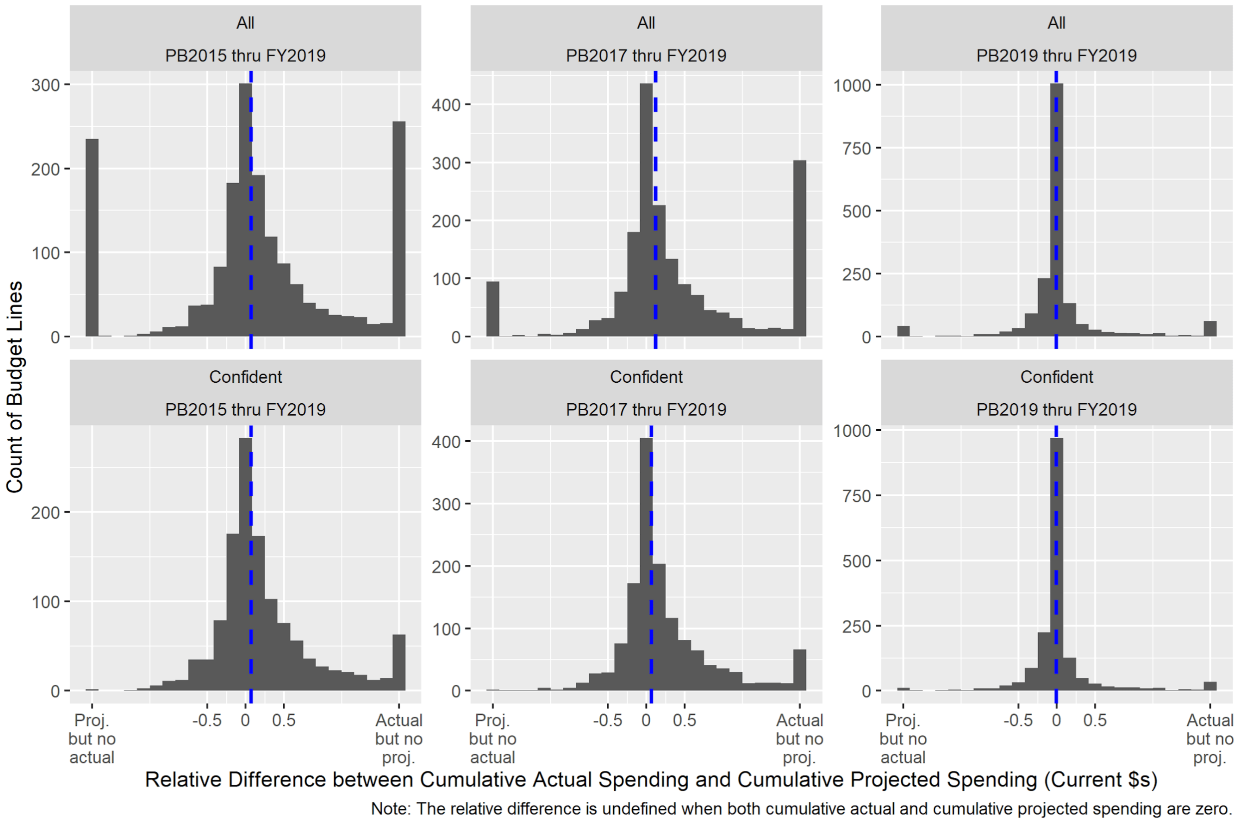


Figure 4. Budget Line Distribution by the Relative Difference Between Cumulative Projection and Actual Spending

For those points in the center of each histogram, the difference between actual and projected values is small. The chart's left side indicates budget line items where the projected value was much higher than what was eventually spent, and the right side of the chart shows budget line items where the estimated value was much lower than what ended up being spent. The groupings all display normal distribution, with the same previously mentioned 0 spend outliers showing on the edges, with the left edge representing cases where spending was projected but none took place and the right edge representing cases where zero dollars were projected, or the budget line was unanticipated by the PB, but spending took place nonetheless. The first row of the graph shows all data, and the second row shows only confident budget lines; note that far fewer budget lines fall at either end of the scale in the confident dataset.

Blue dashed lines indicate the X-Axis median. The closer the estimations are to the actual spending, the more accurate and center-oriented the estimations end up being. There is a distinct drop-off in overall accuracy from the PB year estimation to the out-year estimations, which results in significantly wider tails for the histograms that include more out-years.

---

[8] Relative difference, for the purposes of these charts, is calculated by dividing the numerator of (Cumulative Actual Spending - Cumulative Projected Spending) by the denominator of (Cumulative Actual Spending - Cumulative Projected Spending) / 2. Relative difference is not defined in those cases where both the actual spending and the projected spending values are zero.

Which Services and Budget Categories Have the Most and Least Reliable Projections?

This section examines whether the reliability measures discussed previously vary between OCO and base budget lines and when looking across military departments. For context, the count of P-40 and R-2 budget lines for each category is shown in Figure 5. The following analysis is limited to the confident budget lines, shown in dark blue. While most budget lines merit the confident description every year, the semi-confident and not confident categories can account for a reasonable amount of spending. For example, for the Navy, shipbuilding procurement line items, complicated by advanced procurement, have been a regular source of data import problems. While the number of items is small, those budget lines are of high value.

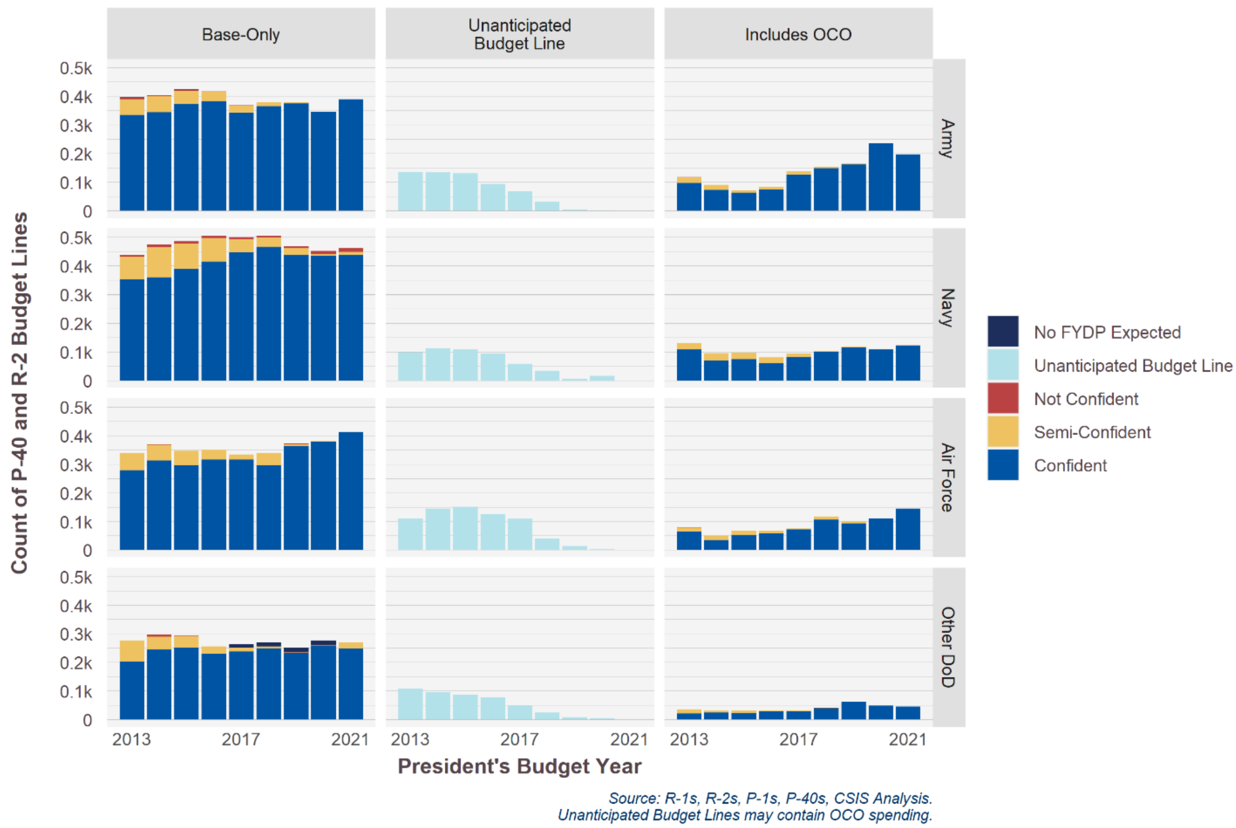


Figure 5. Budget Lines by Military Department and OCO

Influence of Overseas Contingency Operations

When considering the influence of OCO, this paper classifies budget lines based on whether they contained any enacted OCO spending or whether the President's Budget included any OCO spending. This method assumes that budget lines with planned or recent OCO spending may be more likely to have it in the future. A limitation of this method is that sometimes a budget line will add OCO spending even if it had not included it in the past. As is shown in Figure 5, the number of OCO line items experienced a small decline after 2013. However, since PB2016, the number of budget lines including OCO has increased across the military departments.

When examining changes in individual budget lines, as shown in Figure 6, the first year of estimates, PB2019 versus PB2019, shows little difference in distribution between base-only lines and other lines. However, in the 3-year and 5-year projections, shown in the middle and left columns, respectively, the difference is much starker. The first row's base-only budget lines have peaks centered around 0 proportional difference between projected totals and actual spending. In contrast, that is still a high point for the budget lines including OCO, but the frequency is notably lower, resulting in a more rounded dome shape. Instead, a greater portion of budget lines are spread to the right, indicating growth. This can also be seen in the dashed blue lines, which show the median growth and are shifted to the right, indicating a higher median increase in budget lines. Note that unlike in Figure 5, this graph is simplified by only displaying the confident budget lines due to their superior reliability for analysis.

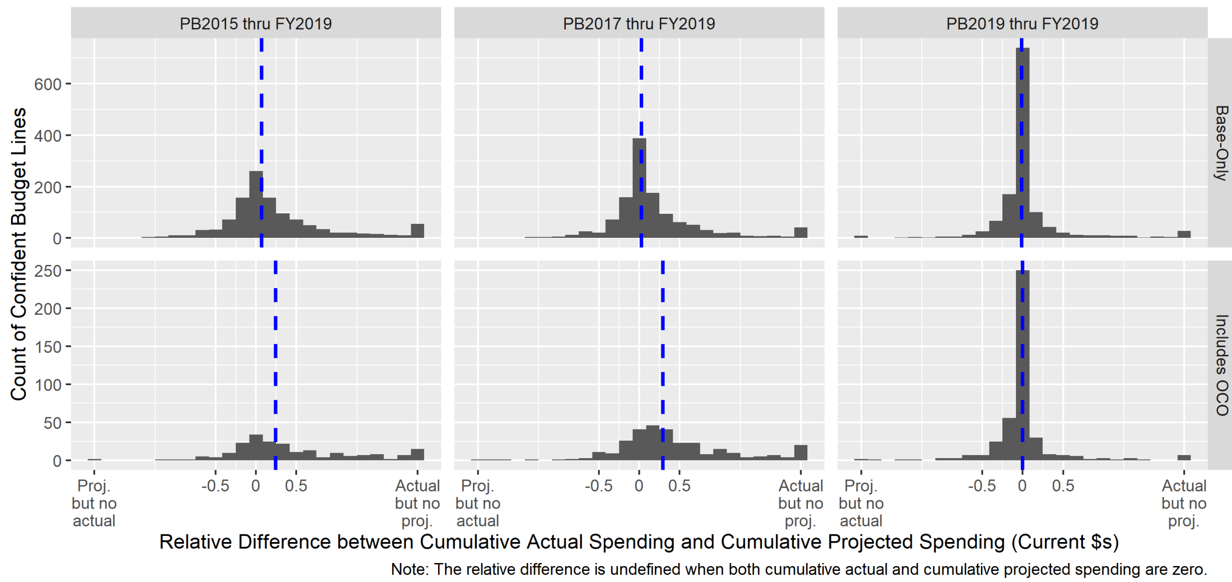


Figure 6. Budget Line Distribution by the Relative Difference Between FY 2019 Actual and FYDP Projection by OCO

Military Services

The differences between Military Services are more subtle than those between base-only and OCO including budget lines, and Figure 7 thus compares the median relative difference between projected and actual spending. Positive values are associated with more spending than projected, while negative values indicate projections exceeded spending. Each of the graphs covers a complete PB year, working from the President's Budget only on the left to the 5-year FYDP on the right. The Military Departments show similar trends within each PB, as assumptions about topline spending levels have widespread influence. PB2013 was devised under the explicit assumption that a budget deal would be reached that would have reduced the strictness of budget caps, a belief that proved optimistic, and as a result, actual spending was consistently below projections. PB2014 still proved optimistic in the early years, though by the fifth out-year, FY2018, Army and Air Force budget lines began to spend more than had been projected. The last three PBs of the Obama administration all proved to underestimate eventual spending. In the latter two budgets, the administration transition for the FY2018 budget leads to particularly stark changes. Finally, the PB2018 projections also proved to underestimate future spending, which may be in part the result of the limited attention top-level leadership in the incoming administration was able to give the FYDP.

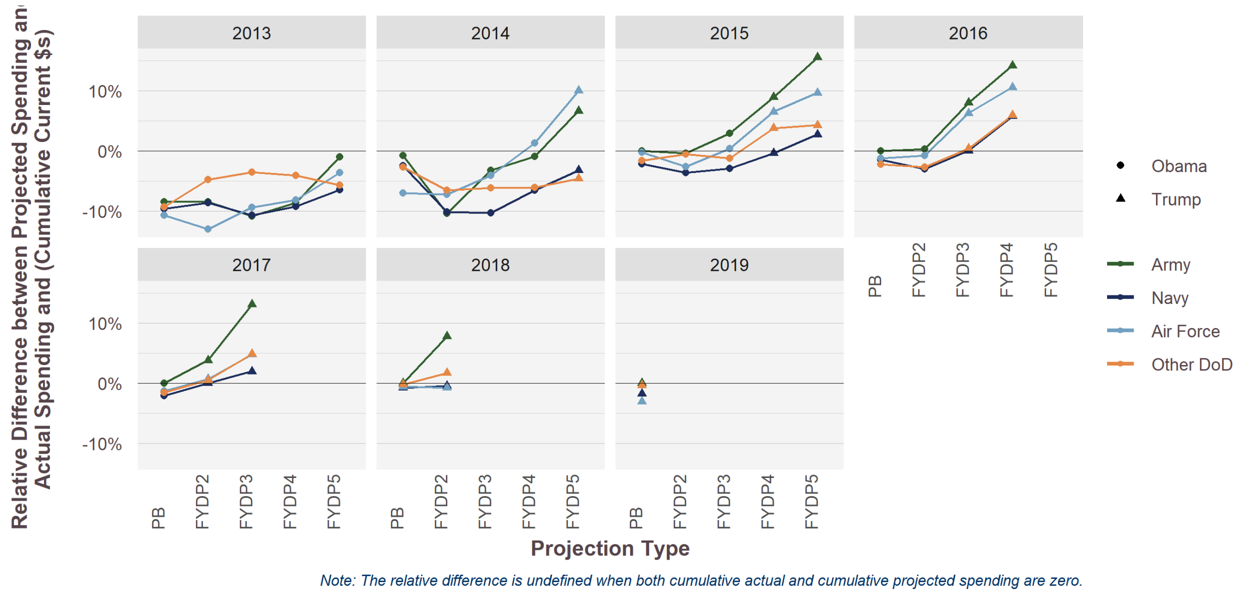


Figure 7. Median Relative Difference by Military Department in Confident Budget Lines

The Army, since PB2015, has had the highest relative difference, and is also the service that most relies on OCO spending. The Air Force has also had relatively higher shifts in median spending, although these Air Force estimates are more reliable in dollar terms, suggesting that Air Force changes are concentrated in lower value budget lines. The pairing of the two services is somewhat surprising, as the Army and Air Force make the most and least use of OCO, respectively.

To better understand the interplay between OCO inclusive budget lines and military department projection reliability, Figure 8 examines both. It shows that across the services, budget lines including OCO, shown in red, tend to underestimate out-year spending to a much greater extent than base budget lines. Indeed, the median relative difference for OCO including lines is repeatedly more than 20%. This unpredictability aligns with expectations, as OCO spending is only reported through the first year of the FYDP. So any OCO spending in subsequent years inherently results in the base budget figures underestimating the eventual funding. There are exceptions to this trend; for example, for the Army and Navy, the PB2013 OCO budget lines consistently overestimated the spending to a greater extent than did base budgets. This spending reduction may be due to the greater variability of OCO-related budget lines, making them a more likely target for cuts once the budget caps arrived in full force.

For both the Army and Air Force, their comparatively high medians shown in Figure 9 can be traced to those years in which OCO inclusive budget lines rocket above 20%. Both the Army and Air Force have been increasing the number of OCO involved budget lines. Interestingly, for the Air Force, this expansion has correlated with a reduction in the degree to which OCO budget lines underestimate actual future spending, suggesting that these new OCO-involved budget lines may be easier to predict. On the other hand, the Army stands out in PB2015–PB2017, as the base-budget lines show larger underestimation than any of the other services.

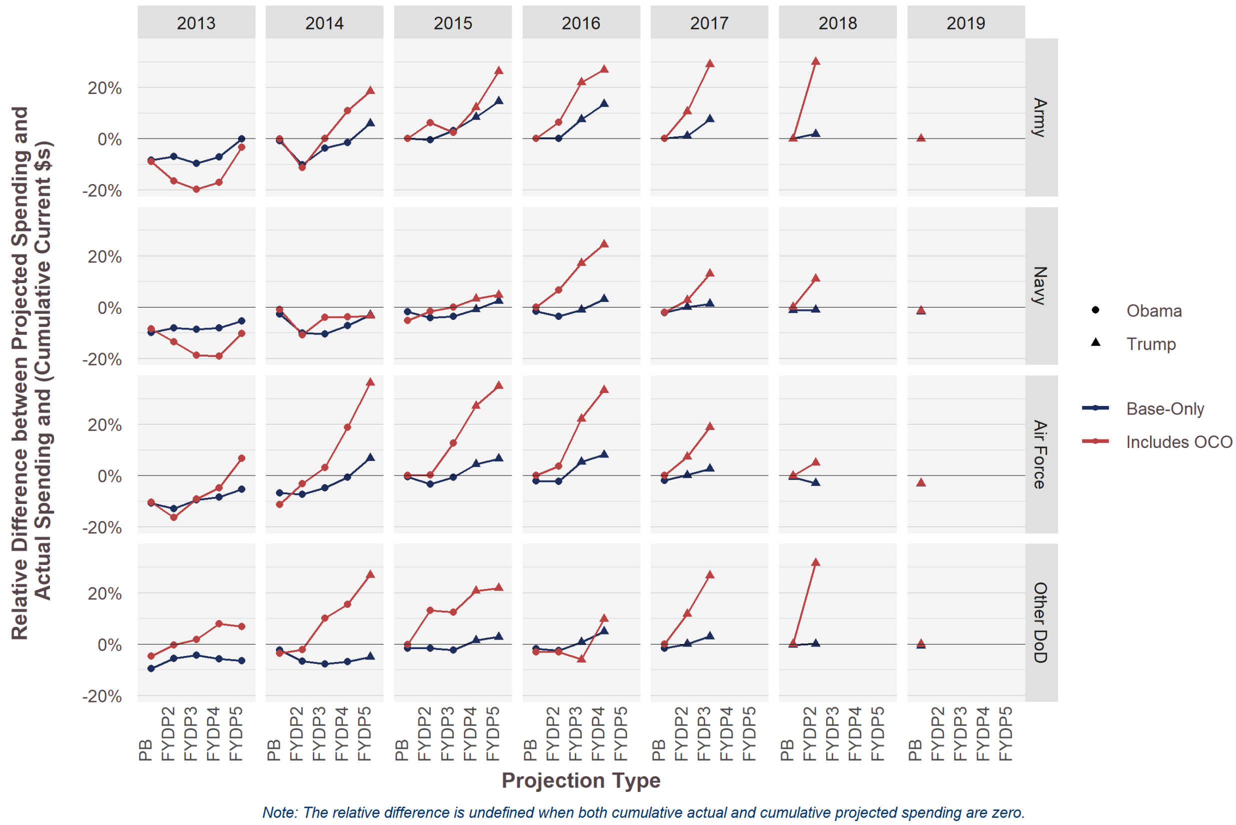


Figure 8. Investment Military Department Topline by OCO Cumulative Percent Difference

Discussion and Conclusions

How Reliable Are Projections Within the FYDP as an Indicator for Actual Spending?

Cumulative investment FYDP projections and actual spending highly correlate, albeit with generally reducing accuracy further out into the out years. Nonetheless, the bottom-up data does replicate previous findings regarding the top-level FYDP: There are notable differences between presidential budgets that with the relationship administration's approach to the budget caps and the level of support from Congress pushing toward overestimation of future funding in PB2013 and PB2014 switching to underestimation in the later years of the Obama administration.

How Does OCO Spending Relate to Projection Reliability?

In keeping with expectations, budget lines that include OCO had less reliable out-year projections compared to projections by base-budget lines. As Figure 4 shows, OCO budget lines typically underestimate future spending, with greater than projected actual expenditures occurring in the out years for which OCO does not project. That said, in PB2013, Army and Navy OCO budget lines were particularly optimistic and experienced more significant funding shortfalls than other budget lines. In part, this reflects OCO serving its intended purpose of allowing for more rapid changes and thus signifying budget lines that are more difficult to predict.

Which Military Departments Have the Most and Least Reliable Projections?

The Army, across multiple measures, has had the most substantial difference between total projected spending and actual spending, as shown in Figure 7. Separate analysis has confirmed that the correlation between Army projections and actual spending is lower than for the other services. As is shown in Figure 5, the Army is the predominant user of OCO-related budget lines.

What Are the Unclassified FYDP's Biggest Data Quality Problems?

The process of building the dataset highlighted problems that the FYDP presently faces that undercut the applications of data science approaches to develop better forecasts and for analysts to track spending plans to strategy. Most noteworthy is the absence of unique identifiers that make it easy to speak the same language across years and sources. Budget lines regularly appear, disappear, and change labels. This is not inherently a problem; budget lines are sometimes abruptly cut and a new, and perhaps unexpected priority, receives funding. More challenging are the cases where funding may have shifted to a different line or the Justification Books or Comptroller documents have budget lines not present in the other source. The FYDP is a tool intended to aid defense planning, which often means changing priorities, and thus shifts are not just expected, but desirable. However, the data's usability would be enhanced if the newly appearing budget lines reported predecessor lines in a machine-readable manner or indicated if they were a genuinely new initiative.

This project takes a step to increasing the usability of FYDP data in analysis by deriving unique identifiers from labels in the budget documents and managing discrepancies. This dataset's confidence labels are meant to be a stopgap that shows where contradictions are present between sources and track transitions between budget lines where identified.

How Could FYDP Projections be Improved?

The unclassified FYDP has the potential to be used as the basis for better projections. The study team held a workshop with leading practitioners and analysts to present initial results from this dataset. Participants made various valuable suggestions, including separately modeling base and OCO budget lines and looking for known budgeting foibles, such as zeroing out of accounts in the President's Budget that are subsequently restored in congressional enactments. Modeling experiments also found that it was helpful to separately model if a budget line was likely to be funded at all. Unfortunately, the bottom-up approach to modeling is only half of the challenge, as exogenous fiscal factors can matter more than the given budget line's specifics.

Is the FYDP Fulfilling its Purpose?

The period from PB2013 to PB2021, which covers both a significant downturn in spending and a large upswing, offers lessons with relevance beyond the budget caps' expiration. These results demonstrate a critical strategic communication limitation of the FYDP during this period. OCO allows greater flexibility but is not well suited to demonstrate commitment, as the PB2013 experience shows that future funding assumptions may not manifest and budget lines that included OCO spending took more significant cuts than those using the traditional FYDP process.

For those seeking budgetary tools offering both flexibility and fiscal discipline, Light et al. (2017) may offer a better alternative: reserve funds. For example, "funds might be held in reserve to address anticipated further growth within an MDAP portfolio (rather than for any single program)" (p. 44). Their proposal focuses on insulating the overall budget from project uncertainty. Still, if the reserve portfolio's scope matches an area of strategic focus, setting reserves aside could be a powerful priority-setting mechanism. Such dedicated funding could insulate a key strategic priority from cost increases or topline pressure and leave room for innovation even in the face of fiscal headwinds.

Well-designed reserve funds may offer a way to provide flexibility while mitigating uncertainty but would certainly need support from key congressional committees. A more minor change would be to embrace Travis Sharp's (2019) approach to measuring strategic shifts and to specifying priority areas within the FYDP that align with existing or new publicly available categories of FYDP budget lines. Updated and clearly labeled strategic capability areas or other exhaustive and mutually exclusive categorization systems could overcome many of the data quality limitations documented in this paper and would make it easier to mark a strategy to a budget and serve what Enthoven and Smith (2005) describe as key to the FYDP's value: the way the FYDP forces "the Secretary to make controversial decisions explicitly" (p. 52).

References

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs* (Vol. 343). RAND Corporation.

Blakeley, K., & Korb, L. (2014). *The war chest: War funding and the end of the War in Afghanistan*. Center for American Progress.

DoD Comptroller. (2020). *DoD budget request*. https://comptroller.defense.gov/Budget-Materials/

Domke, W. K. (1984). Waste, weapons, and resolve: Defense posture and politics in the defense budget. *Policy Sciences*, *16*(4), 371–390.

Enthoven, A. C., & Smith, K. W. (2005). *How much is enough?: Shaping the defense program, 1961-1969*. RAND Corporation.

Harrison, T. (2016). *Defense modernization plans through the 2020s: Addressing the bow wave*. Center for Strategic and International Studies.

Harrison, T., & Daniels, S. (2017). *Analysis of the FY 2018 defense budget*. Center for Strategic and International Studies.

Harrison, T., & Daniels, S. (2020*). Analysis of the FY 2020 defense budget and its implications for FY 2021 and beyond*. Center for Strategic and International Studies.

Hunter, A. (2019, December 13). *Bad idea: Blaming OCO for our defense budget blunders*. Defense 360. https://defense360.csis.org/bad-idea-blaming-oco-for-our-defense-budget-blunders/

Jordan, L. G. (2015). Systemic fiscal optimism in defense planning. *Defense Acquisition Research Journal*, *22*(3), 272–293.

Lewis, K. N. (1994). The discipline gap and other reasons for humility and realism in defense planning. In *New challenges for defense planning: Rethinking how much is enough* (MR-400-RC, pp. 101–134). RAND Corporation.

Light, T., Leonard, R. S., Pollak, J., Smith, M. L., & Wallace, A. (2017). *Quantifying cost and schedule uncertainty for major defense acquisition programs (MDAPs)*. RAND Corporation.

McGarry, B. W., & Epstein, S. B. (2019). *Overseas contingency operations funding: Background and status* (CRS Report No. R44519). Congressional Research Service.

Sharp, T. (2019). *Did dollars follow strategy? A review of the FY 2020 defense budget*. Center for Strategic and Budgetary Assessments.

Woodward, F. M., & Arthur, D. (2019). *Long-term implications of the 2020 Future Years Defense Program*. Congressional Budget Office.

Young, T. D. (2018). Questioning the "sanctity" of long-term defense planning as practiced in Central and Eastern Europe. *Defence Studies*, *18*(3), 357–373.

Evaluating the Impacts of Federal Improvement and Audit Readiness (FIAR) Compliance

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.

Dylan Hunt—is a Master of Public Policy student (2020) at the University of Maryland, specializing in international security and economic policy. His work with the Center for Public Policy and Private Enterprise includes research on Department of Defense (DoD) audits and the DoD's application of Modular Open System Architecture. He completed his undergraduate degree in economics and political science, also at the University of Maryland. Following graduation, Dylan plans to pursue a career at the intersection of economics, finance, and government consulting.

Abstract

Over the last 30 years, the Department of Defense (DoD) slowly became compliant with the Chief Financial Officers Act of 1990, which required federal agencies to undergo an annual financial audit. In 2018, the DoD finally completed its first audit and continued this trend in 2019. This paper seeks to understand the benefits of producing auditable financial statements, their costs, and any impacts on the DoD's acquisition system. We describe the several forms of accounting and review the academic accounting literature that examines the value of audits. We describe the DoD's preparation for and analyze the results of the two completed audits to look more broadly at the benefits. These include uncovering previously unaccounted inventory and improvements to internal accounting systems, even as no instances of massive waste or fraud were identified. Finally, the utility of management cost accounting is discussed.

Introduction

In September 2018, the Department of Defense (DoD) completed its first full audit, which analyzed over $2.7 trillion in assets, about 70% of the federal government's assets (DoD, 2018a). The audit was the synthesis of 24 separate audits of the DoD's components and was monitored by the Office of the Secretary of Defense's Financial Improvement and Audit Readiness (FIAR) Directorate, which was established by the DoD comptroller over a decade ago. The FIAR Directorate was created to improve the department's accounting practices and to plan for the successful audit of the DoD as a whole. This move was in response to the Chief Financial Officers (CFO) Act of 1990, which required that all executive agencies have their financial statements independently audited. Since 2011, the DoD was the sole agency that continually failed to audit its financial statements in their entirety (Miller, 2011).

The first DoD audit cost taxpayers between $918 million and $972 million, accounting for direct costs to independent auditors, indirect costs of government support during the audit process, and remediation costs to improve errors uncovered by the audit (Browne & Starr, 2018; Fine, 2019). Additional costs not accounted for in the $900+ million are compliance costs associated with changing internal DoD processes, hiring additional labor, and other changes to become ready for a complete audit.

Analyzing the costs and benefits will provide a fuller understanding of the value of auditing the DoD. FIAR compliance costs are likely made up of fixed and variable costs. Initial costs to hire and change processes will incur large start-up, fixed costs during the first few audits. For example, over $500 million was spent in remediation following the first audit. This number will likely decrease as accounting and cost reporting processes improve, but the number of improvements required is currently growing faster than the DoD can remediate them. The cost of maintaining FIAR compliance will likely fluctuate, requiring a reallocation of scarce DoD resources.

As the DoD continues to update its financial management practices and alters its cost reporting processes in hopes of receiving an unqualified opinion, this report seeks to answer the questions:

- What are the impacts associated with FIAR?
- What are the benefits to taxpayers of government entities producing financial statements?

Further, we hope to identify the broader implications of FIAR compliance on the DoD.

Other important questions considered relate to the effectiveness of public audits: Do public audits provide valuable information to agency executives to improve processes? More generally, what are the benefits of auditing the DoD?

At large, this report seeks to better understand the value provided to the DoD and its stakeholders. After the completion of the second DoD audit in November 2019, over $2 billion dollars were spent in hopes of improving the financial management of the DoD. While this amount is a drop in the bucket relative to the DoD budget, it is important to constantly question and improve the way the federal government is spending taxpayer money. Further, we will examine impact may have on DoD's acquisition system.

Background

For the past several decades, Congress has struggled to obtain better visibility into the government's financial position. A Government Accountability Office (GAO) report in 1985 put a spotlight on the problems; it concluded that these were numerous issues that called for an overhaul of the government's financial management system (Bowsher, 1985). The GAO believed that successful reform would require a major initiative with a comprehensive, integrated approach (Bowsher, 1985). As a result, Congress—in its oversight role of federal agencies and programs—passed a series of laws and mandates designed to improve the accountability and management of appropriated resources and to form the conceptual foundation of a new financial management structure, as well as additional conditions, requirements, and due dates for the DoD's efforts to become auditable. These included the CFO Act, the Government Performance and Results Act (GPRA), the Government Management Reform Act (GMRA), the Federal Financial Management Improvement Act (FFMIA), and a series of provisions in the National Defense Authorization Acts (NDAA) of Fiscal Years (FY) 2010, 2014, 2016, and 2018.

Perhaps the most impactful of these was the CFO Act of 1990, described by the GAO as "the most comprehensive and far-reaching financial management improvement legislation … since 1950. [It] will lay a foundation for comprehensive reform of federal financial management" (Bowsher, 1991). The most noteworthy part of the act was the requirement for every executive agency to be audited annually.

Prior to the CFO Act, "Government reports found that agencies lost billions of dollars through fraud, waste, abuse, and mismanagement" (GAO, 2020b, p. 3) These concerns continued to grow as Americans and congressmen began to doubt the government's ability to properly manage programs, protect its assets, or wisely use taxpayer dollars in an effective and efficient manner (GAO, 2020b). In 1988, the GAO reported numerous internal control problems, specifically in the DoD, which resulted in hundreds of millions of dollars unaccounted for. In response to these growing concerns, the CFO Act hoped to introduce some accountability for, and effective tracking of, how the federal government spent money.

The CFO Act enjoyed widespread congressional support. Within 2 months of its introduction in the House, the act was amended, voted on by both chambers, and signed into law in late October 1990. Thirty years later, the Senate Committee on the Budget held a hearing looking at the impacts of the CFO Act of 1990. In a show of bipartisanship, members complimented the effectiveness of the act, anticipating continued and improved federal financial management (Dodaro, 2019). Unfortunately, while anecdotal evidence was used throughout the hearing to support the act, quantitative evidence of its effects was absent.

In 1991, the Federal Accounting Standards Advisory Board (FASAB) was created by the GAO, the Treasury Department, and the Office of Management and Budget (OMB) to develop the necessary accounting systems. The GMRA of 1994 required the FASAB to develop a system that would produce government-wide financial statements and required the first statements to be published for FY1997 (Anthony, 2005).

As a result, FASAB developed and published its standards in the FASAB Handbook of Accounting Standards (FASAB Handbook). The FASAB Handbook outlines the objectives for producing the federal government's financial statements and their audits and is the most authoritative source of generally accepted accounting principles (GAAP)[1] for federal entities. Federal government agencies, contractors working with federal government agencies, and accounting firms auditing federal government agencies all consult the FASAB standards on a regular basis. The FFMIA of 1996 strengthened the requirements of the 1994 act.

By FY2003, 20 of the 24 federal CFO Act agencies had been able to produce financial statements backed up with unqualified opinions from auditors. However, the DoD was not one of them. There was increasing significant pressure from the President, the OMB, and Congress for DoD to achieve auditability (Candreva, 2004).

The CFO Act and the other associated legislation ushered in an era of improved financial management of the federal government. Today, however, it may be that the CFO Act of 1990 was too wide-reaching, causing the DoD to spend nearly a billion dollars annually without any fraud, waste, or abuse found. While the exact effects of the large-scale changes imposed are next to impossible to quantify, it is important to understand the value of continuing this process.

Types of Accounting

The CFO Act of 1990 requires executive agencies to conduct financial audits based on financial accounting procedures. The DoD maintains the position that financial auditing is improving their internal business processes and saving money (Cronk, 2019c). Although the government generally uses budgetary accounting, and in some cases managerial accounting, it is important to understand the differences between the accounting types when considering the value of the DoD audit.

[1] GAAP refers to a common set of accounting principles, standards, and procedures issued by the Financial Accounting Standards Board (FASB). Public companies in the United States must follow GAAP.

Financial accounting is required to produce the statements needed to comply with the CFO Act. Widely used in the private sector, it is the type of accounting used to produce a corporation's annual report: balance sheets, income statements, and statements of cash flow and owner's equity. It accounts for assets, liabilities, and cash flows. With financial accounting, revenue is recognized when realized, and expenses are recognized when incurred; this is known as accrual basis (Gibson, 2011). With private sector corporations, the audience is potential lenders and investors (i.e., the capital market). On the other hand, the audience for audits of government agencies is legislators and taxpayers—stakeholders with a financial interest (Candreva, 2004).

The objective of financial accounting is to capture and accurately present past events. There are strict rules, and the statements produced have governmental oversight. Public companies will publish results of their financial audits to comply with regulations but also to assure the public—beyond managements' own assertions—that a company's financial statements are accurate and can be relied upon. Financial accounting, therefore, looks at the big picture of a company or organization over the last year or more. This distinction makes financial accounting backward-looking. Critics of financial accounting argue that the backward-looking nature of financial accounting makes it inadequate to inform and support future decisions. Finally, financial accounting must follow GAAP, which is a combination of standards that are commonly accepted for presenting financial information.

Although financial accounting is required to produce auditable financial statements, government agencies primarily use *budgetary accounting* to manage their finances. Budgetary accounting is used to justify and account for appropriations; this type of accounting is not used in the private sector. There are rigid rules stipulated in laws and guidance from the comptroller general. The objective is to ensure that the government spending complies with the associated restrictions; there is significant oversight to ensure this is the case. The focus is on ensuring that appropriated funds have been spent in accordance with the purpose, time, and amount to meet the terms of the restrictions attached to the appropriation and is used by both internal and external audiences (Candreva, 2004).

Finally, there is *managerial-cost accounting*, sometimes referred to as cost accounting. This type of accounting is used for internal analysis conducted by corporations to evaluate different options, such as whether to lease or buy a facility. Managerial-cost accounting is intended for internal stakeholders. Moreover, the forward-looking nature of managerial-cost accounting makes it attractive to managers looking to make real-time decisions. For example, managerial-cost accounting may have current information on the cost of production for a certain good to determine if continued production is worthwhile, whereas financial counting may have more accurate data but would only be able to look at the historical cost of production during a past time span. Finally, since the focus is on internal management decisions about the organization's mission and scope of operations, there are no set rules or government oversight; consequently, the management decides what to count and the basis for accounting. This type of accounting enables DoD's working capital fund activities to set their rates based on unit cost (Candreva, 2004).

Auditor Opinions

When an organization's financial statement is audited, a formal report is provided by the auditing entity. This auditor's report is a formal assessment of the financial statement, resulting from their independent examination of the information provided, using a formal set of rules (Gibson, 2011). The audits of federal agencies are conducted using generally accepted government auditing standards (GAGAS; Comptroller General, 2018). When the audit is complete, the auditors can render one of four opinions; these are summarized below (Gibson, 2011).

- **Unqualified:** This opinion states that the financial statements represent fairly, in all material respects, the financial position of the organization and are in keeping with the appropriate principles (Gibson, 2011). Within the federal government, these are sometimes referred to as unmodified opinions.
- **Qualified:** This opinion states that—except for the effect of matters pertaining to qualifiers—the financial statements represent fairly, in all material respects, the financial position of the organization and are in keeping with the appropriate principles (Gibson, 2011).
- **Adverse:** This opinion states that the financial statements *do not* represent fairly the financial position of the organization due to nonconformance with appropriate principles (Gibson, 2011).
- **Disclaimer of opinion:** When the scope of the audit is not sufficient to provide enough information to render an opinion, this opinion is rendered (i.e., the auditor does not express an opinion on the examined financial statements; Gibson, 2011).

Audits can also identify weaknesses and inefficiencies in the financial management and control systems based on the severity of the weakness; these classifications include material weakness and significant deficiency (Public Company Accounting Oversight Board [PCAOB], 2004),

Literature Review

The value of audits within the private sector is long established and well documented. Some of the benefits often cited are increased accountability to stakeholders and investors, feedback to improve business processes, and ensured compliance with financial regulations. While there are other benefits, these three encapsulate much of the benefit auditing has for public companies. Similarly, private companies also benefit from auditing.

First, auditing may reduce the likelihood of fraud by management and others because it introduces additional accountability to management. Second, auditing may reduce agency conflict between owners, managers, and banks. Third, audits may be used to evaluate managerial performance given the lack of market measures to the firm's and manager's performance (Van Tendeloo & Vanstraelen, 2008; Vanstraelen & Schelleman, 2017). Perhaps most importantly, audits may reduce the cost of capital for companies being audited by anywhere between 1% and 3% (Elliot, 1994). Other empirical studies suggest that this number may be overstated, but the general effect does exist (Hay & Cordery, 2018). Additional benefits include reducing the likelihood of fraud by management and others because it introduces additional accountability to management, and it may be used to evaluate managerial performance (Vanstraelen & Schelleman, 2017).

While there are numerous benefits for public and private businesses, it may be that auditing government organizations includes different calculations between costs and benefits. Research on public sector accounting is now also a well-established field with publications in numerous academic journals (Goddard, 2010; Hay & Cordery, 2018). Even though government bureaucracies are not accountable to investors or stakeholders, audits of government organizations have also been examined by looking at different principal–agent relationships present between the legislature, the government, and the electorate. Principal–agent relationships are defined by the agent's ability to take actions on behalf of the principal that ultimately affect the principal. In the case of the DoD, the legislature (Congress) is the principal, and the government is the agent (Streim, 1994). Streim argues that auditing can help improve these principal–agent relationships because external accountability is introduced. With this new external accountability, the agent is discouraged from acting in a self-interested way and instead works in a manner more in line with the objectives of the principal (i.e., the legislature). Auditing can also reduce the associated agency costs. Based on this analysis, requiring an annual DoD audit could incentivize the DoD to be more transparent and communicative with Congress.

Within the private sector, as the complexity of business transactions and accounting standards grows, the potential of audits to add value increases (DeFond & Zhang, 2014). It may be that the legislation requiring audits saves money for some agencies but increases costs for others. Accordingly, the DoD's complexity makes it likely that auditing could save money. Its large and multifaceted environment increases "auditing's potential to add value" (DeFond & Zhang, 2014, p. 275) because it protects against possible financial mistakes that can quickly add up. As DoD Comptroller David Norquist often mentions, the first DoD audit was likely the largest audit in history, making it a prime example for possible cost savings (DoD, 2018b).

Studies also suggest that auditing and other financial reporting requirements help add credibility to the organization. In the case of the DoD, credibility may be valuable to both Congress and the public. With numerous anecdotal cases of valuable DoD equipment going missing, improved public trust surrounding how the DoD spends taxpayer dollars could be incredibly valuable (Hay & Cordery, 2018). Additionally, although regulations may already protect against fraud and mismanagement of money, a recent study finds that U.S. municipalities that conduct audits are associated with fewer internal control problems (Rich & Zhang, 2014).

Challenges with DoD's financial management, and the federal government at large, are not new. Former DoD Comptroller and Harvard professor Robert N. Anthony[2] reviewed the history of the federal government's accounting practices, which originally developed with a focus on obligations that aligned with the budgeting and appropriation process rather than on expenses (used for financial statements), which he believed would provide more useful for both planning and control purposes (Anthony, 2000). Anthony described the tension between these two approaches, since both systems are used to some degree in federal departments and agencies. Anthony believed that neither accountants nor managers would pay attention to the information in the expense-based accounts and, consequently, that system would simply atrophy (Anthony, 2000, 2005).

On the other hand, others also question the benefits of financial accounting. Robert Kaplan and Robin Cooper, Harvard University professors, have asserted that financial accounting systems are "completely inadequate" for either "estimating the costs of activities and business processes" or for "providing useful feedback to improve business processes" (Kaplan & Cooper, 1998, p. 14). Further, they argue that financial statements are used primarily to demonstrate to shareholders that a firm is operating profitably. In the case of a government agency, which is neither a business nor does it earn a profit, one may question the value of an audit.

Finally, it is important to understand the limitations of any audit. Soon after the first DoD audit in 2018, numerous articles claimed the audit was a failure due to the DoD's disclaimer of opinion. While the DoD quickly refuted this claim, the DoD did have to deal with numerous questions about the benefits of any such audit (Cronk, 2019b). The public generally has a different expectation of the results of audits. There is a widespread belief that "a person who has any interest in a company … should be able to rely on its audited accounts as a guarantee of its solvency, propriety and business viability" (Koh & Woo, 1998, p. 147). Consequently, when

[2] Anthony was a Harvard faculty member from 1940 to 1982. When requested by Defense Secretary Robert S. McNamara, he took public service leave from the school in 1965 to serve as the DoD comptroller (Hevesi, 2006).

the public has different beliefs about the auditors' duties and responsibilities and what the audit reports really mean, a gap in understanding is created. Koh and Woo refer to this gap between the limitations of an audit and the public expectation of auditing as the "expectation gap" (Koh & Woo, 1998). This gap may still persistent today, and it is important to recognize it when contextualizing the usefulness of the DoD audit.

The DoD Prepares

The DoD is significantly different from the other executive agencies. The DoD is a large, complex organization with annual budgets approaching $700 billion in FY2021 and total assets that exceeded $2.7 trillion in FY2018 (House Committee on Appropriations, 2020; Office of the Under Secretary of Defense [Comptroller; OUSD(C)], 2019). In FY2010, for example, the DoD processed more that 11 million commercial invoices and approximately 198 million payment-related transactions, disbursing over $578 billion (Khan, 2011).

Financial management reform has been an issue the DoD has struggled with for many years. While the audit looks at the department as a whole, the DoD is made up of 24 component parts. Most of these had been audited at different times and at different frequencies due to a variety of factors. Therefore, the DoD had not been ignoring the mandate from the CFO Act of 1990 for 30 years. Instead, its component parts made progress toward fulfilling the audit requirement, but the sheer size of the organization made progress slow. The DoD audit discussed in this paper refers to the complete DoD audit, or a compilation of all component audits.

The NDAA FY2010[3] required the DoD to have a financial statement audit-ready no later than September 30, 2017. The minimum requirements for audits of federal financial statements are contained in OMB Bulletin No. 07-04; it implements audit provisions from the CFO Act of 1990, the GMRA of 1994, and the FFMIA of 1996. OMB Bulletin No. 07-04 requirements include performing an audit annually (Office of Management and Budget [OMB], 2007).

Additionally, the NDAA FY2010 also contained a requirement for the DoD to develop a FIAR Plan; the DoD Comptroller established the FIAR Directorate to lead that effort and manage and integrate department-wide financial improvement efforts and help the DoD get audit-ready. The Office of the Undersecretary of Defense (Comptroller; OUSD[C]) developed and revised the FIAR Guidance. This handbook was intended to serve as a guide for all the organizations involved in the department's audit readiness initiatives. It is updated periodically to ensure it aligns with all applicable federal and departmental financial management requirements. The guide outlined the FIAR strategy, developed to serve as the roadmap for the department to become audit-ready. The guide also defined audit readiness "as having the capabilities in place to allow an auditor to scope and perform a full financial statement audit that results in actionable feedback" (OUSD[C], 2017).

The guide presented a phased methodology for the DoD to become audit-ready by FY2018. The initial three waves were performed concurrently, focused on OUSD(C)'s initial priorities, that is, budgetary information and mission critical asset information. For Wave 4, the DoD's components incorporated the expanded priorities, proprietary information, and valuation into their audit readiness efforts and focused on full financial statement audits. This methodology defined the key tasks, the underlying activities, and the work products required from reporting organizations to become audit-ready. It considered the methods financial statement auditors use to assess financial statement accuracy in accordance with auditing

[3] Section 1003 of the FY2010 NDAA required the DoD's chief management officer, in consultation with the Under Secretary of Defense (Comptroller), to develop and maintain a plan to be known as the "Financial Improvement and Audit Readiness Plan" (NDAA, 2010).

standards, in order to maximize the potential for successful financial statement audits (OUSD[C], 2017).

The FIAR Plan also described specific corrective actions to achieve reliable, accurate, and complete financial data for use in key management decisions. It focused on problems such as weak internal controls, incomplete or inaccurate information, and systems that cannot properly process data and information. By establishing and monitoring critical milestones for resolving these problems, the FIAR Plan gave decision-makers better information and more options. Implementing this plan, after decades-long changes to internal processes, the DoD became compliant with the requirement of the CFO Act in September 2018 and had become audit-ready. The audit, however, returned a disclaimer of opinion.

On November 14, 2018, the DoD released its audit-ready Agency Financial Report for FY2018 (DoD, 2019), which presented the consolidated financial information for 63 DoD entities. The DoD report contained the following major sections:

- **Management's Discussion and Analysis:** This section summarized the DoD's mission and structure and the current state of financial management systems. This section also included a discussion regarding the DoD's compliance with certain laws and regulations. There was also a short discussion of improvements to internal controls that resulted in cost savings and increases in efficiency and effectiveness.
- **Financial Statements:** This section provided consolidated financial information on the DoD's financial operations, condition, and position for all DoD entities. Note 1 acknowledged that, due to the limitations of financial and nonfinancial processes and systems, the department was unable to fully comply with all of the required elements of U.S. GAAP and OMB Circular No. A-136. Many of the reported values for major asset and liability categories were derived largely from nonfinancial systems, such as inventory and logistics systems.
- **Required Supplementary Stewardship Information:** This section identified significant DoD investments that have long-term benefits to the public, such as investments in research and development, which may include the development and testing of prototypes for weapon systems.
- **Required Supplementary Information:** This section provided information on other topics to improve the understanding of the DoD's financial operations, condition, and position, such as delayed or deferred maintenance on real property.
- **DoD OIG Audit Report:** This report includes the DoD OIG's overall audit opinion on the basic financial statements.

The FY2019 Agency Financial Report for FY2019 was released on schedule, on November 14, 2019 (DoD, 2019).

The Audits

FY2018

The audit of the DoD Financial Statement for FY2018, which identified $2.8 trillion in total assets, is almost certainly one of the largest and most complex financial statement audits ever undertaken. The comprehensive audit included 24 standalone audits that were conducted by independent public accounting firms; the DoD OIG performed the overarching consolidated audit. More than 1,200 auditors were involved in the effort. The results were mixed; six organizations received the highest grade—unmodified audit opinions—and two received qualified opinions. All of the other organizations received a disclaimer of opinion. Perhaps the most reassuring finding was that no fraud was identified, and no organization received an adverse opinion (Fine, 2019).

For the issues identified, the auditors issued more than 2,300 Notices of Findings and Recommendations (NFRs) and identified 20 material weaknesses. Almost half of these addressed findings with departments' financial management systems and information technology. To track and respond to the identified NFRs, DoD established a NFR database to capture, prioritize, assign responsibility for, and develop corrective action plans for the audit findings (OUSD[C], 2019).

The direct audit-related costs exceeded $973 million, which includes supporting the audits and responding to auditor requests; achieving an auditable systems environment; and the cost of remediating audit findings. The remediation cost,[4] approximately $559 million of the total, included government and contractor costs for correcting findings and the costs of achieving and sustaining an auditable systems environment (OUSD[C], 2019).

FY2019

The audit of the DoD's FY2019 Financial Statement was completed in November 2019, the second full financial audit. Although some progress was made, no audit opinions changed from FY2018 for the 23 DoD reporting entities that received audits overseen by the DoD OIG. The overall result, a disclaimer of opinion on the Agency-Wide Basic Financial Statement, also remained unchanged (Fine, 2020). The cost of the FY2019 DoD audit once again approached $1 billion, which included the costs of the DoD personnel who prepared for the audit and remediated deficiencies identified during the previous audit (about $770 million) and about $190 million for the five independent public accounting firms that performed stand-alone audits of DoD components (CSPAN, 2019). Between 2018 and 2019, the DoD made progress in many areas, even while auditors found additional issues elsewhere.

Summary of Material Weaknesses and NFRs

Material weaknesses are the largest issues that need to be addressed, defined as "weaknesses in internal controls that result in a reasonable possibility that management will not prevent, detect or correct, a material misstatement in the financial statements in a timely manner" (Fine, 2020). In 2018, auditors identified 20 material weaknesses, which subsequently increased to 25 in the 2019 audit. This uptick is mostly due to the auditors being able to conduct a deeper financial analysis of the DoD during the second audit. Hopefully, uncovering issues like this will help improve long-term financial management within the DoD.

The material weaknesses are large, systemic financial management issues. Thus, the DoD prioritizes only a few of them annually (Fine, 2020). In 2018, the DoD prioritized six different material weaknesses. Of the six, all saw significant progress, and only two of them were reissued during the 2019 audit. Further, reissuing of any material weakness does not mean progress was not made. Any reissuing or the addition of new material weaknesses simply means that additional progress is needed before it is at a satisfactory level for the auditors.

Two examples of material weaknesses are General Property, Plant, & Equipment and Fund Balance With Treasury. The first material weakness means that the DoD could not accurately value its property, plant, and equipment in accordance with GAAP. Second, the DoD has an ineffective process for reconciling its fund balance with the Treasury Department. In 2019, there were 23 other material weaknesses. These examples are given to the scope and seriousness of different material weaknesses. It will take the DoD many years to improve the

---

[4] The remediation cost totals do not include enterprise resource planning deployment or maintenance costs.

existing material weaknesses, and that is not accounting for additional findings by auditors. Given all this information, material weaknesses will likely be a helpful metric for defining the success or failure of the audits in the long-term.

Looking at a more granular level, an important metric for improvement of the DoD's financial management is the NFRs published in each audit. In 2018, auditors found a total of 2,595 NFRs; of these, 23% were closed by the FY2019 audit. The acting DoD comptroller, Elaine McCusker, remarked that this was "solid progress for our first year" and that the NFR number will grow as the auditors continue to delve into DoD's systems and processes (Mehta & Judson, 2019). In 2019, a total of 3,472 NFRs were identified (1,300 were new), showing a significant uptick (Fine, 2020). McCusker believed that this increase was not all bad news. With each audit, auditors can better understand and analyze the DoD, which is reflected through the uptick in NFRs (Mehta & Judson, 2019). In a larger context, this improvement and constantly uncovering problems has nothing to do with managerial decisions. The issues the NFRs identified were generally limited to problems of financial management and reporting issues. Fixing these problems shows movement toward improved financial management but are significantly smaller in nature compared to material weaknesses. Therefore, positive movement in both material weaknesses and NFRs will be a good predictor of the progress of DoD's financial management.

Analysis

> *All business systems have a balance sheet; therefore, the government should have one. I think this assumption is unfortunate.* (Anthony, 2005, p. 9)

It took 30 years after the enactment of the CFO Act of 1990 for the DoD to finally become compliant. As previously stated, both audits received disclaimers of opinion (i.e., no opinion could be given due to the financial statements not providing adequate information). Elaine McCusker, the Pentagon's acting comptroller, reassured reporters that the overall disclaimer of opinion was expected (Mehta & Judson, 2019), and she expected that the DoD will likely continue to receive disclaimers of opinion for some time to come.

If the first two audits are any signal for what lies ahead, then the DoD will spend billions within the next decade for a very slow, steady progress toward achieving an unqualified opinion on its audit. While many of the identified NFRs and material weaknesses from the first audit were improved, with the second audit the auditors found far more issues, often dwarfing the improvements. There is no sign that this trend will slow down; the next few years will help give a clearer idea of the depth of financial management problems within the department.

Supporters of the CFO Act of 1990 anticipated numerous benefits would result from the audits. First, since the act was in response to numerous wasteful spending problems that were uncovered in the 1980s, auditing promised a way for the DoD to reduce waste, fraud, and abuse of taxpayer dollars. Another anticipated benefit was improved taxpayer transparency for how their money was being spent. Finally, many supporters argued that the information gained from the audit would help inform decisions made by managers in the DoD to improve processes (Hanks, 2009).

Benefits

Deputy Secretary of Defense David Norquist, the former DoD comptroller, oversaw the initial audit and constantly emphasized numerous benefits associated with improving the financial management of the DoD at large. He believed the most important benefit was accountability to the taxpayer, to ensure that their money was spent appropriately, without any going missing. While the audits finally meant that the DoD was compliant with a specific regulation, it did not mean that the work was finished. During multiple trips to congressional committees, Norquist defended the value of the audit and gave examples of different benefits and cost savings that the DoD already recognized. However, even with these savings, there is no clear evidence that the recognized benefits of the audit have outweighed the annual costs.

The principal supporter within the government of the CFO Act of 1990, however, is the GAO. The GAO cites five areas that have improved since its inception, including leadership, financial reporting, federal workforce, internal control, and financial management systems (GAO, 2020b). All these benefits were realized as executive agencies became audit-ready. Further, the GAO often uses issues within the DoD to highlight the problems that can be attributed, in part, to not being able to be audited successfully. While the GAO discusses the benefits of the CFO Act broadly, the implication is that similar benefits are likely to be realized by the DoD.

Despite the failure to receive an unqualified opinion, the supporters of the financial audits believe that the effort and expense will, in the end, be beneficial. The DoD's auditable financial statements have already made progress and have produced some benefits. The audits have provided insight into areas where processes and procedures are working well, as well as areas that need to improvements. As these improvements are made, the cost of the audits should decrease. The DoD comptroller testified that "we don't have to wait for a clean opinion to see the benefits of the audit. The financial statement audit helps drive enterprise-wide improvements to standardize our business processes and improve the quality of our data" (*Testimony*, 2018). The benefits include improved internal control, financial management, and inventory control. These are summarized below.

Improved Internal Control

Internal control is a policy or procedure put in place by management to safeguard assets, stop fraudulent behavior, promote accountability, and increase efficiency. One of the key objectives is to put in place a process to prevent employees from misappropriating assets or committing fraud. Effective internal control also provides reasonable assurance that an organization is complying with all the applicable laws and regulations and can perform its mission efficiently and report its finances reliably (which requires cybersecurity). Effective internal controls are essential to achieving and sustaining an efficient and effective organization. As previously stated, the two audits found no indications of fraud or abuse.

During the FY2019 audit, 20 agency-wide material weaknesses in internal controls were identified. These errors created the potential for errors in the financial statement to not be detected (Fine, 2020). As a result of the audits, the DoD initiated remediation efforts that have resulted in significant improvements. For example, enhanced internal controls in the U.S. Pacific fleet resulted in freeing up purchasing power to fund the $4.4 million repair costs of the USS *Paul Hamilton* (DoD, 2019).

Moreover, a significant part of the audit involves reviewing information technology and cyber security. Since many of the same systems used for financial management are also used for operational purposes, identifying vulnerabilities in these systems will also result in recommendations that improve the DoD's overall cybersecurity posture across different networks and systems (Fine, 2019). Mitigation for these shortfalls is underway with the transition to cloud architecture with the ongoing Joint Enterprise Defense Infrastructure procurement (Williams, 2018). The DoD will also work to improve cybersecurity by updating the policy on shared file and drive protection to include encryption use, authentication, and minimum password protection requirements and stringent password protection (Williams, 2018). The audits also identified five instances of noncompliance with laws and regulations across the DoD.[5]

Improved Financial Management

Within the DoD's complex structure, financial transactions often involve several information technology (IT) reporting systems to go from the initial transaction to the point where they are captured in the component's financial statement; often the components do not own and operate all of the IT systems that are used to process these transactions. In 2016, 400 separate IT systems were used to process the DoD's accounting data. The audits identified wide-ranging weaknesses in these systems that prevented the accurate, reliable, and timely reporting of financial data (Fine, 2019). Some of these were identified years prior (Williams, 2018).

A significant benefit is the savings generated from improved efficiencies and better financial operations since real-time improvements were made. For example, in response to the FY2018 audit, the Army implemented a new automated solution for data entry into the U.S. Standard General Ledger, saving 15,500 labor hours (Mehta & Judson, 2019).

Another benefit from the audit requirement is the initiation of financial management improvements that would otherwise take longer or not occur at all. Then the subsequent public reporting enables the tracking by both management and oversight agencies (Brook, 2011). As a result of the audits, several initiatives are being pursued by the DoD to address the weaknesses related to the IT systems. For example, the DoD has plans to eliminate 26 legacy IT systems by FY2022. Additionally, the DoD has established a database to identify IT applications that impact DoD financial statement audits and to track the auditor feedback regarding the system controls (Fine, 2019). During the FY2019 financial statement audits, additional improvements were made by the DoD's components. They were to provide more segments of transactions for testing, along with better supporting documentation for those selected for testing. This meant auditors were able to expand testing to new areas (Fine, 2020).

For example, for the first time ever, the Army was able to provide auditors with transactions for Army Working Capital Fund inventory work in progress, which consists of raw materials that are used to make a finished product, valued at $952 million. Auditors were able to reconcile these transactions to the balances in the accounting system and will also be able to test these in the FY2020 audit. Auditors also found the Army's IT controls over its Logistics Modernization Program system to be effective; no exceptions were identified by the auditors during testing (Buble, 2019). In another case, the Army created a computer application to store and analyze its transactional data for audit, increasing its visibility into its cost drivers and enabling its leadership to commit resources to the highest priority programs (Fine, 2020).

In response to the issues identified, other improvements are planned. The DoD will develop and implement a plan for an integrated pay and personnel system to report financial management data, capture and store key documentation, and determine pay and benefits (Williams, 2018).

Improved Inventory and Property Management

The military services and other DoD components own $291.5 billion in Inventory and Related Property Inventory, which then must be reported on their financial statements. The auditors identified material weakness related to Inventory and Related Property in both audits. Auditors found that numerous DoD components were unable to provide assurance over the existence, completeness, and valuation of inventory. For example, items may have been moved

---

[5] Specifically, the audits identified that the DoD did not comply with the Federal Managers' Financial Integrity Act of 1982, the FFMIA of 1996, the Antideficiency Act, the Federal Information Security Modernization Act of 2014, and the Debt Collection Improvement Act of 1996 (Fine, 2019).

or used but were still in the inventory records; other items were found in the warehouse but not listed in the inventory records; and some items were recorded as in good condition but were unserviceable (Fine, 2020). An accurate accounting is necessary. For example, it is important that the number of spare parts in inventory is accurate to ensure the ability to support operational requirements, as well as to preclude the ordering of unnecessary inventory.

As a result of the FY2018 audit, the Air Force redesigned the process for validating the condition of assets in property systems at Hill Air Force Base in Utah, enabling the accurate capture of approximately $53 million in assets that would have otherwise been misstated. In another example, $280 million of items that were not properly tracked were at Naval Air Station Jacksonville. As a result, $81 million of material, that the service had no idea it had on hand, was identified as available for immediate use. Additionally, the inefficient use of assets was identified, and getting rid of old, unusable material freed up approximately 200,000 square feet of storage space (Mehta & Judson, 2019). With the FY2020 audit, the Navy was able to complete a full inventory of real property assets, resulting in a 98% accuracy rate; and the Air Force completed floor-to-book and book-to-floor inventories of over 96% of its buildings (Buble, 2019).

Audit Opponents

However, this effort is not without its detractors. Challengers of the benefits of auditing DoD's financial statements contest the value of financial statements and the audits, citing numerous issues. Some have argued that the DoD's financial accounting processes have been flawed from their inception. Others argue that financial audits will incur significant costs but do not provide the information necessary to effectively make better managerial decisions. They believe the DoD would be better served with a shift toward managerial-cost accounting. These positions are reviewed below.

Once the FASAB developed the federal governments accounting processes and standards, Professor Anthony was one of the earliest critics. He assessed that "few managers in the executive branch and few legislators or their staffs will use the accounting information developed in the new system" (Anthony, 2005). Since there is little evidence, they used information provided by previous systems.

As discussed previously, the accounting system, as developed by FASAB, was composed of two "separate, uncoordinated systems" for budgeting and accounting in the federal government. The House Appropriations Committee makes its appropriations on an obligation basis, and most other committees also accept the obligation format. The Senate Governmental Affairs Committee, on the other hand, mandated an expense-based format; however, this is in addition to, rather than instead of, the obligation basis. Anthony believed the obligation system could be easily manipulated. When an obligation authority exists, contracts can be charged to it, even if the goods or services are not actually needed. Since funds continued to be appropriated on an obligation basis, government leaders and managers would not pay much attention to the expense-based financial information. He concluded that since the FASAB system continued this separation, the financial management systems would not assist managers and other decision makers in making decisions the way a good accounting system should (Anthony, 2000).

A 2011 Institute for Defense Analyses (IDA) assessment of the DoD's Enterprise Resource Planning (ERP) Business Systems judged that making the department an auditable enterprise was a "wicked problem."[6] When it comes to financial statements, the IDA highlighted

[6] The term *wicked problem* was coined by Rittel and Webber. In their paper, they detail 10 characteristics that describe a wicked problem (Rittel & Webber, 1973)—these complex policy and planning problems

the differences between the DoD and those of commercial businesses, the principal users of audited financial statements. The DoD's primary stakeholders are not shareholders, but taxpayers. Moreover, the DoD is not concerned with making a profit, remaining solvent, limiting risk/liability, and developing tax incentive-based strategies involving the valuation and depreciation of assets, since these have minimal operational value. The IDA concluded that the DoD should discontinue their attempt to achieve comprehensive financial statement audits, and the operational value of audit readiness activities should be assessed before additional resources were expended. The IDA believed that a much more meaningful accounting of the DoD would be a managerial-cost accounting approach (Ketrick et al., 2012).

Many other critics of the DoD's financial statement audit believe that managerial-cost accounting would better provide many of the benefits promised by the supporters of the auditing of DoD financial statements. Consequently, the differences between financial and managerial accounting are important for understanding the value of the DoD audit and the CFO Act of 1990.

As describe above, financial accounting is concerned with income statements, balance sheets, and journal entries. Financial accounting is used to serve external stakeholders. For example, public companies will publish results of their financial audits to comply with regulations but also to assure the public—beyond management's own assertions—that a company's financial statements are accurate and can be relied upon. Financial accounting, therefore, looks at the big picture of a company or organization over the last year or more. This distinction makes financial accounting backward-looking. Critics of financial accounting argue that the backward-looking nature of financial accounting makes it inadequate to inform future decisions. Finally, financial accounting must follow GAAP, which are commonly accepted standards for presenting financial information. Managerial-cost accounting, on the other hand, is intended for internal stakeholders. Additionally, the forward-looking nature of managerial accounting makes it attractive to managers looking to make real-time decisions, and the forward-looking nature would greatly help the DoD plan, rather than recounting the past like financial accounting.

It is important to note that many experts consider managerial-cost accounting the best way to improve businesses practices. While both play an important role, managerial accounting provides the information necessary to create the change that is intended by auditing the DoD, according to many experts. David Norquist, the DoD comptroller who oversaw the first consolidated DoD audit, consistently used language similar to supporters of managerial accounting, even though the DoD is conducting a financial audit. It is unclear who is correct. Experts expect financial audits to affect business decisions little, but Norquist consistently reported improvements in the DoD's business practices.

Consequently, improving financial accounting practices within the DoD may not achieve the accountability that Congress often espouses in support of the CFO Act of 1990 and the DoD audit. Given their rhetoric, it seems that managerial accounting may be more in line with their objectives. Financial accounting is central to an audit, but audits only certify the organization's financial statements are accurate. Managerial-cost accounting, on the other hand, uses financial information to inform internal managerial decisions. If Congress is concerned with the wastefulness of the DoD at large, shifting focus to managerial-cost accounting may provide senior leader managers with information needed to make better decisions.

Resolving this conflict will be central to understanding the costs and impacts of financially auditing the DoD. In October 2019, the GAO released a report discussing the progress in financial management for the federal government (GAO, 2020b). One of the

lack clarity in both their goals and solutions. IDA concluded that achieving auditability in the DoD meets most or all of the 10 characteristics of wicked problems.

recommendations moving forward was that the executive agencies still needed to "better link performance and cost information for decision-making." This recommendation suggests that the DoD may be improving internal financial management systems, along with other executive agencies, but struggling to monitor program performance and implement appropriate changes.

Finally, the cost to achieve DoD auditability may be understated. A major part of the DoD's strategy to achieve auditability has been to modernize its business systems. The GAO has designated the department's business systems modernization efforts as high-risk since 1995 and continues to find weaknesses in the DoD's implementation of business systems. Since 2005, the GAO has issued 12 reports and has made 29 recommendations, and as of June 2019, only 15 of the 29 recommendations contained in the 12 reports have been implemented by the DoD (GAO, 2020a). As part of its business modernization efforts, the DoD has fully implemented several new ERP systems, while others have been cancelled by the DoD or the military service after multibillion-dollar investments, and other projects have been plagued with delays and cost overruns. In FY2012, the DoD invested almost $18 billion to operate and modernize its business systems (Crawford, 2015). The DoD continues to make investments in its business systems; in FY2019 the figure was almost $9 billion (GAO, 2020a). It may not be unreasonable to question if these investments have produced the envisioned value.

Does Auditing the DoD Increase Costs to the Contract Acquisition System?

There was some concern that FIAR requirements would increase the cost of goods and services that the DoD procures. By default, these the financial management requirements flow down to vendors selling to the DoD. To adequately support the DoD requirements, the vendor must have a system of record in place capable of maintaining appropriate controls and processes and be able to produce the necessary supporting documentation to validate the cost charged to the DoD. However, the DoD already had in place an extensive financial management regulation regime prior to the CFO Act of 1990, which already imposed a significant additional cost on contractors.

In 1994, the accounting firm Coopers & Lybrand (now PricewaterhouseCoopers, or PwC) completed the most authoritative study of the costs associated with the DoD regulatory burden contractors' costs. They analyzed over 100 different regulations that increased costs to contractors. Out of the 100 regulations analyzed, three of the top 10 that drive costs were related to finance and accounting (see the inset). Specifically, these three accounted for a 2.9% increase in costs for DoD contractors. Property and equipment management added another 1.2%, for a total of 3.4% of the total 23% cost burden, or approximately one-fourth of total compliance costs. These costs may have a type of spillover effect and mask any small cost that may come from providing any additional financial data, not already required, to comply with FIAR.

Cost Accounting Standards. Directed the establishment of cost accounting standards required. FAR Part 30 outlines policies and procedures for applying the Cost Accounting Standards Board (CASB) rules and regulations (48 CFR Chapter 99) to negotiated contracts and subcontracts. DFARS 252.242-7006(a)(1) further defines an acceptable accounting system.

Cost/Schedule Control Systems. The DoD established use of EVM as a requirement for periodically measuring linear programs with firm baselines established prior to starting development.

The Truthful Cost or Pricing Data Act. Commonly referred to by its historical name, the Truth in Negotiations Act (TINA) requires contractors to submit certified cost or pricing data if a procurement's value exceeds the specified threshold and no exceptions apply.

A nonfinancial requirement imposed on contractors by the DoD that trickles down costs, in part to improve the DoD's audit readiness, is the DoD's item unique identification (IUID) system, which requires businesses to provide a unique identification of all delivered end items with a unit acquisition cost of $5,000 or more (DFARS 252.211-7003, 2020). Although this policy increased the cost for contractors, OEMs have presented data to show that identification technology reduces costs through improved data quality and enhanced quality control during product planning, development, life cycle, and inventory control, and the Aerospace Industry Association developed a common supplier flow-down requirement to further expand IUID use as the single identification across industry and the DoD for supply-chain management. IUID was identified as the single best practice for item management across the corporate spectrum for both commercial and government business by industry groups (Bradford, 2012).

Finally, there is no evidence directly linking the nearly $1 billion annual investment in the financial audits to increased costs to contractors. Based on the data collection and reporting requirements already in place, it is unlikely that auditing the DoD increases costs to the contract acquisition system. Further, while there does not seem to be any direct connection between FIAR compliance and increased acquisition costs to the contract system, there is, however, an opportunity cost. Any monies devoted to audit readiness and compliance are not available to fund other initiatives.

Conclusion

The DoD has spent billions in meeting the congressional mandate to have audit-ready financial statements by September 30, 2017 and must now continue to produce annual financial statements and undergo audits as required by law. With two audits completed, the DoD has identified many shortcomings in the process and has corrected some of these. It will take some time to tell if an unqualified opinion can ever be achieved. In the interim, the DoD's interest and commitment may fade, since within the department, financial management will always cede priority to the operational mission.

The DoD has recognized some benefits as a result of the audits, primarily uncovering inventory that was previously unaccounted for, including helicopters, buildings, and munitions (Cronk, 2019a). Uncovering this additional inventory can help managers be better informed about resources on hand, but it is unlikely that these discoveries will continue beyond the first few audits. More importantly, the DoD is also improving internal accounting systems so that they will represent the state of their finances more accurately and is strengthening these systems against cyberattacks. These improved accounting and financial management systems will be able to provide better financial information to decision-makers and may reduce the cost of future audits. Even though no instances of massive waste or fraud have been identified, the audits will potentially provide a political benefit; they can help provide the desired transparency and demonstrate that the DoD is exercising sound financial management.

Finally, although one of the objectives for the audits is to improve management decisions, the financial accounting and auditing literature rarely, if ever, mentions a connection to these. As could be expected, the information derived from the audits has not impacted the rationale for, or the management of, any major DoD program. Since financial accounting and auditing has demonstrated little practical value in making future decisions, it is unlikely that the DoD will be able to improve its decision-making concerning its programs with financial accounting and audits alone. Greater emphasis on improved managerial-cost accounting would provide longer-term benefits in this regard, by better informing future decisions. In this challenging budgetary and national security environment, the nation and its leaders need help to ensure that they do the right things, and not only do things right.[7]

[7] "Leadership is doing the right things; management is doing things right," is a quote often attributed to Peter Drucker.

References

Anthony, R. N. (2000, Winter). The fatal defect in the federal accounting system. *Public Budgeting and Finance*.

Anthony, R. N. (2005, Fall). The FASAB's dilemma. *Journal of Government Financial Management*. http://www.agacgfm.org/publications/downloads/AuthorAward96-2006.pdf (Reprinted from "The FASAB's dilemma," 1996, *The Government Accountant's Journal*, 32–39.)

Bertuca, T. (2017). Pentagon comptroller: Massive DOD audit effort to drive reform. *Inside the Pentagon*, *33*(44).

Bowsher, C. A. (1985). *Managing the cost of government building an effective financial management structure* (GAO/AFMD-85-35). Government Accountability Office.

Bowsher, C. A. (1991). *The Chief Financial Officers Act: A mandate for federal financial management reform* (GAO/AFMD-12.19.4). Government Accountability Office.

Bradford. N. J. (2012). *History and implementation of item unique identification (IUID): Has it improved asset visibility?* [Civilian research project, Army War College]. https://apps.dtic.mil/dtic/tr/fulltext/u2/a568454.pdf

Brook, D. A. (2011). Exploring the value financial statement audits. *Journal of Government Financial Management*, 38–43.

Browne, R., & Starr, B. (2018, November 16). *Pentagon spends $500 million to fix problems found in audit*. CNN. https://www.cnn.com/2018/11/16/politics/pentagon-audit-500-million/index.html

Buble, C. (2019, November 18). Defense department fails its second audit, yet is making progress. *Government Executive*. https://www.govexec.com/defense/2019/11/defense-department-failed-its-second-audit-yet-making-progress/161367/

Candreva, P. J. (2004, Fall). *Accounting for transformation*. The Armed Forces Comptroller.

Comptroller General. (2018, July). *Government auditing standards: 2018 revision* (GAO-18-568G). Government Accountability Office. https://www.gao.gov/assets/700/693136.pdf

Coopers & Lybrand. (1994, December). *The DoD regulatory cost premium: A quantitative assessment*. Defense Technical Information Center.

Cronk, T. M. (2019a, November 16). The Pentagon completed its second audit. What did it find? *Defense News*. https://www.defensenews.com/pentagon/2019/11/16/the-pentagon-completed-its-second-audit-what-did-it-find/

Cronk, T. M. (2019b, November 18). *Audit for 2019 shows DOD's progress*. Department of Defense. https://www.defense.gov/Explore/Features/story/Article/2018453/audit-for-2019-shows-dods-progress/source/GovDelivery/

Cronk, T. M. (2019c, November 20). *Norquist: With 2 audits done, DOD still has work to do*. Department of Defense. https://www.defense.gov/Explore/News/Article/Article/2022270/norquist-with-2-audits-done-dod-still-has-work-to-do/

CSPAN. (2019, November 22). *Senate hearing on Department of Defense audit*. https://archive.org/details/CSPAN2_20191122_130400_Senate_Hearing_on_Department_of_Defense_Audit/start/960/end/1020

DeFond, M., & Zhang, J. (2014). A review of archival auditing research. *Journal of Accounting and Economics*, *58*, 275–326.

DoD. (2018a, November 14). *United States Department of Defense agency financial report, Fiscal Year 2018*. https://comptroller.defense.gov/Portals/45/Documents/afr/fy2018/1-Foreword_About_the_DoD_Agency_Financial_Report_and_Table_of_Contents.pdf

DoD. (2018b, November 16). *The largest audit in history ... really?* Department of Defense. https://www.defense.gov/Explore/Features/story/Article/1690550/the-largest-audit-in-history-really/

DoD. (2019, November 14). *United States Department of Defense agency financial report, Fiscal Year 2019*. Department of Defense. https://comptroller.defense.gov/Portals/45/Documents/afr/fy2019/1-Foreword_About_the_DoD_Agency_Financial_Report_and_Table_of_Contents.pdf

Dodaro, G. L. (2019, October 30). *Hearing on the Chief Financial Officers Act of 1990: Achieving the vision*. Senate Budget Committee. https://www.budget.senate.gov/chief-financial-officers-act-of-1990-achieving-the-vision

Elliot, R. K. (1994, September). The future of audits. *Journal of Accountancy*, *178*(3), 74–82.

Federal Financial Management Improvement Act of 1996, Pub. L. No. 104-208, 110 Stat. 3009 (1996). https://www.congress.gov/104/plaws/publ208/PLAW-104publ208.pdf

Fine, G. (2019, January 8). *Understanding the results of the audit of the DoD FY 2018 financial statements* [DoD Inspector General report].

Fine, G. (2020, January 28). *Understanding the results of the audit of the DoD FY 2019 financial statements* [DoD Inspector General Report].

GAO. (2020a, March). *DOD has made progress in addressing recommendations to improve IT management, but more action is needed* (GAO-20-253). https://www.gao.gov/assets/gao-20-253.pdf

GAO. (2020b, August). *Substantial progress made since enactment of the 1990 CFO Act; refinements would yield added benefits* (GAO-20-566). https://www.gao.gov/assets/710/708666.pdf

GAO. (2020c, September). *DOD needs to implement comprehensive plans to improve its systems environment* (GAO-20-252). https://www.gao.gov/assets/710/709884.pdf

Gibson, C. H. (2011). *Financial reporting analysis: Using financial accounting information* (12th ed.). Thomson South-Western.

Goddard, A. (2010). Contemporary public sector accounting research: An international comparison of journal papers. *The British Accounting Review*, *42*(2), 75–87.

Hanks, C. H. (2009, July). Financial accountability at the DoD: Reviewing the bidding. *Defense Acquisition Research Journal*, *16*(2).

Hay, D., & Cordery, C. (2018, June). The value of public sector audit: Literature and history. *Journal of Accounting Literature*, *40*, 1–15.

Hevesi, D. (2006, December 18). Robert N. Anthony dies at 90; made accounting easier to learn. *New York Times*.

House Committee on Appropriations. (2020, July 14). *Appropriations committee approves Fiscal Year 2021 defense funding bill*. https://appropriations.house.gov/news/press-releases/appropriations-committee-approves-fiscal-year-2021-defense-funding-bill.

Kaplan, R. S., & Cooper, R. (1998). *Cost and effect—Using integrated cost systems to drive profitability and performance*. Harvard Business School Press.

Khan, A. A. (2011, September 23). *Improved controls, processes, and systems are needed for accurate and reliable financial information* (GAO-11-933T) [Testimony before the Subcommittee on Government Organization, Efficiency and Financial Management, Committee on Oversight and Government Reform, House of Representatives]. Government Accountability Office. https://www.gao.gov/assets/gao-11-933t.pdf

Koh, H. C., & Woo, E. (1998, March 13). The expectation gap in auditing. *Managerial Auditing Journal*.

Mehta, A., & Judson, J. (2019, November 16). The Pentagon completed its second audit. What did it find? *Defense News*. https://www.defensenews.com/pentagon/2019/11/16/the-pentagon-completed-its-second-audit-what-did-it-find/

Miller, J. (2011, November 21). *23 of 24 agencies get financial audit opinions*. Federal News Network. https://federalnewsnetwork.com/all-news/2011/11/23-of-24-agencies-get-financial-audit-opinions/

National Defense Authorization Act for Fiscal Year 2010, Pub. L. No. 111-84, 123 Stat. 2190 (2009). https://www.congress.gov/111/crpt/hrpt288/CRPT-111hrpt288.pdf.

Office of Management and Budget. (2004, December 21). *Management's responsibility for internal control* (OMB Circular No. A-123). http://www.whitehouse.gov/sites/default/files/omb/assets/omb/circulars/a123/a123_rev.pdf

Office of Management and Budget. (2007, September 4). *Audit requirements for federal financial statements* (OMB Bulletin No. 07-04). http://www.whitehouse.gov/sites/default/files/omb/bulletins/fy2007/b07-04.pdf

Office of Management and Budget. (2010, September 29). *Financial reporting requirements* (OMB Circular No. A-136). http://www.whitehouse.gov/sites/default/files/omb/circulars/a136/a136_revised_2010.pdf

Office of the Under Secretary of Defense (Comptroller). (2017, April). *Financial Improvement and Audit Remediation (FIAR) guidance*. https://comptroller.defense.gov/Portals/45/documents/fiar/FIAR_Guidance.pdf

Office of the Under Secretary of Defense (Comptroller). (2019, June). *Financial Improvement and Audit Remediation (FIAR) report*. https://comptroller.defense.gov/Portals/45/documents/fiar/FIAR_Report_June_2019.pdf

Public Company Accounting Oversight Board. (2004). *Auditing standard no. 5: An audit of internal control over financial reporting that is integrated with an audit of financial statements* (PCAOB Release No. 2003-025). https://pcaobus.org/Standards/Archived/PreReorgStandards/Pages/Auditing_Standard_5_Appendix_A.aspx

Rich, K. T., & Zhang, J. X. (2014). Does audit committee monitoring matter in the government sector? Evidence from municipal internal control quality. *Journal of Governmental & Nonprofit Accounting*, *3*(1), 58–80. http://aaajournals.org/doi/10.2308/ogna-50832

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*, 155–169.

Streim, H. (1994). Agency problems in the legal political system and supreme auditing institutions. *European Journal of Law and Economics*, *1*(3), 177–191. http://link.springer.com/10.1007/BF01552469

Testimony before the House Armed Services Committee: Department of Defense update on the Financial Improvement and Audit Remediation (FIAR) plan of defense. (2018, January 10) (testimony of D. L. Norquist). https://docs.house.gov/meetings/AS/AS00/20180110/106760/HHRG-115-AS00-Wstate-NorquistD-20180110.pdf

Van Tendeloo, B., & Vanstraelen, A. (2008). Earnings management and audit quality in Europe: Evidence from the private client segment market. *European Accounting Review*, *17*(3), 447–469.

Vanstraelen, A., & Schelleman, C. (2017, June 8). Auditing private companies: What do we know? *Accounting and Business Research*, *47*(5), 565–584.

Williams, L. C. (2018, November 16). DoD's failed audit shines a light on IT woes. *Federal Computer Week*. https://fcw.com/articles/2018/11/16/dod-audit-tech-woes.aspx

PANEL 13. ENHANCING ACQUISITION WITH ARTIFICIAL INTELLIGENCE AND CYBERSECURITY

Wednesday, May 12, 2021	
12:15 p.m. – 1:30 p.m.	**Chair: Rear Admiral Lorin Selby, USN,** Chief of Naval Research ***Lessons Learned in Building and Implementing an Effective Cybersecurity Strategy*** Carol Woody, Software Engineering Institute Rita Creel, Software Engineering Institute ***Using Interdependence Analysis (IA) to Derive Adequate Requirements and Testing for Artificial Intelligence (AI) Enabled Critical Functions within a System*** Scot Miller, Naval Postgraduate School Bruce Nagy, Naval Air Warfare Center Weapons Division

Rear Admiral Lorin Selby, USN—is a native of Baltimore, Maryland and graduated from the University of Virginia with a Bachelor of Science in Nuclear Engineering and earned his commission through the Navy's Reserve Officers Training Corps program. He also holds a Master of Science in Nuclear Engineering and a Nuclear Engineer degree from the Massachusetts Institute of Technology.

His shipboard tours include USS Puffer (SSN 652), USS Pogy (SSN 647) and USS Connecticut (SSN 22). From July 2004 to May 2007 he commanded USS Greeneville (SSN 772) in Pearl Harbor, Hawaii. During these assignments, Selby conducted several deployments to the Western Pacific, Northern Pacific, Northern Atlantic and Arctic Oceans.

Ashore, Selby's staff assignments include duty as a company officer and instructor at the U.S. Naval Academy, service as the deputy director of the Navy's liaison office to the U.S. House of Representatives and duty as the Submarine Platforms and Strategic Programs branch head in the Submarine Warfare Directorate on the Navy Staff. Following selection as an acquisition professional, he served as the program manager for both the Submarine Imaging and Electronic Warfare Systems Program Office (PMS 435) and the Advanced Undersea Systems Program Office (PMS 394).

As a flag officer, Selby served as commander, Naval Surface Warfare Centers (NSWC) from October 2014 to August 2016. In this position, he led more than 17,000 scientists, engineers, technicians and support personnel, both civilian and active duty, within eight NSWC divisions located across the country.

From June 2016 until May 2020, he served as the Navy's chief engineer and the Naval Sea Systems Command (NAVSEA) Deputy Commander for Ship Design, Integration and Naval Engineering (SEA 05), where he led the engineering and scientific expertise, knowledge and technical authority necessary to design, build, maintain, repair, modernize, certify and dispose of the Navy's ships, aircraft carriers, submarines and associated combat and weapons systems.

In May of 2020, he assumed command of the Office of Naval Research as the 26th Chief of Naval Research.

Selby is authorized to wear the Distinguished Service Medal, the Legion of Merit (three awards), Meritorious Service Medal (four awards), the Navy and Marine Corps Commendation Medal (six awards) and the Navy and Marine Corps Achievement Medal (three awards) in addition to various unit awards.

Lessons Learned in Building and Implementing an Effective Cybersecurity Strategy

Carol Woody—a principal researcher for the CERT Division of the Software Engineering Institute at Carnegie Mellon University, is building capabilities and competencies for measuring, managing, and sustaining cybersecurity for highly complex software intensive systems. She has successfully implemented solutions in many domains, including banking, mining, manufacturing, government, and finance. She co-authored the book Cyber Security Engineering: A Practical Approach for Systems and Software Assurance, published by Pearson Education as part of the SEI Series in Software Engineering. The CERT Cybersecurity Engineering and Software Assurance Professional Certificate, released in March 2018, is based on the research she led. [cwoody@cert.org]

Rita Creel—is acting deputy director for the CERT Division of the Software Engineering Institute at Carnegie Mellon University. She has more than 25 years of experience in software-intensive systems engineering and acquisition, cybersecurity, and systems and software measurement and analysis. [rc@cert.org]

Abstract

Today's missions rely on highly integrated and complex technology that must be protected from a wide range of adversaries in a very dynamic and contested cyber environment. The predominant response to the growing, shifting cyber threat has been to apply cyber hygiene best practices and focus on satisfying compliance mandates for an authority to operate. While necessary, these steps alone are not sufficient, given the pace of technology change and the increasing abilities of our adversaries.

For organizations developing or acquiring complex, software-enabled technologies, a cybersecurity strategy provides a critical set of guidelines that enable intelligent, risk-based decisions throughout the life cycle. The strategy identifies planning, design, monitoring, and enforcement considerations for integrating cybersecurity into all products, processes, and resources. As such, it defines expectations for how the individual technology components, their assembled configurations, and their interactions will meet the security requirements of a mission.

Effective cybersecurity requires the application of engineering rigor to the process of defining security requirements in the context of other system imperatives. Cybersecurity engineering is a discipline focused on analyzing and managing mission and system cyber risk and trade-offs across the life cycle. Cybersecurity engineers evaluate interactions, dependencies, and system response to attacks. They identify security practices and mechanisms that need coordination across the life cycle, spanning components, people, processes, and tools. They prepare the technology to handle the operational environment where it will ultimately reside. In this paper, we introduce the purpose of a cybersecurity strategy and describe the role of cybersecurity engineering in implementing it. We identify six key cybersecurity engineering activities and share observations on how these activities can be used to address the challenges acquisition programs face as they work to improve cybersecurity under resource and time constraints.

A Strategy for Cybersecurity in Acquisition

Most acquisition programs currently address cybersecurity by relying on a small cadre of security experts that is not adequately integrated into the acquisition and development life cycle. These security experts are often tasked to evaluate compliance with mandates and to assure good cyber hygiene. Called upon only to address security, they may have limited systems and software domain knowledge.

While compliance mandates must be met for an authority to operate and good cybersecurity hygiene is necessary, these alone will not ensure sufficient security for a mission. Today's missions call for a comprehensive cybersecurity strategy, built and implemented with extensive collaboration among participants in a variety of roles throughout the life cycle.

What Is an Effective Cybersecurity Strategy?

A cybersecurity strategy is a critical element of an acquisition program's Program Protection Plan (DoD, 2020), establishing a foundation for full life cycle cybersecurity assurance. An effective cybersecurity strategy defines acquisition and development life cycle activities, roles, and responsibilities essential to ensure the system can be fielded with an acceptable level of risk. The cybersecurity strategy identifies planning, design, monitoring, and enforcement considerations for integrating cybersecurity into technology at and across all levels of a system with a goal of exposing and removing

- gaps that allow unexpected and unwanted interactions among components and between components and external systems.
- weaknesses in the design and implementation that make the system vulnerable.
- opportunities for processes or people to bypass controls, allowing attackers to impact mission success.

To accomplish this goal, the cybersecurity strategy expands activities, roles, and responsibilities for a program's cybersecurity experts far beyond compliance and hygiene, to owning and leading execution of the strategy itself. The strategy's owner is responsible for defining how a system's cybersecurity must perform to meet the system's mission, even when under attack. These responsibilities include activities that achieve the following:[1]

- Plan and design trusted relationships.
- Negotiate appropriate security requirements to ensure confidentiality, integrity, and availability with sufficient monitoring in systems and software to identify problems.
- Plan and design a system with sufficient resiliency to be able to recognize, resist, and recover from attacks.
- Plan for operational security under all circumstances, including designed-in methods of denying critical information to an adversary to avoid or minimize mission impact.
- Evaluate alternatives to select an acquisition option with an acceptable level cybersecurity risk.

As is evident from this list of activities, building a cybersecurity strategy is a challenging effort that touches every aspect of the life cycle. Participants contributing to the strategy come with a variety of backgrounds and frequently, different vocabularies. They must build a shared understanding of the mission, how security risks can impact mission success, and the cybersecurity activities, roles, and responsibilities needed to mitigate the risks. More details about defining the cybersecurity strategy are available in the paper "Building a Cybersecurity Strategy" by Carol Woody and Robert Ellison, published in the Special Issue on Rigor and Inter-Disciplinary Communication of the *Journal of Systemics, Cybernetics and Informatics* (Woody & Ellison, 2020a).

---

[1] Key elements of an effective cybersecurity strategy were introduced in a webcast delivered in November 2020 (Woody & Creel, 2020).

To effectively execute a cybersecurity strategy, a high level of collaboration is needed across acquisition and development roles. *Cybersecurity engineering* is a full life-cycle discipline that operationalizes the cybersecurity strategy and orchestrates collaboration across acquisition program stakeholders. The next section of this paper describes challenges in six key cybersecurity engineering activities related to strategy execution.

Executing the Strategy: Cybersecurity Engineering Challenges

Cybersecurity engineers are experts who own and lead execution of the cybersecurity strategy. They articulate the roles, responsibilities, and collaboration touchpoints across the life cycle needed to assure system security for the mission. Their knowledge and experience, a combination of operational security, systems engineering, and software engineering, is difficult to find, but increasing demand is stimulating interest and growth in the discipline.

Cybersecurity engineers working on an acquisition program must understand systems, software, and cybersecurity. In addition, they must be sufficiently committed to the program to have impact. Attending occasional meetings is insufficient; rather, they must engage early in the life cycle to affect requirements and design trade-off analyses, risk assessments, and where possible, contract language. Focusing on code and other downstream artifacts—which most security analyses and verification tools target—is too little, too late. Flaws and weaknesses can be introduced early in the life cycle even before code is introduced.

Finally, cybersecurity engineers must consider the overall engineering and operational context, the system's engineering, operational, and sustainment environments. In addition to operational threats, sources of risk include software-enabled tools used to build, test, or maintain hardware, software, and firmware; communications between suppliers, and between acquirers and suppliers; and people who may unintentionally or intentionally compromise the acquisition program. Cybersecurity engineers develop and analyze scenarios for what could go wrong; recommend modifications to products, processes, and resources to mitigate risks; and monitor system development and verification to confirm that the system can withstand operational misuse and abuse.

The following six key cybersecurity activities are important to executing the cybersecurity strategy:

- determining risk
- defining and monitoring system and component interactions
- evaluating trusted dependencies
- anticipating and planning responses to attacks
- coordinating security throughout the life cycle
- measuring to improve cybersecurity

What Makes Effective Cybersecurity Engineering Challenging?

Cybersecurity engineering is challenging because it attempts to change existing patterns in the way security's role is perceived and executed on acquisition programs, and because it requires a combination of knowledge and expertise that is rare. Where cybersecurity engineers are found, they often represent a lone voice in a sea of technical experts and managers where they struggle for inclusion.

In the remainder of this section, we review the six key cybersecurity engineering activities, exploring challenges and lessons learned to date in changing existing patterns and increasing cybersecurity engineers' involvement in early acquisition activities. We also share effective approaches we have seen in each of these areas that can help address the challenges.

Challenges in Determining Risk

Perceptions of risk drive assurance decisions, and the lack of cybersecurity expertise in acquisition and development risk analysis has led to poor assurance choices. Involving participants who have knowledge about successful attacks and how threats can impact a system's operational mission can be very useful in the decision-making steps for determining appropriate prioritization. However, to effectively communicate cybersecurity issues to leadership and systems and software engineers in the program office, these participants must have more than a rudimentary understanding of the system and software acquisition and development life cycle.

Risk considerations must start with good cybersecurity requirements, but simply providing a list of standards and policies—an approach too frequently used by programs—does not ensure that risks to the specific mission and technology are addressed. New versions of standards and policies are now appearing so fast that programs have little time to respond to them; they then continue working with outdated versions until contract modifications can be made. We have also seen programs fail to establish even minimum cybersecurity or compliance requirements in the contract and the vendor plans to deliver a system that will not be implementable without contract modifications and further development. When developing good cybersecurity requirements, participants must do the following:

1. understand how the many layers of technology to be implemented to address a mission can be compromised;
2. define requirements that enable consideration of the broad and ever-increasing threat environment;
3. clearly define how success will be verified.

Requirements for many aspects of a system contribute to cybersecurity. The structure of these requirements affects how the vendor will respond. When requirements for anti-tamper, supply chain, information security, software assurance, cybersecurity, and safety controls are developed by independent teams that do not collaborate, the contractor cannot clearly propose a system that meets what could be conflicting needs. The vendor will make choices based on their perspective, which will likely differ from the intent of the program. Adjustments must be made later in the life cycle as time and funding permit, and the results will vary widely depending on the knowledge of the decision makers involved at each subsequent stage. Security teams are usually too small to support all of the interactions needed for this approach to cybersecurity risk decision-making.

In many cases, the cybersecurity risks and potential mission impacts of choices made in design and engineering are not well understood by system and software engineers who (1) were not trained in cybersecurity and (2) do not understand the capabilities of attackers in the contested environment where the systems are fielded. The program too frequently accepts risks without understanding the mission impact.

Challenges for Defining and Monitoring System and Component Interactions

In systems design, we see engineers rush to decompose the system into technology components and delegate requirements to these various pieces. Interface protocols and application programming interfaces (APIs) are defined to provide mechanisms for data sharing among components. Cybersecurity controls are selected from a wide array of sources and sprinkled like “fairy dust” among the components. However, nowhere is there a coherent plan for how the overall system cybersecurity will operate across the attack surface of the system. Views of the hardware, network interfaces, software interfaces, and mission capabilities are provided, but the operational aspects of cybersecurity are fragmented.

To address this void, programs must define how the technology will support mission execution and evaluate how information flows and data exchanges can be attacked. To conduct a cyber-threat analysis of a system, we applied the Security Engineering Risk Analysis (SERA) Method (Alberts et al., 2014). The SERA Method defines a scenario-based approach for analyzing complex security risks in software-reliant systems and systems of systems. The SERA Method incorporates a variety of models that can be analyzed at any point in the life cycle to (1) identify security threats and vulnerabilities and (2) construct security risk scenarios. An organization can use those scenarios to evaluate whether planned controls will be sufficient and focus its limited resources on enhancing how to meet the needs of mitigating the most significant remaining security risks.

For the most part, we see programs failing to assemble the information they have in ways that enable their overall cybersecurity to be evaluated. System and software engineering analysis does not consider the real threats and the ways their designs and selected controls can be compromised. In many cases, these acquisition and development technology teams do not include resources that understand how technology can be attacked, and they do not sufficiently consult external resources with this knowledge to help support the design analysis. By the time the separate security team sees the design, development is typically already underway. Design weaknesses are left in place because they are discovered too late to be corrected without major cost and schedule impacts.

Challenges for Evaluating Trusted Dependencies

Security dependencies relate to components that some build for others to use or connect with (i.e., inherited risk). The following are known areas of concern for inherited/dependency risk:

- Each dependency represents a risk (e.g., reuse, open source, collaboration environments).
- Dependency and trust decisions should be based on realistic assessments of the dependency’s inherent risks and opportunities.
- Dependencies are not static, so trust decisions must be re-evaluated regularly to identify changes that warrant reconsideration.
- Using shared components to build technology applications and infrastructure increases the dependency on and potential mismatch of others’ decisions (i.e., supply chain risk).

Layer on layer of reused software is integrated into systems based on the way the software is built and fielded. Code libraries allow developers to quickly assemble functionality, reusing language constructs and functionality constructs for frequently applied program actions. Tools for automating many of the development and testing steps are assembled into integrated development platforms (IDPs) for ease of use.

It is cost effective to reuse components and code previously developed for other projects to save time in development. However, that material was not created to meet the specifications of the current system and could have vulnerabilities introduced by the original developer; in these cases, the new system will inherit these vulnerabilities. Reuse practices that minimize the impact of this risk must be established. Since the current developers may not be familiar with the code, addressing vulnerabilities may be more time-consuming than rewriting. In many cases, reuse is not cost effective when total costs are considered.

The reuse challenge also applies to commercial-off-the-shelf products and other third-party code used to address needed system functionality. Shared infrastructures, such as cloud environments, fall into this category. In some cases, the responsibility for addressing cybersecurity risk falls to programs when they are incorporating reused components into the system; sustainment planning for the system must reflect these issues. Using patching to address known vulnerabilities must be scheduled in a timely manner to minimize attacker opportunities—during both development and operations. In other cases, the responsibility of handling the risk must be spelled out in contracts or service level agreements (SLAs) established with the provider of services. We see lack of clarity in who is responsible for addressing risks from these sources and the timeliness in which cybersecurity issues will be addressed. Programs consider cost savings for services without establishing the critical mechanisms needed to ensure that the trust relationships that need to be established will operate with sufficient cybersecurity.

A cybersecurity strategy should provide plans for implementing and monitoring cybersecurity risk for each type of reuse implemented by a program. (Our experience is that, in practice, there is little planned, and response is primarily reactive.) Each plan involves establishing a vendor relationship that must be managed. Recent increases in supply chain risk show that this is an area where programs need to improve greatly.[2] Further information about inherited risk is available in an SEI white paper on reducing the attack surface for a system (Woody & Ellison, 2020b).

Challenges for Anticipating and Responding to Attacks

A growing and diverse population of adversaries uses both simple and increasingly sophisticated capabilities to compromise the confidentiality, integrity, and/or availability of technology assets. Adversaries often use a combination of technology, processes, standards, methods, and practices to craft a successful attack (i.e., socio-technical responses). Attacks are designed to take advantage of the ways users normally use technology or to contrive exceptional situations where users' defenses are circumvented. Attackers' profiles, capabilities, and methods are continually evolving.

To be successful, an attacker must bring together three key pieces:

- a weakness in the design, coding, or implementation of the system that can be triggered to promote unexpected and unwanted execution capabilities that can compromise confidentiality, availability, and/or integrity of the system;
- access to that weakness through connected trusted systems, compromised authentication and authorization capabilities, or other attack vectors;

---

[2] See the SEI webcast, *SolarWinds Hack: Fallout, Prevention, and Recovery* (Software Engineering Institute, 2021).

- tools that enable the attacker to exploit the weakness, bypass security controls, and gain unauthorized system capabilities to trigger unwanted system behaviors.

Much attention is focused on potential vulnerabilities in code, and extensive effort has been expended to improve and implement a wide range of code analysis capabilities to reduce the availability of weaknesses for attackers to exploit. However, design weaknesses introduced in architecture and design implementation represent well over half of the top 25 common weaknesses and enumerations (CWEs; Common Weakness Enumeration, 2020).

These weaknesses, which are prevalent in today's systems and software engineering, result from insufficient security requirements and insufficient understanding of how that technology can be compromised. Knowledge about the ways that technology can be compromised is still rarely taught in college and university curriculums; therefore, it may not be a part of the technical knowledge of someone with a background in computer science or system engineering.

As reuse continues to expand (see the "Challenges for Evaluating Trusted Dependencies" section above), broad availability of code increases through open source and acquisition sources as well as theft. The same tools that can be used to remove weaknesses can also be used by attackers to identify unmitigated weaknesses to exploit. There are also tools for reverse engineering operational modules to recover the source. Versions of these tools are free, so the capabilities available to an attacker to discover and exploit weaknesses are extensive.

With both the level of software and the pool of potential exploits expanding, it is critical to update code as soon as new exploits are identified. Too much software was not designed and implemented to be effectively maintained for cybersecurity. Costly and cumbersome certification processes have not kept up with the realities of attacker capabilities, and some system engineers may not understand the attack potential of the software they implement. As a result, the cybersecurity strategy is too limited to address the critical realities of system vulnerabilities.

Challenges for Coordinating Security Throughout the Life Cycle

Assuring the security and resiliency of a system's mission-critical functions requires the following activities:

- Plan for what might go wrong; develop requirements for misuse/abuse cases with corresponding system and software requirements for secure, resilient response and operation.
- Build security and resiliency into the system.
- Verify that expected security and resiliency characteristics were actually built into the system.
- Establish clear authority and responsibility for coordination at an appropriate level in the organization to ensure effective participation and coverage.

The lack of attention on misuse and abuse cases is increasingly a major gap in system requirements. Some programs leverage reusable components and shared platform capabilities as they rush to define functionality and identify ways that cost and schedule can be reduced to deliver results as quickly as possible. However, in their rush, they often fail to sufficiently specify how the system should perform under attack and compromise. Even when threat modeling is well established, system requirements lack the sufficient integration of threat knowledge to drive planned capabilities. Reliance is placed on selected compliance controls, but these are not tightly coupled with system operational capabilities, leaving gaps where unexpected behaviors and attempts to bypass controls are supported. Completeness of system capability is assigned to the system engineering team to ensure the system meets its requirements. However, from what we have seen, no one is formally given the responsibility for developing a complete set of misuse and abuse cases to ensure all reasonable threats are addressed.

With the transition to iterative development and an increased adoption of DevSecOps, the development approach and environments now become a permanent part of the operational system. These must be in place to support the creation and fielding of new functionality and updates for long periods of time. Hence, they must be treated as part of the operational environment and hardened sufficiently to ensure they do not become conduits for attack of the system and its components. Traditionally, development environments do not have this level of rigor applied to them since they were thought of as temporary. For many programs, this perception of transience still remains.

For every system, cybersecurity is not the only critical quality. Reliability, maintainability, usability, safety, anti-tamper, and other system attributes, as well as cost and schedule, are equally important. Program management must establish how cybersecurity experts will collaborate with those addressing these other important attributes to define how trade-offs will be managed and identified. Choices that impact mission risk should not be made without careful thought and consideration. We see too many programs that fail to establish coordination mechanisms among the various groups, managing each attribute and determining trade-offs considering only cost and schedule. As a result, mission-critical capabilities can be jeopardized.

Challenges for Measuring to Improve Cybersecurity

Cybersecurity attributes of critical products, processes, and resources must be identified, measured, and monitored throughout the life cycle. The best means for identifying metrics is by using the Goal-Question-Indicator Measure (GQIM) method, which identifies and defines meaningful indicators that align management, engineering, and improvement with business goals.[3]

The cybersecurity strategy should describe how a program will measure improvement. Opportunities would include the following:

- Use risk assessments that consider both current and potential threats, vulnerabilities, and impacts to identify critical goals for attributes to be measured and monitored.
- Develop and consistently apply well-formed measurement definitions and procedures to establish the credibility of the measurement and analysis results.
- Include all elements of the socio-technical environment that touch engineering and acquisition activities (e.g., processes, procedures, products, and resources).

---

[3] SEI training materials are available for GQIM. For more information, see *Goal-Driven Measurement (IGDM) SEMA Course* on the SEI website (https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=635664).

- Support measurement with robust engineering planning; define a security measurement plan that spans the life cycle, and develop requirements for any needed instrumentation.

Every activity within the life cycle as well as each process and tool can produce data to feed a measurement effort. The challenge is determining where improvements in current practice would be useful for mission success and identifying the information that can help determine if a change has been successful.

We see programs measuring lots of individual activities, but many programs are not defining how these low-level metrics are useful to management to determine if expected cybersecurity results will be achieved. As a start, the strategy must define a vision of successful cybersecurity for the program. If the team assembled to address cybersecurity is sufficiently diverse, there will be several visions of success. The plan must consider ways to measure against each of these visions to provide a broader view of the potential goal. To date, we have seen many metrics collected, but little planning for what is needed. Common activities we have seen include using checklists for compliance and counting vulnerabilities; however, these activities alone will not provide a means for measuring cybersecurity improvement.

Summary

Building a cybersecurity strategy is challenging, and applying it to a system under acquisition and development is even more so. We have been working with a broad range of acquisition programs to tackle this important need, but changing the existing patterns of cybersecurity neglect is extremely difficult.

A cybersecurity strategy is a critical planning document for program security. It must define the actions a program plans to perform to address cybersecurity. The strategy must clearly establish who is responsible for its success and the range of personnel resources that should be used to support its success. Those who implement the plan are a key element to its success. They must be knowledgeable in cybersecurity and system and software engineering and sufficiently focused on the need to provide results. They must be able to navigate the program's acquisition and development life cycle and keep the plan current with trade-offs that are made across the life cycle that expand potential risk. These plan implementers must be invested in ensuring that cybersecurity is addressed for the program; they should not just be brought in periodically to glance at materials and sit through a few meetings.

The cybersecurity strategy will not remain static. Recognizing when changes are needed is a critical aspect of monitoring change and improvement. Metrics can play a valuable role in this aspect.

The strategy must cover more than just compliance; it must address the system's attack surface and make every effort to limit risks from all forms of technology (e.g., hardware, software, and firmware) and all sources of technology (e.g., reuse, third-party components, and external services).

On the whole, programs continue to react to cybersecurity instead of building it into all aspects of the system and its life cycle. A well-planned cybersecurity strategy can help bridge this gap to support improved mission success.

We hope this paper will motivate cybersecurity strategy owners to refocus their efforts on delivery of effective cybersecurity results. They should ensure the right resources are empowered to drive cybersecurity improvement throughout the acquisition and development life cycle.

References

Alberts, C. J., Woody, C., & Dorofee, A. J. (2014). *Introduction to the security engineering risk analysis (SERA) framework* (CMU/SEI-2014-TN-025). Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?AssetID=427321

Common Weakness Enumeration. (2020, August 20). *2020 CWE top 25 most dangerous software weaknesses.* https://cwe.mitre.org/top25/archive/2020/2020_cwe_top25.html

DoD. (2015, November 10). *Violence in the workplace. Memorandum: Outline and guidance for acquisition programs' cybersecurity strategies.* https://www.acqnotes.com/wp-content/uploads/2014/09/DoD-CIO-Cybersecurity-Strategy-Outline-and-Guidance-10-Nov-15-1.pdf

DoD. (2020, July 20). *Technology and program protection to maintain technology advantage* (DoD Instruction 5000.83). Office of the Undersecretary of Defense for Research and Engineering. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500083p.pdf?ver=2020-07-29-120425-050

Mead, N. R., Shoemaker, D., & Woody, C. (2013). Principles and measurement models for software assurance. *International Journal of Software Engineering*, *4*(1). https://doi.org/10.4018/jsse.2013010101

Software Engineering Institute. (2021, February 10). *SolarWinds hack: Fallout, recovery, and prevention* [Webcast]. https://www.sei.cmu.edu/news-events/events/event.cfm?customel_datapageid_5541=308529

Woody, C., & Creel, R. (2020, November 15). *What is cybersecurity engineering and why do I need it?* [Webinar]. Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?assetID=650076

Woody, C., & Ellison, R. (2020a). Building a cybersecurity strategy. *Special Issue on Rigor and Inter-Disciplinary Communication of the Journal of Systemics, Cybernetics, and Informatics (JSCI)*, *18*(1), 206–216. http://www.iiisci.org/journal/sci/Contents.asp?var=&next=ISS2001

Woody, C., & Ellison, R. (2020b). *Attack surface analysis—Reduce system and organizational risk.* Software Engineering Institute. https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=650193

Interdependence Analysis for Artificial Intelligence System Safety

Bruce Nagy is a Research Engineer at the Naval Air Warfare Center, Weapons Division at China Lake. His research focuses on advanced game theory techniques, artificial intelligence, and machine learning applications for tactical decision aids. Nagy has earned four degrees: one in mathematics, two in electrical engineering, and one in biology from The Citadel and the Naval Postgraduate School. He led the development of advanced algorithms and metrics that resolved national defense issues in satellite communications for the DoD. At UCLA, during postgraduate work, he investigated modeling brain stem communication with muscle groups at the cellular level, in cooperation with the NIH. [bruce.nagy@navy.mil]

Scot Miller is a Faculty Associate Researcher at the Naval Postgraduate School. Prior to that, he served 29 years in the Navy as a P-3 pilot and Information Professional. He holds bachelor's and master's degrees in Operations Research from the U.S. Naval Academy and Naval Postgraduate School, and other graduate degrees from the University of West Florida and the Naval War College. [scot.miller@nps.edu]

Abstract

Engineers responsible for evaluating tactical and weapons systems for system safety will need a new approach for evaluating emerging artificial intelligence (AI)-enabled systems, since these systems leverage machine learning (ML) techniques. For many reasons, ML algorithms are often difficult to diagnose for safety purposes. For instance, they did not lend themselves easily to codebase inspections, thus necessitating the reduction in "autonomy" of the ML-enabled component. By modifying Interdependence Analysis (IA) techniques, a more rigorous approach to evaluating AI/ML-enabled weapons can be found. The IA process produces a rigorous exploration based on observability, predictability, and direct ability, highlighting the key requirements that encapsulate all interactions between human and machine. This paper explores using IA to define the interaction requirements for human–machine teaming, employs those results to identify key critical functions, and leverages those findings to reveal how "autonomy" reduction might be employed.

Introduction

Artificial intelligence (AI) increasingly is used in many regimens. When AI is used in weapons systems, are there new considerations the tester, certifier, and operators should consider? If so, how should they handle them? Let us define AI. AI is "the theory and development of computer systems able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages" ("Artificial Intelligence," 2021).

Johnson (2021) describes two types of AI. The first is handcrafted knowledge systems that use traditional rules-based software to codify subject matter knowledge of human experts into a long series of programmed rules (Johnson, 2021). These are simply just what we have always called computer applications. Various test and certification agencies execute well-developed procedures for assessing these applications and are not considered.

The second kind of AI becoming more prevalent is machine learning (ML) and this demands investigation. In ML, test data trains an algorithm to identify something or some pattern. Once the system has learned, it is ready to identify similar or the same things or patterns from more incoming data. There are variations on exactly how this works. One important variation is that if one is trying to find a specific thing or pattern in the real world, the system must have "seen" that data in the training data set.

These types of ML capabilities have improved accuracy, often in many cases achieving 90% success in identifying the object. That is a great accomplishment for the ML algorithm, if it is designed to recommend possible work shoes when one is shopping for new shoes. However, if the algorithm is distinguishing between a church steeple, the tower of a hospital, and a medium range ballistic missile launcher, 90% does not inspire user confidence, when destroying the missile launcher is the goal. Moreover, the ML algorithms work in mysterious ways, processing a myriad of complex mathematical equations, meaning it is often unrealistic to determine how the algorithm succeeds. For ensuring the safety of weapons that include ML-enabled components, evaluating ML functions in detail is important. This paper identifies the potential challenges of employing ML and how human–machine teaming engineering design techniques contribute to a rigorous ML system safety assessment.

Safety Failure Modes of Artificial Intelligence Machine Learning

The Navy organization responsible for certifying weapons as system safe is the Naval Ordnance Safety and Security Activity (NOSSA). They employ the MIL STD 882E as their rule book (DoD, 2012), and the Joint Software System Safety Guide as their compass (Joint Services-Software Safety Authorities, 2016). As part of that review, evaluation, and certification process, they conduct hazard analyses of the weapons system under test and associated supporting elements. Their intent is to identify all possible outcome likelihoods and their consequences for every system function.

An example from the previous section highlighted a 10% likelihood that the AI ML algorithm would identify a hospital or church as the target, with the loss of many innocent lives, waste of an otherwise good weapon, and a black eye for the United States. That is just one example of possible AI ML failures. Others are listed in Table 1 (Faria, 2017). They break into three categories: system failure, human–machine interaction issues, and active sabotage. Examples of each might be: 1. ML training data set has inherited biases that skew results; 2. humans assume ML algorithm is always right and do not apply critical thinking to results; and 3. an adversary manipulates the training data set.

TABLE 1. EXAMPLES OF AI SYSTEM FAILURE MODES (Faria, 2017).

Failure Category	Failure Mode Examples
System Produces Faulty/Poor Decision Recommendation	Biased outcomes/predictions
	Skewed outcomes/predictions
	Uncertain outcomes/predictions
Human–Machine Operation Issues	Operators have lack of trust in the system
	Operators are overly trusting (overreliant) in the system
	Operators ignore the system
	Operators misunderstand the system recommendations/predictions
	Operators introduce errors into the system
System Under Attack (Cyber attack)	System is overtaken by adversary/adversary is controlling system
	System and its outcomes are corrupted by adversary
	Adversary jams or shuts down system
	Adversary gains access to system; decision information/knowledge is compromised

The consequences of these failures depend on which part of the kill process of the weapons was affected. Table 2 provides a functional view of general weapons.

TABLE 2. GENERAL WEAPONS FUNCTIONS

Functions	Details
Transport	Truck, Rail, Ship, Aircraft
Storage	Bunker, Warehouse, Environment
Load on Platform	Pier side, Unrep, Cranes, Vertrep
Maintenance	Power, BIT, Lubricants, Fueling
Readied	Power on, Check out
Placed on Launcher	Movement, Rails, Hoists, Handled
Launched	Fire, Released, Sent
Navigate	Waypoints, Terrain
Avoid	Terrain, Weather, ESM
Deceive	ECM, ESM, Cloaking, IR
Identify	Target Recognition, Human Input, Sense Laser Illumination
Fuze/Arm	Timing, Mechanical, Human Input, Sensing
Effects	Detonation, Dispersal, Virus, Programs

Observation shows that many of the functions in the table could be supported by ML tools. Navigation and target recognition could be aided by an ML computer vision tool, as well as avoiding other platforms or specific topographical entities. It stands to reason that NOSSA needs to consider all of these failure modes for each weapon function that uses a ML component.

Each failure mode is comprised of possible root causes, as seen in Table 3 (Johnson, 2021). With 31 root causes and at least 15 functions, there are at least 465 possible failure combinations to assess in NOSSA's failure analyses.

TABLE 3. EXAMPLES OF ROOT CAUSES OF AI SYSTEM FAILURES (Johnson, 2021)

Type of Root Cause	Root Cause Examples
Issues within the training datasets	Biased training datasets
	Incomplete training datasets
	Corruption in the training datasets
	Mis-labeled data
	Mis-associated data
	Lack of rare examples—data doesn't include unusual scenarios
	Unrepresentative datasets
Issues with the process of data validation	Poor data collection methods
	Poor data validation methods
	Improper data validation criteria
	Insufficient data validation
Issues with the ML algorithms	Underfitting in the model—when the model does not attain sufficiently low error on the training data
	Overfitting in the model—when the model presents very small error on the training data, but fails to generalize to new data
	Cost function algorithm errors—when the trained model is optimized to the wrong cost function
	Wrong algorithm—when the training data is fit to the wrong algorithmic approach or mathematical model
Issues with the operational datasets	Uncertainty/error in the operational datasets (Epistemic uncertainty)
	Corruption in the operational data
	Introduction of datatypes that the AI system is not designed to handle
	The pace of the situation overwhelms the human–machine decision process

Type of Root Cause	Root Cause Examples
Operational complexity	The decision space overwhelms the decision process (the number of options is too large or a viable option does not exist)
Operator trust issues	Lack of explainability
	Lack of confidence
	Overreliance
	Insufficient operator training or experience with system
Operator induced error	Inverse trust issues in which the AI system loses "trust" in the human operator or identifies operator problems
	Operator misuses the system accidentally or intentionally
	Operator fails their part in the decision process accidentally or due to being overwhelmed, negligent, or confused
Adversarial attacks	Hacking
	Deception
	Inserting false or corrupt data
	Gaining control of the AI system

A weapon uses radar computer vision to identify a turn point as part of its navigation. A possible root cause for failure might be mislabeled data. In this case, the turn point is mislabeled as 1,500 feet in elevation, when it should be labeled 1,500 meters in elevation. The likelihood that the weapons might fly into the terrain is high, with the consequence that the weapons effects are not delivered.

Here is a more insidious example. A weapon requires a system update, for instance, a new load of training data. The weapon is already loaded onboard a ship, so the ship's weapons maintenance team is assigned the task. Their communications connection to the shore base where the file will be downloaded is sketchy. Despite best efforts, the download is interrupted several times due to latency and signal jitter (common occurrences underway). Even the checksums that ensure a good transmission are misplaced, leading the ship to think it has a good download, when it really does not. Now there is an unknown likelihood the weapons could malfunction, and unknown consequences, and no way to check.

Understanding every function in the life cycle of a combat or weapons system is critical. Existing procedures already account for the vast number of functions that are not related to any ML support and are not addressed in this paper. However, there needs to be a way to determine how ML-related techniques impact the weapons and associated operators.

Tables 1–3 list many ways that ML functions cause safety issues. If their likelihoods and consequences are significant, these functions can be designated critical, which by the MIL STD requires applying a scrupulous code check process. This is impractical for ML functions, so reducing the criticality of the ML functions remains paramount.

Criticality is associated with "autonomy," which in this case means that the function will operate automatically without human intervention. Thus, if the function can be modified to include human involvement, then the criticality is not rated as high, and no code check is required, which then resolves the evaluation conundrum of critical ML functions. One proven way to investigate human–machine interaction is through a system engineering technique called Co-Active Design, especially a key component called interdependence analysis (IA). Does applying IA to the ML functional evaluation help?

Co-Active Design

The purpose of IA is understanding how people and an agent (in our case the ML function) can effectively team by identifying and providing insight into the potential interdependent relationships used to support one another throughout an activity (Johnson et al., 2014). One of the IA tool's strengths is that it can guide the initial design and anticipate how the tool or system will be used by the warfighter. In this specific case, though, NOSSA is most interested in how humans can be involved in the application of a ML algorithm so as to ensure that it is not deemed a critical function. As we have seen, critical functions require a software code review, which is problematic, since advanced ML algorithms are complex, especially in the mathematics involved, and a software code review is not practical.

The IA tool is in the form of a table, as shown in Figure 1. IA breaks the planning process down hierarchically into specific tasks. Those tasks are analyzed to determine the necessary capacities needed to conduct each function, using cognitive task analysis techniques (Annett, 2003; Bolte et al., 2003; Chapman et al., 2009; Crandall et al., 2006). Once capacities are identified, the next step is assessing each performer's (ML component or human) ability to provide that capacity. The assessment process uses a color-coding scheme, as shown in Figure 2. The color scheme is dependent on the type of column being assessed. The color assessment is subjective, but the process compels the analyst to carefully consider the algorithm and human interactions. Under the "performer" column, the colors are used to assess the individual's capacity to perform the activity specified by the row. The green color in the "performer" column indicates that the performer can do the task. Yellow indicates less than perfect reliability. Orange indicates some capacity, but not enough for the task. Red indicates no capacity.

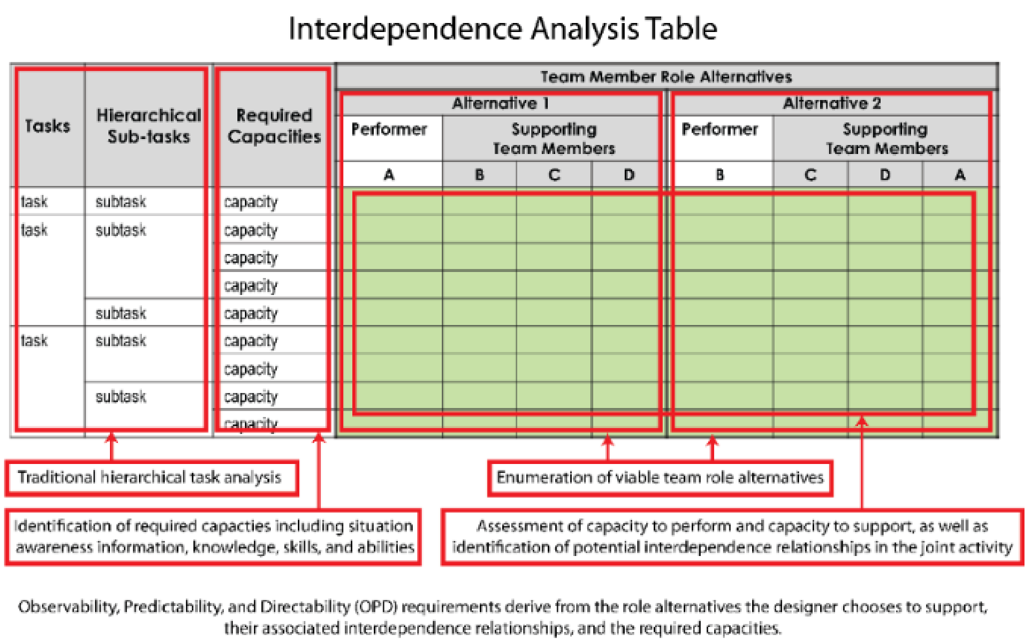


Figure 1. Explanation of the Different Areas of the Interdependence Analysis (IA) Table (Johnson, 2014)

Color Key for Team Member Role Alternative Capability Assessment	
Performer	Supporting Team Members
I can do it all	My assistance could improve efficiency
I can do it all but my reliability is < 100%	My assistance could improve reliability
I can contribute but need assistance	My assistance is required
I cannot do it	I cannot provide assistance
Not applicable / Not significant	Not applicable / Not significant

Figure 2. Color Dey for Team Member Role Alternative Capability Assessment (Johnson, 2014).

Once completed, careful analysis of the results displays or infers several key design inputs for the human–machine teaming involved in the robot delivery process:

1. From the color schemes, analysts determine how many possible paths exist between the users and the systems to accomplish the tasks. This identifies areas where multiple paths are availability (improving reliability), and where single points of failure may compromise success (and require further design attention). And indeed, may be showstoppers to the NOSSA evaluators.
2. Analysis shows where the system is required to either ensure effectiveness or improve efficiencies. Teaming is achieved by practicing three actions between the performer and the supporting team members: staying observable, predictable, and directable (OPD). Where there are opportunities for a team member to assist the performer, and vice versa, engineers and designers can brainstorm ways to achieve that teaming through the lens of OPD. In this case, NOSSA evaluators seek ways for the AI-enabled function to be either supported by or monitored by humans. This reduces "autonomy" and the criticality of the system function.
3. Combining these first two results helps designers and developers prioritize efforts on the machine's development. Which capacity executed by the performer adds the most value? This establishes a context for organizing work efforts and resources. From a NOSSA perspective this analysis helps focus on potential ML-enabled trouble areas. As described in Table 1, three areas of potential issues exist: system failure, human–machine interaction issues, and active sabotage. As evaluators use the IA, when they evaluate the OPD between the ML algorithm and the humans, they should also consider these three possible failure mode areas, associated possible root causes from Table 3, and which general functional area is impacted from Table 2.

Applying IA to ML Safety Use Case

Nagy (2021) created an unclassified use case for exploring system safety aspects of ML-enabled systems (see Figure 3). We use that scenario as a use case for the IA. The results are depicted in Table 4. In this scenario, a truck is loaded with two robots, which are then driven to a point, then unloaded. For simplicity, assume predecessor functions such as storage and maintenance (from Table 2) have no ML components. The robots proceed via another route to deliver packages to one or two recipients. As will be introduced in the analysis, at several stages in the process ML-enabled techniques are used to conduct activities or make decisions. Understanding how the human–machine teaming is associated with those activities and decisions is critical to the evaluation process that NOSSA uses. Many of the other functions normally considered in an IA were skipped as not relevant to the AI safety discussion.

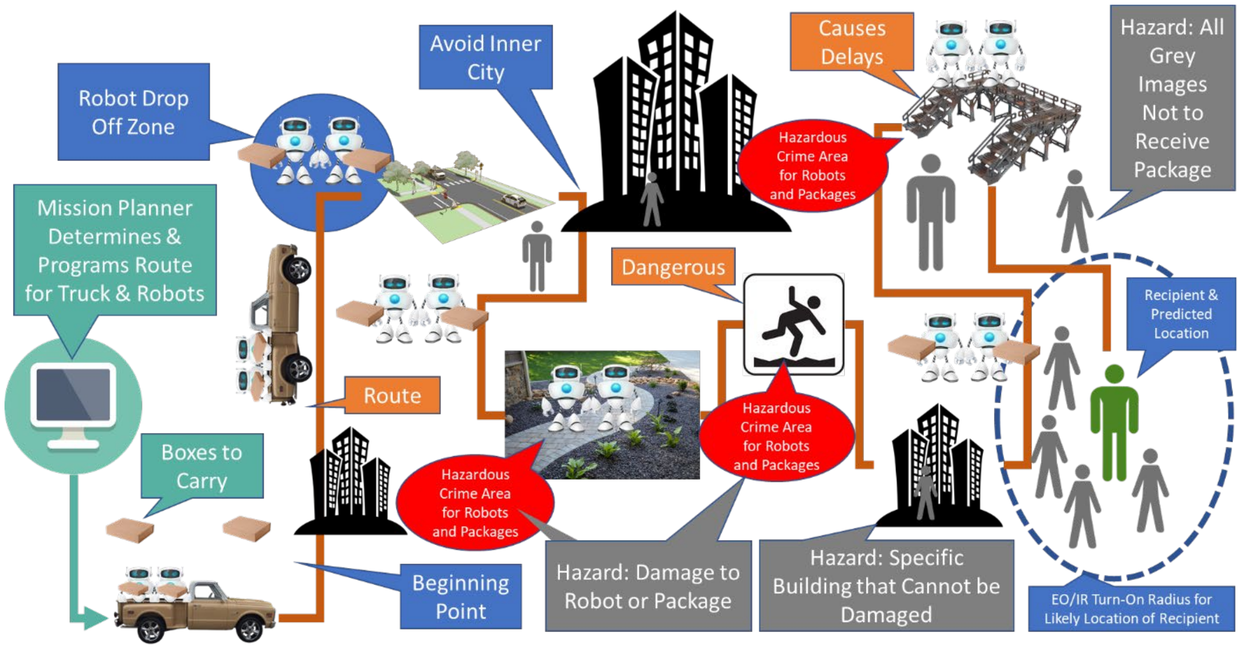


FIGURE 3. AI-ENABLED ROBOT DELIVERY SAFETY USE CASE (Nagy, 2021)

The above use case supports two functions. The first involves mission planning, defining the route and robots to use. The second function involves the management of the robot. There are two performers of interest in the scenario, the robot and the user. For the first system, the user must agree to both the route and robots to use. In the second system, the user interfaces with the robot through a graphical user interface (GUI). In most IA, two options are available. One where the unmanned system is chosen to perform functions, and humans provide support, and vice versa. In this case there is just one option, that the robot traverses a route and delivers the packages. This simplifies the IA.

In column D, we designate which part of the robot is executing that capacity. For example, route movement is the role of the stabilizer and propulsion mechanisms, while identify customer is a segment of the computer vision ML module. These distinctions are crucial, since if an ML module is involved, NOSSA evaluators need to up their awareness. The supporting performer is either the mission user, but in NOSSA's case, the evaluator. Column F describes the interactions between the robot and human, described by the three OPD characteristics. In this column we evaluate how the human uses those techniques to work with the robot. As one can see by examining the Column F, OPD in this scenario is accomplished through the GUI. The GUI is well designed and enables specific windows for observing location and robot state, provides a route details window which predicts the robots next moves, and specific robot and route approval buttons to provide direction. An additional geography map provides back up situation awareness for users. A final feature, crucial for NOSSA evaluators, is that the geographic map on the GUI can also be toggled to a statistical display.

Details associated with the GUI's capabilities are relevant to the interdependence possibilities for the ML-enabled components of the overall system. They enable the GUI to reduce the autonomy ML related functions within the delivery robot. These specific GUI functions are listed in Table 5.

TABLE 5. SCENARIO GUI FUNCTIONS

Number	Scenario GUI Function Details
1	Enables manual take over control, e.g., direction, speed, avoidance
2	Provides awareness of autonomous system's actions—and change through "Affiliation" designation—who receives package
3	Option to cancel delivery actions
4	Verify recipient visually
5	Abort operations if needed
6	Change the success threshold (a statistic identifying the likelihood of success) when selecting recipient and related package delivery approach
7	Modify methods for recipient recognition or navigation method, from a ML function to traditional approach

Figure 4 pictures the GUI and highlights its Markov Decision Process (MDP) State design. A MDP is a way to model discrete, stochastic-based actions, popularly used in engineering and data science disciplines. In the Figure 4 example, the GUI allows the user to observe each state of the robot (and the actions associated with the state), displays the route prediction of where the robot is planning to go (actions that move to a different state), and ways to move from state to state by directing the robot to make task changes, if necessary.

IA examines the process defined between the human user and the automated system for analysis. The MDP GUI in Figure 4 shows how a user can track the progress of the robots (following a process), as they take various actions from state to state (the process flow). The list of seven ways to reduce autonomy, described above, emphasizes how IA can analyze process states and their related actions, as well as defining where user intersession (monitoring and decision override, etc.) can occur to ensure autonomy reduction of the robots' critical functions during mission execution.

FIGURE 4. MARKOV DECISION STATE GUI DESIGN (Nagy, 2021)

To understand this analysis process, let us examine several capacities and the associated analysis. In capacity row 1 in Table 4, the evaluator learns that OPD is executed through the GUI and is straight forward. The function reflects a standard database look up call, and it is clear that it is not ML-related. Normal NOSSA procedures will apply. In capacity row 4, the capacity is the application of a naive Bayes algorithm to determine best input attributes. This sounds very ML-like, so the analyst should enlist two sets of questions. First, what are the OPD interactions between this algorithm and the humans? The GUI provides a view into the statistical data. This enables evaluators to ascertain whether the algorithm is working properly. The second set of questions addresses the three areas of ML failure modes. From Table 1, the first is “Does the algorithm produce faulty/poor decision recommendations?” The second is “Are there human–machine operation issues?” and the last is “Is the system under (cyber) attack?” Our goal is not to resolve these questions directly, though at some point that will be required. Rather, are there human interventions that prevent these issues that by reducing the “autonomy” of the algorithm? Because if there are, then the function is not so critical as to require a code base check.

TABLE 4. IA FOR SYSTEM SAFETY USE CASE

. TASKS	. SUB TASKS	C. CAPACITIES	D. PERFORMING ROBOT COMPONENT	E. SUPPORTING HUMAN	F. OBSERVABILITY, PREDICTABILITY, AND DIRECTABILITY ASSESSMENT WRT NOSSA EVALUATIONS, PLUS SPECIFIC GUI FUNCTIONS FROM ABOVE
-thru GUI	ap obstacles	1. Use leg route & obstacle DB	Data Loader Manager	User	OPD-thru GUI and MDP (1, 7)
		2. Use wx DB	Data Loader Manager	User	OPD-thru GUI
		3. Use police intel DB	Data Loader Manager	User	OPD-thru GUI
	haracterize legs	4. Use naive Bayes (nB) to determine best input attributes	nB, DB Farm, DB Manager	Evaluator	OPD-thru GUI Leverage statistical output part of GUI to verify inputs for attributes make sense. (2, 3, 6)
		5. User Random Forest (RF) to estimate probability and missing attributes	RF, DB Farm, DB Manager	Evaluator	OPD-thru GUI While RF is a black box to evaluators, in this case techniques exist to prove that the results are useful. Evaluators need to understand this proof and how to apply. (2, 3, 6)
	elect robot/route pairs	6. Apply temporal greedy search (TGS) to create robot /route candidates	TGS, Business Rule Manager, DB Farm, DB Manager	Evaluator	OPD-thru GUI While TGS is an algorithm, it is not ML, no special attention required (1, 2, 3, 6, 7)
		7. Use non-linear optimization (NLO) to determine combos that provide highest likelihood of mission success			OPD-thru GUI While NLO is an algorithm, it is not ML, no special attention required (1, 2, 3, 6, 7)
nload robot in delivery Zone	emove from truck	8. Activate robots	Processor, Power Regulator, and Power Supply	Truck Driver	OPD-thru GUI (1)
obot navigation	etermine lead	9. Select robot as lead	Main Navigation and Guidance Controller	User	OPD-thru GUI (1)
	avigate	10, Access planned waypoint DB	Main Navigation and Guidance Controller	User	OPD-thru GUI and MDP (1, 2, 7)
		Update status	Main Navigation and Guidance Controller	User	OPD-thru GUI and MDP (1, 2, 3, 5, 6, 7)
elivery	nter delivery zone	11. Compare up date to plan	Main Navigation and Guidance Controller	User	OPD-thru GUI and MDP (1, 2, 3, 5, 6, 7)
		12. Adjust location as necessary	Main Navigation and Guidance Controller	User	OPD-thru GUI and MDP (1, 3, 7)
	dentify customer	13. Use computer vision (CV) to identify customer	Image DB and CV	User, Recipient	OPD-thru GUI and MDP; CV is ML, so a human on the loop checking the identity as seen by the robot reduces “Autonomy.” In other systems, this may not be feasible. May have to assume risk here. (1, 4)
		14. Check time so delivery can be synchronous	GPS Signal & SATCOM Transceiver, GPS Translator	User	OPD-thru GUI and MDP (2, 3, 5)
		15. Deliver package	Robot arms	Recipient	(1, 4, 7)

If we review these questions from that lens, what do we learn? First, because the evaluator employs the seven GUI functions to explore the algorithm outputs, that means the evaluator or user could be involved in directing the rerunning of the algorithm so that it could succeed (see functions 2, 6, and 7 in Table 5). Alternatively, the user may be able to insert data directly (functions 1 and 7). Because the user, using the GUI, can modify robot data and inputs, question number two comes into view. Can the human introduce errors into the statistics so that the system is not functioning properly? Yes, this is possible, but can be addressed through proper training, something Navy systems scrupulously employ. Finally, the fact that humans can view the algorithm output means that they could detect adversary or insider activity disrupted the data (in this case, all the GUI functions apply). Yes, there are procedural implications for using this part of the system. One can also surmise, though, that because of the human ability to address all three types of failure modes, that an evaluator could conclude that the software criticality of this particular algorithm is relatively low. That is, the algorithm's "autonomy" is not high.

Capacity row 5 is similar to row 4, but there are interesting differences with regard to "explainable" AI—explaining the AI decision. For example, when using a Random Forest (RF) algorithm (an advanced ML technique based on many variations of a single decision tree), the user interface might consider a design that supports the strength of the algorithm—this is referred to as a symbiosis. For example, the GUI might be designed to show the user when there is a gap in the existing input data to the RF (a potential "real world issue") and how this data gap is being estimated/compensated by the RF algorithm, along with providing the user with a statistical explanation of the algorithm's estimation approach and solution. Further, Nagy (2021) shows that an RF algorithm can be used to estimate missing statistical data, including statistical data describing highest success in supporting a final decision (e.g., routes and robot selection). This type of statistical explanation for critical decision-making data provides the user with increased confidence as to whether to accept or reject the estimations of an RF algorithm.

This raises another key point. If possible, GUI design and algorithm selection should be symbiotically determined, facilitated by the IA results. For the RF example, selection of an ML algorithm's capability to compensate for "real world issues" through statistical understanding of how the data was estimated should influence the GUI design with the end goal of reducing the autonomy of an ML algorithm. The design supporting the expandability strength of the ML should provide visibility into potential "real world issues" and how the selected algorithm compensates with statistical insight for the user. This IA driven, symbiotic GUI to ML algorithm design ensures that the user has the knowledge needed to make a final approval regarding any data estimated, thereby assisting in answering the first question, "Does the algorithm produce faulty results?" When symbiosis of design and algorithm occur, as provided in the RF example, there is greater confidence that the answer will be "no."

When symbiosis does not occur, there is a greater chance that system safety risks increase. Again, using the RF algorithm as an example, through its estimation process and statistical explanation, the user is provided with the necessary data to have confidence in approving or rejecting the route and robot selection during a "real world issue" lack of available data. Providing the necessary "knowledge" for a user to have confidence in the final approval becomes a key requirement necessary to maintain overall system performance and adequately reduce autonomy of an ML algorithm. Without the user having confidence in how well the algorithm is addressing "real world issues," final approval authority is meaningless. This is because the man-in-the-loop is not achieving the designed result of autonomous reduction of the ML algorithm. When things go wrong during deployment, it is about making sure that the user has the necessary "knowledge" provided by the ML algorithm (or other means) to make the final approval decision.

In this example we use the three general fault areas for ML. Table 3 adds root causes. In this particular example, the possible root causes that might require deeper examination would include training dataset corruption, mis-validation of data set, overfitting of the ML model, operator overreliance, or adversarial deception. These are just two examples of how to use IA to support NOSSA critical function analysis for ML-enabled functions. A complete IA would yield far more insights into potential problem areas and root causes.

Next Steps and Recommendations

After considering the analysis conducted above, one asks, "Does IA serve to solve the challenges of evaluating ML based components in weapon systems?" "Not completely," the authors argue. However, what IA does do is to add detail to those functional areas that need to be evaluated. IA helps identify specific ML components, which humans might be involved in the process to reduce "autonomy" and suggest which mechanisms amongst the OPD triad might be best employed, such as GUIs and their design. Moreover, not all designers appreciate the interdependence that should exist between user and the algorithm and therefore build no OPD connections. This makes reducing "autonomy" infinitely harder. Conducting IA rapidly speeds that discovery, which might suggest necessary rework by the original developer.

By adding the three main fault areas of ML use to the IA, very specific evaluation details and questions are raised. While it may not solve the emerging ML evaluation conundrum, it does add considerable detail to the kinds of discussions that system developers and NOSSA ought to consider, especially when evaluators use the root cause details to inform their questions. Further, the authors recommend adding a seventh column to the IA table, which would include what root causes were suspect and why. Here are several specific considerations that also emerged from developing this work.

As a nascent requirement for the NOSSA evaluators, recommend careful consideration of each function using these techniques, plus frequent review of deployed system performance. Specifying a special follow up with operational users may be required for the first years of use. While Nagy's scenario benefits from a very capable GUI already informed by a knowledge of IA, NOSSA evaluators should expect that not all systems submitted for such review will be as well developed. In fact, NOSSA may want to consider making an IA a requirement for submission of a system for certification.

We recommend not updating training data sets to the deployed edge, which is to platforms deployed in actual operations. At this point, the processes that support such updates are not well understood. Any updates to training data sets ought to be reexamined by NOSSA. However, as the use of ML devices becomes prevalent, NOSSA will not be able to maintain the pace to evaluate these changes. New procedures will need to be considered, developed, evaluated, and then adopted. We recommend tasking the Naval Postgraduate School, the Joint AI Center, or the Navy's Digital Software Engineering Transformation Working Group to start this work.

In many respects, updating training data sets is like the Navy's development security operations (DEVSECOPS) efforts to update patches to the Fleet in hours, not weeks. NOSSA should learn from those lessons, though recognizing that the size of most training data sets makes over the air updates challenging and unreliable. NOSSA and training data set updaters should consider leveraging standard storage device deliveries. This would include ways to secure the data on those storage devices.

Introducing ML techniques into systems, because of the importance of the training data sets, means that changes to the system engineering processes are necessary. In the past, the SE "Vee" diagram is a set of procedures that are executed to an end state, normally at the end of a variety of test and evaluation events. This includes operational test, security and weapons certification, and in NOSSA's case, safety certification. According to Johnson (2021), this new SE "Vee" may change to be continuous for the entire life cycle of a system. This means, in theory, that NOSSA has a continuous responsibility to monitor system safety. That is a significant change, and worth thinking about. It may be that IA provides at least a way to wrap one's head around this potentially new responsibility. IA could be used to identify those functions that do require continuous evaluation.

References

Annett, J. (2003). Hierarchical task analysis. *Handbook of cognitive task design* (pp. 17–35). Lawrence Erlbaum Associates.

Artificial Intelligence. (2021). In *Oxford online dictionary*. https://en.oxforddictionaries.com/definition/artificial_intelligence

Bolté, B., Endsley, M., & , Jones, D. (2003). *Designing for situation awareness: An approach to user-centered design*. Taylor & Francis.

Chipman, S., Schraagen, J., & Shalin, V. (2009). *Cognitive task analysis*. Lawrence Erlbaum Associates, Inc.

Crandall, B., & Klein, G. (2006). *Working minds: A practitioner's guide to cognitive task Analysis*. Bradford Books.

DoD. (2012). *Department of Defense standard practice: System Safety* (MIL-STD-882E).

Faria, J. (2017, October 23–26). Non-determinism and failure modes in machine learning. *Proceedings of IEEE 28th International Symposium on Software Reliability Engineering Workshops*, 310–316.

Johnson, B. W. (2021, March 3). *Metacognition for artificial intelligence system safety* [Online conference presentation]. Developing Artificial Intelligence in Defense Programs Acquisition Research Symposium, Monterey, CA, United States.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., van Riemsdijk, B. M., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, *3*(1), 43–69.

Joint Services-Software Safety Authorities. (2016). *Software system safety: Implementation process and tasks supporting MIL-STD-882E.* Joint Staff.

Nagy, B. N. (2021, March 25). *Using event-verb-event (EVE) constructs to train algorithms to recommend a complex mix of tactical actions that can be statistically analyzed* [Online conference presentation]. Fifth Annual Naval Application of Machine Learning, San Diego, CA, United States.

Varshney, K. (2016, January 31–February 5). Engineering safety in machine learning. *Proceedings of Information Theory and Applications Workshop*, 1–5

PANEL 14. BETTER DECISION-MAKING THROUGH TECHNOLOGY

Wednesday, May 12, 2021	
12:15 p.m. – 1:30 p.m.	**Chair: Ms. Jill Boward,** Executive Director, Program Executive Office for Integrated Warfare Systems (PEO IWS) ***Applying Multi-Criteria Decision-Making (MCDM) to the Technology Investment Decision-Making Process*** Tom Ivan, The MITRE Corporation Ed Ille, The MITRE Corporation ***Topological Data Analysis in Conjunction with Traditional Machine Learning Techniques to Predict Future MDAP PM Ratings*** Brian Joseph, Data Analytics, OUSD (A&S) ***Rethinking Government Supplier Decisions: The Economic Evaluation of Alternatives (EEoA)*** James Fan, Naval Postgraduate School Francois Melese, Naval Postgraduate School

Ms. Jill Boward— As the Executive Director, Program Executive Office for Integrated Warfare Systems (PEO IWS), Ms. Boward directs the acquisition and Fleet support of the Surface Navy's 155 combat systems, weapons, radars, and related international and foreign military sales programs. The mission of PEO IWS is to develop, deliver and sustain operationally dominant naval combat systems for Sailors and Marines. Ms. Boward is responsible for an organization of over 400 civilian and military personnel and 128 programs and projects with an annual acquisition budget of $6 billion.

Ms. Boward has served over 32 years as a civilian in the Department of Defense and entered the Senior Executive Service in September 2015. Prior to this position Ms. Boward served as the Director, Cost Engineering and Industrial Analysis at the Naval Sea Systems Command (NAVSEA 05C).

Ms. Boward spent four years from 2011-2015 as the Director for Amphibious and Auxiliary Ship Programs for the Deputy Assistant Secretary of the Navy, Ships (DASN Ships), supporting the Assistant Secretary of the Navy for Research, Development and Acquisition (ASN RDA). She spent over 25 years with the Naval Sea Systems Command's Cost Engineering and Industrial Analysis Group (NAVSEA 05C; previously NAVSEA 017) in various capacities, to include Director from 2015-2019; Deputy Director from 2007-2011; Industrial Planning and Analysis Division Director from 2002-2006; and as the Surface Combatants Division Director in 2002.

She also served as a cost team leader and analyst supporting the DDG 51 program and as a cost analyst for various combat and weapons systems. From 1991-1994, Ms. Boward worked as the staff engineer for the Contracts Management Division in the U.S. Army Garrison - Directorate of Engineering and Housing, Republic of Panama.

Ms. Boward earned both a Bachelor of Science in Industrial Engineering and Operations Research and a Master of Engineering Administration from Virginia Tech, and a Master of Science in National Resource Strategy from the Industrial College of the Armed Forces, National Defense University.

Ms. Boward was awarded the Department of the Navy Meritorious Civilian Service Award in addition to numerous other recognitions and performance awards throughout her career.

Applying Multi-Criteria Decision-Making to the Technology Investment Decision-Making Process

Thomas R. Ivan, D.Eng—is a Principal Systems Engineer with The MITRE Corporation and a former U.S. Marine Corps communication-electronic officer. Ivan's current research interest is applying Multi-Criteria Decision-Making methods to help solve complex system engineering and technology investment decision challenges. Ivan has a Bachelor of Science from the U.S. Naval Academy, a Master of Science from the Naval Postgraduate School, and a Doctor of Engineering from George Washington University. [tivan@mitre.org]

Edward E. Ille—is a Principal Systems Engineer with The MITRE Corporation. Ille has over 35 years of experience in aerospace and defense in roles ranging from design engineering to operations and executive leadership. Ille has interest in applying Multi-Criteria Decision-Making methods to help solve complex system engineering and technology investment decision challenges. Ille has a Bachelor of Science in electrical engineering from Carnegie-Mellon University. [eille@mitre.org]

Abstract

Department of Defense (DoD) activities are pursuing innovative solutions to communication and electromagnetic warfare challenges, especially for contested electromagnetic (EM) environments. Candidate technologies suitable for this environment are in a state of near-constant change. Current approaches for selecting these technologies for investment of limited resources can be inconsistent and based on subjective assessments and biased decision-making processes. Decision-makers require a structured and objective approach to technology selection and investment decision-making. Multi-Criteria Decision-Making (MCDM) methods provide this structured, consistent, and repeatable approach (Georgiadis et al., 2013).

To improve return on investment (ROI) of increasingly limited resources, candidate technologies and associated enabling products must be objectively evaluated against relevant measurable criteria. Implementing MCDM methods as part of the technology investment decision-making process increases consistency, objectivity, and repeatability of the process, leading to increased ROI of limited resources.

To research applicability and value of MCDM to technology investment decisions, this paper focuses on technology evaluation within a single domain, Private Cellular Networks, for use in U.S. and mission partner tactical operations. A decision framework for Private Cellular Network technology investment decision-making is presented to serve as a model for larger and more complex technology investment decisions using MCDM methods.

Introduction

Organizations within the DoD are engaged in capability modernization initiatives where potential candidate technologies and their enabling products are in a state of near-constant change. These technologies span multiple technical and operational domains to include wireless and cellular communications, Electro-Optical (EO), Free Space Optics (FSO), Low Earth Orbit (LEO) satellite communications, and other means of electronic information transmission, receipt, processing, and storage. In this highly complex, dynamic environment, often the approach to identifying, evaluating, and selecting technologies for investment of scarce resources is inconsistent and largely based on subjective assessments and decision-making processes. In this environment characterized by rapid technology evolution and capability enhancement, decision-makers require a structured and objective approach to technology selection and resource investment decision-making.

To improve return on investment (ROI) of limited resources, candidate technologies and associated enabling products should be objectively evaluated against relevant measurable criteria. A structured decision-making model enables development of relevant criteria against which candidate technologies and enabling products can be assessed. Implementing multi-criteria decision-making (MCDM) methods as part of the technology investment decision-making process increases consistency, objectivity, and repeatability of the process leading to increased ROI of limited resources.

Research Issue Statement

Current approaches to identifying, evaluating, and selecting emerging technologies for investment of scarce resources can be inconsistent and based largely on subjective assessments and unstructured decision-making processes. To ensure that technology investment decisions are properly aligned to organizational goals and objectives, decision-makers require a structured and objective approach to technology selection and investment decision-making.

Research Questions

The research introduced in this paper focuses on answering the following questions regarding technology investment decision-making:

- Can appropriate technology investment decision criteria be identified?
- What technology investment decision criteria can be appropriately quantified?
- How can technology investment decision criteria be quantified?
- Can statistical analysis be used to accurately develop weighted decision criteria?
- Can the technology investment decision-making process be made more objective, consistent, and repeatable?

Multi-Criteria Decision-Making

Multi-Criteria Decision-Making (MCDM) is a method to help decide when there are multiple alternatives, with each alternative having several characteristics and attributes (Yoon & Hwang, 1995). MCDM methods are sufficiently flexible for application to simple decision-making scenarios with binary decision alternatives as well as complex, multi-criteria, multi-alternative scenarios. MCDM is well-suited for technology investment decisions because those decisions routinely include criteria that must be identified, weighted, and compared to generate priority rankings for each decision alternative.

Methodology

The five phases in the proposed research methodology presented in this paper are briefly described in the following paragraphs and shown in Figure 1.

Phase 1, Preliminary Investigation, focuses on capability gaps identified by a Capability-Based Assessment (CBA) or similar activity. Phase 1 produces the problem statement and research objectives. It also produces the research questions whose answers help mitigate or close gaps in technology investment decision-making processes.

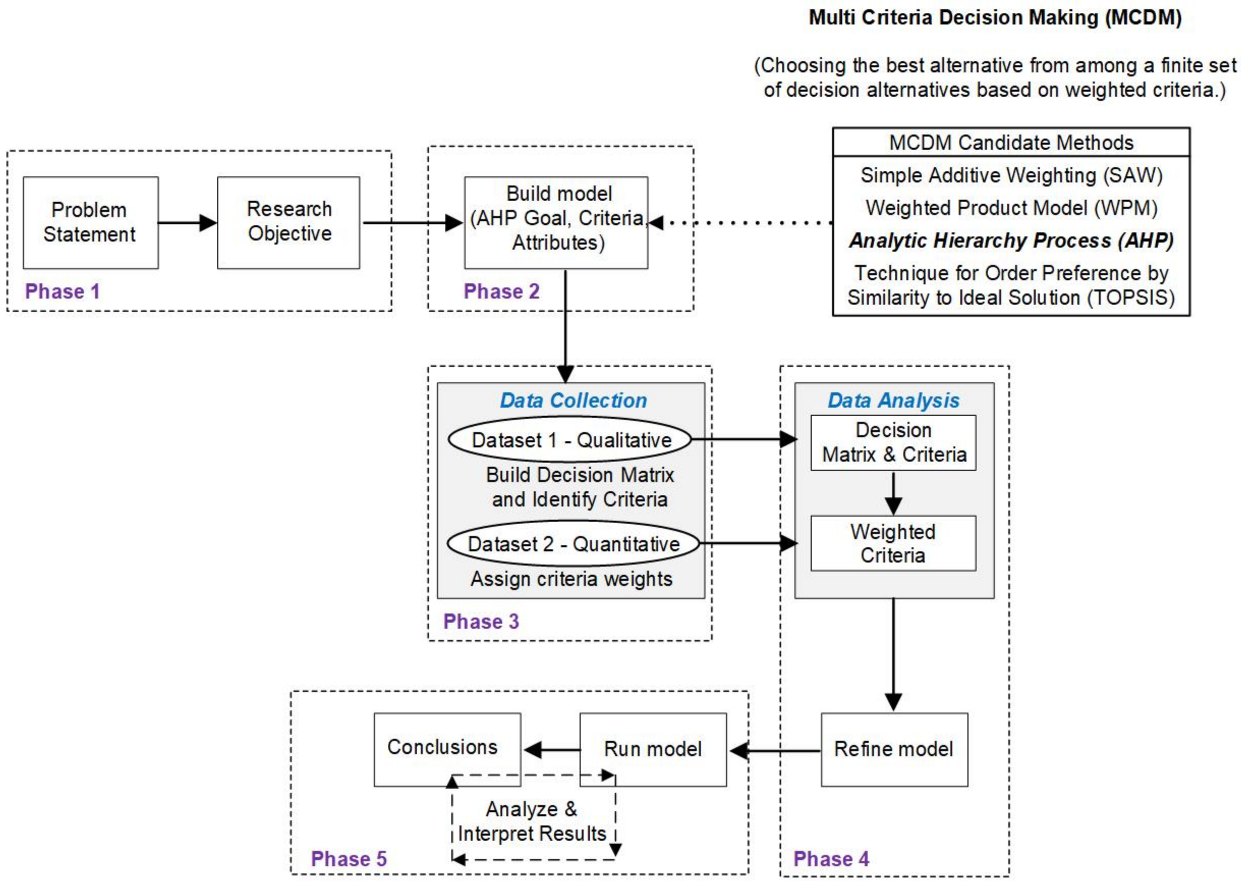


Figure 1. Methodology Map

Phase 2, Initial Model Development, focuses on the problem statement, research objectives, and research questions formulated during Phase 1. Outputs from Phase 1 research are used to build the initial version of the MCDM model to include identifying the Goal, Criteria, and Criteria Attribute components of the MCDM model. The decision model goal is the selection of the most appropriate technology and enabling a product consistent with organizational objectives, areas of greatest need, and highest possible ROI.

During Phase 2, several candidate MCDM methods were considered for applicability to the technology investment decision-making environment. Analytic Hierarchy Process (AHP) was selected as an appropriate MCDM method for the technology investment decision-making environment because of its ability to consider a set of evaluation criteria and a set of alternative options, from which the most appropriate decision can be made. Another factor in selecting AHP is that it is a “fast and understandable method for people who are not familiar with multi-criteria decision support methods” (Tscheikner-Gratl et al., 2017).

Phase 3, Data Collection, focuses on qualitative and quantitative data relevant to the decision environment. Qualitative and quantitative data on candidate technologies and enabling products must be collected to support development of the decision matrix, decision criteria, decision weights, and decision alternatives used in the model.

Phase 4, Data Analysis and Model Refinement, focuses on analyzing collected data to refine model components and relationships proposed in the initial model.

Phase 5, Model Run and Results, is where the decision model is run using as input the qualitative and quantitative data analyzed in Phase 4. The analyzed data is input into the model to calculate decision criteria weights using AHP. Decision criteria weights are tested for consistency using Saaty's Random Index before use as the basis for generating a ranked order of decision alternatives (Saaty, 1980).

Multi-Criteria Decision-Making Methods

Several MCDM methods exist and have been applied to a wide range of decision-making scenarios. Each MCDM method has strengths and weaknesses relative to the decision-making environment for which its application is considered.

A short description of frequently used MCDM methods is provided in the following paragraphs.

- Weighted Sum Model (WSM): WSM is designed for single-dimensional decision-making problems. WSM is a value-measurement MCDM method with criteria assigned a weight to represent relative importance for comparison to other decision criteria (Tscheikner-Gratl et al., 2017). Simple in design and application, WSM is not well-suited for multi-dimensional decision-making problems such as technology investment decisions because of its limitations for addressing multiple decision criteria scenarios (Triantaphyllou et al., 1998).
- Weighted Product Model (WPM): WPM is like WSM but uses multiplication rather than addition to calculate the sum for each decision alternative. WPM is suitable for single dimensional decision-making problems where the solution sought is that with the largest value relative to the other candidate solutions. WPM is not well-suited for multi-dimensional decision-making problems such as technology investment decisions because of its limitations for addressing multiple decision criteria scenarios (Triantaphyllou et al., 1998).
- Simple Additive Weighting (SAW): SAW is based on simple addition of scores assigned to different criteria associated with accomplishing the goal of the decision-making process. Though SAW calculations are simple and do not require complex computer algorithms, the results produced with SAW do not always realistically reflect the complexity of the multi-criteria technology investment decision-making process (Adriyendi, 2015).
- Technique for Order Preference by Similarity to Ideal Solution (TOPSIS): TOPSIS is an MCDM method based on the concept that the alternative chosen should be the one having the shortest distance from the ideal solution and furthest distance from the negative ideal solution. The ideal and negative ideal solutions are unique to the decision that must be made and can be based on quantitative or qualitative factors. TOPSIS is a calculation-intensive MCDM method whose proper application is dependent on accurate identification of ideal and negative ideal solutions (Gavade, 2014).
- Analytic Hierarchy Process (AHP): AHP decomposes complex MCDM problems into hierarchical relationships between decision criteria and their attributes. AHP allows the decision-maker to view the decision in terms of a hierarchy of criteria. AHP is a value-measurement MCDM method with criteria assigned a weight to represent its relative importance for comparison to other decision criteria (Tscheikner-Gratl et al., 2017). AHP considers qualitative and quantitative criteria when building decision criteria and developing criteria weights. AHP's framework for establishing relationships between goals, criteria, criteria attributes, and decision alternative weights makes it well-suited to the technology investment decision-making process.
- The MCDM methods briefly described previously share common components for intent, design, and application discussed in the following section.

Multi-Criteria Decision-Making Components

Though different in design and application to decision-making scenarios, MCDM methods share several common characteristics, as described here (Mocenni, 2018):

- The decision context must be identified. The decision that must be made, the decision-maker, and key contributors are identified as the first step in the MCDM process (Department for Communities and Local Government, 2009).
- Decision options are identified before being assessed or evaluated for suitability. Decision options are eventually assessed against weighted decision criteria as the MCDM process is applied. Decision options can be the selection of the most appropriate solution or elimination of one or more candidate solutions from further evaluation and consideration.
- Independent decision criteria are identified. The decision criteria should be independent of each other to avoid introducing inconsistencies in the criteria weighting results. The relevance of the criteria to each identified option is also identified. The decision criteria are those factors against which potential solutions are evaluated for acceptance or rejection (Vargas, 2010).
- Weights are developed for each of the decision criteria to establish the relative importance of the criteria when evaluated against each other. The weights can be developed through qualitative methods, quantitative methods, or a combination of both methods. Decision criteria weights are applied during the evaluation of potential solutions to create a rank order of all evaluated solutions. Subject Matter Expert input is a critical component of developing credible criteria weights.
- Each potential solution, or decision alternative, is assessed against criteria and assigned a value using criteria weights.
- Options are rank ordered to help the decision-maker either select one option or eliminate one or more options from further consideration.

Figure 2 shows MCDM components and their relationships in helping the decision-maker reach an evidence-based selection from available alternatives.

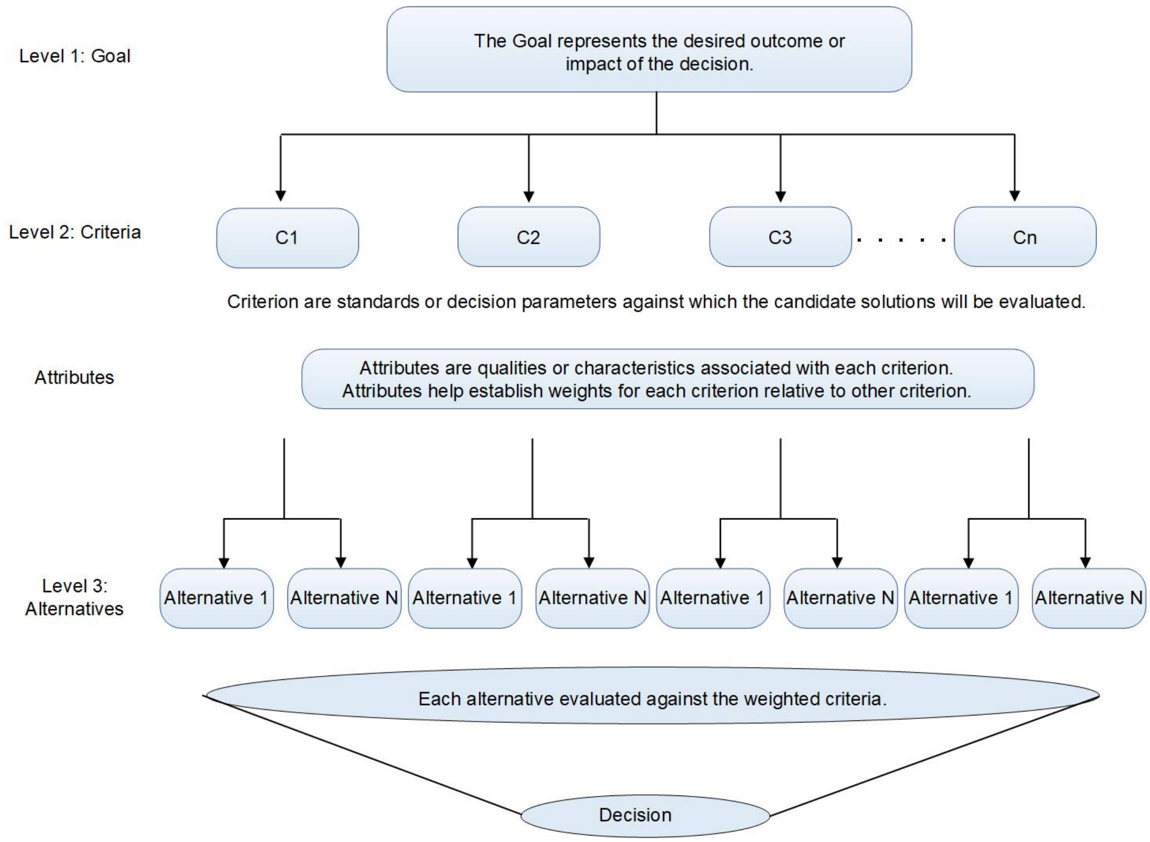


Figure 2. MCDM Components, Levels, and Relationships

Applying Multi-Criteria Decision-Making Methods to Decision-Making Processes

Since MCDM methods incorporate computational tools to weight and assess decision options, they are well-suited for application to complex, multi-dimensional decision-making scenarios such as technology investment decision-making.

The goal of a decision-making process is to provide the decision-maker with the ability to investigate the future and to make the best possible decision based on past and present information and future predictions (Georgiadis et al., 2013). MCDM methods accomplish this goal by considering quantitative data, qualitative data, and expert judgment to assess and rank decision options from which the decision-maker can select the most appropriate option.

MCDM methods integrate quantitative and qualitative data analysis as components of the decision-making process. MCDM methods align closely, and are appropriately applied, to a decision-making process with the following characteristics (Foote, 2010):

- Decision objectives are identified. A decision results from comparing one or more alternatives against one or more criteria relevant to the decision-making environment.
- Decision options available to the decision-maker are identified.
- Decision criteria upon which the decision will be made are identified. The decision criteria are independent of one another.
- Decision criteria weights are developed for use in evaluating and comparing decision alternatives. The weights are developed using a combination of quantitative and quantitative data sources. Expert judgment provided by subject matter experts is a critical success factor in determining a criterion’s weight relative to other decision

criterion. Qualitative subject matter expert judgments are transformed into numerical values (quantitative data) using a pairwise comparison scale (Saaty, 1980).

- Decision options are evaluated against the weighted criteria. Each option is evaluated against the weighted decision criteria independent of other decision options.
- Decision options are rank ordered to help the decision-maker either choose the most appropriate alternative or to eliminate one or more alternatives from further consideration.
- A decision is made by selecting the most appropriate alternative from among all ranked alternatives.

By integrating quantitative and qualitative data analysis into the decision-making process, MCDM methods strengthen the technology investment decision-making process by

- Removing subjectivity in decision criteria weight assignments.
- Improving documentation and communication of all components of the decision-making process. MCDM methods facilitate documentation and communication between the decision-maker and stakeholders impacted by the decision.
- Calculating and analyzing the decision criteria weights and alternative rankings to generate an audit trail for future analysis or refinement as appropriate.
- Providing a platform for sensitivity analysis and "What-If" studies

The technology investment decision-making process is a multi-criteria environment where MCDM methods can be applied to identify decision goals, identify decision criteria, develop decision criteria weights, and assess available options against the weighted criteria to help the decision-maker select the most appropriate option.

The next section of the paper presents a case study for applying a selected MCDM method, Analytic Hierarchy Process, to the decision-making process for investing resources to pursue Private Cellular Network Technology and enabling products.

Case Study: Applying Analytic Hierarchy Process to Investment Decision-Making for Private Cellular Network Technology

AHP is a structured MCDM method well-suited to organize and analyze complex decisions. AHP uses qualitative and quantitative data as critical components of the decision-making process. In AHP, decision alternatives are identified along with criteria relevant to each decision alternative. AHP reduces complex decisions to a series of pairwise comparisons, which allow the decision-maker to establish priorities to support the decision-making process. AHP establishes weights for decision criteria, which are then applied to evaluate and prioritize decision alternatives (Mocenni, 2018).

Because of its ability to create a decision hierarchy identifying the decision objective, decision criteria, decision criteria weights, and decision alternatives, AHP is well-suited to technology investment decision-making processes. Relevant qualitative and quantitative data can be collected from technical research, subject matter expert judgment, formal testing, or capability demonstration events.

Figure 3 shows the MCDM components and levels presented in Figure 2 tailored to the decision-making process for determining appropriateness of Private Cellular Network technology as an investment initiative.

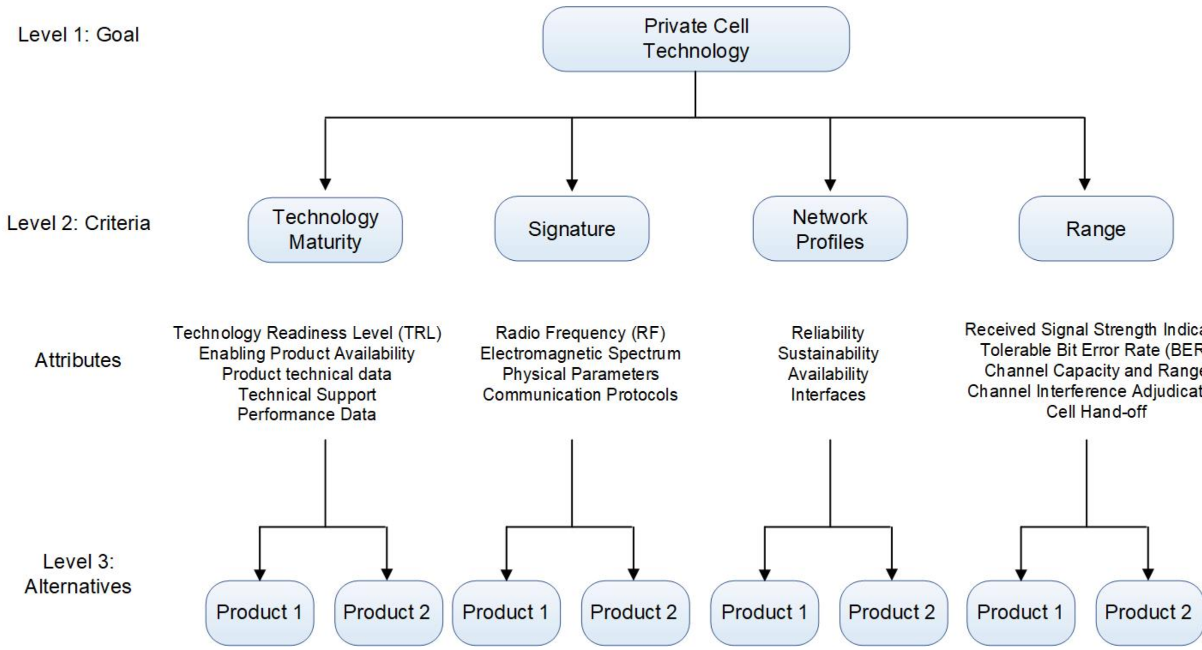


Figure 3. Applying AHP to the Private Cell Technology Investment Decision-Making Process

For the Private Cellular Network Technology case study, the Goal, which represents the desired outcome or impact of the decision, is to make an investment decision on pursuing the technology and, if yes, identifying the most appropriate material solution.

Decision Criteria for the Private Cellular Network Technology Case Study are Technology Maturity, Signature, Network Profile, and Range. These criteria are the standards or measurements against which the decision will be made and candidate solutions will be evaluated. The four Criteria provide sufficient depth and granularity to define the technology and to differentiate candidate enabling products if the decision is made to pursue this technology as an investment initiative.

The Attributes listed for each criterion are qualities or characteristics associated with that criterion. Attributes add depth and identify characteristics of interest to help establish weights for each criterion. Attributes can focus on technical, operational, security, cost, supportability, or other characteristics that help describe the criteria.

In the Private Cellular Network Technology Case Study, Decision Alternatives are a binary yes or no decision to pursue the technology and a multi-criteria decision to evaluate candidate enabling products. Two candidate material solutions, Product 1 and Product 2, are shown as decision alternatives in Figure 3. These two material solutions will be evaluated against the weighted criteria established by AHP’s Pairwise Comparison Matrix shown in Figure 4.

It is important to note that a Doctrine, Organization, Training, Material, Leadership and Education, Personnel, Facilities and Policy (DOTMLPF-P) analysis should be performed prior to pursuing a material solution to mitigate or close a capability gap. The DOTMLPF-P Analysis will determine if a non-material or material approach is required to mitigate or close a capability gap. The technology under consideration, and the products that realize the technology’s capability, are material solutions to mitigating or closing the capability gap.

Qualitative and quantitative data are used to establish criteria weights, which are applied to rank order the decision alternatives of Product 1 or Product 2.

The qualitative data component of the AHP model identifies decision criteria and associated attributes in the Criteria component of the AHP hierarchy structure. Qualitative data can be collected from technical research and subject matter expert judgment. Figure 4 shows how Saaty's (1980) Pairwise Comparison Scale is used to translate qualitative verbal judgments to quantitative numeric values.

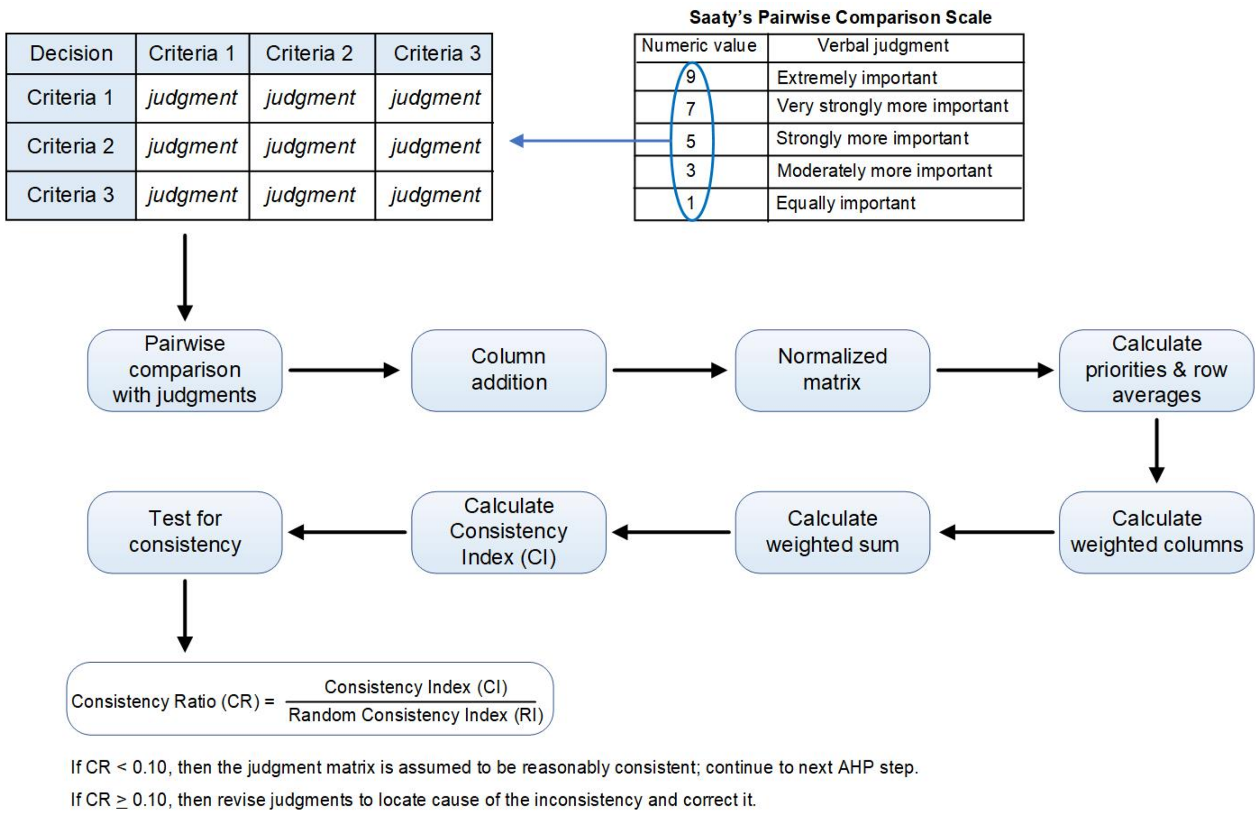


Figure 4. Developing Decision Criteria Weights Using Pairwise Comparison

The criteria weights developed through Saaty's (1980) Pairwise Comparison Scale are tested for consistency as the final step in the pairwise comparison. The Consistency Ratio (CR) compares the Consistency Index (CI) of the Subject Matter Expert–generated judgment matrix against the consistency index of a random matrix (RI). The random matrix is one where judgments have been entered randomly and therefore are expected to be highly inconsistent. The RI is the average consistency index of 500 randomly populated matrices. As shown in Figure 4, if the Consistency Ratio is < 0.10, the judgment matrix is assumed to be reasonably consistent and is acceptable for use in the AHP model (Saaty, 1980).

With criteria weights established and verified for consistency against the Random Index, the Product 1 and Product 2 material solutions can be evaluated against the weighted criteria.

Figure 5 shows the final steps in the Case Study, which produce a rank ordered list of decision alternatives from which the decision-maker can select.

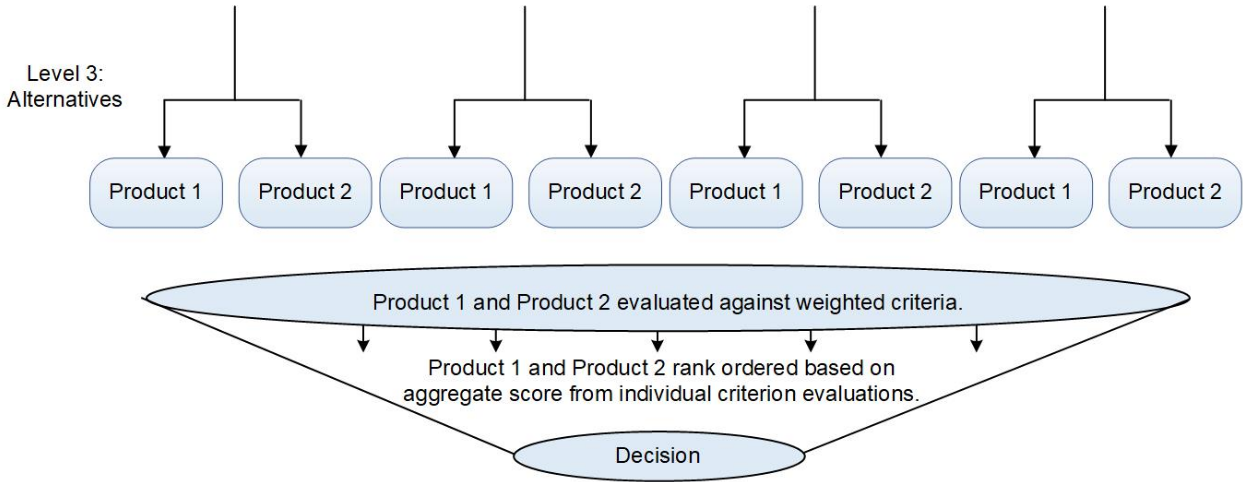


Figure 5. Applying Criteria Weights to Decision Alternatives

Evaluating Product 1 and Product 2 against the weighted criteria produces an aggregate score that can be used to rank order the decision alternatives. In the Private Cellular Network Technology Case Study, there are only two products from which the decision-maker can select. Though the number of decision alternatives can be sufficiently larger than shown in the Case Study, the AHP process steps and application are consistent with the two-decision alternative Private Cellular Network Technology Case Study.

Findings

Key findings from MCDM method research and selection, model development, decision criteria research, and criteria weighting are presented in the following paragraphs.

- Criteria identification and weighting are heavily influenced by the quality of subject matter expert judgment. Pairwise comparisons used to translate qualitative data to quantitative data can be influenced by the breadth and depth of knowledge and experience of subject matter experts providing the qualitative assessments.
- Decision criteria should be limited to those that clearly enable differentiation of decision alternatives. As the number of criteria increases, the difference in their respective weights becomes increasingly smaller and potentially less meaningful.
- The criteria may be sensitive to unique characteristics of the environment in which the technology and enabling product is intended to operate. Criteria and associated attributes must be considered in the context of the environment in which the technology or product will operate. Criteria with absolute measurements, such as weight or cost, are less susceptible to operating environment influences than more dynamic criteria such as electromagnetic spectrum and channel interference criteria.
- An Acceptance Threshold should be established as a value that must be equaled or exceeded for any decision alternative to be selected. Determining an Acceptance Threshold for selecting one of the decision alternatives can help avoid the scenario where even the highest ranked alternative fails to perform in a manner that mitigates the targeted capability gap.
- The mission owner's risk tolerance can impact the choice between decision alternatives. Risk tolerance is based on the mission owner's assessment of the impact if the technology and enabling product fail to perform as expected. Different mission owners

can have different risk tolerances when addressing suitability of similar technologies and products for investment decisions.

- The MCDM approach to technology investment decisions is not a substitute for a formal Analysis of Alternatives (AoA) and subsequent formal source selection evaluations. However, the MCDM approach proposed in this paper is well-suited to rapid capability testing and fielding as a first step in large-scale system evaluations and investment decisions.
- Improvement in the ROI for technology initiatives attributed to applying MCDM methods is difficult to measure until a robust body of data is collected and analyzed. The ROI for applying MCDM methods should include factors such as cost avoidance of pursuing technologies that do not mitigate or close capability gaps, pursuing technologies that have no measurable contribution to mission success, and pursuing technologies for which enabling products do not yet exist.

Conclusion and Practical Application

The MCDM model presented in this paper meets the proposed research objectives and answers the research questions. The MCDM model can produce a consistent and repeatable technology investment decision-making process that removes subjectivity and decision-maker bias, either intended or unintended, from the decision-making process. The structured and objective approach to technology selection and investment decision-making ensures that technology investment decisions are properly aligned to organizational goals and objectives.

The application of MCDM methods to technology investment decision-making researched and summarized in this paper removes inconsistency and subjectivity from that decision-making process. Removing inconsistency and subjectivity results in a reduced level of risk potentially introduced to mission success by adopting technology and enabling products that fail to perform as required in the anticipated operating environment. The decision support model presented in this paper, based on the AHP MCDM method, can be applied by organizations pursuing technology-based solutions to their most critical operational challenges and capability gaps.

References

Adriyendi. (2015). Multi-attribute decision making using simple additive weighting and weighted product in food choice*. International Journal of Information Engineering and Electronic Business*, *6*.

Department for Communities and Local Government. (2009). *Multi-criteria analysis: A manual.* http://www.communities.gov.uk

Foote, K. E. (2010). *An introductory overview to multi criteria evaluation.* https://www.colorado.edu/geography/foote/2010/Methodology_Norlund_final.ppt

Gavade, R. (2014, November 20). Multi-criteria decision making: An overview of different selection problems and methods. *International Journal of Computer Science and Information Technologies*.

Georgiadis, D. R., Mazzuchi, T. A., & Sarkani, S. (2013). Using multi-criteria decision making in analysis of alternatives for selection of enabling technology. *Systems Engineering, 16*(3).

Mocenni, C. (2018, June 7). *The analytic hierarchy process*. http://www.dii.unisi.it/~monecci/Note_AHP.pdf

Saaty, T. L. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw-Hill International Book Company.

Triantaphyllou, B., Shu, B., & Ray, T. (1998). Multi-criteria decision making: An operations research approach*. Encyclopedia of Electrical and Electronics Engineering*, *15*.

Tscheikner-Gratl, F., Egger, P., Rauch, W., & Kleidorfer, M. (2017). Comparison of multi-criteria decision support methods for integrated rehabilitation prioritization. *Water*.

Vargas, R. V. (2010). *Using the analytic hierarchy process (AHP) to select and prioritize projects in a portfolio* [Paper presentation]. PMI® Global Congress 2010—North America, Washington, DC.

Yoon, K. P., & Hwang, C.-L. (1995). *Multiple attribute decision making: An introduction*. Sage Publications.

Topological Data Analysis in Conjunction with Traditional Machine Learning Techniques to Predict Future MDAP PM Ratings

Brian B. Joseph—is a Senior Operations Research Analyst in the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD (A&S)), Acquisition Data Analytics (ADA). In this role, Joseph serves as the Deputy Director, Data Analytics supporting the Principal Deputy to the Assistant Secretary of Defense, Acquisition Enablers regarding acquisition data, data science, metrics, and statistics. Prior to this, Joseph served as the Statistician, Security Policy & Oversight Division (SPOD), Counterintelligence and Security Directorate, Office of the Under Secretary of Defense for Intelligence. In this role, Joseph served as Principal Action Officer to the Director, SPOD regarding DoD personnel security investigations (PSI) metrics. Joseph earned his MS in Cost Estimation and Analysis from the Naval Postgraduate School, MS in Applied Statistics from the Rochester Institute of Technology, BS in Environmental Management from the University of Maryland University College, and an AS in Mathematics from Northern Virginia Community College. He is currently pursuing a PhD in Data Science from Northcentral University. [Brian.b.joseph.civ@mail.mil]

Trami Pham—is a Data Scientist and Senior Consultant at Booz Allen Hamilton, with experience in developing machine learning models using defense acquisition. Currently, Pham supports the Acquisition, Data, and Analytics Division of OUSD(A&S) by building predictive models to forecast program health. Pham earned her BA in Applied Mathematics from the University of California, Berkeley. [Trami.d.pham.ctr@mail.mil]

Christopher Hastings—is a Data Analyst and Consultant for Artlin Consulting, specializing in defense acquisitions. He works within the Acquisition, Data, and Analytics division of OUSD(A&S), providing analyses on Major Defense Acquisition Programs (MDAPs) and Middle Tier of Acquisition (MTA) programs. He earned his BA in mathematics and economics from the University of Virginia. [Christophe.p.hastings2.ctr@mail.mil]

Abstract

Topological data analysis (TDA) is an unconventional machine learning technique that is used to understand the underlying topology of data. The premise is that data has shape. The two methodologies used in TDA are persistent homology and the mapper algorithm. Traditional machine learning techniques include supervised unsupervised methods such as clustering, Bayesian networks, neural networks, support vector machines (SVM), and random forests. The goal of this study is to apply TDA methods in conjunction with traditional machine learning algorithms to Defense Acquisition Executive Summary (DAES) data to determine if TDA helps to improve prediction measures (accuracy, f-measure, sensitivity, and specificity) over using traditional methods only when predicting program manager ratings from Major Defense Acquisition Programs (MDAPs). We show that TDA when used in conjunction with traditional machine learning models at a local level of the DAES data improved the accuracy of predicting PM cost ratings of MDAPs at 80% of all nodes in training and testing as compared to implementing these models without TDA at the global level.

Keywords: Topological data analysis, machine learning, prediction measures

Background/Research/Business Need

The Data Analytics Division of Acquisition Enablers (AE) within the Office of the Under Secretary of Defense for Acquisition and Sustainment OUSD(A&S) has been developing machine learning model minimal viable products (MVP) to assist in prioritizing analysts' focus on which major defense acquisition programs (MDAPs) may become problematic. Human resources have been reduced in A&S to perform analytic tasks of determining problematic programs in the program assessment process in the Acquisition Data Analytics Division of the

AE Directorate since the reorganization of OUSD Acquisition, Technology, and Logistics into OUSD(A&S). As such, prioritizing problematic programs using machine learning models efficiently assists analysts in performing program assessment for executive leadership. There is anecdotal evidence that has shown that TDA, when used in conjunction with traditional machine learning models, improves overall accuracy of these machine learning models at localized sections of the data. SymphonyAI (2021) in a white paper discusses how traditional machine learning models use global optimization that assumes/guesses the shape of the data to derive parameters to approximate the dataset which often produces errors in some regions of the data. TDA in contrast creates separate models of the underlying data based on the output network topology that is responsible for different local sections of the data. This technique produces a better representation than a single globalized model. Therefore, we wanted to test whether this localized modeling methodology of TDA is more efficient and improves accuracy of predicting program manager ratings in DAES data.

Machine Learning

Machine learning is binned into unsupervised and supervised learning. Unsupervised learning uses methods such as clustering to segment data into smaller datasets and dimensionality reduction to make it easier to visualize data that are high dimensional (e.g., 25 or more features). Clustering models include hierarchical and K-Means. Supervised learning consists of regression and classification models. The classification models used to assist in the prioritization effort are neural networks, random forests and single tree models, and SVM.

Supervised Learning Classification Models

Random forests are an ensemble technique analogous to bagging trees. It works by collecting a bootstrapped sample of identical and independently distributed trees and conducting recursive partitioning on them. Classification is based on a majority vote of the aggregated trees. The beauty of this technique is that it obtains an estimate of the misclassification error and also performs random feature selection to estimate the relative importance of the explanatory variables (Friedman et al., 2009).

Support vector machines are large-margin powerful predictive models that can be utilized for classification or regression. They are a class of distance-based classifiers that attempt to use hard margins for stability in classification. They can be linear or nonlinear in form. The beauty and utility of SVM is the implementation of kernel methods that transform vectors from the input space and calculate their inner products in the feature space therefore bypassing the calculation of the function Φ in the input space, which would be untenable. This allows the SVM to perform classification of datasets in which the underlying boundaries of the classes are not readily clear. Some examples of kernels are the Gaussian radial basis, Laplace radial basis, and the hyperbolic tangent kernels. The use of kernels offers a rich model class to essentially tune the SVM (Clarke et al., 2009).

Neural Networks are extremely powerful classifiers as they can be tuned by many different parameters. They are also heavily nonlinear classification models. The sigmoid function ψ that defines the neural net may be modeled using the logistic, hyperbolic tangent, or heavy side step sigmoid functions. These sigmoid functions in conjunction with the size of the hidden layers offer ways to tune the neural network as a more robust classifier (Clarke et al., 2009).

TDA

TDA is an emerging and exciting form of unsupervised learning. Georges (2019) states that TDA is based on topology, a branch of mathematics that examines the notion of shape. TDA attempts to analyze highly complex data and draws on the notion that all data has a fundamental shape and that shape has meaning. Figure 1 below is an illustration of some common shapes of data, which include regressions, clusters, flares, and loops.

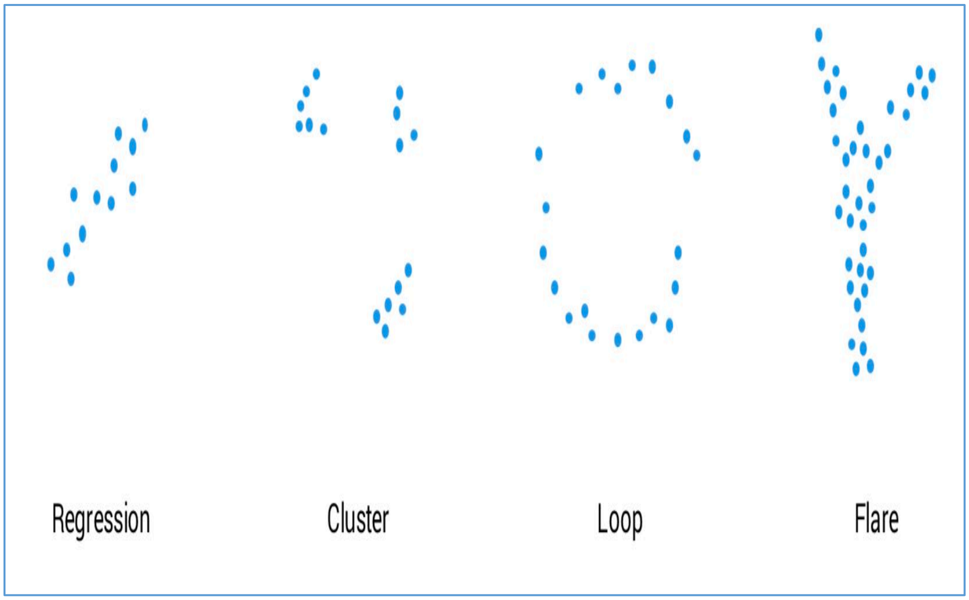


Figure 1 . Common Shapes of Data (Ayasdi, 2020)

The two methodologies used in TDA are persistent homology and the mapper algorithm. Persistent homology provides a framework and efficient algorithms to quantify the evolution of the topology of a family of nested topological spaces. Persistent diagrams are used to capture and visualize the birth and death of homological features over a specific period of time (Fasy et al., 2015). The mapper algorithm is a tool used to visualize the topology of the data under consideration. This method of TDA will be used for this research. The inputs to the algorithm are a point cloud of data, a filter function, a covering of a metric space, a clustering algorithm, and tuning parameters. Figure 2 depicts an illustration of the mapper algorithm and filter function.

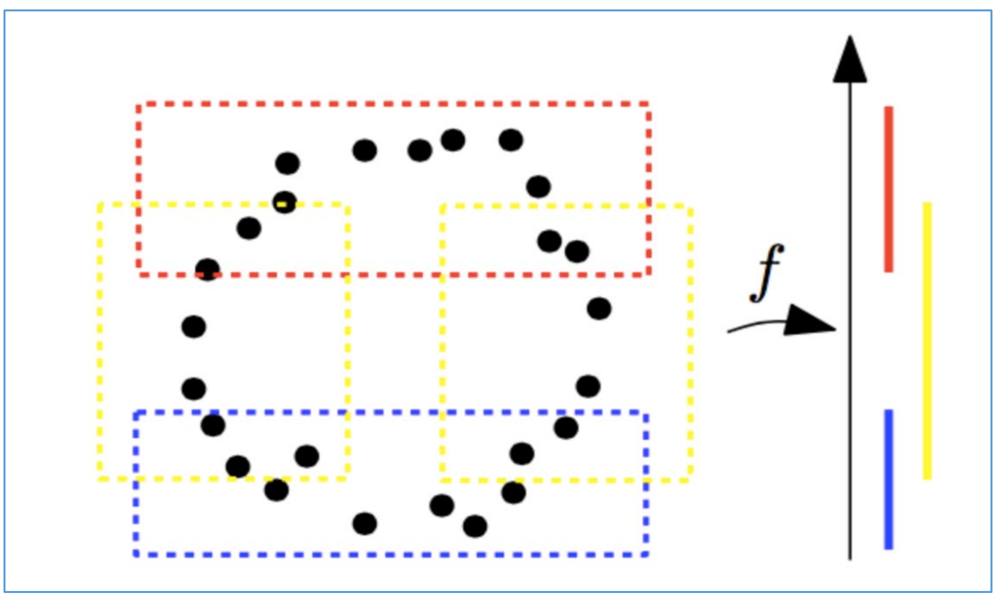


Figure 2. Mapper Algorithm and Filter Function (Chazal & Michel, 2016)

The output is a network graph that represents the topology of the data (Herring, 2018). Figure 3 below illustrates the steps to implement the mapper algorithm. It shows that the notional data to be mapped is a hand. Next, a filter function is identified. In this research, a kernel distance estimator will be used as the filter function. Third, determine the number of overlapping bins to map the input data. In this case, six bins are selected. Finally, create a network topology representation of the original dataset using nodes and edges (Lum et al., 2013). The nodes represent the clusters of local regions created by the binning. It is important to note that information from one node can be contained in another node as a result of overlapping bins. The edges connect clusters to display the overall topology.

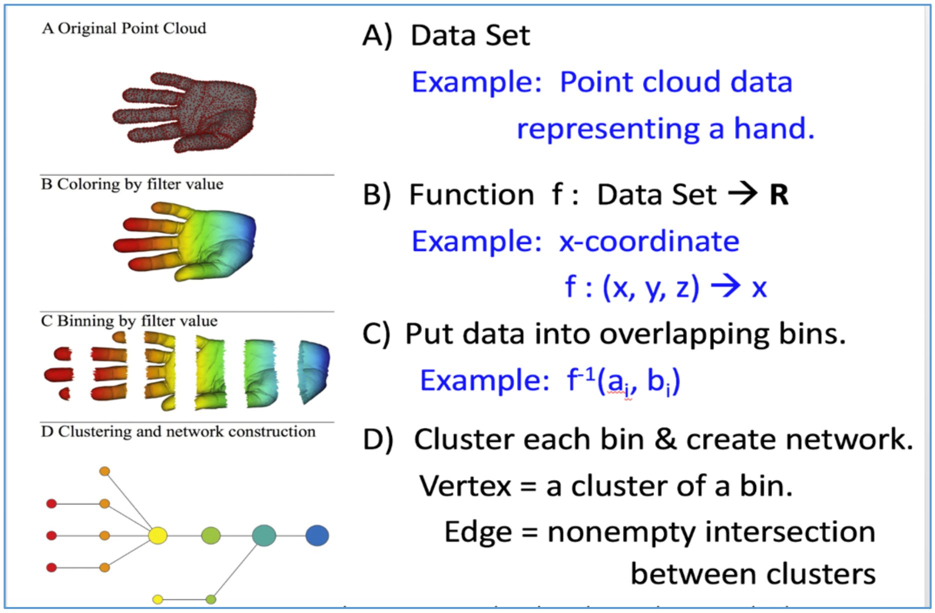


Figure 3. Implementation Steps of Mapper Algorithm (Lum et al., 2013)

Research Question

Can TDA in conjunction with traditional machine learning models improve the accuracy of the predictions of those machine learning models when used without TDA?

Hypothesis

H_0: Traditional machine learning algorithms (neural network, random forest, recursive partitioning, and SVM) have higher predictive accuracy when combined with TDA in at least 70% of nodes for training and testing sets.

H_a: Traditional machine learning algorithms (neural network, random forest, recursive partitioning, SVM) have higher predictive accuracy when not combined with TDA in at least 70% of nodes for training and testing sets.

Related Work

Chazal & Michel (2016) demonstrate how to use the mapper algorithm in R's TDAMapper package to construct topologies of any data set into network graphs, as well as how to label the categories of each node by a specific color to assist with understanding the data's topology better. Riihimaki et al. (2020) used a TDA classifier to determine if it provided better accuracy than a SVM classifier when modeling repeated measures data. The results of this experiment are that their TDA classifier outperformed the SVM classifier in accuracy 96.8%

to 68.7% respectively in one of three use cases. Kindelan et al. (2021) used persistent homology to build a TDA classifier that provided superior accuracies on eight separate data sets than traditional k-NN classifiers. Wu and Hargreaves (2020) implemented a TDA classification model on mixed data (numerical and categorical) using persistent homology of heart disease data. The results were that the TDA classification model performed better in accuracy than traditional state-of-the-art machine learning models such as decision trees, logistic regression, naïve Bayes, neural networks, single trees, and SVM in predicting heart disease. Joseph and Sconion (2020) used sentiment analysis to extract average sentiment of selected acquisition report executive summaries to determine if the average sentiment was highly correlated to be viable as a predictor feature/variable in predicting unit cost growth of MDAPs. Joseph and Hastings (2020) derived new schedule features/variables (months to threshold, difference from current to next DAES, difference from previous to current DAES, and previous milestone slips) from schedule milestone and APB schedule data gathered from DAES data to predict and understand the factors that may cause schedule slips in MDAPs.

Methodology

Four traditional machine learning classification models are initially applied to DAES data in order to predict future program manager cost ratings. PM cost ratings are the target variable and 10 other attributes (consisting of schedule, unit cost, and average sentiments of DAES executive summaries) are used as features for these models. The classification models used in this research are neural network, random forest, recursive partitioning (single tree based), and SVM. The accuracies of these models are recorded. Next, TDA is applied to the same DAES data using the mapper algorithm in R programming language to create a network topology of the data. This is an implementation of the localized modeling discussed above. The contents of each resulting network node of the TDA model are then modeled using the previous traditional machine learning classification models, and the resulting accuracies of each model are compared to the results of the globally optimized machine learning models when not used in conjunction with TDA to determine if accuracies improve more at the local node level over the global level of the DAES data. The null hypothesis is tested, and conclusion is drawn to answer the research question.

Data Collection and Preprocessing

Data for this research was collected from Defense Acquisition Management Information Repository (DAMIR) and the Defense Acquisition Visibility Environment databases. Unit cost, schedule, PM rating, and DAES executive summary data was extracted separately from the database and then joined by PNO, Schedule URI. Next, the data was cleansed to remove missing values. The next step was to remove unnecessary html tags from the executive summary and PM rating explanation text variables. The average sentiment variable was derived from previous research conducted by Joseph and Sconion (2020). Schedule slip features were derived from research conducted by Joseph and Hastings (2020). Further cleaning of text was done using R programming language's TM package to remove punctuations, stop words, conduct stemming, and convert all words to lower case to remove duplication during future text classification analysis. Average sentiment was extracted from DAES executive summaries using the sentimentR package and R programming language. The final dataset contained 10 feature variables, one target variable (PM cost rating), and 4,000 rows of non-missing entries of DAES data.

Analysis

Classification without TDA

Globally optimized supervised machine learning using the four classification models discussed above were implemented on the DAES data set with PM rating for cost as the target variable. Tables 1 and 2 show the confusion matrix outputs for the SVM model. The training set produced an accuracy of 79.3% while the test set provided a 73.7%. This is consistent with typical training and test sets. The accuracies of the training set are usually higher than those of the test set. The training accuracies for the random forest, recursive partitioning, and neural network models are 99.1%. 64.1%, and 60.6% respectively. The testing accuracies for the random forest, recursive partitioning, and neural network models are 98.3%, 62.6%, and 56.7% respectively.

Table 1. Confusion Matrix SVM Training

	Green	NoRating	Red	Yellow
Green	444	20	56	76
NoRating	63	618	17	19
Red	22	16	503	24
Yellow	130	16	92	551
Accuracy = 79.3%				

Table 2. Confusion Matrix SVM Testing

	Green	NoRating	Red	Yellow
Green	195	19	30	40
NoRating	55	293	17	10
Red	16	7	241	26
Yellow	75	11	44	254
Accuracy = 73.7%				

Classification with TDA

The TDA mapper algorithm was implemented on the data set in R programming language using the following parameters: a sample size of 4,000 rows of data with 10 features, a Euclidean distance similarity function, the kernel distance estimator (KDE) filter function, and bins with 10 intervals overlapping at 50%. Figure 4 illustrates the resulting network graphing output from the mapper algorithm in R programming language. Figure 4 also depicts that the network shape of the underlying original DAES data is a regression type. Other renderings were flare shaped in some iterations prior to this final rendering. The node numbers are from left to right.

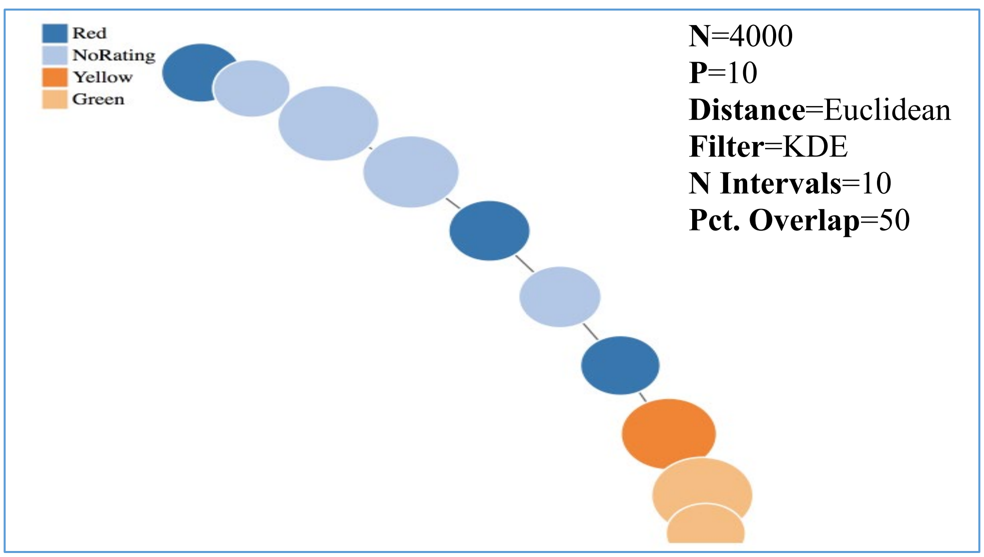


Figure 4. Network Topology Output of Mapper Algorithm in R of DAES Data (Shape Regression)

Figure 5 illustrates the number of rows of data assigned to each node from the mapper algorithm. There are 10 nodes because 10 bins were requested in the input of the mapper algorithm parameters. We notice that the sum of the row contents does not sum to the 4,000-sample size. This is due to the 50% overlap in the binning where some row IDs of one node may be included in other nodes. An extraction of that row ID information can give context to how each node can be described by an analyst and subject matter expert of the data.

```
Nodegroup Nodesize PM_Rating_Cost.maj.vertex  filter.kde
        1      561                       Red 0.001011081
        2      529                  NoRating 0.004631899
        3     1028                  NoRating 0.005935219
        4      922                  NoRating 0.007293477
        5      607                       Red 0.009355363
        6      625                  NoRating 0.011357024
        7      575                       Red 0.013117226
        8      891                    Yellow 0.015866479
        9     1030                     Green 0.016917012
       10      570                     Green 0.018434909
```

Figure 5. R output of TDA Mapper Network Graph Nodes

Tables 3 and 4 depict the confusion matrix and accuracy produced by implementing the SVM machine classification model on the contents of node 1 of the resulting TDA mapper algorithm network topology output data. Tables 5 and 6 depict the confusion matrix and accuracy produced by implementing the SVM classification model on the contents of node 10 of the resulting TDA mapper algorithm network topology output data. In both cases, the accuracy results of SVM when used with TDA at the local level is an improvement over the accuracy of the SVM model when implemented globally on the data set. The results of the other classification models can be found in Table 7.

Table 3. Confusion Matrix SVM TDA Training Node 1

	Green	NoRating	Red	Yellow
Green	54	1	3	2
NoRating	3	42	0	0
Red	8	0	179	10
Yellow	10	2	2	58
Accuracy = 89.0%				

Table 4. Confusion Matrix SVM TDA Testing Node 1

	Green	NoRating	Red	Yellow
Green	22	1	1	76
NoRating	2	27	1	19
Red	5	0	87	24
Yellow	3	3	0	551
Accuracy = 85.6%				

Table 5. Confusion Matrix SVM TDA Training Node 1

	Green	NoRating	Red	Yellow
Green	133	4	10	11
NoRating	1	18	0	0
Red	3	0	41	2
Yellow	11	0	15	131
Accuracy = 85.0%				

Table 6. Confusion Matrix SVM TDA Testing Node 10

	Green	NoRating	Red	Yellow
Green	64	3	6	11
NoRating	1	7	0	0
Red	1	0	16	0
Yellow	9	0	9	63
Accuracy = 78.9%				

Results

Table 7 shows the results of implementing machine learning classification models with and without TDA to predict future PM cost ratings. It can be seen that

- 80% of all training and testing models have improved accuracy when used in conjunction with TDA
- 85% of the training models from traditional machine learning methods produced improved accuracy when used in conjunction with TDA vice using the traditional methods independently
    - Random Forest model improved in 40% of the training nodes
    - All other models improved in 100% of the training nodes
- 75% of the testing models from traditional machine learning methods produced improved accuracy when used in conjunction with TDA
    - Random Forest model improved accuracy 0% of the TDA produced testing nodes
    - All other models improved accuracy 100% of the TDA produced training nodes
- Weaker learners improved in training and testing accuracy while the strongest learner (Random forest) decreased by 0.4%-6.2% accuracy in testing performance when used with TDA.
- There may be a point of diminishing returns on increased accuracy if traditional models already perform at 98% accuracy
    - Further research needed to unpack this phenomenon.

Table 7. Accuracy Results of Machine Learning With and Without TDA

Accuracy Results of Using TDA with Machine Learning Vs Machine Learning Only						
Node	**Accuracy**	**Sample Size**	**Recursive Partitioning**	**Support Vector Machine**	**Random Forest**	**Neural Network**
	Without TDA					
	Training	2,667	64.1	79.3	99.1	60.6
	Testing	1,333	62.6	73.7	98.3	56.7
	With TDA					
Node 1	Training	374	85.0	89.0	99.7	79.1
	Testing	187	83.4	85.6	96.3	80.7
Node 2	Training	353	87.2	92.4	98.0	84.3
	Testing	176	85.7	90.3	97.2	80.5
Node3	Training	685	88.5	88.3	98.1	83.1
	Testing	343	86.3	85.7	96.8	79.3
Nde 4	Training	615	87.5	84.9	98.7	86.8
	Testing	307	85.7	81.8	95.4	86.7
Node 5	Training	405	84.9	90.1	100.0	86.7
	Testing	202	76.2	82.2	92.1	83.7
Node 6	Training	417	89.9	89.2	99.8	92.1
	Testing	208	85.6	83.2	92.8	84.1
Node 7	Training	383	84.6	88.8	99.0	72.8
	Testing	192	81.8	87.0	94.3	70.8
Node 8	Training	594	84.0	84.7	97.8	79.6
	Testing	297	78.1	84.2	92.6	69.3
Node 9	Training	687	85.7	86.5	98.7	80.0
	Testing	343	81.9	76.7	94.2	67.9
Node 10	Training	380	86.1	85.0	100.0	88.6
	Testing	190	83.1	78.9	94.7	80.0
	Accuracy Increase With TDA Over Without TDA					
Node 1	Training	NA	20.9	9.7	0.6	18.5
	Testing	NA	20.8	11.9	-2.0	24.0
Node 2	Training	NA	23.1	13.1	-1.1	23.7
	Testing	NA	23.1	16.6	-1.1	23.8
Node3	Training	NA	24.4	9.0	-1.0	22.5
	Testing	NA	23.7	12.0	-1.5	22.6
Nde 4	Training	NA	23.4	5.6	-0.4	26.2
	Testing	NA	23.1	8.1	-2.9	30.0
Node 5	Training	NA	20.8	10.8	0.9	26.1
	Testing	NA	13.6	8.5	-6.2	27.0
Node 6	Training	NA	25.8	9.9	0.7	31.5
	Testing	NA	23.0	9.5	-5.5	27.4
Node 7	Training	NA	20.5	9.5	-0.1	12.2
	Testing	NA	19.2	13.3	-4.0	14.1
Node 8	Training	NA	19.9	5.4	-1.3	19.0
	Testing	NA	15.5	10.5	-5.7	12.6
Node 9	Training	NA	21.6	7.2	-0.4	19.4
	Testing	NA	19.3	3.0	-4.1	11.2
Node 10	Training	NA	22.0	5.7	0.9	28.0
	Testing	NA	20.5	5.2	-3.6	23.3

Conclusion and Recommendations

Based on the results of the analysis in 80% of training and testing cases, we can fail to reject the null hypothesis and conclude that traditional machine learning algorithms (recursive partitioning, SVM, and neural networks) have higher predictive accuracy when combined with TDA at least 70% of all nodes. The random forests improved accuracy 40% of time in training instances and is the only model that did not improve with TDA Mapper implementation in all cases, although it does at nodes 1, 5, 6, and 9 for the training set

Machine learning at the local network group level appears to improve classifier performance than if done solely at the global level in this use case and from literature on TDA. It is recommended that TDA be used in conjunction with traditional machine learning models when predicting targets for other acquisition-related use cases.

Continuing and Future Work

Based on the above research, my data analytics team in ADA (lead by Trami Pham) has implemented a random forest model with and without TDA to predict future PM ratings in the DoD Comptroller's Advana environment. This model has more feature variables than the MVP discussed above, so accuracy results are slightly different. Additionally, the team is working to implement long-short-term-memory (LSTM) neural network and SVM models in conjunction with

TDA. Table 8 depicts the results of the comparison of the random forest model with and without the use of TDA. The use of TDA has improved the accuracy of the random forest model by over 6.5% in all prediction periods.

Table 8. Prediction Accuracy comparison of A&S' Advanced Analytics MVP App in Advana

	30/60/90 Day Model Predictions		
	30 days/1 time step	60 days/ 2 time steps	90 days/3 time steps
Random Forest	90.9%	89.1%	90.0%
TDA + Random Forest	97.9%	97.4%	97.9%

Figure 6 is an illustration of TDA used in conjunction with a random forest classification algorithm implemented as part of the Advanced Analytics application housed in the OUSD(Comptroller) Advana environment and displays the network graph produced by the mapper algorithm. It is interactive so if one clicks on a node in the application, the contents of that node can be displayed. The confusion matrix, prediction accuracies, and other model prediction performance scores such as precision, recall, and f-measures are presented for each node.

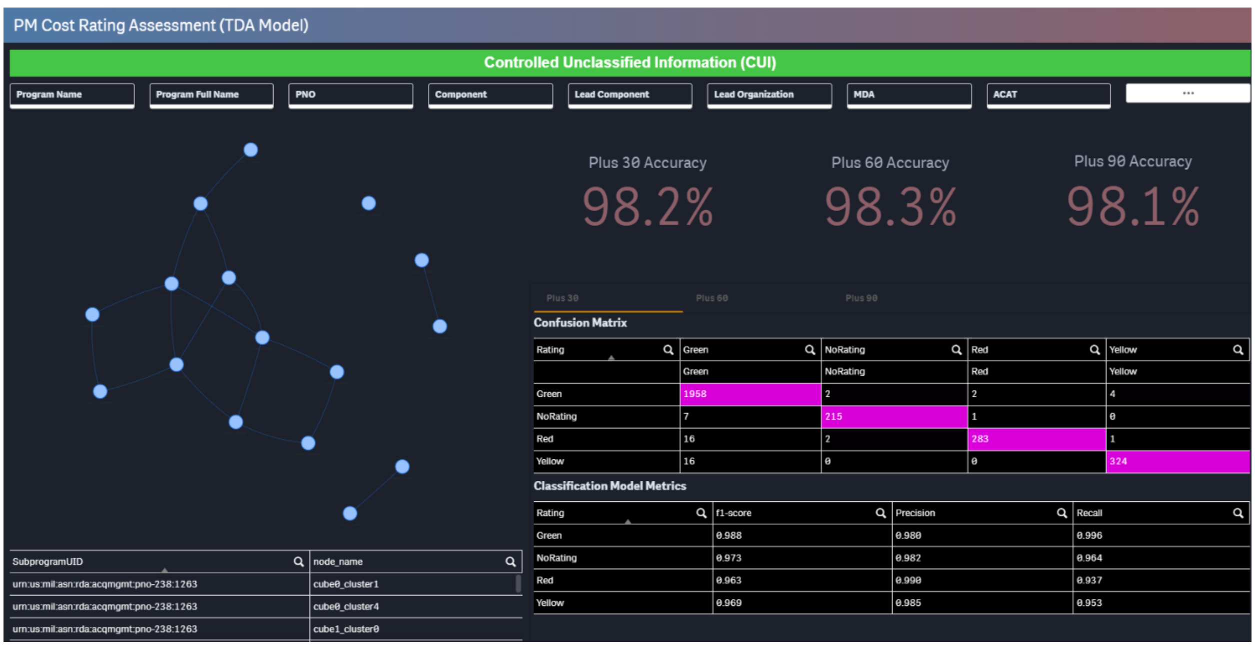


Figure 6. TDA With Random Forest Model Confusion Matrix and Network Graph Application in the USD(C) Advana Environment (Advana, 2021)

Figure 7 displays the predictions of future MDAP PM cost ratings in 30/60/90-day intervals for individual MDAPs that are currently reporting in the DAMIR/DAVE databases. As an example, it can be seen that the MQ-4 Triton is currently reporting a red PM cost rating but is predicted to turn green in 60 to 90 days. The analyst may decide based on current red and 30-day red predictions that this program may need some attention. Leadership, however, may determine that since the program is set to trend green in 60 to 90 days that it does not require attention. As another example, if programs are currently rated green and are projected to trend green over the 30/60/90-day time horizons, there is no need for the analyst or leadership to waste valuable time in conducting a program assessment for that MDAP. Better use of their time can be used prioritizing those programs that are green and yellow and trending to red.

Figure 7. Actual PM Cost Rating 30/60/90-Day Predictions for MDAPs in the Advanced Analytics Application of the Acquisition Analytics Dev Stream in Advana (Advana, 2021)

Finally, we are investigating the use of TDA to predict duration lengths in MTA programs. Besides improving the accuracy of machine learning models, we also plan to use the TDA to understand the relationships and topology of MTA program data.

References

Almgren, K., Kim, M., Lee, J. (2017). Extracting knowledge from the geometric shape of social network data using topological data analysis. *Entropy 2017*, *19*(7), 360. https://doi.org/10.3390/e19070360

Ayasdi. (2020). *AyasdiAI vs. open source TDA*. https://www.ayasdi.com/wp-content/uploads/2020/07/Ayasdi-TDA-vs.-Open-Source-TDA-06.03.20-1.pdf

Ayasdi. (2020). *Understanding Ayasdi: What we do, how we do it, why we do it* [White paper]. https://s3.amazonaws.com/cdn.ayasdi.com/wp-content/uploads/2018/04/04142230/UnderstandingAyasdi_WP_061617v011.pdf

Chazal, F., & Michel, B. (2016). *Mapper algorithm with the R-package TDAmapper*. http://bertrand.michel.perso.math.cnrs.fr/Enseignements/TDA/Mapper.html

Clarke B., Fokoue E., & Zhang, H. H. (2009). *Principles and theory for data mining and machine learning*. Springer.

Elyasi, N., & Moghadam, M. (2019). *An introduction to a new text classification and visualization for natural language processing using topological data analysis*. https://arxiv.org/pdf/1906.01726v1.pdf

Farrelly, C. (2018). *Topological data analysis for data professionals: Beyond Ayasdi*. https://www.kdnuggets.com/2018/01/topological-data-analysis.html

Fasy, B., et al. (2015). *Introduction to the R package TDA*. https://cran.r-project.org/web/packages/TDA/vignettes/article.pdf

Friedman J., Hastie T., & Tibshirani, R. (2009). *The elements of statistical learning. data mining, inference, and prediction* (2nd ed.). Springer.

Georges, A. (2019). *Topological based machine learning methods* [Dissertation]. https://escholarship.org/uc/item/4vr8963d

Guo, W., & Banerjee, A. G. (2017). Identification of key features using topological data analysis for accurate prediction of manufacturing system outputs. *Journal of Manufacturing Systems*.

Herring, A. W. (2018). *The Mapper algorithm*. https://jdc.math.uwo.ca/TDA/Herring-Mapper.pdf

Joseph, B., & Hastings, C. (2020). Using ANOVA and multinomial logistic regression to analyze defense acquisition executive summary (DAES) and acquisition program baseline (APB) milestone estimates to determine contributing factors to schedule slips. *Proceedings of the 17th Annual Acquisition Research Symposium*. 1–20.

Joseph, B., Hastings, C., & Houston, K. (2020). *Updates to selected analyses from the performance of the defense acquisition system series: 2019 SARs update*. https://www.acq.osd.mil/aap/assets/docs/PDAS%202019%20Excerpts_Final%20-cleared.pdf

Joseph, B., & Sconion, D. (2020). Using natural language processing, sentiment analysis, and text mining to determine if text in selected acquisition report executive summaries are highly correlated with major defense acquisition program (MDAP) unit costs and can be used as a variable to predict future MDAP costs. *Proceedings of the 17th Annual Acquisition Research Symposium*. 1–11. https://www.acq.osd.mil/aap/assets/docs/SYM-AM-20-061_Panel11_Joseph_Paper_4-13-2020.pdf

Kim, J., et al. (2015). *Tutorial on the R package TDA*. http://www.stat.cmu.edu/topstat/topstat_old/Talks/files/Jisu_150623_TDA_tutorial.pdf

Kindelan, R., Cerda, M., Frias, J., & Hitchsfeld, N. (2021). *Classification based on topological data analysis*. https://arxiv.org/pdf/2102.03709v1.pdf

Levine, P. (2018). *Annual weapons system assessment. GAO needs to step up its game*. https://www.ida.org/-/media/feature/publications/a/an/annual-weapons-acquisition-assessment---gao-needs-to-step-up-its-game/p-10692.ashx

Lum, P. Y., Singh, A., Lehman, T., Ishkanov, M., Vejdemo-Johansson, M., Allagapen, M., Carlsson, J., & Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports*, *3*(1236). http://www.nature.com/srep/2013/130207/srep01236/full/srep01236.html

Riihimaki, H., Chacholski, W., Theorell, J., Hilbert, J., & Ramanujam, R. (2020). A topological data analysis based classification method for multiple measurements. *BMC Bioinformatics*, *21*(336). https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-020-03659-3

SymphonyAI. (2021). *Topological data analysis and machine learning: Better together*. https://www.symphonyai.com/wp-content/uploads/2020/07/SAI-Topological-Data-Analysis-and-Machine-Learning-Better-Together-vf.pdf

Wu, C., & Hargreaves, C. A. (2020). *Topological machine learning for mixed numeric and categorical data*. https://arxiv.org/pdf/2003.04584.pdf

Rethinking Government Supplier Decisions: The Economic Evaluation of Alternatives (EEoA)

Dr. Francois Melese—Naval Postgraduate School [fmelese@nps.edu]

Dr. James Fan—Naval Postgraduate School [james.fan@nps.edu]

Abstract

This paper offers an economic model to assist public procurement officials in ranking competing vendors when benefits cannot be monetized. An important defense application is "source selection"—choosing the most cost-effective vendor to supply military equipment, facilities, services, or supplies. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature that underpins widely applied decision tools. The bulk of the literature, and most government-mandated decision tools, focuses on the demand side of a public procurement. The "economic evaluation of alternatives" (EEoA) extends the analysis to the supply side. A unique feature of EEoA is to model vendor decisions in response to government funding projections. Given a parsimonious set of continuously differentiable evaluation criteria, EEoA provides a new tool to rank vendors. In other cases, it offers a valuable consistency check to guide government supplier decisions.

Keywords: defense acquisition, decision analysis, multi-attribute auction

Introduction

As nations struggle to recover from a global pandemic that devastated lives and destroyed economic activity, massive government spending aimed at limiting the damage has shattered fiscal balance sheets. Record deficits and debt will place nations under enormous pressure to trim defense expenditures. To preserve capabilities, hard choices lie ahead that require a sober assessment of security challenges, and robust methodologies to prioritize defense and other public investments.

Defense procurement is big business. Recently the U.S. Department of Defense (DoD) spent over $300 billion on acquisition, research, development, test, and evaluation, most of it sourced to the private sector (Schwartz et al., 2018). The Organization for Economic Cooperation and Development reports member countries spend more than 12% of their cumulative GDP on public purchases (OECD, 2016). Significant academic effort has been focused on the defense acquisition process through an economic lens; these include theoretical studies (Cavin, 1995), experimental studies (Davis, 2011; Kirkpatrick, 1995), and empirical studies (Horowitz et al., 2016). Indeed, understanding and improving the efficiency and effectiveness of public procurement is of utmost practical and academic importance.

One of the biggest challenges for public procurement officials is to rank vendors when benefits cannot be monetized. Indeed, government benefits are often depicted as bundles of desirable characteristics or attributes that cannot easily be combined with costs into a single overall measure such as profitability. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature generally referred to as multi-criteria decision-making (MCDM). A proliferation of applications of decision tools derived from this literature has appeared in the fields of management science, operations research, and decision sciences (prominent examples include Keeney & Raiffa [1976]; Kirkwood [1995, 1997]; Clemen [1996]; Parkes & Kalagnanam [2005]; Ewing et al. [2006]).

Today, widespread application of MCDM tools and techniques is mandated through various laws, rules, and regulations that govern public procurement, though the specific approach is not prescribed. For example, the main guide for federal procurement officials in the United States is the Federal Acquisition Regulation (FAR).[1]

> Evaluation criteria are the factors an agency uses to determine which of several competing proposals submitted in response to an RFP [Request for Proposal] would best meet the agency's needs. In establishing effective evaluation criteria, an agency must clearly identify the factors relevant to its selection of a vendor and then prioritize or weight the factors according to their importance in satisfying the agency's need in the procurement. …This allows the agency to rank the proposals received. (FAR, Proposal Development, Section M-Evaluation Factors for Award)

Similar source selection techniques are frequently applied in the United States at state and local levels, and in the private sector.

While demand side developments of MCDM models have been extensively studied in the academic literature, the literature is mostly silent about the supply side (vendor) problem. Vendor decisions (bid proposals) are generally treated as exogenous in the Decision Sciences and Operations Research literature. In contrast, the economic evaluation of alternatives (EEoA) captures both demand side—procurement official decisions—and supply side—vendor optimization decisions. Our model formulation is in the spirit of Lancaster's (1966, 1971) "Characteristics Approach to Demand Theory" as modified by Ratchford (1979), and closely corresponds to the third of six approaches to structure an EEoA introduced in Chapter 4 of "Military Cost–Benefit Analysis: Theory & Practice" (Melese, 2015, p. 96).

EEoA encourages public procurement officials to carefully consider the impact on vendor proposals of announced priorities (i.e., desired criteria, characteristics, or attributes for solicited quantities of products, services, or projects, such as computer systems, vehicles, weapon systems, logistics packages, and buildings). Officials should also consider the impact of anticipated future budgets. In response to government-issued priorities—evaluation criteria, quantities, and funding—competing vendors, with different input costs and technologies (described using "engineering production functions")[2] maximize their production offers—bid proposals that consist of bundles of non-price characteristics or attributes.

EEoA models public procurement official decisions in two stages. In the first stage, along with the requirement (quantity demanded), and funding guidance, the procurement official reveals desired evaluation criteria (characteristics or attributes) of the product or service (but not the relative importance/weights). Given this information, competing vendors engage in constrained optimizations based on their respective production technologies ("engineering production functions"), and input costs, to generate proposals that match anticipated future funding. Since input costs and production functions vary among vendors, they play a critical role in their bid proposals—interpreted as bundles of non-price characteristics or attributes embedded in each identical unit offered by a particular vendor. In the second stage, the

[1] Note the exclusive focus on the demand side in the FAR (i.e., ranking exogenously-determined bids received from vendors; see https://www.acquisition.gov/browse/index/far). Also note that standard practice for U.S. military (and other procurement officials) is to: 1) announce factors ("evaluation criteria") relevant to the selection, but then only after receiving vendor proposals, 2) assign specific relative importance/weights to those factors to rank vendors. This practice is modeled in the economic evaluation of alternatives (EEoA).

[2] For interesting discussions of "engineering production functions," see Chenery (1949), Kurtz & Manne (1963), Wibe (1984), Charnes et al. (1991), and Hildebrand (1999).

procurement official ranks competing vendors according to the government's utility function over the evaluation criteria[3] (see Figure 1).

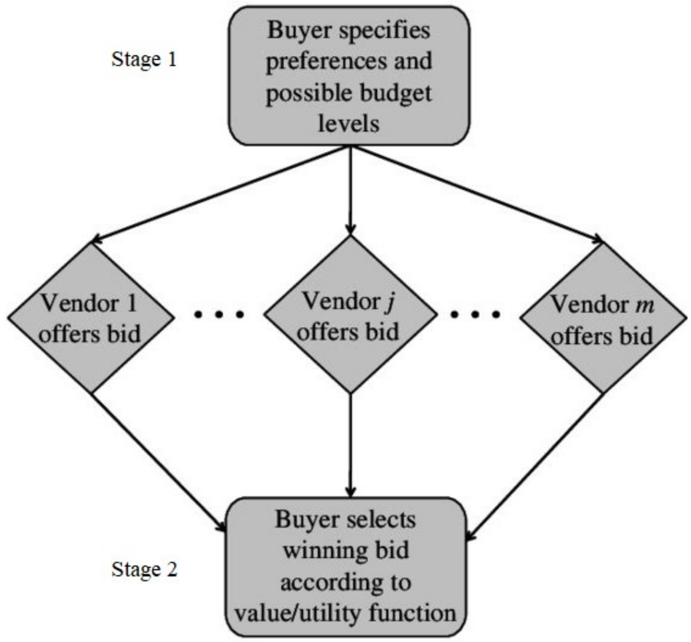


FIGURE 1. The Two-Stage Procurement Process

The dual objective of EEoA is to encourage governments: 1) to consider the supply side (i.e., to recognize the importance of modeling vendor responses to information provided or inferred in public procurements); and 2) to offer an alternative to the standard MCDM approach when benefits cannot be monetized. An attractive feature of EEoA is that it offers a novel technique to measure "benefits" that serves as a valuable consistency check for MCDM preference trade-offs among key attributes.[4] We explore assumptions under which the two decision models (MCDM and EEoA) are isomorphic from a procurement official's perspective. In practice, however, we demonstrate how EEoA can yield significantly different solutions (rank orderings of vendors) than the standard MCDM approach.

The paper proceeds as follows. The next section develops the two-stage economic evaluation of alternatives (EEoA) model. On the supply side, two cases are presented to illustrate the model: 1) where vendors have identical attribute costs, but different production technologies ("engineering production functions"); and 2) where vendors have different attribute costs, but identical production technologies. A simple example serves to integrate procurement official (demand) considerations, with vendor (supply) decisions, under varying (probabilistic) scenarios. The next section contrasts an application of EEoA, with the standard textbook application of MCDM. The last section concludes with recommendations for future research.

---

[3] Note this is analogous to steps mandated in the FAR, except that, since funding is fixed in EEoA (i.e., the unit price is the same for each vendor), the second step involves the submission by vendors of sealed non-price bids for the announced level of funding, interpreted and evaluated by procurement officials as bundles of characteristics and attributes that respond to previously announced evaluation criteria (for example, see FAR 14.5).

[4] Both Australian and Canadian Ministries of Defence are considering implementing this consistency check for the MCDM component of their portfolio decision models. (Personal correspondence with fellow NATO SAS-134 Defence Official Panel Members studying Defence Portfolio Management for NATO; emails received 11/2018)

The Economic Evaluation of Alternatives (EEoA) Model

The challenge for our public procurement official is to select a competing vendor that delivers the best performance (combination of desired non-price attributes) for each identical unit of a requirement (e.g., 100 ventilators, or 50 computers, or 20 drones, or 2 hospital ships), at affordable funding levels. The EEoA framework can be thought of as a multi-attribute sealed bid procurement auction that extends traditional price-only auctions to one in which competition among $j \in [1, m]$ vendors (bidders) takes place exclusively over bundles of $i \in [1, n]$ non-price characteristics or attributes (a_{ij}).[5]

The EEoA model structures the problem as a two-stage optimization (see Figure 1). In the first stage, the public procurement official provides j competing vendors with the evaluation criteria, available funding, and the requirement (quantity demanded).[6] Given the anticipated budget, **B**, and their respective production technologies ("engineering production functions") and input costs, competing vendors offer their best possible non-price attribute packages bundled into each identical unit required.[7] Note that the greater the funding available, the greater the available funding per unit, which allows vendors to bundle more of the desired attributes into each identical unit (e.g., better ventilators, computers, drones, ships).[8]

The vendor (supply side) problem is formulated in the section titled First Stage EEoA: The Vendor's Problem (Supply Side). Competition takes place exclusively over non-price bid proposals from each vendor, evaluated by procurement officials as bundles of attributes offered by each vendor for a standard unit of the requirement. Whereas attributes for each unit of the requirement are identical for each vendor, the proposed bundles differ among vendors. Competing vendors' bid proposals (bundles of attributes) depend on a vendor's specific costs to generate each attribute, their individual engineering production function to combine those attributes, and anticipated future funding.

In the second stage, the procurement official's objective is to select the vendor j that maximizes the government's utility function, $U_j = U_j(a_{1j}, a_{2j}, \dots, a_{nj})$, subject to projected funding (i.e., the per unit affordability or budget constraint), **B**. For analytic tractability we assume the utility function is quasi-concave, and that attributes are continuous, non-negative, monotonic increasing variables (i.e., the domain of the buyer's utility function, and sellers' production functions and attribute cost functions) are the nonnegative real numbers. Non-satiation in the relevant range of attributes is also assumed, such that, $\partial U_j / \partial a_{ij}$>0, or the greater the score of the $i \in [1, n]$ desired attributes, a_{ij}, the more value (utility/benefit) for the buyer, but the more costly it is for sellers to produce.

---

5 For example, in the case of military medical transportation of patients to receive emergency treatment, safe transport may require the use of a reliable ventilator. In evaluating ventilators, some key attributes include battery duration, gas consumption, and levels of leakage (L'Her et al., 2014, Blakeman & Branson, 2013).

6 Since there is a fixed requirement (quantity demanded), the budget, **B**, can be interpreted as the unit funding/budget available to vendors to produce a unit of the required product or service. For example, if we anticipate $25,000 of funding is available for 50 computers, the budget (**B**) used by competing vendors to build their proposals would be $500 per unit.

7 For example, suppose we have $25,000 of funding for 50 computers, or a budget, **B**=$500/unit. Then, for example, each of 50 identical Apple notebook computers offered at $500/unit would satisfy the basic evaluation criteria (screen size, memory, battery life, software), but consist of a somewhat different bundle of those characteristics/attributes, than each of 50 identical Microsoft (or Dell, or HP) notebook computers.

8 The greater the funding available, the greater the funding per unit, allowing vendors to offer more of the desired attributes for each identical unit demanded by the buyer. For example, suppose for our 50 computers, instead of $25,000 (**B**=$500) of funding, it turns out $50,000 (**B**=$1000) will be available. Then each of the 50 identical notebook computers offered by Apple will have more and/or better characteristics/attributes, and so will each of the 50 identical notebook computers offered by Microsoft (e.g. bigger screens, more memory, longer battery life).

Following the literature, we allow the buyer's utility function (scoring/ranking rule) to be linear, additive, and separable across attributes (see Keeney & Raiffa, 1976; Kirkwood, 1997). The public procurement official's problem is to select a vendor $j \in [1, m]$ that maximizes the government's utility function:

$$(1)\ \boldsymbol{U_j} = \mathrm{U_j}\left(\boldsymbol{A}_j^T\right) = \boldsymbol{W}\boldsymbol{A}_j^{\mathrm{T}},$$

where desired attributes are known to sellers, and the bundle of attributes in vector $\boldsymbol{A_j}$ = $[a_{1j}\ a_{2j} \ldots a_{nj}]$ represents each vendor's offer (bid proposal) for each unit required. The relative weights for each attribute are the procurement official's private information, given by the vector:

$$\boldsymbol{W} = (w_1, w_2, w_3, \ldots, w_n \mid w_i \in \mathbb{R}^+, i \in [1, n]).$$

The procurement official maximizes (1) subject to a funding/affordability constraint:

$$(2)\ TC_j \ \leq\ \mathbf{B},$$

such that the total unit cost (price) of any vendor's bid proposal, TC_j, must fit within forecasted future funding (i.e., the per unit budget). **B.** Note that whereas the set of non-price attributes in the buyer's utility function are revealed to the $j \in [1, m]$ competing vendors, the ***relative*** (preference or "trade-off") ***weights***, w_i, are not.[9] This reflects practical application of the FAR:

> In government acquisition, procuring commands have their own best practices and priorities … but they all follow the [Federal Acquisition Regulation]. And in their selection of suppliers, they assign weights to their parameter criteria in accord with their priorities. … These weights for scoring of proposals do not have to be specifically revealed as an algorithm, but are typically communicated to offerors in terms of [rank ordering of] importance.
>
> Colonel John T. Dillard, U.S. Army (Retired),
> Past Program Manager for Advanced Acquisition Programs

In this formulation of the procurement problem, both buyer and seller suffer from imperfect and asymmetric information. While the seller does not know the specific relative importance/weights assigned to desired attributes (or "evaluation criteria"), the buyer (procurement official) does not know the vendors' costs of producing a particular attribute, nor the technology (engineering production functions) that combines those attributes into vendor proposals.[10] The supply side vendor problem is examined in detail in the next section, followed by the demand side procurement problem.

First Stage EEoA: The Vendor's Problem (Supply Side)

The first stage of the two-stage EEoA optimization framework focuses on the vendor's problem. The economic approach assumes vendors are strategic players, so that the anticipated/forecasted (per unit) funding/budget, **B**, for the procurement, impacts vendors'

---

[9] For example, consider the following summary of Federal Acquisitions Regulations (FAR) Sections 15.1 and 15.3 "Evaluating proposals under the RFP [Request for Proposal] best value trade-off analysis criteria": In a negotiated bid there are factors [evaluation criteria] with varying weights assigned. The solicitation tells you the weight of each factor. However, government contracting agencies are not required to publicize the actual source selection plan [it is an internal document]. The agency has broad discretion on what it believes to be the best value. Note, however, the agency must be consistent in following their source selection plan in evaluating every vendor, or risk bid protests—e.g., see Melese (2018).

[10] "Seller costs can be expected to depend on [the] local manufacturing base, and sellers can be expected to be well informed about the cost of (upstream) raw materials" (Parkes & Kalagnanam, 2005, p. 437).

formulation of their competing bid proposals. Vendor bid proposals consist of optimal attribute bundles, $\boldsymbol{A_j}$, from competing firms that maximize overall performance (output, $\boldsymbol{Q_j}$) given their respective engineering production functions, costs, and constraints.[11]

Specifically, given *n* desired attributes (a_{ij}), and anticipated future funding (the per unit budget, **B**), the m vendors each offer competing bid proposals (bundles of attributes), $\boldsymbol{A_j}$, based on their production technology,[12] and their unit costs of producing each attribute, $c_{ij}(\boldsymbol{B})$.[13] For any fixed requirement (quantity demanded) and funding level (per unit budget, **B**), a representative vendor's problem is to maximize the attribute output/performance of each (identical) unit required, subject to the vendor's costs of producing each attribute. Wise & Morrison (2000) observe that a multi-attribute auction allows competing vendors to differentiate themselves in the auction process and bid on their competitive advantages. Competing vendors offer their best possible non-price attribute bundle for the projected per unit funding/budget, **B,** given their idiosyncratic technology reflected in their respective "engineering" production functions given by Equation 3.

The vendor's problem can be expressed as selecting an attribute vector (bid proposal), $\boldsymbol{A_j}$ = $[a_{1j}, a_{2j}, \ldots, a_{nj}]$ that maximizes output or "product performance" given by their engineering production function:

$$(3)\ \boldsymbol{Q_j} = Q_j(\boldsymbol{A}_j^T),$$

subject to total unit costs (TC) not exceeding anticipated per unit funding (**B**) for the project,

$$(4)\ \boldsymbol{TC_j} = \sum_{i=1}^{n} c_{ij}\,(B)\,a_{ij} \leq \mathbf{B}.$$

Hollis Chenery was the first economist to introduce "engineering production functions" similar to Equation 3. In his pioneering article in the *Quarterly Journal of Economics*, he observes: "The engineer must usually resort to testing various sizes and combinations of equipment to determine the effect of such variables as size, speed, temperature, etc., upon total performance" (Chenery, 1949).[14] A detailed survey by Soren Wibe (1984) in *Economica* offers a useful contrast between "engineering" and "economic" production functions.[15] Similarly, but in a different context, Hildebrandt (1999) in this journal introduced what he calls a "technological military production function" derived from underlying technical relationships that relate military inputs to measures of effectiveness/performance. In his study, alternatives are scored along their important attributes to estimate a measure of effectiveness that reflects capabilities required to complete a mission. Although their theoretical foundations differ, engineering production functions parallel traditional MCDM approaches in the use of so-called "value" functions to estimate effectiveness.

For ease of exposition, the remainder of the study focuses on two vendors and two (non-price) attributes. We assume each vendor has a different technology (engineering production

---

[11] Note the supply-side development in this section generalizes a special case of the multi-attribute auction found in Simon and Melese (2011).
[12] Each vendor's bundle is a technologically-determined combination of attributes: for instance, a computer is a combination of screen size, memory, battery life, and others with unit costs associated with each attribute.
[13] For instance, with bigger budgets, a vendor's costs to provide more of a particular attribute (say computer memory) might enjoy increasing returns to scale because of quantity discounts.
[14] Chenery (1949) connects his engineering approach to production functions to studies that helped motivate Lancaster (1966), stating: "The use of multi-dimensional products has already been suggested in the field of consumption" (p. 514).
[15] Also see the extension of this survey offered by V. Kerry Smith (1986).

function) to combine the two attributes, and different attribute costs. The Lagrangian function for the vendor's problem is given by:

$$(5)\ \mathcal{L}_j = Q_j(a_{1j}\ , a_{2j}, \boldsymbol{B}) + \lambda_j[\boldsymbol{B} - \sum_{i=1}^{2} c_{ij}\ (\boldsymbol{B})\ a_{ij}], \text{ for j=1,2.}$$

Since vendors compete on their product's quality/performance, we assume they will use the maximum expected per unit funding, **B**, to develop their bid proposals, so that Equation 4 is an equality. First order necessary conditions for an optimum are given by:

$$(5a)\ \partial\mathcal{L}_j/\partial a_{1j} = \partial Q_j/\partial a_{1j} - \lambda_j c_{1j}(\boldsymbol{B}) = 0,$$

$$(5b)\ \partial\mathcal{L}_j/\partial a_{2j} = \partial Q_j/\partial a_{2j} - \lambda_j c_{2j}(\boldsymbol{B}) = 0,$$

$$(5c)\ \partial\mathcal{L}_j/\partial \lambda_j = \boldsymbol{B} - \sum_{i=1}^{2} c_{ij}\ (\boldsymbol{B})\ a_{ij} = 0.$$

Solving Equations 5a–5c, yields optimal attribute bid proposals (performance outputs) for each vendor $j = 1,2$, for each identical unit required, for any given per unit budget, ***B***:

$$(6a)\ a_{1j}^{*} = a_{1j}^{*}(y_{1j}(\boldsymbol{B}), y_{2j}(\boldsymbol{B}), c_{1j}(\boldsymbol{B}), \boldsymbol{B}),$$

$$(6b)\ a_{2j}^{*} = a_{2j}^{*}(y_{1j}(\boldsymbol{B}), y_{2j}(\boldsymbol{B}), c_{2j}(\boldsymbol{B}), \boldsymbol{B}).$$

For purposes of illustration, we assume a standard Cobb-Douglas (see Cobb & Douglas, 1928; Douglas, 1976) engineering production function in the spirit of Charnes et al. (1991) and others, with two attributes $(a_{1j}\ , a_{2j})$ as inputs:

$$(6)\ Q_j\left(a_{1j}\ , a_{2j}\right) = \ a_{1j}^{y_{1j}} a_{2j}^{y_{2j}};$$

where the elasticities, y_{nj} (i.e., the % change in output/performance from a % increase in an attribute), are assumed to be independent of available funding (the budget, $\boldsymbol{B}$), and sum to 1. [16]

As stated by Charnes et al. (1986), the Cobb-Douglas engineering production function given by Equation 6 is "the simplest … case of static production with a single output [bundle of attributes] to be produced with a single function—one to a firm [vendor] or plant—from factors [yielding attributes] which are acquired at fixed prices per unit [i.e., fixed unit costs to produce each attribute]." In terms of our model, this suggests starting with the assumption that the unit costs for each vendor, j, are independent of available funding, or that: $c_{1j}(\boldsymbol{B}) = c_{1j}$, and $c_{2j}(\boldsymbol{B}) = c_{2j}$.

Two special cases help illustrate our model: 1) where vendors share common attribute costs, but have different production technologies (engineering production functions), and 2) where vendors share the same production technology, but have different attribute costs.

Vendors with Common Costs and Different Technologies

In the first case (illustrated in Figure 2), vendors $j = 1,2$ have identical, attribute costs (i.e., $c_{1j}(\boldsymbol{B}) = c_1$ and $c_{2j}(\boldsymbol{B}) = c_2$), but different, constant engineering production functions (i.e.,

---

[16] Interestingly, an article by Marsden et. al. (1972) in *Applied Economics* shows how a Cobb-Douglas production function for waste treatment plants can be derived from chemical and biological laws. Another notable engineering production function study by Kurtz & Manne (1963) in the *American Economic Review* estimates a Cobb-Douglas production function from engineering data for various metal machining processes. They also emphasize "it is the *characteristics* [or attributes] of the task that determine the input-output relationship" (p. 667).

$y_{11} \neq y_{12}$ and $y_{21} \neq y_{22}$). From the first order necessary conditions for an optimum ((5a) – (5c)), and (6), competing vendors' optimal attribute bundle bid proposals, for the expected per unit funding/budget level **B**, are given by:

$$(6a')\ a^*_{1j} = [y_{1j}/(y_{1j} + y_{2j})\ c_1]\ \mathbf{B}, \text{ and}$$

$$(6b')\ a^*_{2j} = [y_{2j}/(y_{1j} + y_{2j})\ c_2]\ \mathbf{B}.$$

Figure 2 illustrates optimal attribute bundle bid proposals for each vendor for a specific unit funding/budget level, **B**: $A_1 = (a^*_{11}, a^*_{21})$ and $A_2 = (a^*_{12}, a^*_{22})$. The optimum for each vendor is determined graphically by the tangency of each vendor's isoquant (derived from their separate production functions), with the common budget constraint.

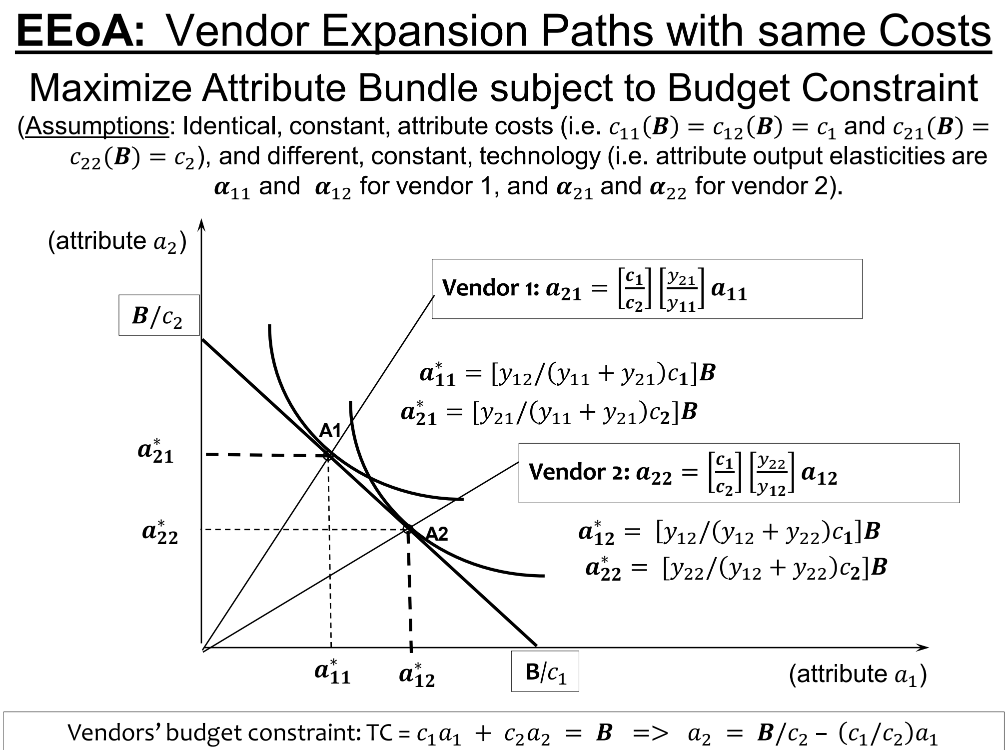


FIGURE 2. Common Attribute Costs but Different Technologies

Suppose instead of a single funding forecast, the buyer (procurement official) reveals a range of possible budget estimates for the procurement (say optimistic, pessimistic, and most likely).[17] Then Equations 6a' and 6b' can be combined to yield each vendor's expansion path, given by:

$$(7)\ a_{2j} = [(c_{1j}(B)/c_{2j}(B))\,(y_{2j}/y_{1j})]\ a_{1j}\ , \text{ for } j = 1{,}2.$$

The two expansion paths defined by Equation 7 reveal optimal attribute bundles offered by each vendor at different possible funding levels, *B*. Each point on the expansion paths derived for each vendor reveals optimal attribute bundle offers (bid proposals) for each identical unit required, over different possible budgets.

[17] For example, see Simon & Melese 2011.

Given this formulation, if attribute costs and technology parameters are constant (i.e., independent of funding levels), then the expansion paths are linear.[18] Expansion paths for the first case, where vendors' share common costs but different technologies, are given by:

$$\text{(7a) } a_{21} = [c_1/c_2][y_{21}/y_{11}]\, a_{11}, \text{ for vendor 1, and}$$

$$\text{(7b) } a_{22} = [c_1/c_2][y_{22}/y_{12}]\, a_{12}, \text{ for vendor 2.}$$

This is illustrated as two straight lines from the origin in Figure 2. For the specific per unit budget level, **B**, the two competing attribute bundle bid proposals offered by each vendor (from Equations 6a' and 6b') appear as points $A_1 = (a_{11}^*, a_{21}^*)$ and $A_2 = (a_{12}^*, a_{22}^*)$ on the competing vendors' expansion paths.

Vendors With Common Technologies and Different Costs

Turning to the second example (illustrated in Figure 3), suppose vendors have different (constant) attribute costs, but identical (constant) engineering production functions (i.e., in Equation 6: $y_{1j} = y_1$ and $y_{2j} = y_2$ for j=1,2), together with constant returns to scale (such that: $y_1 + y_2 = 1$; $i.e. if\ y_1 = y\ then\ y_2 = 1 - y$). In this case the two vendors' optimal bid proposals for unit funding/budget level, **B**, are given by:

$$\text{(6a'') } a_{1j}^* = [y/c_{1j}]\, \mathbf{B}, \text{ and}$$

$$\text{(6b'') } a_{2j}^* = [(1-y)/c_{2j}]\, \mathbf{B}, \text{ (j=1,2).}$$

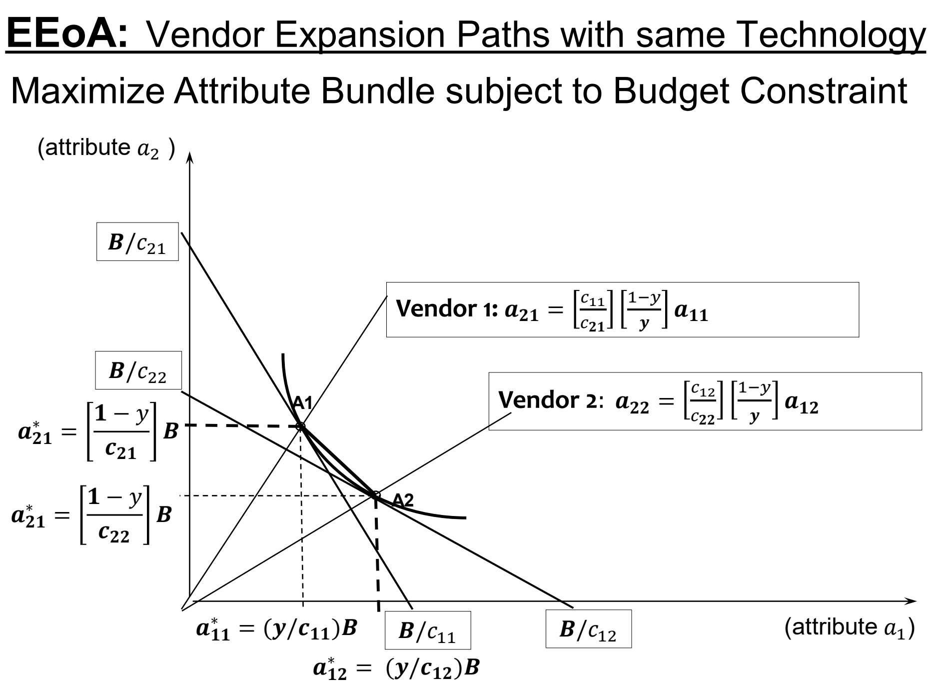


FIGURE 3. Common Technology but Different Attribute Costs

Similar to the first case, Figure 3 illustrates competing optimal attribute bundle bid proposals for each vendor, for the unit funding/budget level, **B**: $A_1 = (a_{11}^*, a_{21}^*)$ and $A_2 =$

[18] For example, see Nicholson & Snyder (2017), pp. 330–333.

(a^*_{12}, a^*_{22}). Now the optimum for each vendor occurs at the point where their respective budget constraints are tangent to their common isoquant. If vendors' technology and attribute cost parameters are constant (i.e., independent of funding levels), both expansion paths are again linear. Expansion paths for this second case (where vendors share a common technology, but have different attribute costs), are illustrated as two straight lines from the origin in Figure 3, given by:

$$\text{(7a')}\ a_{21} = [c_{11}/c_{21}][(1-y)/y]\, a_{11}, \text{ for vendor 1, and}$$

$$\text{(7b')}\ a_{22} = [c_{12}/c_{22}][(1-y)/y]\, a_{12}, \text{ for vendor 2.}$$

Focusing on this second case (where vendors share a common technology, but have different attribute costs), for any unit funding/budget level, **B**, connecting the two optimal vendor attribute production points (A_1 and A_2) creates an attribute "production possibility frontier" (PPF), illustrated in Figure 3. The slope of this PPF reflects attribute trade-offs possible in the marketplace by switching from one vendor to another. This technical (or engineering) trade-off is given by the slope: $\Delta a_2/\Delta a_1 = (a^*_{21} - a^*_{22})/(a^*_{11} - a^*_{12})$.

The first stage vendor optimization problem in the two-stage EEoA framework highlights the importance of modeling the supply side (i.e., vendor decisions in response to anticipated future funding). The second stage focuses on the demand side (i.e., the procurement official's source selection problem).[19]

Second Stage EEoA: Procurement Official's Problem (Demand Side)

For any given requirement (quantity demanded), and forecasted per unit funding/budget, **B**, the procurement official (decision-maker) must rank the vendors' (optimized) bid proposals. For example, consider attribute bundles such as those illustrated in Figure 3: Vendor 1=>(a^*_{11}, a^*_{21}) and Vendor 2=>(a^*_{12}, a^*_{22}). Recall the lens through which the government evaluates competing vendors is the utility function given by Equation 1.[20] In EEoA, the government supplier decision ("source selection") depends on the public procurement official's (decision-maker's) preferences revealed through explicit trade-offs for any pair of attributes that leave decision-maker's indifferent in any given scenario. These explicit pair-wise comparisons elicited from a public procurement official (or expert decision-makers) generate relative weights assigned to the desired attributes.

The public procurement official's problem is to select a vendor $j \in [1, m]$ with a bid proposal (per unit attribute bundle) $\boldsymbol{A}_j = [a_{1j}, a_{2j}, \ldots, a_{nj}]$) that maximizes the government's utility function given by Equation 1. Recall, following the standard assumption in the literature (see Keeney & Raiffa [1976]; Kirkwood [1997]), the utility/benefit provided by any vendor j is given by the linear, separable utility function:

$$\text{(1')}\ \boldsymbol{U_j} = \mathrm{U_j}(\boldsymbol{A}_j^T) = \boldsymbol{W}\boldsymbol{A}_j^{\mathrm{T}} = \sum_{i=1}^{n} w_i\, a_{ij},$$

where the vector $\boldsymbol{A_j} = [a_{1j}\ a_{2j} \ldots a_{nj}]$ represents the bundle of attributes (performance) of each unit, offered by each of the $j \in [1, m]$ competing vendors. As discussed earlier, specific relative trade-off weights for every attribute are the procurement official's private information, given by the vector:

---

[19] Note this second stage demand-side problem is the exclusive focus of most textbooks, the majority of the related decision sciences and operations research literature, and standard support tools and algorithms.

[20] An interesting extension of Equation 1 is developed later to address uncertainty when different possible scenarios (states of nature) impact the government's utility function (for example, due to possible future changes in the political, economic, or threat environment).

$$\boldsymbol{W} = (w_1, w_2, w_3, \ldots, w_n \mid w_i \in \mathbb{R}^+, i \in [1, n]).$$

The procurement official is also fiscally informed, with a forecasted funding/budget (affordability) constraint for the procurement given by Equation 2. So the per unit price (total unit costs) of any vendor proposal, TC_j, must fit within forecasted future funding (the anticipated per unit budget, $\mathbf{B}$), or $TC_j \leq \mathbf{B}$. The next step is to combine demand and supply (i.e., the procurement official's source selection problem) with vendors' (optimization-generated) bid proposals. The following simple source selection example demonstrates how EEoA integrates demand and supply.

Demand and Supply: A Two Scenario, Two Vendor, Two Attribute Example

For purposes of illustration, suppose a public procurement official responsible for UN peacekeeping missions is asked to select a vendor for a new fleet of Autonomous Electric Off-road Light Armored Transport Vehicle (AEOLATV). Assume the anticipated (per unit) budget, **B**, for the program allows two competing vendors to offer the required set of vehicles, and that there are only two evaluation criteria in the government's utility function: ***Top Speed*** of each vehicle measured in miles per hour (a_1), and ***Range*** measured in miles (a_2).[21] In Figure 3, this involves a choice between vendor 1 that offers less speed but more range (a_{11}^*, a_{21}^*), and vendor 2 that offers more speed, but less range (a_{12}^*, a_{22}^*).

In EEoA, the source selection decision (vendor ranking) depends on the procurement official's (decision-maker's) preferences revealed through pair-wise comparisons (i.e., explicit acceptable trade-offs between pairs of attributes within a particular scenario). This generates relative weights assigned to the desired attributes within a particular scenario.

A straightforward modification of Equation 1' allows us to extend the analysis to address different possible scenarios (states of nature) that could impact the procurement official's pair-wise comparisons.[22] Equation 8 accounts for k possible scenarios (or "states of nature"), Ns, $\forall s \in$ [1,k], with corresponding probabilities, $P(Ns)$. This linear, separable ***expected*** utility function captures the differing relative weights, derived from explicit preference trade-offs among pairs of attributes that depend on specific scenarios (states of nature). Now the procurement official's problem is to select the vendor (e.g., bidder or investment alternative), $j \in$ [1,m], that maximizes the government's ***expected*** utility given by:

$$(8)\ \mathbf{E}(\boldsymbol{U_j}) = \sum_{s=1}^{k} P(Ns) \sum_{i=1}^{n} w_{is}\, a_{ij}.$$

Consider a simple case with two possible states of nature N1 & N2, (e.g. Scenario s=1 a High Tech Threat environment, vs. Scenario s=2 a Low Tech Threat Environment), with corresponding probabilities, P(N1) and P(N2).[23] From Equation 8, the government's expected utility function (scoring rule) for the two scenario, two attribute case is:

$$(8')\ \mathbf{E}(\boldsymbol{U_j}) = P(N1)[w_{11}a_{1j}+w_{21}a_{2j}] + P(N2)[w_{12}a_{1j}+w_{22}a_{2j}].$$

Totally differentiating the procurement official's (government's) utility function (8') and setting the result equal to zero in each scenario (N1 & N2), generates two sets of relative weights (or indifference curves). In general, relative weights for any two pairs of attributes (a_1,a_2) in each of the k scenarios in Equation 8 are given by:

---

[21] For example, we could assume all other characteristics (or attributes) of the vehicles offered by the vendors are the same, so top speed and range are the only differentiating factors.

[22] For example, different possible threat environments in which the United Nations might operate.

[23] In the AEOLATV example, scenario N1 could represent the possibility of facing a fast adversary with limited range with probability P(N1), and scenario N2 a slower adversary with greater range with probability P(N2); where P(N1)+P(N2)=1.

$$(9)\ \partial a_2/\partial a_1 = -(w_{1s}/w_{2s}) = -X_s,\ \forall s \in [1,\mathrm{k}].$$

The last term in Equation 9, $X_s > 0$, represents the acceptable trade-off determined by a decision-maker (procurement official) between any pair of attributes (a_1,a_2) for a specific scenario: $w_{1s} = (w_{2s})$x(X_s). It reflects acceptable pair-wise trade-offs for the government over the relevant range of attributes in each scenario. These preference trade-offs define linear indifference curves between any two pairs of attributes in each scenario (or piecewise linear approximations over specific ranges of attributes). The slopes of these indifference curves are the relative weights for each pair of attributes, in each state of nature, over relevant ranges of each attribute.

Optimal vendor rankings in EEoA can be determined by comparing the slope of the government's (buyer's) revealed preferences (indifference curves), with the competing vendor-proposed bundles of attributes (production possibility frontiers). For example, Figure 4 illustrates two different sets of indifference curves (dashed lines) that reflect two different scenarios. In turn, these yield two different vendor rankings.

For a given per unit budget, **B**, if the slope of the indifference curve is steeper than the slope of the production possibility frontier (where the PPF reflects technical/engineering trade-offs available between competing vendors), or if from Equation 9, -X = $-(w_1/w_2) < -(a_{21}^* - a_{22}^*)/(a_{11}^* - a_{12}^*)$, then vendor 2 is selected, since U_2^***>**U_1. If the reverse is true, then vendor 1 wins, since U_1^***>**U_2 (see Figure 4).

Suppose a government decision-maker (public procurement official) is willing to trade off relatively more range (a_2) for the same incremental increase in top speed (a_1) in scenario N1, than in scenario N2. For example: 20 miles of range for an extra 10 mph top speed in N_1, versus only 10 miles for an extra 10 mph in N_2. In this case, $-X_1 = -2 < -X_2 = -1$, implies the slope of the indifference curve is steeper (more negative) in Scenario N_1 than in N_2.[24] From Figure 4, vendor 2 is ranked higher (offers greater utility) in scenario N1, and vendor 1 in scenario N2. This is consistent since the decision-maker revealed a stronger relative preference for top speed in scenario N1 (i.e., was willing to trade off more range), and vendor 2 offers relatively higher top speed (a_{12}^*) than vendor 1 (a_{11}^*).

[24] In this case, under scenario N1 vendor 2 ranks higher (offers greater utility) than vendor 1, and there is a rank reversal under scenario N2.

EEoA: Procurement Agency Choice

Maximize Utility subject to Budget Authority Constraint

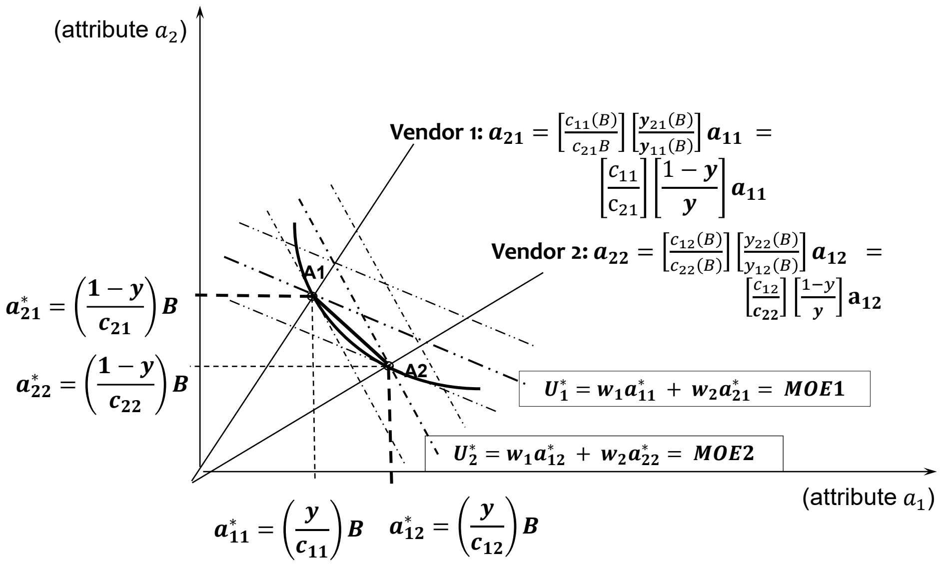


FIGURE 4. Procurement Agency Vendor Selection

In general, probabilities assigned to each scenario in Equations 8 or 8' generate an ***Expected Utility*** vendor ranking metric that consists of a probability-weighted average of pair-wise attribute trade-offs (-Xs) that define expected utility functions in each of the $s \in$ [1,k] scenarios. For example, in the two scenarios, two vendors, two attributes case, this determines the slope of a new indifference curve that is a combination of the two indifference mappings illustrated in Figure 4. For any specified budget, the tangency (or corner point) of this new indifference curve with the PPF reveals the optimal Expected Utility ranking of the two vendors. The next section contrasts this EEoA with the standard textbook MCDM model commonly applied by public procurement officials to guide government supplier decisions.

Comparison of EEoA and MCDM Models

The topic of multi-criteria decision-making (MCDM) has spawned a rich literature with many variations to account for decision-making in complex scenarios. This section presents a standard textbook MCDM model frequently applied to guide government supplier decisions as a baseline (see Keeney & Raiffa [1976]; Kirkwood [1997]). We contrast this MCDM model with the EEoA approach within a single scenario. The MCDM additive value function typically used to rank vendors is given by:

$$(10)\ \boldsymbol{V_j} = \mathrm{V_j}\left(\boldsymbol{A}_j^T\right) = \boldsymbol{\lambda} v_i(a_{ij}) = \textstyle\sum_{i=1}^{n} \lambda_i\, v_i(a_{ij}).$$

This value function is the sum of individual value functions, $v_i(a_{ij})$, defined over relevant ranges of each attribute $i \in [1, n]$, for any vendor j. The vector of preference weights is given by:

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3 \ldots, \lambda_n \mid \lambda_i \in \mathbb{R}^+,\ i \in [1, n]).$$

The individual value functions $v_i(a_{ij})$ are typically monotonic and scaled (normalized), while the preference weights (λ_i) reflect the importance of each attribute. While these weights $(\boldsymbol{\lambda})$ are analogous to the relative weights $(\boldsymbol{W})$ in EEoA, they are only equivalent if raw attribute measures are used in MCDM instead of normalized values to determine pair-wise trade-offs (i.e., iff $v_i(a_{ij}) = a_{ij}$). For purposes of comparison with EEoA, it is convenient to assume procurement officials (decision-makers) are subject to the same funding/affordability constraint given by (2): $TC_j \leq \mathbf{B}$. Implications of this MCDM model are explored in the next section under the usual assumption that attribute measures are normalized using individual value functions with preferential independence.

Implicit Trade-Offs in MCDM vs. Explicit Trade-Offs in EEoA

From Equation 10, the only theoretical difference between the procurement official's objective function (1) or (1') in EEoA, and MCDM is an additional step in Equation 10 that involves normalizing attribute measures through individual value functions. In fact, the demand side of EEoA can be thought of as a special case of MCDM, where $v_i(a_{ij}) = a_{ij}$.

In theory, any value function, v_i, in conjunction with the appropriate attribute weights λ_i, can recover the EEoA utility function for any given vector of attributes $\boldsymbol{A_j}$.This is clear when we consider a procurement official's value function with two attributes as before:

$$(10')\ \boldsymbol{V_j} = \sum_{i=1}^{n} \lambda_i v_i(a_{ij}) \Rightarrow [\lambda_1 v_1(a_{1j}) + \lambda_2 v_2(a_{2j})].$$

Totally differentiating Equation 10 or 10' and setting the result equal to zero yields ***implicit*** trade-offs in the MCDM approach between any two pairs of attributes (a_1, a_2) (i.e., the first two terms in Equation 11). For sake of consistency given a particular decision-maker's preferences, this should precisely correspond to the explicit trade-offs (revealed preferences) obtained from that decision-maker in EEoA (i.e., represented by the last two terms in Equation 9).

$$(11)\ \partial a_2/\partial a_1 = -[\lambda_1 v_1{}'(a_1)]/[\lambda_2 v_2{}'(a_2)] = -\frac{w_1}{w_2} = -X_s.$$

While the MCDM approach adds a degree of freedom for procurement officials and expands the decision space, it risks obscuring explicit trade-offs between attributes revealed in the EEoA approach. From Equation 11, we see that:

$$\lambda_1 / \lambda_2 = X_s[v_2{}'(a_2)/v_1{}'(a_1)], \text{ or}$$

$$Z = [v_2{}'(a_2)/v_1{}'(a_1)],$$

where the constant Z= $\lambda_1 / (\lambda_2 X_s)$. So in general, for any pair of attributes, and alternatives (i.e., vendors $j \in$ [1,m]),

$$(12)\ Z v_1'(a_{1j}) = v_2{}'(a_{2j}).$$

Integrating both sides of (12) yields:

$$(13)\ v_2(a_{2j})/v_1(a_{ij}) = Z = \lambda_1/(\lambda_2 X_s).$$

That is to say, if the goal is to ensure EEoA and MCDM approaches generate the same rank ordering, procurement officials must set individual attribute value functions v_i's and attribute weights λ_i's in the precise ratio specified in Equation 13.

In practice, there is no reason to assume this happens, and reconciling the two approaches to generate the same rank ordering is non-trivial. While a procurement official may have a certain trade-off in mind between pairs of measurable attributes when developing the MCDM value function, normalizing each attribute with individual value functions, and selecting appropriate weights to assign to those value functions, can easily yield *implicit* pairwise trade-offs among attributes that generate different rank orderings than the *explicit* pairwise trade-offs determined in EEoA.[25] Which decision support model best elicits public procurement officials' (decision-makers') preferences remains an important empirical question and warrants further research. We now turn to another important contribution of EEoA: the importance of modeling the supply side; specifically, accounting for vendor responses to anticipated future funding.

Accounting for Vendor Responses to Anticipated Future Funding

Traditionally, MCDM models focus on the demand side of a public procurement and treat supply side vendor decisions as exogenous. This section demonstrates the potential value of explicitly accounting for vendor responses to anticipated future funding (affordability or procurement official's budget constraints).

Since each vendor's expansion path represents their optimal attribute bundle bid proposals for any given budget (see Figures 2, 3, and 4), these expansion paths can easily be converted, through the buyer's utility function (1'), into cost-effectiveness (or Budget-Utility) ***functions*** for each vendor. For example, substituting each vendor's optimal attribute bundle (6a'') & (6b'') into Equation 1' for any specific scenario yields two points in cost-effectiveness space that represent the utility of each vendor's bid proposal for the per unit funding/budget, **B**: $(\mathrm{U}_1^*, \mathbf{B})$ and $(\mathrm{U}_2^*, \mathbf{B})$. Different budgets represented along the expansion paths generate different utility. For example, the cost-effectiveness/utility relationships illustrated in Figure 6 reflect the value to the government of each vendor's offers at different funding levels.

There is an important contrast between the endogenously derived EEoA cost-effectiveness ***functions*** for each vendor, and the exogenous cost-effectiveness ***points*** generally used to represent vendor offers in MCDM.[26] This becomes especially apparent when vendor costs depend on anticipated future funding. For instance, with bigger budgets, a vendor's costs to provide more of a particular attribute (say computer memory) might enjoy increasing returns to scale because of quantity discounts, learning curves, the ability to employ just-in-time inventory techniques, or the possibility of adopting other process improvements that reduce a vendor's costs of incorporating/producing a desired attribute.

Consider the case illustrated in Figure 5, where vendor 1's costs of producing attribute 1 are assumed to depend on the funding level or anticipated per unit budget, **B** (i.e., $c_{11}(\boldsymbol{B})$). For ease of exposition, suppose both vendors $j = 1, 2$ have identical, constant production technologies (i.e., $y_{1j} = y_1 \text{ and } y_{2j} = y_2$), and constant returns to scale $y_1 + y_2 = 1$. The difference between them is in their individual attribute costs. As before, let $c_{12}(\boldsymbol{B}) = c_{12}; c_{22}(\boldsymbol{B}) = c_{22}; \text{ and } c_{21}(\boldsymbol{B}) = c_{21}$, but now suppose vendor 1's costs for attribute 1 depends on the budget. For example assume the following relationship: $c_{11}(B) = c_{11} - k\boldsymbol{B} > 0$. Also let $\boldsymbol{B} < c_{11}/k$, $c_{11} > c_{12}$, and $k \in [0,1)$.[27] In this case (from (6a'') and (6b'')), each vendors' optimal attribute bundle proposals for a unit funding/budget level $\boldsymbol{B}$ is given by:

---

[25] Note: Linear normalization combined with careful swing weighting in MCDM could recover similar trade-offs to those explicitly revealed in EEOA (see Equation 9), resulting in an identical rank ordering of competing vendors. (An example is available upon request.)

[26] For an example of the latter, see the U.S. Defense Acquisition Guidebook, which states: "Cost-effectiveness comparisons in theory would be best if the analysis structured the alternatives so that all the alternatives have equal effectiveness (the best alternative is the one with lowest cost) or equal cost (the best alternative is the one with the greatest effectiveness). Either case would be preferred; however, in actual practice, in many cases the ideal of equal effectiveness or equal cost alternatives is difficult or impossible to achieve due to the complexity of AoA [Analysis of Alternatives] issues. *A common method for dealing with such situations is to provide a scatter plot [of points representing competing vendor proposals] of effectiveness versus cost*" (emphasis added; DoD, n.d., ch. 2–2.3., 2.7).

[27] These simple assumptions help illustrate our point. A model with quadratic costs could add another dimension (a "knee of the curve," i.e., monotonic increasing with a single inflection point) to the cost-effectiveness function, which could offer an interesting extension of the model.

$$(14a)\ a_{11}^{*} = [y/c_{11}(\boldsymbol{B})]\ \boldsymbol{B} = [y/(c_{11} - k\boldsymbol{B})]\boldsymbol{B},$$

$$(14b)\ a_{21}^{*} = [(1-y)/c_{21}]\boldsymbol{B}, \text{ and}$$

$$(15a)\ a_{12}^{*} = [y/c_{12}]\boldsymbol{B},$$

$$(15b)\ a_{22}^{*} = [(1-y)/c_{22}]\boldsymbol{B}.$$

Figure 5 illustrates each vendor's optimal attribute bundle bid proposals (given by Equations 14a and 14b and Equations 15a and 15b) for a specific budget, **B** (i.e., points $A_1: (a_{11}^{*}, a_{21}^{*})$ and $A_2: (a_{12}^{*}, a_{22}^{*})$).

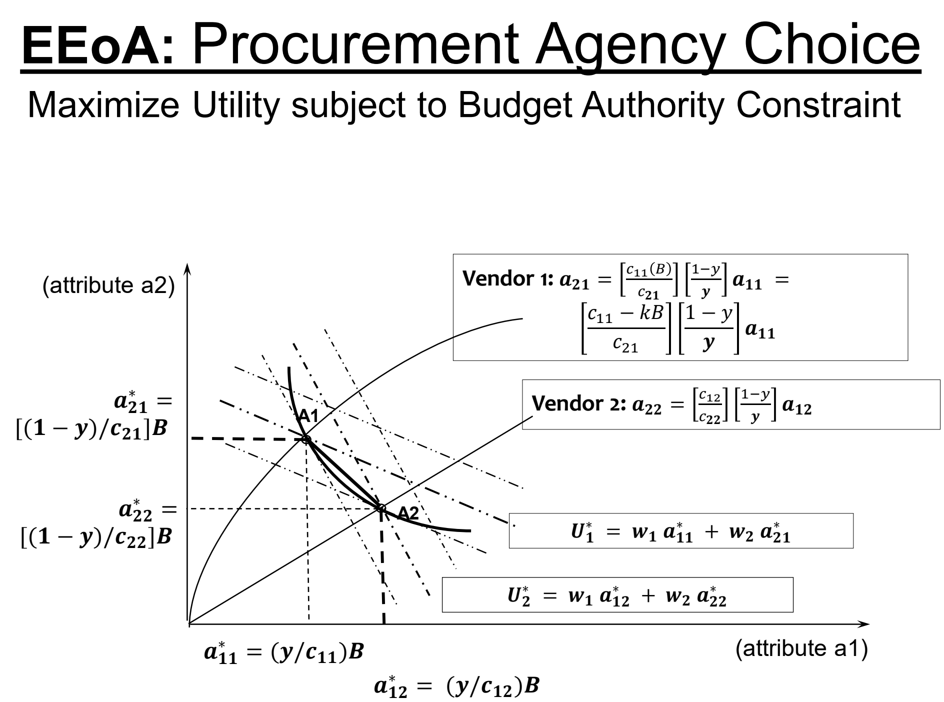


FIGURE 5. Vendor Selection When Vendor 1's Attribute Costs Depend on Budget

The expansion path for vendor 2 is again linear, with the same positive, constant slope for any budget (i.e., identical to (7b')). However, since vendor 1's attribute costs now depend on the anticipated per unit funding/budget, B, vendor 1's expansion path is nonlinear, increasing at a decreasing rate as illustrated in Figure 5 and given by:[28]

$$(16)\ a_{21} = [c_{11}(B)/c_{21}][(1-y)/y]\ a_{11} = [(c_{11} - kB)/c_{21}][(1-y)/y]\ a_{11},$$

where the slope (first derivative) is given by:

$$(16')\ \partial a_{21}/\partial a_{11} = [c_{11}(B)/c_{21}][(1-y)/y] = [(c_{11} - kB)/c_{21}][(1-y)/y] > 0,$$

[28] The illustration of the two expansion paths assumes that throughout the relevant range of budgets (funding levels), $(c_{11}(B)/c_{21}) > (c_{12}/c_{22})$.

and change in slope with a change in the budget (second derivative) given by:

$$(16'')\ \partial(\partial a_{21}/\partial a_{11})/\partial B = [c_{11}'(B)/c_{21}][(1-y)/y] < 0.$$

Substituting vendor 1 and 2's optimal attribute bundle offers Equations 14a and 14b and Equations 15a and 15b into the procurement official's (buyer's) utility function for any given scenario in Equation 8' yields:[29]

$$(17)\ U_1^* = w_1 a_{11}^* + w_1 a_{21}^* = w_1\,[y/c_{11}(B)]\,B + w_2\,[(1-y)/c_{21}]\,B$$

$$(18)\ U_2^* = w_1 a_{12}^* + w_1 a_{22}^* = w_1\,[y/c_{12}]\,B + w_2\,[(1-y)/c_{22}]B.$$

Equations 17 and 18 represent functions that can be plotted in cost-effectiveness (Budget-Utility) space over a relevant range of funding scenarios (see Figure 6). In this case, assuming identical, constant costs for attribute 2 (i.e., $c_{21} = c_{22} = c_2$), from Equations 17 and 18,

$$(19)\ U_1^* \gtreqless U_2^* \text{ as } c_{12} \gtreqless c_{11}(\boldsymbol{B}) = c_{11} - k\boldsymbol{B} \text{ or as } \boldsymbol{B} \gtreqless (c_{11} - c_{12})/k = B'.$$

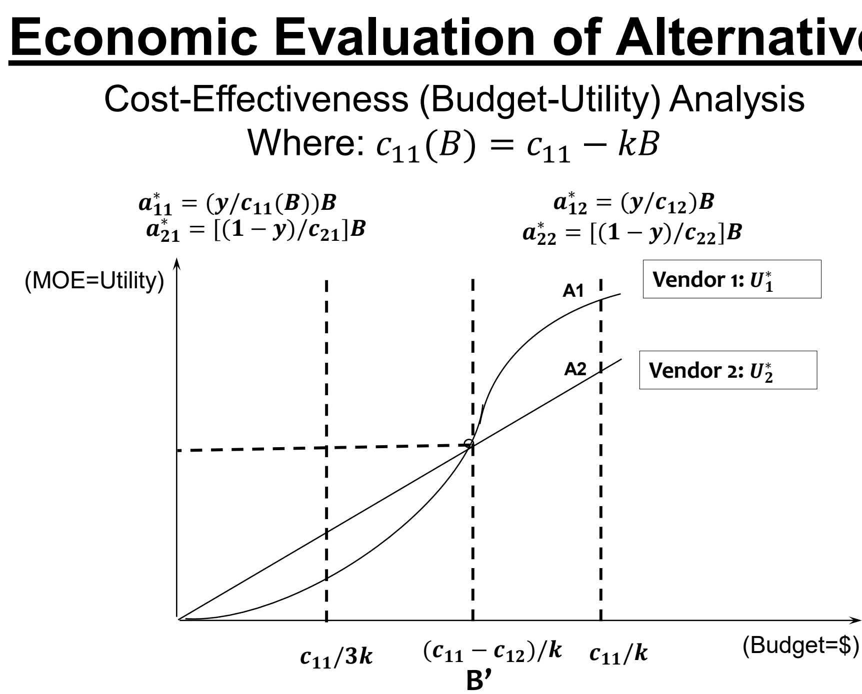


FIGURE 6. Vendor Selection in Cost-Effectiveness (Budget-Utility) Space

What is revealed in Figure 6 is an *optimal rank reversal*. The relation given by Equation 19 indicates it is optimal for the buyer to switch vendors at B'. For any unit funding/budgets **B**>B', vendor 1 is ranked higher than vendor 2. The two are ranked the same for the budget, **B**=B', and for budgets **B**<B', vendor 2 is ranked higher than 1. As expected, evaluating the slopes of the two vendors' cost-effectiveness functions at the switch point, B'=$(c_{11} - c_{12})/k$, yields:

$$(20)\ \partial U_1^*/\partial \boldsymbol{B} > \partial U_2^*/\partial \boldsymbol{B} \text{ or } (c_{11}(\boldsymbol{B}) - c_{11}'(\boldsymbol{B})\boldsymbol{B})/c_{11}(\boldsymbol{B})^2 > 1/c_{12} \text{ since } c_{11} > c_{12}.$$

[29] For a specific unit funding level **B**, this represents two optima that can be compared that represent the maximum utility a buyer can obtain from each vendor. This is illustrated in Figure 4 as the highest indifference curve attainable given the corresponding point on the attribute production possibility frontier.

> This highlights the importance of modeling the supply side. Specifically, this example emphasizes the importance for public procurement officials to obtain realistic budget forecasts for government programs, and to offer those as guidance to vendors. As two pioneers in defense economics Hitch and McKean (1967) wisely counseled: “As a starter ... several budget sizes can be assumed. If the same [vendor] is preferred for all … budgets, that system is dominant. If the same [vendor] is not dominant, use of several … budgets is nevertheless an essential step, because it provides vital information to the decision maker.”

Instead of plotting procurement alternatives (vendor bid proposals) as single points in cost-effectiveness (budget-value) space, EEoA encourages procurement officials in fiscally constrained environments to solicit bids over a range of possible budget scenarios.[30]

From a practical standpoint, the biggest limitation of the EEoA approach is that as the number of attributes (n) under consideration expands, it is increasingly burdensome to generate required pairwise comparisons. For example, assuming each alternative (vendor proposal) includes a set of n attributes, applying EEoA requires $\frac{n(n-1)}{2}$ pairwise comparisons to fully flesh out the decision-maker’s preferences. Interestingly however, EEoA could be applied in combination with MCDM as a consistency check for important attributes. That is to say, if $\partial a_2/\partial a_1 = -(w_{1s}/w_{2s}) = -X_s$ is the explicitly determined trade-off that a public procurement official is comfortable with in a particular scenario, given specific ranges for each attribute, then weights developed in MCDM should reflect this relative preference (trade-off).[31] The test simply involves application of Equation 11.

In other words, procurement officials can generate pairwise comparisons for the most critical attributes as a consistency check. For example, when comparing options for AEOLATV procurement, it may be the case that *Top Speed* and *Range* are the most important attributes to consider among the dozens or even hundreds of other attributes. After carefully applying traditional MCDM techniques to develop measures of effectiveness for each AEOLATV alternative, use EEoA’s explicit trade-off determination to ensure that the decision-maker is indeed willing to trade X amount of *Top Speed* for Y amount of *Range*, and vice versa, in the specified attribute ranges. If the explicit trade-off determination is one that the decision-maker is uncomfortable with, it is crucial to revisit the value functions and weighting schemes used to generate the measures of effectiveness for each option. While this can be a time-consuming process, it ensures that the best alternative is chosen for large procurements to satisfy the mission.

Conclusion and Avenues for Future Research

This paper offers an economic model to assist public procurement officials to rank competing vendors when benefits cannot be monetized. The problem of ranking public investment alternatives when benefits cannot be monetized has spawned an extensive literature

---

[30] In this case, the standard technique of eliminating “dominated alternatives” could lead to sub-optimal decisions. For example, see Melese (2015) or the specific example of the EEoA model developed in Simon & Melese (2011).

[31] If the extra burden of normalization and swing weighting required in MCDM causes a decision-maker to “misevaluate” their trade-off preferences, then EEoA offers an alternative framework/perspective that can help to realign their weighting. Note that in theory a rational decision-maker with perfect information and infinite computational capability would never need to do this. Since in practice it is difficult to define “correct” weighting within scenarios, contrasting the development of weights in MCDM and EEoA is an empirical question worth investigating.

that underpins widely applied decision tools. The bulk of the literature, and most government-mandated decision tools, focus on the demand side of a public procurement. The EEoA extends the analysis to the supply-side.

Introducing the supply side offers multiple avenues for future research. Notably, it provides fertile ground to apply both auction and game theory literatures. An interesting extension would be to leverage auction theory and introduce strategic shading of bids by vendors. Another is to consider the risk of collusion among vendors, or allow some vendors to enjoy economies of scale (or to make engineering production function parameters a function of the budget). Whereas EEoA models vendors as proposing bundles of characteristics to win a prize (i.e., funding), alternative optimization assumptions and strategic behaviors could be assumed.

A rich opportunity also exists for both experimental and qualitative research to significantly improve public procurement. An important empirical question is whether procurement officials and managers would have an easier time using EEoA or MCDM (or some combination)? Consistency tests could be conducted in experimental settings to explore when the two techniques converge (offer identical vendor rankings), and when (and why) they diverge?

In conclusion, the EEoA captures both demand side—government procurement official decisions—and supply side—vendor optimization decisions. A unique feature of EEoA is to model vendor decisions in response to government funding projections. Given a parsimonious set of continuously differentiable evaluation criteria, EEoA provides a new tool to rank vendors. In other cases, it offers a valuable consistency check for MCDM models to guide government supplier decisions.

References

Blakeman, T. C., & Branson, R. D. (2013). Evaluation of 4 new generation portable ventilators. *Respiratory Care*, *58*(2), 264–272.

Cavin, E. (1995). How should modeling and simulation be used to support defense acquisition? *Defence and Peace Economics, 6*(2), 77–88.

Charnes, A., Cooper, W. W., & Schinnar, A. P. (1986). An extended Cobb-Douglas form for use in production economics (Research report CCS 555). *Center for Cybernetic Studies,* 1–31.

Charnes, A., Cooper, W. W., & Schinnar, A. P. (1991). An extended Cobb-Douglas form for use in production economics. In T. K. Kaul & J. K. Sengupta (Eds.), *Economic models, estimation and socioeconomic systems* (pp. 99–124). Elsevier Science Publishers.

Chenery, H. B. (1949). Engineering production functions. *The Quarterly Journal of Economics, 63*(4), 507–531.

Clemen, R. T. (1996.) *Making hard decisions: An introduction to decision analysis* (2nd ed.). Duxbury Press.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *American Economic Review*, *8*(1), 139–165.

Davis, D. (2011). Designing a viable prediction market to forecast defense acquisition cost and schedule outcomes. *Defence and Peace Economics*, *22*(3), 351–366.

DoD. (n.d.). *Defense acquisition guidebook.*

Douglas, P. H. (1976). The Cobb-Douglas production function once again: Its history, its testing, and some new empirical values. *Journal of Political Economy*, *84*(5), 903–916.

Ewing, P. L., Jr., Tarantino, W., & Parnell, G. S. (2006). Use of decision analysis in the army base realignment and closure (BRAC) 2005 military value analysis. *Decision Analysis*, *3*(1), 33–49.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2019). https://www.acquisition.gov/sites/default/files/current/far/pdf/FAR.pdf

Hildebrandt, G. (1999). The military production function. *Defence and Peace Economics*, *10*(3), 247–272.

Hitch, C., & McKean, R. (1967). *The economics of defense in the nuclear age*. Harvard University Press.

Horowitz, S., Harmon, B., & Levine, D. (2016). Inflation adjustments for defense acquisition. *Defence and Peace Economics*, *27*(2), 231–257.

Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value trade-offs*. John Wiley and Sons.

Kirkpatrick, D. (1995). The rising unit cost of defence equipment—The reasons and the results. *Defence and Peace Economics*, *6*(4), 263–288.

Kirkwood, C. (1997). *Strategic decision making*. Duxbury Press.

Kirkwood, C. W., & Sarin., R. K. (1980). Preference conditions for multi-attribute value functions. *Operations Research*, *28*(1), 225–232.

Kurz, M., & Manne, A. (1963). Engineering estimates of capital–labor substitution in metal machining. *The American Economic Review*, *53*(4), 662–681.

Lancaster, K. J. (1966a). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157.

Lancaster, K. J. (1966b). Change and innovation in the technology of consumption. *American Economic Review*, *56*(2), 14–23.

Lancaster, K. J. (1971). *Consumer demand: A new approach*. Columbia University Press.

L'Her, E., Roy, A., & Marjanovic, N. (2014). Bench-test comparison of 26 emergency and transport ventilators. *Crit Care*, *18*, 506.

Marsden, J., Pingryu, D., & Whinston, A. (1972). *Production function theory and the optimal design of waste treatment facilities: Vol. 4. Applied economics.*

Melese, F. (2015). The economic evaluation of alternatives. In F. Melese, A. Richter, & B. Solomon, *Military cost–benefit analysis: Theory & practice* (Ch. 4). Routledge.

Melese, F. (2018). Cost–benefit analysis of bid protests: A representative bidder model. *Defence & Peace Economics.*

Nicholson, W., & Snyder, C. M. (2017). *Microeconomic theory*. South-Western Press.

OECD. (2016). *OECD public procurement* [Dataset]. https://qdd.oecd.org/subject.aspx?Subject=GOV_PUBPRO_2016

Parkes, D., & Kalagnanam, J. (2005). Models for iterative multi-attribute procurement auctions. *Management Science*, *51*(3), 435–451.

Ratchford, B. (1979). Operationalizing economic models of demand for product characteristics. *Journal of Consumer Research*, *6*(1), 76–85.

Schwartz, M., Sargent, J., & Mann, C. (2018). *Defense acquisitions: How and where DOD spends its contracting dollars*. Congressional Research Service. https://fas.org/sgp/crs/natsec/R44010.pdf

Simon, J., & Melese, F. (2011). A multiattribute sealed-bid procurement auction with multiple budgets for government vendor selection. *Decision Analysis*, *8*(3), 170–179.

Smith, V. K. (1986). Another view of the state of engineering production functions. *Economica*, *53*(212), 529–532.

Wibe, S. (1984). Engineering production functions: A survey. *Economica*, *51*(204), 401–411.

Wise, R., & Morrison, D. (2000). Beyond the exchange—The future of B2B. *Harvard Business Review*, *78*(6), 86–96.

PANEL 15. LOGISTICS IN CONTESTED ENVIRONMENT

Wednesday, May 12, 2021	
1:45 p.m. – 3:00 p.m.	**Chair: Mr. David R. Kless,** Executive Director, Operations, DLA Logistics Operations ***Framework for Augmenting Current Fleet with Commercially Available Assets for Logistics Support in Contested Environment*** Aruna Apte, Naval Postgraduate School Ken Doerr, Naval Postgraduate School Uday Apte, Naval Postgraduate School ***Capability-Based Planning in Humanitarian Operations: A Hybrid Optimization and Simulation Framework for Strategic Acquisition in the Armed Forces*** Gonzalo Barbeito, Universitaet der Bundeswehr Stefan Pickl, Universitaet der Bundeswehr Max Krueger, Hochschule Furtwangen Dieter Budde, Universitaet der Bundeswehr ***Understanding the Challenges of Providing PPE in the U.S. during COVID-19*** Shaghayegh Arangdad, North Carolina State University A. Blanton Godfrey, North Carolina State University

Mr. David R. Kless—serves as the Executive Director, Operations (J31), in DLA Logistics Operations (J3). He assumed this responsibility upon his appointment to the Senior Executive Service in 2018. In this role, Kless is responsible for coordinating and synchronizing day-to-day current operations of DLA's worldwide logistics, acquisition and technical support services in support of combatant commands (CCMD); military services; federal, state, and civil agencies; and foreign governments. J31 is comprised of the DLA Joint Logistics Operations Center (JLOC) Division, Military Services Support (MSS) Division, DLA-TRANSCOM Support (DLA-T) Division, Whole of Government Support Division, Mission Assurance Support Team, and the Portfolio Programs Support Team. J31 also identifies customer requirements from contingency planning and readiness posture of the combatant commands and the services, and manages the overseas customer interface network of DLA Europe & Africa, DLA Pacific, and DLA-CENTCOM & SOCOM.

Kless received his Bachelor of Science degree in Political Science from the U.S. Naval Academy. He then graduated with a Master of Science degree in Operations Research from the Navy Postgraduate School. He later obtained a Master of Arts degree in National Security and Strategic Studies from the U.S. Naval War College. Kless currently holds DAWIA Acquisition Workforce Improvement Act (DAWIA) Level II certification in Acquisition Logistics.

After graduating from the U.S. Naval Academy, Kless served in the Navy performing multiple logistics and operation assignments within the Navy logistics and acquisition community. After numerous positions of increasing responsibility with a distinguished career as a Naval officer, he retired as a Navy captain in 2011.

He then joined the Defense Logistics Agency as the division chief for Whole of Government Support in DLA Logistics Operations. He later served as DLA deputy chief of staff, Office of the Director.

Over his career, Kless has received many awards, including a Meritorious Civilian Service Award. His military awards include the Defense Superior Service Medal; Meritorious Service Medal, two awards; Best Retail Sales and Service Award; Navy and Marine Association Peer Selected Leadership Award; Navy Battle Efficiency Award; Admiral Stanley R. Arthur Award for Logistics Excellence; and Navy & Marine Corps Commendation Medal, four awards.

Framework for Augmenting Current Fleet with Commercially Available Assets for Logistics Support in Contested Environment[1]

Dr. Aruna Apte—is a tenured Professor of Operations and Logistics Management in the Graduate School of Defense Management, Naval Postgraduate School (NPS), Monterey, CA. Apte received her PhD in operations research from Southern Methodist University (SMU) in Dallas. She joined NPS in 2005 as tenure-track Assistant Professor. [auapte@nps.edu]

Dr. Uday Apte—is Distinguished Professor and Associate Dean of Research and Development at the Graduate School of Business and Public Policy, Naval Postgraduate School, Monterey, CA. Before joining NPS, Dr. Apte taught at the Wharton School, University of Pennsylvania, Philadelphia, and at the Cox School of Business, Southern Methodist University, Dallas. He is experienced in teaching a range of operations management and management science courses in the Executive and Full-time MBA programs. Prior to his career in the academia, Dr. Apte worked for over 10 years in managing operations and information systems in the financial services and utility industries. Since then he has consulted with several major US corporations and international organizations. [umapte@nps.edu]

Dr. Ken Doerr—B.S. (Quantitative Business Analysis) Indiana University Bloomington, 1984; Ph.D. (Management Science) University of Washington Seattle, 1994. Appointed to the faculty at the Naval Postgraduate School in 2001. Other academic posts include faculty appointments with the University of Miami in Coral Gables and Santa Clara University, research fellowships at the University of Waterloo and the University of Cincinnati, and a sabbatical year with the Industrial Engineering faculty at Tel Aviv University. His research has focused on logistics risk, the cost-benefit analysis of technology investments, and behavioral operations management. His papers have appeared in several leading journals, including Management Science, The Academy of Management Review, IIE Transactions and The Journal of Applied Psychology. Prior to joining academia, Dr. Doerr was employed for several years as a Systems and Management Science professional with Shell Oil, Monsanto (now part of Bayer), and Peoplesoft (now part of Oracle). [khdoerr@nps.edu]

Abstract

China believes logistics in the contested environment is an Achilles's heel for the U.S. Navy. It is therefore critical that we explore ways to develop capabilities to replenish potential combating forces through Next Generation Logistics Ships (NGLSs).

The objective in this research is to study and analyze options for rearming, refueling, and resupplying in the contested and distributed environment. The framework created is flexible in terms of the scenarios.

Feedback from subject matter experts (SMEs) helped us gain insight into the complexity of the problem and its vast scope. We developed mathematical models based on the scenarios approved by the sponsor. The sponsor did not wish us to model an objective of minimizing costs or the number of ships required to deliver commodities within a certain deadline or under a certain schedule. Measuring the number of deliveries required in a scenario supplied by SMEs allowed us to determine a mix of NGLS vessels without cost or deadline data. We would like to point out that the number of deliveries needed by vessels of each type, as described in the report, can be interpreted in many ways, in terms of the number of ships required. The summary of our results and analysis suggests certain recommendations.

---

[1] Please note that this is an abridged version of the original technical report. The full report is available as NPS-LM-20-155 at ARP.

Introduction and Background

The U.S. government came out publicly with an explicit statement that the so-called "nine-dash line," which the People's Republic of China (PRC) asserts delineates its claims in the South China Sea, is contrary to international law (Figure 1). China claims that the "nine-dash line" encircles as much as 90% of the contested waters. The line runs as far as 2,000 km from the Chinese mainland to within a few hundred kilometers of the Philippines, Malaysia, and Vietnam. The PRC maintains that it owns any land or features contained within the line, which confers vaguely defined "historical maritime rights" (Liu, 2016). It encircles the area where China demands economic rights. Another interpretation is that the line marks the islands and reefs China wants to control rather than the waters inside its boundaries. The PRC has long favored a strategy of ambiguity. It does not openly go against international law but prefers to leave space for its more ambitious claims (Apte et al., 2020).

Figure 1. The Nine-Dash Line and Surrounding Countries

China defiantly lands planes on artificial islands in the South China Sea while U.S. warships patrol in protest (Figure 2). The string of "unsinkable aircraft carrier" islands is an imminent threat to U.S. allies in Southeast Asia. This, plausibly, is where a war with China will likely be fought. When thinking in a geostrategic sense about China, the island-chain formulation is helpful. Since the 1950s, U.S. planners have described a first island chain, running from the Japanese islands through the Philippines and down to the tip of Southeast Asia. Dominating inside that line has been the goal of China's recent buildup in naval and missile capabilities. But U.S. officials warn that Chinese strategists are becoming more ambitious, set on gaining influence up to the second island chain—running from Japan through the Micronesian islands to the tip of Indonesia (Figure 3).

As with its initial forays into the South China Sea, China is using so-called scientific missions and hydrographic surveying ships as the tip of the spear. Japan and Singapore essentially serve as anchors at the north and south ends of the island chains. These two nations have been integrating their defense capabilities with the United States through training, exercises, and arms purchases. They are exploring better relations with India as the Pacific and Indian Oceans are increasingly viewed as a single strategic entity. This nascent alliance is a crucial element in the U.S. strategy for the region.

China believes logistics in the contested environment is an Achilles's heel for the U.S. Navy (USN). It is therefore critical that we explore ways to develop capabilities to replenish potential allied combatant forces in the Pacific through Next Generation Logistics Ships (NGLSs).

Figure 2. Chinese Dredging Vessels in the Waters Around Mischief Reef in the Disputed Spratly Islands in the South China Sea

Figure 3. First and Second Island Chains

In this research, we offer a framework using mathematical models to refuel, rearm, and resupply for future logistics in such contested environments to support the potential combat operations of the USN. The resupply mission scenarios developed for this research are based on actual data supplied by subject matter experts (SMEs), but those data are disguised by the authors. At the foundation of the framework are the following research questions:

1. Is the current fleet of vessels adequate to carry out the mission?
2. Are there new vessels that can be modified or produced for the purpose of better sustainment through the three vectors of refuel, rearm, and resupply?
3. If so, what type of vessels, and how many of each kind, should be acquired?

In order to answer these questions, we first look at answers from existing literature on logistics, perhaps derived from a different environment. The capabilities of the new vessels mentioned in question 2 are based on top-level requirements supplied to the authors by SMEs. The methodology considers the supply chain from controlled zone to contested zone, utilizing only those new vessels. We develop and use different scenarios and methodologies to arrive at answers based on different objectives. We develop a framework for augmenting the current fleet with NGLSs for support in contested logistics. The objective is to study and analyze options for rearming, refueling, and resupplying allied combatants in the contested and distributed environment.

Motivation for the New Vessels

To optimize its future fleet logistics platforms, the USN and United States Marine Corps (USMC) are exploring the concept of a common hull, multi-mission auxiliary ship design. The Commandant of the Marine Corps, General David Berger, explained his perspective on amphibious forces, including the need for more small ships, at an Amphibious Warship Industrial Base Coalition event:

> I think our amphibious fleet has great capability. It is not enough for 2030. It is not enough for 2025. We need the big decks, absolutely. We need the LPD-17, that is the mothership, the quarterback and the middle. But we need a light amphibious force ship, a lot of them that we don't have today. (Abott, 2020, para. 8)

Abott (2020) continues that the Navy said this non-acquisition program will be one "that designs, develops, and tests the Integrated Naval Force Structure Assessment, to evaluate next generation medium platform solutions for logistics mission requirements in support of Distributed Maritime Operations (DMO) and Littoral Operations in Contested Environment (LOCE)" (para. 11).

The USN and USMC announced that they will seek a medium amphibious ship that can support the kind of dispersed, agile, constantly relocating force described in the LOCE and Expeditionary Advanced Base Operations (EABO) concepts the Marine Corps has written, as well as the overarching DMO from the Navy (Eckstein, 2020). Marine Corps planners described the features they need on this medium amphibious ship. They not only wanted a ship that could move Marines around with some range, but they also wanted the ship to be able to beach itself, like a landing craft, to help offload gear and vehicles as needed. Presently, there is a new focus on the stern landing vessel designed by Australian company Sea Transport, which could serve as the new inspiration for the medium amphibious vessel as requirements development, EABO wargaming, and simulations take place.

Future Surface Combatant Force is developing alternate surface ship force structure concepts and evaluating their cost and effectiveness, performing force-wide warfighting and mission effectiveness studies, identifying capabilities and characteristics needed to meet future threats, and developing a Technology Investment Strategy to help guide investments for an effective future fighting force. Our research supports this concept.

Some of the vessels, NGLSs, will be commercial ship designs tailored to fit the top-level requirements that can conduct logistics missions in a contested environment. Through these new NGLS vessels, the USN will enable refueling, rearming, and resupply of naval assets, afloat and ashore, in support of LOCE and EABO (Katz, 2020). In a memorandum signed by the Chief of Naval Operations (CNO) and Commandant of the USMC (O'Rourke, 2020), Force Structure Assessment (FSA) morphed into Integrated Naval FSA (INFSA), where *Naval* refers to Navy *and* Marine Corps. Acting Secretary of the Navy Modly announced that

> there are certain ship classes that don't even exist right now that we're looking at that will be added into that mix, but the broad message is, it's going to be a bigger fleet, it's going to be a more distributed fleet, it's going to be a more agile fleet. And we need to figure out what that path is and understand our topline limitations. (O'Rourke, 2020)

He added that the service is also considering new amphibious ships, as well as new kinds of supply ships and "lightly manned" ships that are "more like missile magazines that would accompany surface action groups" (O'Rourke, 2020).

General David H. Berger, the commandant of the Marine Corps, states,

> We must also explore new options, such as inter-theater connectors and commercially available ships and craft that are smaller and less expensive, thereby increasing the affordability and allowing acquisition at a greater quantity. We recognize that we must distribute our forces ashore given the growth of adversary precision strike capabilities, so it would be illogical to continue to concentrate our forces on a few large ships. The adversary will quickly recognize that striking while concentrated (aboard ship) is the preferred option. We need to change this calculus with a new fleet design of smaller, more lethal, and more risk-worthy platforms. (O'Rourke, 2020)

We now offer a summary of certain requirements for these vessels that lead to their capabilities since we base our assumptions underlying the developed models on their top-level requirements (TLRs). TLRs are design specifications of performance requirements for future ships.

Description of the New Vessels: Next Generation Logistics Ships (NGLS)

These vessels do not necessarily exist yet but have TLR thresholds defined for each performance dimension.

Platform Supply Vessel (PSV)

In summary, the vessel should have a sustained speed of about 11–12 knots. The range of travel for the platform supply vessel (PSV) is about 3,500 nm. Its fuel capacity needs to be about 20,000 bbl. Ammunition and cargo capacity needs to be adequate for replenishing cargo, ammunition, and fuel at sea from Combat Logistics Force (CLF), specifically, about 800–900 short tons and deck area being about 10,000 sq ft. A major capability planned for the PSV is to deliver about 5,000 bbl of fuel in under about 2 hours at sea. In addition, it needs to be able to deliver 15 loads/hour of ammunition and/or cargo in parallel with refueling. This vessel will be unmanned throughout the operational cycle with organic support only when necessary. Autonomously executing the mission is a required capability of PSV.

Fast Supply Vessel (FSV)

Much smaller than the PSV but much faster, the sustained speed of a fast supply vessel (FSV) is 23 knots, and the range of travel is about 800–1000 nm. The fuel-storage capacity is required to be about 1,000 bbl. Deck area for ammunition and dry cargo is about 2,500 sq ft. A major capability planned for the FSV is to replenish the PSV in littorals. It also needs to do water transfers with hose reel with roll-on/roll-off capabilities. On shore, the FSV needs to be able to refuel at a minimum of about 500 gallons/minute with a 2,000-ft hose reel. It also needs to be capable of conducting missions for 2–3 days without replenishment. Finally, it needs to be able to transfer cargo to a pier or ashore.

Light Amphibious Warships (LAW)

These lighter ships will help the Navy and Marine Corps meet new challenges, including sea-control-and-denial operations. The light amphibious warships (LAW) will serve as maneuver and sustainment vessels to confront the changing character of warfare. The LAW will have beachability and the ability to maneuver shore to shore. It will also be able to provide transfer of fuel and cargo from T-ships on beaches and ports (developed and undeveloped) to forces within contested environments. The idea is to have a risk-worthy vessel (defensible enough that risks are not excessive or cheap enough that we can afford to lose it) with priority for personnel survivability. Being an amphibious vessel, the LAW should deal with 1:40 to 1:100 beach gradients. The loaded LAW should have a speed of about 18 knots. Thus, its speed is between the speeds of the PSV and FSV. Its minimum operating range is to be about 5,000 nm. It is to be capable of transferring 500 gallons/minute of fuel at sea or to shore. The LAW is to be capable of conducting up to 11-day missions without replenishment. It is expected to receive, store, and transport up to 90,000 gallons of fuel in port as well as at sea. This fuel will be transferred at the rate of 150 gallons/minute in port as well as at sea. It can have four fueling stations around its cargo deck for filling trucks and vehicles. It has a crane with maximum outreach of about 14 T. It has a cargo area of about 10,000 sq ft and deck loading capacity of about 500 lb/sq ft.

Our methodology derives a mathematical model based on capabilities for resource optimization for humanitarian missions (Apte & Yoho, 2018). However, we bear in mind the distinction between a contested environment and an uncontested environment, since the PSV and the FSV cannot defend themselves, but the LAW can. Therefore, if the NGLS ships do not encounter combat, the missions are similar. If they do encounter combat, the PSV and FSV will simply be lost, while the LAW will face an attrition rate. The attrition rates of these vessels have been estimated elsewhere (Dougherty et al., 2020). Since our results can simply be adjusted to account for those already estimated attrition rates, we do not model combat attrition in this study; we merely note that it is a factor which favors the LAW over the alternatives. In short, the humanitarian assistance and disaster relief (HADR) inference is relevant in the contested environment.

Methodology

We include the capacities of the vessels and offer an objective that minimizes the ***number*** of deliveries by the appropriate vessels on corresponding route. There exists a time constraint for total unload time to ships in the Weapon Engagement Zone (WEZ). To model this unload-time constraint, we modified capacities of the vessels per the time constraint in order to represent the time constraint of delivery. For example, if unload time was constrained to 1 hour, and an NGLS vessel could only unload 10 pallets in an hour or 5,000 BBL of fuel, we modified its cargo-capacity accordingly.

These transportation/transshipment models consider controlled and contested zones. An assumption of the scenarios we were given is that most of the supplying vessels and Combat Logistics Force (CLF) are in the controlled zone, so there is transfer of commodities in the contested zone from NGLSs to the SAG and transshipment nodes. The transshipment node provides supplies for the different Expeditionary Advance Bases (EABs) on the shore in the contested zone. We developed models and analyzed scenarios for the NGLS vessels. The modes of transportation in the models and scenarios are the PSV, FSV, and LAW. Each of these vessels have certain preferred routes and requirements for capacities, loading/unloading, and platforms. These translate into restrictions and constraints for the models.

We define a delivery as the carrying of commodities from a supply node to a demand node on the given route by a vessel designated to travel on that route. The models are executed using plausible but hypothetical numbers in order to maintain the unclassified nature of the report. The tables list the supply and demand at the nodes of the network. Though the results are based on hypothetical numbers, these numbers can be scaled up or down by using an appropriate multiplier.

In the process of developing these models, given that the fuel (F) is stored separately from ammunition and supplies (A-S), we separated the models for F and A-S. Fuel capacity is measured in barrels, whereas ammunition and supplies are in pallets. In the case of A-S, both potentially occupy the same square footage of the vessels. Therefore, we combined these two commodities (A-S) when we developed square footage constraints for the models. The models for both F and A-S are very similar except for the supply, demand, and capacities.

Scenario: Capability Restricted Transportation

We offer two scenarios and the corresponding models based on feedback from SMEs. Vessels allowed on the respective routes are shown in Figure 4-1. This scenario first looks at the entire network (Figure 4-1). We offer a different perspective by splitting the transshipment network into two separate transportation networks (Figure 4-2). In these scenarios, the model treats SAG (node 2) as one entity.

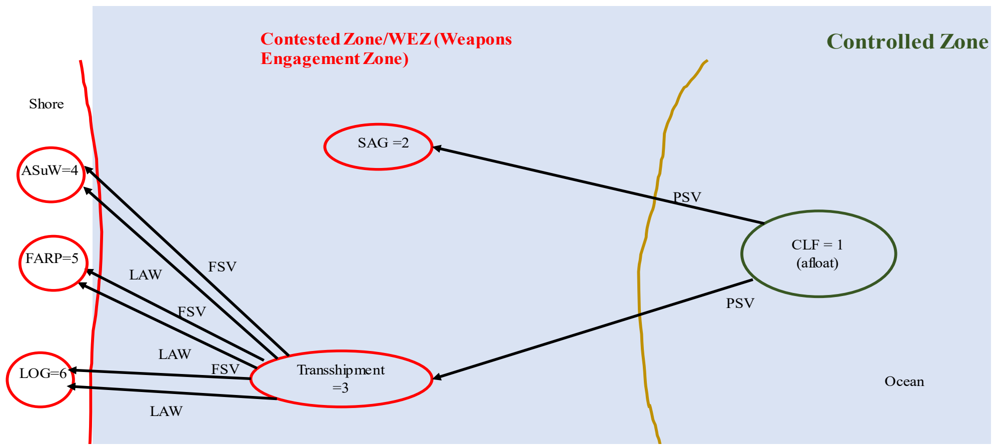


Figure 4-1. Scenario Based on Subject Matter Expert Feedback

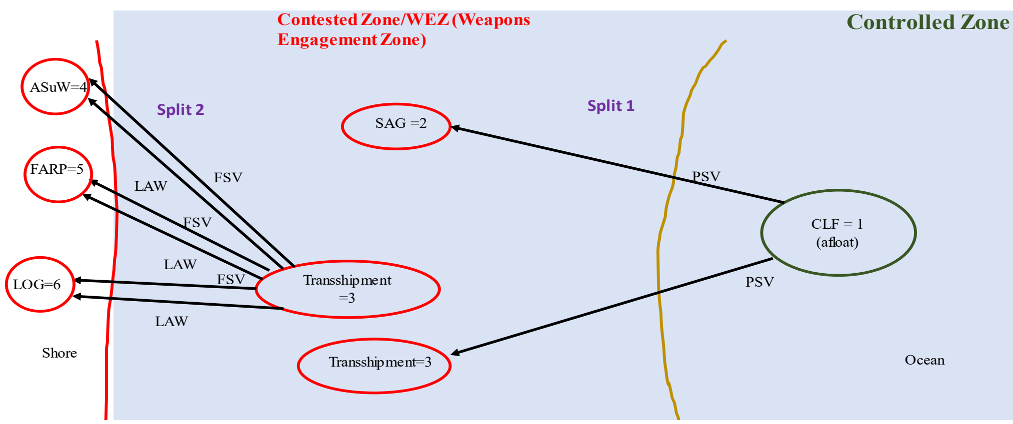


Figure 4-2. Scenario Based on SME Feedback With Split 1 and Split 2

As shown in Figure 4-2, Split 1 transports commodities from CLF to SAG and Transshipment, whereas Split 2 transports commodities from Transshipment to ASuW, FARP, and LOG. The advantage of splitting the entire/combined/transshipment network into two transportation networks is twofold. First, Split 1 focuses on the USN, whereas Split 2 focuses on the USMC. This helps in maintaining the needs of Marines ashore and Navy forces afloat. Second, transshipment of the commodities is assumed to be done sequentially, and though the two transportation networks have the same assumption, they can be executed in parallel, thus reducing total time. In the split model, of course, the supply from the transshipment node is set equal to the demand to the transshipment node. The corresponding model is given next.

Model (Split) for Capability Restricted Transportation based on Scenarios in Figures 4.1 and 4.2 for Fuel

Total supply *at node i for fuel =* S_{Fi},, ***Total demand*** *at node j for fuel =* D_{Fj}

Shared volume capacity *for fuel on vessel k enroute from node i to node j =* cF_{kij}

Modes of transportation, k*: PSV =1, FSV = 2, LAW = 3*

Split 1

Decision Variables:

X_{Fkij} *= flow of fuel from source i to node j on vessel k, i=1, j= 2, 3, k = 1*

Y_{kij} *= # of deliveries by vessels of type k from node i to j and l*

Objective Function: *Minimize Number of Deliveries*

$$\min(y_{112} + y_{113})$$

Constraints:

Supply at CLF = 1, $(x_{F112} + x_{F113}) \leq SF_1$

Demand at SAG = 2, $x_{F112} \geq DF_2$

at Transshipment = 3, $x_{F113} \geq DF_3$

Capacity Fuel Volume

$$(cF_{112})Y_{112} - X_{F112} \geq 0$$
$$(cF_{113})Y_{113} - X_{F113} \geq 0$$

Y_{kij}*'s integer and ≥ 0,* X_{kij}*'s ≥ 0*

Split 2

Decision Variables:

X_{Fkjl} *= flow of fuel from transshipment node j to sink l on vessel k, j=3, l=4, 5, 6, k = 2, 3*

Y_{kjl} *= # of deliveries by vessels of type k from node j to l*

Objective Function: *Minimize Number of Deliveries*

$$\min(y_{234} + y_{235} + y_{236} + y_{334} + y_{335} + y_{336})$$

Constraints:

Supply at Transshipment = 3, $x_{F234} + x_{F235} + x_{F236} + x_{F334} + x_{F335} + x_{F336} \leq SF_3$

Demand at ASuW = 4, $x_{F234} + x_{F334} \geq DF_4$

at FARP = 5, $x_{F235} + x_{F335} \geq DF_5$

at LOG = 6, $x_{F236} + x_{F336} \geq DF_6$

Capacity Fuel Volume

$$(cF_{234})Y_{234} - X_{F234} \geq 0$$
$$(cF_{235})Y_{235} - X_{F235} \geq 0$$
$$(cF_{236})Y_{236} - X_{F236} \geq 0$$
$$(cF_{334})Y_{334} - X_{F334} \geq 0$$
$$(cF_{335})Y_{335} - X_{F335} \geq 0$$
$$(cF_{336})Y_{336} - X_{F336} \geq 0$$

Y_{kjl}'s integer and ≥ 0, X_{kjl}'s ≥ 0

Results

We further evaluated the number of deliveries by the vessels by incorporating the restrictions on the vessels due to their capabilities. Top-level requirements for NGLSs informed us of the inability of certain vessels for transportation between certain nodes. These were incorporated in the structure of the scenarios and corresponding models. The supply and demand at the nodes and capacities of vessels in these scenarios are given in Table 1 for F and for Ammunition and Supplies. In Table 1, the capacity of vessels for ammunition and supplies is constrained from CLF to SAG by the length of time of 1 hour, the maximum time any ship in the SAG can be engaged for fueling. The assumption is that since the pallets are delivered at the rate of 60 pallets/hour, only 60 pallets can be delivered to SAG, though the true capacity of the PSV is 800 pallets.

Table 1. Supply, Demand, and Capacities: Fuel in BBL and Ammunition and Supplies in Pallets

	Fuel in BBL	Ammunition and Supplies in Pallets
Nodes	Supply/Demand	Supply/Demand
Supply at CLF 1	100000	100000
Supply at Trans 3	6500	750
Demand at SAG 2	22000	100
Demand at Trans 3	6500	750
Demand at ASuW 4	100	50
Demand at FARP 5	6300	350
Demand at LOG 6	100	350
Routes	Capacity	Capacity
PSV from CLF 1 to SAG 2	5500	60
PSV from CLF 1 to Trans 3	5500	800
FSV from Trans 3 to ASuW 4	1000	250
FSV from Trans 3 to FARP 5	1000	250
FSV from Trans 3 to LOG 6	1000	250
LAW from Trans 3 to ASuW 4	2200	1000
LAW from Trans 3 to FARP 5	2200	1000
LAW from Trans 3 to LOG 6	2200	1000

In Table 2, results for the scenario in Figure 4-1 and 4-2 show that it is necessary to have a total of 11 deliveries for fuel and 6 for ammunition and supplies. It can be seen from

Table 2 that the number of deliveries in the combined network and the total from Split networks are the same, however.

Table 2. Minimum Number of Deliveries for Transportation of Fuel in BBL and Ammunition and Supplies in Pallets

	Combined	Split 1	Split 2	Combined	Split 1	Split 2
	Deliveries	Deliveries	Deliveries	Deliveries	Deliveries	Deliveries
PSV from CLF 1 to SAG 2	4	4		2	2	
PSV from CLF 1 to Trans 3	2	2		1	1	
FSV from Trans 3 to ASuW 4	1		0	1		0
FSV from Trans 3 to FARP 5	0		0	0		0
FSV from Trans 3 to LOG 6	1		1	0		0
LAW from Trans 3 to ASuW 4	0		1	0		1
LAW from Trans 3 to FARP 5	3		3	1		1
LAW from Trans 3 to LOG 6	0		0	1		1
Total	11	6	5	6	3	3

In order to offer another perspective, we further expanded the scenarios where we separate SAG 2 into three DDGs and one LCS (Figure 5-1) in one case and three DDGs and one FFG (Figure 5-2) in the other. It needs to be noted that splitting SAG in corresponding vessels offers better insight into the situation since it offers delivery numbers for each demand node. We believe this will further help decision-makers.

In both scenarios, the demand node afloat is SAG, and demand nodes ashore are EABs, specifically ASuW, FARP, and LOG. It should also be noted that we used a period of 8 days in this scenario, since it is the maximum period for a DDG between refueling events. This assumption forces an LCS in the first case and an FFG in the other to be refueled twice; therefore, we assumed the demand at LCS in the first case and FFG in the other to be twice as much. The increased demand for these vessels increases the deliveries to these ships and not to DDGs. Given the capacities of the NGLS vessels and the demand at LCS and FFG, the delivery numbers were different. In case of ammunition and supply replenishment, this scenario changes. DDGs can only be engaged for at most 1 hour for delivery of fuel. Therefore, we assumed that a corresponding A-S delivery, since it is done by the same vessel, can also be done in parallel for only 1 hour at the rate of 60 pallets per hour. If we remove this restriction for A-S delivery, the capacities change. The idea here is that the refueling can be done for one DDG at a time independently or consecutively (like a milk run). Though we separated SAG into different demand nodes, we did not execute the model for minimizing delivery time since approximate distances from CLF to each node in SAG would be similar. The model for these is:

Models based on scenarios in Figure 5-1 and 5-2 for Fuel

Total supply *at node i for fuel =* S_{Fi},, ***Total demand*** *at node j for fuel =* D_{Fj}

Shared volume capacity *for fuel on vessel k enroute ij =* cF_{kij}

Modes of transportation: *PSV =1, FSV = 2, LAW = 3*

Decision Variables:

X_{Fkij} *= flow of fuel from source i to node j on vessel k, i=1, j= 2-1, 2-2, 2-3, 2-4, and 3, k = 1*

X_{Fkjl} *= flow of fuel from transshipment node j to sink l on vessel k, j=3, l=4, 5, and 6, k = 2, 3*

Y_{kij} *= # of deliveries by vessels of type k from node i to j and l*

Objective Function: *Minimize Number of Deliveries*

$$\min(y_{112-1} + y_{112-2} + y_{112-3} + y_{112-4} + y_{113} + y_{234} + y_{235} + y_{236} + y_{334} + y_{335} + y_{336})$$

Constraints:

Supply — *at CLF = 1,* $(x_{F112-1} + x_{F112-2} + x_{F112-3} + x_{F112-4} + x_{F113}) \leq SF_1$

at Transshipment = 3, $x_{F234} + x_{F235} + x_{F236} + x_{F334} + x_{F335} + x_{F336} \leq SF_3$

Demand — *at DDG = 2-1,* $x_{F112-1} \geq DF_{2-1}$

at DDG = 2-2, $x_{F112-2} \geq DF_{2-2}$

at DDG = 2-3, $x_{F112-3} \geq DF_{2-3}$

at LCS or FFG = 2-4, $x_{F112-4} \geq DF_{2-4}$

at Transshipment = 3, $x_{F113} \geq DF_3$

at ASuW = 4, $x_{F234} + x_{F334} \geq DF_4$

at FARP = 5, $x_{F235} + x_{F335} \geq DF_5$

at LOG = 6, $x_{F236} + x_{F336} \geq DF_6$

Transshipment (Flow Balance)

$$x_{F113} - (x_{F234} + x_{F235} + x_{F236} + x_{F334} + x_{F335} + x_{F336}) \geq 0$$

Capacity Fuel Volume

$$(cF_{112-1})Y_{112-1} - X_{F112-1} \geq 0$$
$$(cF_{112-2})Y_{112-2} - X_{F112-2} \geq 0$$
$$(cF_{112-3})Y_{112-3} - X_{F112-3} \geq 0$$
$$(cF_{112-4})Y_{112-4} - X_{F112-4} \geq 0$$
$$(cF_{113})Y_{113} - X_{F113} \geq 0$$
$$(cF_{234})Y_{234} - X_{F234} \geq 0$$
$$(cF_{235})Y_{235} - X_{F235} \geq 0$$
$$(cF_{236})Y_{236} - X_{F236} \geq 0$$
$$(cF_{334})Y_{334} - X_{F334} \geq 0$$
$$(cF_{335})Y_{335} - X_{F335} \geq 0$$
$$(cF_{336})Y_{336} - X_{F336} \geq 0$$

Y_{kij}*'s integer,* X_{kij}*'s ≥ 0*

Scenario: Separated SAG: Three DDGs and One LCS or Three DDGs and One FFG

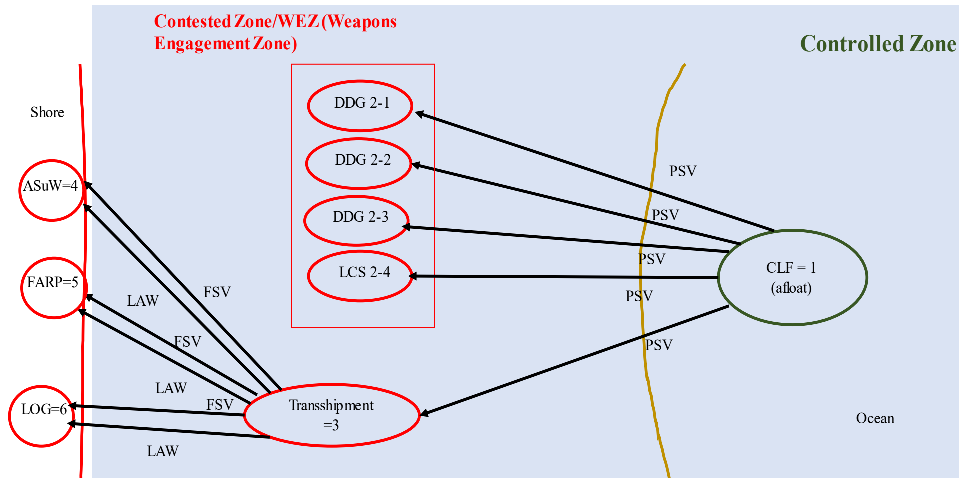


Figure 5-1. Scenario Based on Separated SAG: Three DDGs and One LCS

Separating SAG into DDGs and LCS or FFG creates a unique difficulty in replenishment. However, we adjusted ship capacity to accommodate the differences between resupply periods.

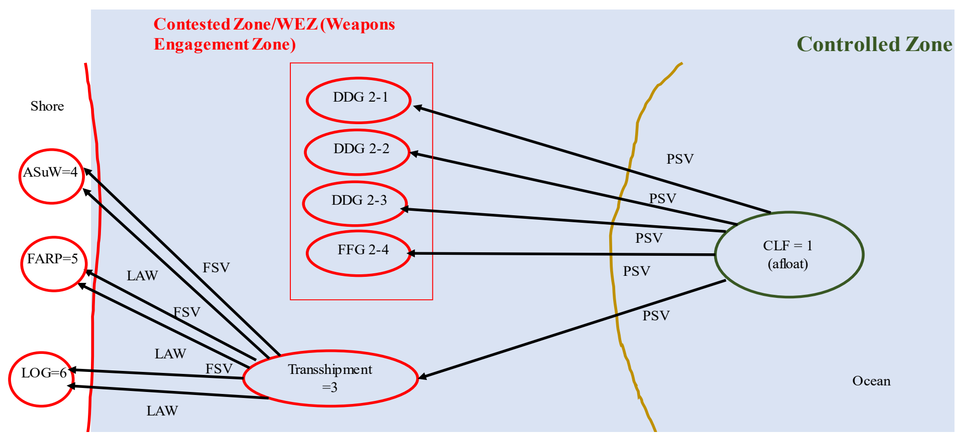


Figure 5-2. Scenario Based on Separated SAG: Three DDGs and One FFG

The supply and demand at the nodes, and capacities of vessels, in these scenarios are given in Table 3 in case of LCS and in Table 4 in case of FFG. Assuming that DDG can sustain for 8 days without refueling, FFG must be refueled every seven days and LCS every four days. We incorporated this by adjusting the demands for the given period in the models based on Figures 5-1 and 5-2.

Table 3. Supply, Demand, and Capacities: Three DDGs and One LCS

	Fuel in BBL	Ammunition and Supplies in Pallets
	Supply/Demand	Supply/Demand
Supply at CLF 1	1000000	100000
Supply at Trans 3	6500	750
Demand at DDG 2-1	5000	25
Demand at DDG 2-2	5000	25
Demand at DDG 2-3	5000	25
Demand at LCS 2-4	3000	20
Demand at Trans 3	6500	750
Demand at ASuW 4	100	50
Demand at FARP 5	6300	350
Demand at LOG 6	100	350
	Capacity	Capacity
PSV from CLF 1 to DDG 2-1	5500	60
PSV from CLF 1 to DDG 2-2	5500	60
PSV from CLF 1 to DDG 2-3	5500	60
PSV from CLF 1 to DDG 2-4	5500	60
PSV from CLF 1 to Trans 3	5500	800
FSV from Trans 3 to ASuW 4	1000	250
FSV from Trans 3 to FARP 5	1000	250
FSV from Trans 3 to LOG 6	1000	250
LAW from Trans 3 to ASuW 4	2200	1000
LAW from Trans 3 to FARP 5	2200	1000
LAW from Trans 3 to LOG 6	2200	1000

Table 4. Supply, Demand, and Capacities: Three DDGs and One FFG

	Fuel in BBL	Ammunition and Supplies in Pallets
	Supply/Demand	Supply/Demand
Supply at CLF 1	1000000	100000
Supply at Trans 3	6500	750
Demand at DDG 2-1	5000	25
Demand at DDG 2-2	5000	25
Demand at DDG 2-3	5000	25
Demand at FFG 2-4	7000	25
Demand at Trans 3	6500	750
Demand at ASuW 4	100	50
Demand at FARP 5	6300	350
Demand at LOG 6	100	350
	Capacity	Capacity
PSV from CLF 1 to DDG 2-1	5500	60
PSV from CLF 1 to DDG 2-2	5500	60
PSV from CLF 1 to DDG 2-3	5500	60
PSV from CLF 1 to DDG 2-4	5500	60
PSV from CLF 1 to Trans 3	5500	800
FSV from Trans 3 to ASuW 4	1000	250
FSV from Trans 3 to FARP 5	1000	250
FSV from Trans 3 to LOG 6	1000	250
LAW from Trans 3 to ASuW 4	2200	1000
LAW from Trans 3 to FARP 5	2200	1000
LAW from Trans 3 to LOG 6	2200	1000

Results

The results with LCS are given in Table 5. Table 6 describes the results with FFG.

Table 5. Minimum Number of Deliveries for Transportation of Fuel in BBL and Ammunition and Supplies in Pallets: Three DDGs and One LCS

	Fuel	Ammunition and Supplies
	Deliveries	Deliveries
PSV from CLF 1 to DDG 2-1	1	1
PSV from CLF 1 to DDG 2-2	1	1
PSV from CLF 1 to DDG 2-3	1	1
PSV from CLF 1 to LCS 2-4	1	1
PSV from CLF 1 to Trans 3	2	1
FSV from Trans 3 to ASuW 4	1	1
FSV from Trans 3 to FARP 5	0	0
FSV from Trans 3 to LOG 6	1	0
LAW from Trans 3 to ASuW 4	0	0
LAW from Trans 3 to FARP 5	3	1
LAW from Trans 3 to LOG 6	0	1
Total	11	8

Table 6. Minimum Number of Deliveries for Transportation of Fuel in BBL and Ammunition and Supplies in Pallets: Three DDGs and One FFG

	Fuel	Ammunition and Supplies
	Deliveries	Deliveries
PSV from CLF 1 to DDG 2-1	1	1
PSV from CLF 1 to DDG 2-2	1	1
PSV from CLF 1 to DDG 2-3	1	1
PSV from CLF 1 to LCS 2-4	2	1
PSV from CLF 1 to Trans 3	2	1
FSV from Trans 3 to ASuW 4	1	0
FSV from Trans 3 to FARP 5	0	0
FSV from Trans 3 to LOG 6	1	0
LAW from Trans 3 to ASuW 4	0	1
LAW from Trans 3 to FARP 5	3	1
LAW from Trans 3 to LOG 6	0	1
Total	12	8

Summary, Analysis, and Conclusion

Feedback from the SME helped us gain insight into the complexity of the problem and its vast scope. We used this input to refine our scenarios. We developed mathematical models based on these scenarios. We have listed those scenarios that will offer decision-makers with a choice based on their requirement. We constrained the capacity based on the maximum time a ship can be engaged in a supply event to reflect the delivery time.

The top-level requirements of the vessels under consideration, as we understood, incorporate capability of a vessel on a certain route based on speed, platform, and capacity. The fuel storage tanks are separate from the storage for ammunition and supplies. Hence, we kept these two commodities separate. Fuel has its own issues, and so do ammunition and supplies. Note that the separate trips for these two commodities could be combined when trying to operationalize these results into a schedule involving a particular number of ships.

The sponsor did not wish us to model an objective of minimizing costs (which were not available) or the number of ships required to deliver commodities within a certain deadline or under a certain schedule (because deadlines and schedules change based on operational priorities). Measuring the number of deliveries required allowed us to determine a mix of NGLS vessels without addressing cost, deadline, or scheduling restrictions.

In our model, *number of deliveries* are the deliveries made by a specific vessel, from a supply node to a demand node, on a specific route for a specific commodity. We would like to point out that deliveries can be interpreted in many ways. For example, a LAW making 13 deliveries of fuel to FARP can be (a) 13 LAWs (making one delivery each), or (b) seven LAWs (six LAWs making two deliveries each and one LAW making one delivery), or six LAWs (making one delivery each, and one LAW making seven of the 13 deliveries). Thus, it is up to the decision-makers to determine how they would like to interpret and implement the results. A decision-maker may go for 13 LAWs if the cost is reasonable and the environment is highly contested. But if it is not, perhaps seven LAWs will be adequate. Again, the number of deliveries may be interpreted by the decision-makers based on their preference and available budget, and there could be many such interpretations. Similar statements can be made about PSVs or FSVs. For example, if there are five deliveries made by PSVs, it could mean that (a) there are five PSVs making one delivery each, or (b) two PSVs making two deliveries each and one PSV making one. One must note, however, that the deliveries will be constrained by overall capacity of the vessel. If one PSV tops out after four deliveries, then the interpretation would change. It would be entirely up to the decision-makers to decide how they would want to interpret the solution. In Table 7 we summarize the results of the scenarios.

Table 7. Summary of Scenario Results

	Fuel				Ammunition and Supplies			
Scenarios	Number of Deliveries by PSV	Number of Deliveries by FSV	Number of Deliveries by LAW	Total Deliveries	Number of Deliveries by PSV	Number of Deliveries by FSV	Number of Deliveries by LAW	Total Deliveries
Scenario Based on Figure 4-1 and 4-2								
Combined	6	2	3	11	3	1	2	6
Split 1 and Split 2	6	1	4	11	3	0	3	6
Scenario Based on Separated SAG: 3 DDGs and LCS (Figure 5-1)	6	2	3	11	5	1	2	8
Scenario Based on Separated SAG: 3 DDGs and FFG (Figure 5-2)	7	2	3	12	5	0	3	8

The models we have developed are scalable. The scenarios can be expanded as per the requirement of number of demand nodes. For example, if there are three SAGs that must be supported, the demand of one SAG in our scenario can be multiplied by three. Of course, in that case, the number of deliveries will increase. Or there may be more than one ASuW Strike EAB, say two, or both of these cases may exist. In that case, the demand for that demand node can be doubled. Such adjustments can be also be made to distances or when minimum time for deliveries needs to be known. The corresponding results are given in Table 8.

Table 8. Minimum Deliveries With Increased Demand Nodes: Fuel in BBL and Ammunition and Supplies in Pallets

Scenarios	Three SAGs		Two ASuWs		Three SAGs, Two ASuWs	
	Deliveries		Deliveries		Deliveries	
	Fuel	Ammunition and Supplies	Fuel	Ammunition and Supplies	Fuel	Ammunition and Supplies
PSV from CLF 1 to SAG 2	12	5	4	2	12	5
PSV from CLF 1 to Trans 3	2	1	2	1	2	1
FSV from Trans 3 to ASuW 4	0	0	0	0	0	0
FSV from Trans 3 to FARP 5	0	0	0	0	0	0
FSV from Trans 3 to LOG 6	0	0	1	0	0	0
LAW from Trans 3 to ASuW 4	1	1	1	1	1	1
LAW from Trans 3 to FARP 5	3	1	3	1	3	1
LAW from Trans 3 to LOG 6	1	1	0	1	1	1
Total	19	9	11	6	19	9

As stated earlier, we did not incorporate load and unload time. Incorporating load and unload time might increase the total time for deliveries. This may lead to acquisition of more vessels so the actual transportation and delivery can be done in parallel to reduce the time. For example, in case four PSVs are needed to deliver required fuel to SAG (based on the assumptions about distance and speed of the PSV), and that a warship can only be engaged for at most 1 hour for this delivery, our model shows it takes a total of 7 days. However, given that DDGs can sustain for 8 days after one refueling event and there are three DDGs in a SAG, an acquisition strategy for acquiring four PSVs so each PSV takes less than 2 days to deliver may be a better solution than one PSV making four deliveries in 7 days. Again, this is a choice the decision-makers can make based on the flexibility of these models.

Based on our analysis, we recommend the following to negotiate battlespace constraints. We suggest that the time constraint for a PSV engaging with SAG in WEZ should be investigated, since that is the binding constraint on capacity to transfer. The capacity of the PSV for carrying fuels is much larger than that, and the same is true for transferring the pallets of ammunition and supplies. It will be necessary to increase the rate of transfer if the time spent in the WEZ cannot be altered. Our capacity assumptions were based on threshold as opposed to objective TLRs. Hence, objective TLRs may be the direction to go. This may need tweaking at the TLRs and some platform modification so that sustainment can be made much faster and with fewer deliveries. We summarize the number of deliveries made by FSV and LAW for each of the scenarios in Table 9.

Based on this summary, one can see that the most FSVs needed for each of these scenarios to transport fuel are *two,* whereas for the same scenarios, *five* LAWs are also needed. Similarly, the most FSVs needed for each of these scenarios to transport ammunition and supplies is *one*. However, *three* LAWs are also needed for those scenarios. These results and our analysis therefore suggest that acquisition of LAWs is preferred to FSVs, since it may be prohibitively expensive to maintain a separate maintenance support infrastructure for FSVs when their range of usefulness is relatively narrow. A closer examination of those instances in which the model recommended FSVs on a route reveals that, in every case, a LAW could have accomplished the resupply mission in an equal number of trips. That is, the model recommendation to acquire an FSV, in every case, is merely an alternate optima. Although the

FSV does not look very useful in these scenarios, these scenarios did not require the TLRs (especially speed, since deadlines were not given) in which that ship dominated the others.

Table 9. Deliveries by FSV and LAW

	Fuel		Ammunition and Supplies	
Scenario	Number of Deliveries by FSV	Number of Deliveries by LAW	Number of Deliveries by FSV	Number of Deliveries by LAW
Scenario Based on Subject Matter Expert Feedback (Figure 4-1 and 4-2)				
Combined	1	3	1	2
Split 1 and Split 2	2	4	0	3
Scenario Based on Separated SAG: 3 DDGs and LCS with Sustainment (Figure 5-1)	2	3	1	2
Scenario Based on Separated SAG: 3 DDGs and FFG with Sustainment (Figure 5-2)	2	3	0	3
Scenario Based on Increased Demand Nodes				
Three SAGS	0	5	0	3
Two ASuWs	1	4	0	3
Three SAGs, Two ASuWs	0	5	0	3

References

Abott, R. (2020, February 2). FY 2021 request starts work on future amphibs and logistics ships. *Defense Daily*. https://www.defensedaily.com/navy-starts-work-future-amphibious-logistics-ships/navy-usmc/

Apte, A., Doerr, K., & Apte, U. (2020). *Framework for augmenting current fleet with commercially available assets for logistics support in contested environment* (NPS-LM-20-155). Naval Postgraduate School, Acquisition Research Program.

Apte, A., & Yoho, K. (2018). Resource selection in support of humanitarian operations: A case of the United States Navy. *Journal of Humanitarian Logistics and Supply Chain Management*, *8*(2), 184–198.

Dougherty, S. R., Garcia, R. J., Lemenager, K. D., Nye, B. S., Rego, J., Sandridge, B. E., & Shofner, M. G. (2020). *Logistics in contested environment* [Capstone report, Naval Postgraduate School].

Eckstein, M. (2020, February 20). *Navy researching new class of medium amphibious ship, new logistics ships*. USNI News. https://news.usni.org/2020/02/20/navy-researching-new-class-of-medium-amphibious-ship-new-logistics-ships

Katz. (2020, May). Navy announces "Next Generation Logistics Ship" program with June 25 industry day. *Inside Defense*. https://insidedefense.com/insider/navy-announces-next-generation-logistics-ship-program-%C2%A0june-25%C2%A0industry-day

Liu, Z. (2016, July 11). Here's what's behind the "nine-dash line" that sparked the South China Sea conflict. *South China Morning Post*. https://www.businessinsider.com/the-nine-dash-line-at-the-heart-of-the-south-china-sea-conflict-2016-7

O'Rourke, R. (2020, September 8). *Navy force structure and shipbuilding plans: Background and issues for Congress* (CRS Report No. RL32665). Congressional Research Service. https://fas.org/sgp/crs/weapons/RL32665.pdf

Capability Based Planning in Humanitarian Operations: A Hybrid Optimization and Simulation Framework for Strategic Acquisition in the Armed Forces

Professor Dr. Stefan Wolfgang Pickl—studied mathematics, electrical engineering, and philosophy at the Technical University of Darmstadt (diploma in 1993; ERASMUS-scholarship at the EPFL Lausanne) and earned a doctor's degree at the TU Darmstadt in 1998, followed by his habilitation at the University of Cologne in 2005. He is Counsellor to the Center for Information Technology and Algorithms (CITA) at the University of Nevada, Las Vegas, as well as an Associated Member of the Center for Network Innovation and Experimentation (CENETIX) at the Naval Postgraduate School (NPS) in Monterey, CA. In cooperation with associated sectors for decision finding, he carries out analyses and supervises international experiments there, especially in the field of Maritime Interdiction Operations (MIO), Sensor Allocation, and Complex Harbor Protection. Within the scope of the Data Farming/ CREATE-Project, he works together with the NUS Singapore in the field of Security Scenario Analysis/Intelligent Design of Transport. [stefan.pickl@unibw.de]

Professor Dr. Max Krüger—is a former officer of the German Navy and, since 2011, has been a professor at Furtwangen University of Applied Sciences in the Department of Business Administration and Engineering. He studied computer science and did his dissertation on game theory at the Bundeswehr University in Munich. From 2006 to 2011, he worked as a Scientific Expert for the Industrieanlagen-Betriebsgesellschaft (IABG mbH) in Ottobrunn in the areas of OR and applied AI. His particular research interests are in Bayesian Networks, classification problems and optimization in general, as well as concrete applications (e.g., in aerial and maritime surveillance). [max.krueger@email.de]

Major General (Ret.) Dr. Dieter Budde—joined the Bundeswehr in 1966 as a paratrooper; 1969–1977, pilot training and assignment in Army Air Units; 1975, Dipl. sc. pol. (political science), Ludwig Maximilian University Munich; General Staff training; General Staff Officer MoD (military policy, arms control); troop assignments; Arms Control Branch, Policy Division, Mons, BE; Chief of Staff 1. Gebirgsdivision; Head of Student Section UniBw Munich; Higher Command and Staff Course (HCSC) Camberley, UK; Commander Army Aviation Transport Regiment; Commander Air Mechanized Brigade; General Army Aviation and Commander Army Aviation School; Commander Air Mobile Operations Division; about 4;500 flying hours on helicopters and fixed wings. 2000 Dr. rer. pol. (political and social sciences) University of the German Armed Forces Munich; 2005 Retired as Major General; 2006–2009 Managing Director StratCons Strategic Consulting GmbH; Member of an advisory board of a company as a supplier to the aviation industry; 2010 Master of Arts (MA) Governance; 2013 Master of Arts (MA) Philosophy; 2013–2015 High Level Advisor with the German Society for International Cooperation (GIZ) in the project "Strengthening the Civil Aviation Oversight of Afghanistan"; 2014 until today Senior Advisor to the Chair for Operations Research and Theoretical Mathematics at the Universität der Bundeswehr, Munich; 2015–2017 Key Expert with the Turkish-EU Project "Technical Assistance for Capacity Building of Safety and Security Training Requirements for Turkish Civil Aviation"; 2019 until today, Lectureship at the University of the Bundeswehr, Munich; 2020 Dr. phil. (Philosophy), University of Cologne. [dibu11@gmx.de]

Gonzalo Barbeito—is a DAAD grant holder and Doctoral Candidate in the Department of Computer Science at the Universität der Bundeswehr München, Germany. Barbeito deals with mathematical optimization, stochastic modeling and computer simulation for operational system analysis. In his doctoral research, he develops an optimization and simulation framework aimed at improving last-mile logistics in the context of humanitarian operations. He pursues a shift from traditional knowledge-based modeling towards a more objective hybrid knowledge-data paradigm. [gonzalo.barbeito@unibw.de]

Abstract

This paper describes an innovative Hybrid Optimization and Simulation approach for assessing acquisition strategies in the Armed Forces. This work has been conceptualized and is currently being developed with and for experts in the field of disaster and emergency management in order to tackle the real issues arising during such crises. As the overarching framework, we operate under the umbrella of Capability Based Planning (CBP), a methodology widely used by the defense community and referred to by many as the "gold standard" for strategic planning. Drawing from the general building blocks of CBP, we aim to develop an analytic process to transparently assess and explain humanitarian capabilities acquisitions in the Armed Forces and include them in its traditionally defense-oriented strategic planning.

Keywords: Strategic Acquisition, Humanitarian Logistics, Capability Based Planning, Defense Planning

Introduction

The significance of disasters for societies is tremendous. It has been estimated that, every year, more than 500 disasters strike our planet, with a death toll of over 75,000 people and affecting more than 200 million people (Caunhye et al., 2012). Furthermore, several studies have demonstrated an increasing trend in both the frequency and severity of these events (Behl & Dutta, 2019; Habib et al., 2016; Leiras et al., 2014). In order to lessen their effects, be it the loss of human life or the impact on the economy, and to bring society back to a state of normality, the field of Disaster and Emergency Management (DEM) prepares resources and activities that will deal with the humanitarian aspect of emergencies. Due to the sheer complexity of DEM operations, cooperation between military and civilian actors is often required. Beyond the context of war, military forces are particularly well equipped to deal with certain key areas in DEM, like logistics and engineering, among other support possibilities that could be explored to mitigate the impact of disasters on civilian populations.

This paper proposes the integration of these support possibilities in the strategic analysis of the Armed Forces while still maintaining defense-related investments as a priority.

Preliminaries and Background

Capability Based Planning

Capability Based Planning (CBP) is a general planning framework that aims to provide an organization with capabilities suitable for a wide range of modern-day challenges and risks, simultaneously framing these capabilities within an economic framework (Davis, 2002). This approach relies on goals and functional needs for broadly defined scenarios, making planning more responsive to uncertainty and risk, and provides a rational basis for decisions on future acquisitions. In this sense, CBP differs from threat-based planning, a framework popular during and up to the end of the Cold War, where strategic planning was heavily oriented towards specific threats or scenarios (Hales & Chouinard, 2011).

While CBP has been widely adopted by the Defense community, with notable results in the United States, Canada, Australia, and the United Kingdom, its extension to the field of Safety and Security, including DEM, proved challenging (Hales & Chouinard, 2011). These challenges arise, primarily, from perspective differences between the Defense and the DEM communities, and from within the DEM community itself. Going a step further in the analysis, the internationally funded Technical Cooperation Program (Taylor, 2013; Technical Cooperation Program Joint Systems Analysis Group Technical Panel 3, 2004) identified four main building blocks in the general framework of CBP:

1. High level **capability objectives** derived from government guidance.
2. **Operational concepts** for strategic, operational, and tactical levels to describe the systems and possible interactions.
3. Standard groupings of disparate elements in **capability clusters** to make the analysis process more manageable.
4. **Resource constraints** that define the limits within which the capabilities need to be realized.

While these blocks are almost exclusively built and described around defense concepts, it is possible to partially abstract them as an optimization problem. This, in turn, allows the inclusion of tasks and operations outside the typical concepts in defense but connected at its core.

This research proposes an extension of CBP to include humanitarian capabilities and operational concepts while preserving the resource constraints at strategic levels. In essence, we aim to study and expand existing military capabilities befitting humanitarian operations, acknowledging the existing economic restrictions at strategic levels.

Operations Research in the Context of Capability Based Planning

Operations Research (OR) has long been applied to the higher level of strategic defense planning. The following paragraphs describe the intended approach to meet the requirements of all CBP building blocks using OR methodologies.

The definition of OR has been approached from multiple perspectives by many relevant authors in the field. In the interest of clarity and simplicity, we will adopt the OR definition described by the Association of European Operational Research Societies (EURO):

> [Operations Research] can be described as a scientific approach to the solution of problems in the management of complex systems. In a rapidly changing environment, an understanding is sought which will facilitate the choice and the implementation of more effective solutions which, typically, may involve complex interactions among the elements of the system, for instance, people, materials and money.[1]

Closely related to the building blocks of a CBP framework described earlier, an effective CBP implementation requires the developed plan to meet the capability objectives under the defined operational concepts, minimizing risk and cost and complying with the resource and general constraints (Technical Cooperation Program Joint Systems Analysis Group Technical Panel 3, 2004). In order to optimally distribute resources, CBP depends on OR methods to help make better decisions. In particular, this research leverages two powerful techniques found at the intersection between OR and CBP, **Simulation** and **Optimization**.

On the one hand, different simulation methodologies are used for representing operational concepts and to develop and explore the sandbox in which different scenarios will be tested. These are the different techniques used in this work:

- **System Dynamics** (SD) is a modeling a simulation technique for studying the dynamics of complex systems (Sterman, 2000). It uses a set of simple building blocks and entities, namely *Stocks*, *Material,* and *Information Flows* and *Delays* to describe how these systems change over time. Due to a high level of abstraction in the modeling approach, SD is normally regarded as a strategic modeling methodology.
- **Discrete Event Simulation** (DES) is a method that requires the modeler to divide the studied system into a sequence of operations performed across entities over discrete time (i.e., the model clock only advances when something significant happens in the model). It is generally considered to be a low abstraction modeling technique and is used to model processes in-depth (Borshchev, 2013).
- **Agent Based Modeling** (ABM) is a more recent modeling approach, focusing on the behavior of individual interacting entities (namely, agents) to create emergent behavior

[1] from https://www.euro-online.org/

(bottom up approach), instead of the process affecting those entities (top down; Borshchev, 2013).

- In order to exhaustively explore scenarios under uncertainty, **Monte Carlo Simulation** allows the modeler to assess individual simulation run outcomes by stochastically varying input parameter values (Rubinstein & Kroese, 2016). This technique is well suited to deal with systems where the input-output interactions are too complex to assess analytically.

Architecturally, two main streams can be identified with respect to simulation techniques. The first stream combines SD, DES, and ABM into a single model using multimethod modeling (Borshchev, 2013). The second stream considers exclusively the Monte Carlo simulations, running parallelly to the first stream.

Analytical optimization approaches, on the other hand, are also used for finding, at different strategic levels, the best resource allocation strategy from an economic and operative standpoint. These models have been successfully applied in several DEM problems, as shown in Behl and Dutta (2019) and Habib et al. (2016). In this work, two practical optimization problems will be explored: a modification of the vehicle routing problem (VRP; Lahyani et al., 2015; Toth & Vigo, 2002) and a strategic acquisition problem framed as an extension of the classical Knapsack Problem (Bakirli et al., 2014; Brown et al., 2004). Both problems are solved in a sequential workflow, as explained in the Methodology and Concept Description.

Humanitarian Logistics and Capability Based Planning

In a humanitarian context, the term *logistics* represents the processes and systems involved in mobilizing people, resources, skills, and knowledge to help vulnerable people affected after a disaster. Given the wide array of problems that this field encompasses, we will focus on the **Last-Mile Relief Distribution** (LMRD) problem, a well-known problem that will act as a proxy for humanitarian operations (Balcik et al., 2008; Stapleton et al., 2011). In this problem, a fleet of capacitated vehicles must economically distribute relief resources between local depots and affected areas.

An interesting characteristic of the LMRD problem is the trade-off between economic distribution and the life-saving utility: There is a clear correlation between the number of vehicles and the achievable satisfaction of demand. Due to limited resources, however, it is necessary to correctly assess the optimum supply and transport capacity that can successfully satisfy this demand while reducing transportation costs and idle capacities.

The extension of humanitarian logistics to the CBP framework, while not exactly easy, is conceptually quite straightforward: First, the number of transportation units and supply capacities robust against a broad set of disaster scenarios is calculated and immediately compared with present distribution capacities. If a need for improvement in transportation capacities is detected, it will be included in the much broader strategic investment plan, considering the investment priorities of the Armed Forces as a whole.

Research Question

The concrete question that this research aims to answer is *How can humanitarian capabilities for military support in DEM be characterized when framed within a broader strategic acquisition plan in the Armed Forces?*

By answering this question, we expect to bridge an observed gap in military strategic planning, broadening the application potential of CBP by enriching its current defense-oriented paradigm with humanitarian goals.

Methodology and Concept Description

Conceptually, this work is being developed with experts in the field of DEM, both from civil organizations and with military backgrounds.

Figure 1 shows the broad framework description: On the right, the classical capability and investment approach remains unchanged. On the left, we show the proposed extension to the classical approach by linking humanitarian operations to optimal capability requirements—as described in **problem (a)**—and finally conducting a holistic assessment of the acquisition requirements and transparently supporting decision-makers with optimal economic distribution—tackled in **problem (b)**. The connection shown in this figure between the classical CBP approach (right side) and the proposed extension (left side) represents the multi-purpose existing capabilities in the Armed Forces and one key argument to justify military support in humanitarian operations.

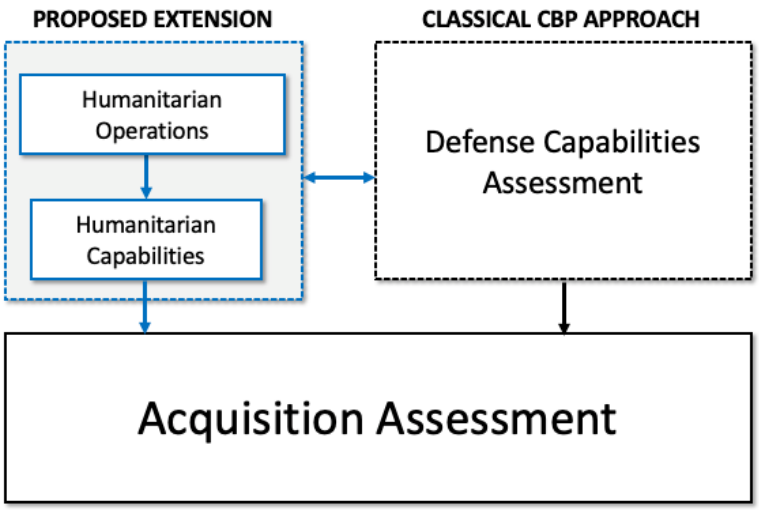


Figure 1. Proposed CBP-Based Acquisition Framework (Left Side)

Architecturally, this work adopts a connected modular approach, developing two different models tackling different organizational levels:

a) the *relief distribution problem*, used to identify a gap in capabilities, and
b) the *acquisition problem*, supporting the acquisition of those capabilities.

Figure 2 shows a conceptual description of the process flow and the connection between both problems, with their corresponding inputs and outputs.

Figure 2: Conceptual Process Flow

Problem (A): Determining Humanitarian Logistics Capabilities

This problem encompasses both the tactical and operational levels observed in the CBP framework and will, for practical purposes, keep a narrow perspective within Humanitarian

Operations. At this level, we developed a multi-method simulation model of a disaster requiring civilian relocation to shelters. The behavior of this model was designed in cooperation with experts in logistics and DEM and follows real constraints and variables to the best extent of the possibilities.

The scenario used to exemplify typical operations observed in this model can be described as follows:

In the aftermath of a disrupting disaster (e.g., earthquake, pandemic with quarantine demands, large scale blackout, among others), the civilian population is relocated to shelters for their safety. Regional military forces are in charge of supplying these shelters using limited structural resources and commodities, with optional civilian support that can be required at higher costs. *The central task at this level is the determination of required vehicles (both military and civilian)—together with corresponding distribution routes—for a timely delivery of relief goods to the population in need.*

There are two main decisions at this level; the first one, of an **operative nature**, involves the cost and time efficient determination of routes and supply schedule for every vehicle, *given specific fleet configurations*.

In order to shift the problem complexity away from the user, a Rich-Vehicle Routing Problem (R-VRP; Lahyani et al., 2015) optimization model is solved parallelly to the simulation, identifying the optimal routing strategy for each vehicle in the fleet. This mathematical model was developed exclusively for this research and captures a complex set of real-world VRP taxonomic features not simultaneously contemplated in previous mathematical VRP models, such as

1. *Split-Delivery*: Multiple vehicles are allowed to visit a single shelter, effectively sharing the supply requirements for that shelter.
2. *Multi-Echelon*: Unlike traditional VRP, in this case, the fleet is not necessarily stationed at the depots and might have their own fleet base (for example, one or more supply regiments)
3. *Multi-Depot*: This problem needs to contemplate multiple sources of relief in the map.
4. *Heterogeneous Fleet*: As observed in real world problems, a fleet normally comprises different types of vehicles, each with its own fuel consumption, fixed usage costs, and load capacities.
5. *Multi-Trip*: Vehicles can travel multiple time between depots and shelters in order to resupply if needed.
6. *Multi-Commodity*: Each vehicle can transport multiple products with different packaging options.

The transportation parameters used in this problem are shown in Figure 3.

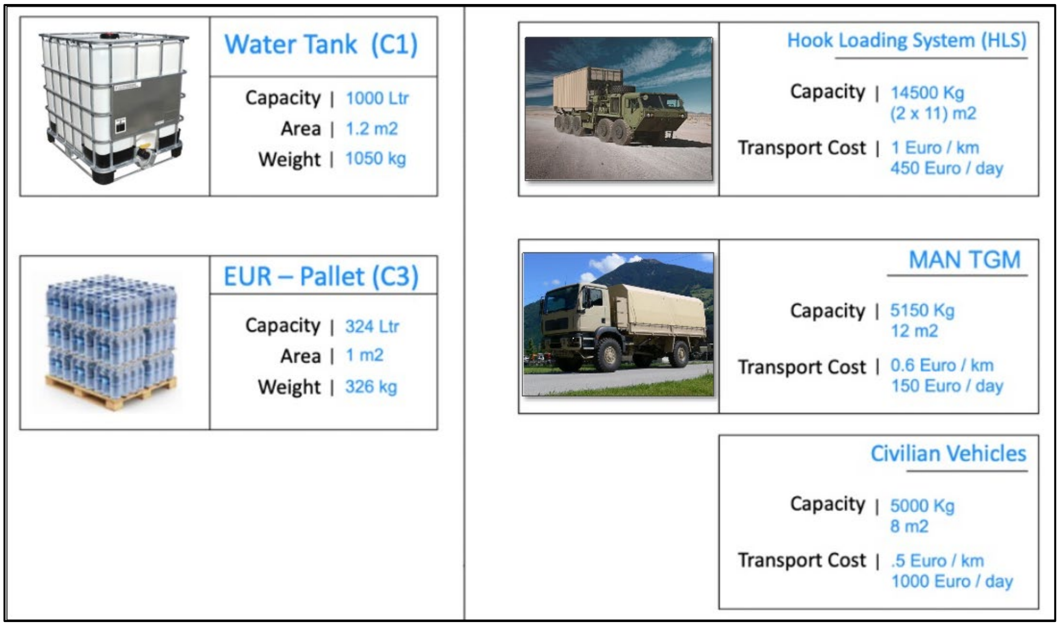


Figure 3. Load Transportation Parameters

(Note: This figure is the authors' own elaboration with image captions from https://commons.wikimedia.org.)

A conceptual description of the user interface designed for this level can be seen in Figure 4. Shelters and Depots show a live feed of their stock levels (red and green bars), and the position of every vehicle in the fleet is updated in a GIS environment. Using this module as a sandbox, users can modify input parameters and observe the success potential of different supplying strategies.

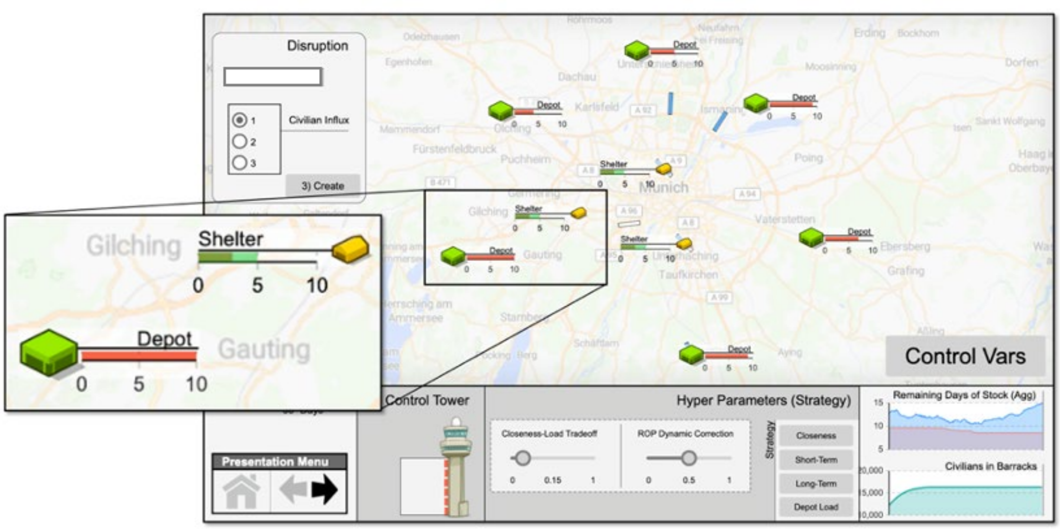


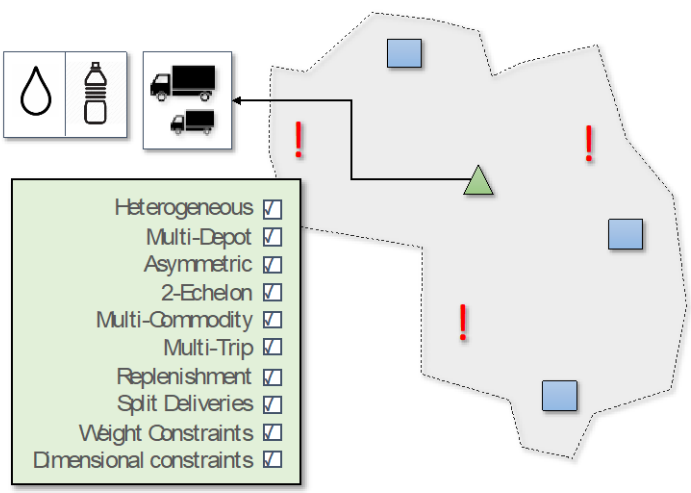
Figure 4. Proposed User Interface for the Humanitarian Logistics Model (Problem [a])

The second type of decision in problem (a) is of a **strategic nature** and aims to explore and identify the best combination of structural and fixed-operational resources to successfully see the delivery plan through to completion.

Essentially, the structural resources to define at this stage are

- Transport capabilities (number and type of vehicles in the fleet)
- Supply capacities (units per day of each product that depots can supply)

In Figure 5, an example is shown based on a real case study that has been anonymized for presentation purposes. The problem includes *a fleet base* (green triangle), *three depots* (blue squares) and *three shelters* (red exclamation marks). The goal is to design the optimal distribution plan for a fleet of heterogeneous vehicles starting from and returning to the fleet base. Each vehicle must pick up two types of commodities at the depots and distribute them to the different shelters. The Rich-VRP features included in this model are shown in the green box in Figure 5. The right side of this figure shows the results of wrapping the optimization model in a Monte Carlo framework, stochastically varying the population sizes at the shelters (first three columns), and defining a new demand composition at each scenario. The last three columns show the optimum number of truck types for each scenario and the cost of that solution, respectively. The row in blue indicates the highest demand scenario.



Input			Results		
Pop S_1	Pop S_2	Pop S_3	Trucks A	Trucks B	Cost
500	500	500	2	0	1634
2000	2000	2000	5	0	4677
5000	5000	5000	8	6	13078
500	5000	500	4	0	4784
5000	500	500	6	0	5975
500	500	5000	4	1	5141
2000	2000	5000	7	0	7646
2000	5000	2000	6	0	7363
5000	2000	2000	6	2	8087

Figure 5. Conceptual Map and Different Scenarios Designed in a Complex Rich-VRP

Problem (B): Strategic Acquisition Plan, Defense Structure Coupled With Humanitarian Operations

The acquisition problem represents the strategic level and takes as input an explicit formulation of capability requirements from multiple defense branches in the Armed Forces and the humanitarian capabilities determined in problem (a). **The goal at this level is to formulate the acquisition strategy that maximizes the capability needs of the army for both defense and humanitarian considerations.** This problem assumes that all inputs provided by the defense branches are the output of similar analytic assessments conducted by experts and, hence, out of the scope of this project. The generic hierarchical structure used in this problem is shown in Figure 6.

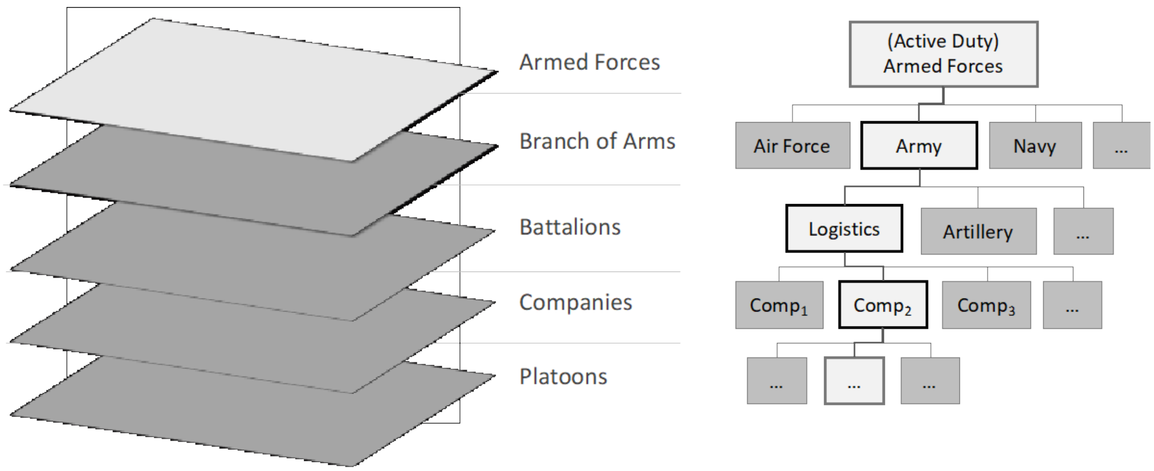


Figure 6. Example of a Generic Hierarchical Structure in the Armed Forces

The implementation of this problem consists of a modified knapsack optimization model aiming to maximize the value of an investment while not exceeding the available budget. Built on top of this optimization model, a user interface allows experts and decision-makers to explore and configure different parameters at each strategic level in the army.

The designed interface, shown in Figure 7, gives an overview of the outputs produced by the model, with the total distribution of funds for each branch of arms and the capability of exploring and parameterizing specific hierarchical levels individually. At this point, it is worth noting that the outputs of this model do not replace by any means the need for expert assessment or decision-makers. These results are exclusively meant to support the decision-making process and reduce the burden of computing complex calculations on the user.

Finally, both problems are merged in a single platform and under a comprehensive management dashboard in order to provide tactical, operative, and strategic information of the different scenarios contemplated for analysis.

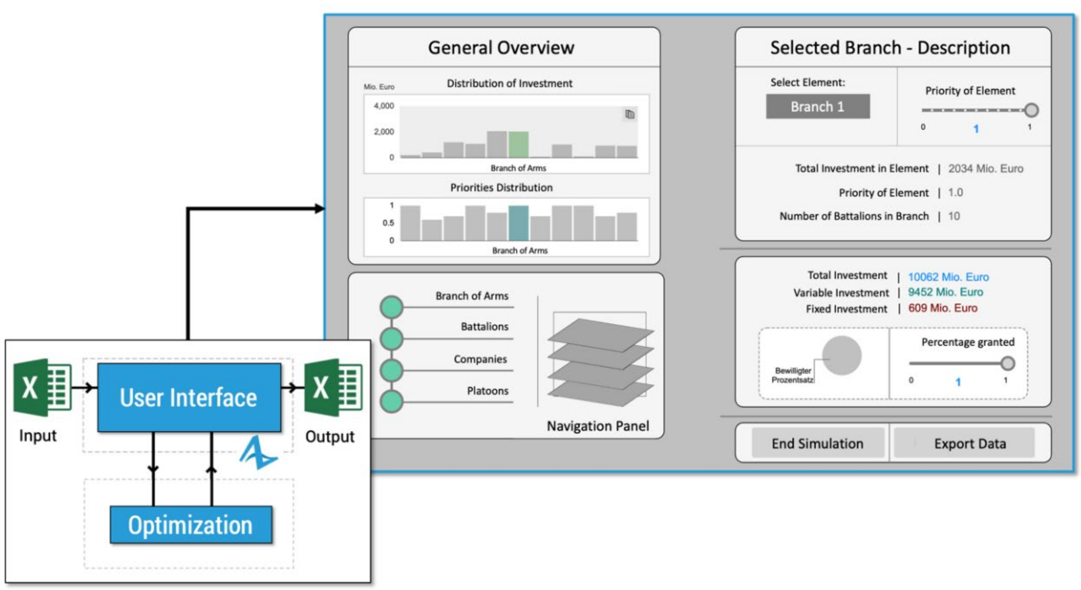


Figure 7. User Interface for the Strategic Acquisition Problem With Detail for Higher Hierarchical Level (Armed Forces Strategic Command)

First Results

Due to the early stage of this project and the sensitive nature of the information used, results will be presented on a qualitative basis, with the goal of assessing feasibility and scalability of the approach. With respect to the former, the approach yields promising results: Figure 4 and Figure 6 showcase not simply conceptual designs, but actual results of this framework, specifically from the management dashboard, in a dynamic environment with which the user can interact. Furthermore, the quantitative values in those figures correspond to real values calculated using the underlying mathematical models. An analysis of those values in these lines would not be of interest since all values shown are generic, given the confidentiality of the information handled.

With respect to scalability, it was observed that the bottleneck of this approach is the *relief distribution problem (a)*, which might struggle when the problem is too expressive (e.g., due to multiple transportation options and complex routing strategies considered) or too large. As a reference, the current optimization model correctly handled maps with simple behavior with up to 75 nodes; however, it struggled with larger instances. For the full expressiveness of the model, instances with up to 16 nodes were solved in realistic time.

Conclusion and Outlook

This paper described the concept behind a streamlined framework for the integration of humanitarian operations within the strategic planning of the Armed Forces. Concretely, this research bridges an observed gap in military strategic planning, broadening the application potential of Capability Based Planning by enriching its current defense-oriented paradigm with humanitarian goals. Even though the project has an already working implementation of the described concept, several validation steps are still needed, as well as a deeper development regarding the hierarchical army sub-structures and the specifics of each disaster tackled.

References

Bakirli, B., Gencer, C., & Aydog, E. (2014). A combined approach for fuzzy multi-objective multiple knapsack problems for defence project selection. *Journal of the Operational Research Society*, *65*(7), 1001–1016. https://doi.org/10.1057/jors.2013.36

Balcik, B., Beamon, B. M., & Smilowitz, K. (2008). Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, *12*(2), 51–63. https://doi.org/10.1080/15472450802023329

Behl, A., & Dutta, P. (2019). Humanitarian supply chain management: A thematic literature review and future directions of research. *Annals of Operations Research*, *283*(1–2), 1001–1044. https://doi.org/10.1007/s10479-018-2806-2

Borshchev, A. (2013). *The big book of simulation modeling. Simulation modeling with anylogic: Agent based, discrete event and system dynamics methods*.

Brown, G. G., Dell, R. F., & Newman, A. M. (2004). Optimizing military capital planning. *INFORMS Journal on Applied Analytics*, *34*(6). https://doi.org/10.1287/inte.1040.0107

Caunhye, A. M., Nie, X., & Pokharel, S. (2012). Optimization models in emergency logistics: A literature review. *Socio-Economic Planning Sciences*, *46*(1), 4–13. https://doi.org/10.1016/j.seps.2011.04.004

Davis, P. K. (2002). *Analytic architecture for capabilities-based planning, mission-system analysis and transformation*. https://www.researchgate.net/publication/228769350_Analytic_Architecture_for_Capabilities-Based_Planning_Mission-System_Analysis_and_Transformation

Habib, M. S., Lee, Y. H., & Memon, M. S. (2016). Mathematical models in humanitarian supply chain management: A systematic literature review. *Mathematical Problems in Engineering*, *2016*. https://doi.org/10.1155/2016/3212095

Hales, D., & Chouinard, P. (2011). *Implementing capability based planning within the public safety and security sector: Lessons from the defence experience*. Defence R&D Canada – Centre for Security Science.

Lahyani, R., Khemakhem, M., & Semet, F. (2015). Rich vehicle routing problems: From a taxonomy to a definition. *European Journal of Operational Research*, *241*(1), 1–14. https://doi.org/10.1016/j.ejor.2014.07.048

Leiras, A., de Brito, I., Queiroz Peres, E., Rejane Bertazzo, T., & Tsugunobu Yoshida Yoshizaki, H. (2014). Literature review of humanitarian logistics research: Trends and challenges. *Journal of Humanitarian Logistics and Supply Chain Management*, *4*(1), 95–130.

Rubinstein, R., & Kroese, D. P. (2016). *Simulation and the Monte Carlo method* (Vol. 10). John Wiley & Sons.

Stapleton, O., Pedraza Martinez, A., & Van Wassenhove, L. N. (2011). Last mile vehicle supply chain in the International Federation of Red Cross and Red Crescent Societies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1437978

Sterman, J. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Irwin/McGraw-Hill. http://web.mit.edu/jsterman/www/BusDyn2.html

Taylor, B. *Analysis support to strategic planning* (TR-JSA-2-2012). The Technocal Cooperation Program. https://cradpdf.drdc-rddc.gc.ca/PDFS/unc194/p801995_A1b.pdf

Technical Cooperation Program Joint Systems Analysis Group Technical Panel 3. (2004). *Guide to capability-based planning*. DoD. https://www.hsdl.org/?view&did=461818

Toth, P., & Vigo, D. (2002). An overview of vehicle routing problems. In *The vehicle routing problem* (pp. 1–26). https://doi.org/10.1137/1.9780898718515.ch1

Understanding the Challenges of Providing Personal Protective Equipment in the United States During COVID-19

Shaghayegh R. Arangdad—North Carolina State University [srezaei@ncsu.edu]

A Blanton Godfrey—North Carolina State University [blanton_godfrey@ncsu.edu]

Abstract

On March 11, 2020, the World Health Organization designated "coronavirus disease 2019" (COVID-19) a global pandemic. The COVID-19 pandemic caused massive shortages in the Personal Protective Equipment (PPE) supplies needed to treat the virus as the disease spread rapidly throughout the world during 2020. Global supply chains suddenly became a new problem in public attention.

In the United States, there were no reliable databases about what was needed in different hospitals, states, or healthcare systems. Also, there was no accurate database of the production capabilities of U.S. suppliers of PPE.

The aim of our research was to investigate the best approaches to determine the needs for PPE during the pandemic and to improve the methods used to predict the PPE needs for hospitals and other healthcare facilities. The second part of our research was to develop an understanding of the capabilities of the U.S.-based companies to produce large quantities of specific PPE for both the current pandemic and future needs. We feel that this work has implications for anyone in the healthcare supply chain space (Department of Defense [DoD] task forces, Defense Logistics Agency [DLA], Defense Health Agency [DHA], Federal Emergency Management Agency [FEMA], Department of Health and Human Services [DHHS], Department of Homeland Security [DHS], state, local, and healthcare systems).

Keywords: COVID-19, Shortage, Supply Chain Challenges, Personal Protective Equipment

Introduction

In November 2019, Coronavirus Disease 2019 (COVID-19) first appeared in China (Scher, 2020). Soon after, Thailand and Japan became the first and second countries outside China confirming COVID-19 cases in mid-January 2020 (Joseph, 2020; Schnirring, 2020). In quick response, screening travelers was started at three U.S. international airports: Los Angeles (LAX), San Francisco (SFO), and New York (JFK) on January 17, 2020 (Centers for Disease Control and Prevention [CDC], 2020). The first COVID-19 case in the United States was detected in Washington State on January 20, 2020 (Holshue et al., 2020). Then the virus started spreading across the country. COVID-19 has hit the United States harder than any other country in the world (WorldOMeter, n.d.). The number of deaths rose quickly, and just in 4 months, late May 2020, the number of deaths reached 100,000 (Winsor et al., 2020). At the end of March 2021, the total number of cases globally was approximately 129 million, resulting in 2.82 million deaths. About 20% (0.552 million) of these were in the United States (Johns Hopkins University & Medicine, n.d.).

One of the impacts of the COVID-19 pandemic was a major disruption to the PPE supply chains in the United States and large PPE shortages across the country. Before the pandemic, more than 70% of the critical respiratory-related PPE used in the United States was sourced from overseas companies, primarily in China. When high global demand drastically reduced available supplies from China during early 2020, major distributors were unable to fill orders. We quickly saw two major problems. In the United States, there was a lack of information about what was needed in different hospitals, states, or healthcare systems. Models were developed using available estimates of infections, hospitalizations, intensive care unit (ICU) usage, and other factors, but most of these models were seriously flawed. Even worse, the data needed in these models were not accurate, timely, or complete (Davenport et al., 2020). Without accurate and timely data, it is difficult to construct objective-based responses that provide lifeline services for communities in emergencies.

The second major problem was the lack of an accurate database of the production capabilities of U.S. suppliers of PPE. Many who produced end-product PPE soon discovered the materials they needed were no longer available. Others discovered that they were missing several critical items in the supply chain or that these items were in short supply. Others found they had the capability to ramp up production quickly, but they wanted guarantees of purchases before committing to major investments. Without a national understanding of demand, companies were hesitant to make large investments.

This research investigates the best approaches to determine the needs for PPE during the pandemic, suggests improvements of the methods used to predict the PPE needs for hospitals and other healthcare facilities, and develops a better understanding of the supply chain capabilities for PPE. Future materials acquisition programs for similar situations need better models as well as new data collection, data management, and data quality strategies. This study also provides a better understanding of the capabilities of U.S.-based companies to produce large quantities of specific PPE for both the current pandemic and future needs.

Supply Chain Challenges in Healthcare

Supply chain management in healthcare is a dynamic process that includes manufacturing facilities, purchasers, and distribution services for commodities and services to health providers and patients (Iyengar et al., 2020). Implementation of modern supply chain management strategies in the healthcare area has been extremely slow. The top three challenges in healthcare supply chain management are costs, including invisible costs, lack of supply chain health information technologies, and lack of quality. Many extra costs within the healthcare supply chain are a result of inefficient and unnecessary processes involved in the transportation and delivery of supplies from suppliers to healthcare providers.

The lack of a modern information technology system for demand and supply chain transparency is another challenge. Exchanging health data is often difficult, and there are many problems with digital sharing. The number of competing health systems and different platforms creates addition complications. These inefficiencies often lead to delays, poor-quality care, and additional costs. Healthcare organizations are often highly dependent on group purchasing organizations and usually do not have access to details about the critical supply chains.

During the pandemic, the demands for PPE were far higher than supplies, leading to rapid increases in prices for available PPE and serious shortages. Many new organizations took this opportunity to start producing needed PPE, often with little regard for standards and adequate quality. Many of these organizations, both manufacturers and distributors, had no experience in producing and/or distributing these products. Promises were made with no certainty the supplies could be produced or sourced, extreme delays in delivery—or sometimes any delivery—were common. It is also clear, even in non-pandemic situations, that supply chain management is more complex in healthcare than in other industries because of the impact on people’s health, requiring adequate and accurate medical supply according to the patient’s needs (Beier, 1995).

Supply Chain of Personal Protective Equipment

According to the Occupational Safety and Health Administration (OSHA), Personal Protective Equipment (PPE) is defined as equipment that is worn with the objective of minimizing the exposure to hazardous substances. The most common items are gloves, foot and eye protection, protective hearing devices, hard hats, respirators, and full body suits (OSHA, 2004).

PPE plays a critical, major role in healthcare environments and many other industries. PPE is used by professionals, especially healthcare workers, who are exposed to viruses, infections, and disease during their daily activities. The use of PPE is essential to protect healthcare workers from disease spread (CDC, 2004). These protective items like gloves, gowns, masks, respirators, googles, and face shields became especially critical products during the COVID-19 pandemic in healthcare systems.

One of the main characteristics of the supply chain of PPE in the United States is that it is similar to many other goods based on demand. Products are manufactured only in sufficient quantities to cover the anticipated normal demand, leaving little ability to increase production if needed (Patel, 2017). As a result, it becomes extremely difficult to accommodate abrupt increases in the demand, as would happen with a public health crisis similar to the current pandemic, where demand increased rapidly.

Impact of COVID-19 Pandemic in Healthcare Supply Chain and Personal Protective Equipment Shortages

As the coronavirus pandemic spread across the world rapidly, healthcare systems quickly faced major shortages in many countries. Production and distribution of PPE to healthcare frontline workers quickly became one of the main public health challenges during COVID-19.

There are a number of factors that contributed to the massive shortages of PPE. The practice of many healthcare organizations and their purchasing organizations of just-in-time delivery was one of these factors (Rondinone et al., 2021). This practice, which works well in normal time, limited excess inventory and unnecessary costs. However, this model did not have the necessary flexibility to meet the current situation with rapid increases in demand. Suppliers who normally produced based on a predictable and regular demand struggled to change their production processes to meet the rapid changes in demand. The inventory management strategies that meet normal demand cycles need a reconsideration to avoid shortages of essential supplies in medical crises (Patrinley et al., 2020).

Another factor that contributed to the current shortages is the global nature of the medical supply industry. Around 50% of masks used in the world are made in one country, China. As the virus hit China first, the lockdowns there stopped production quickly, causing significant shortages of masks throughout the world (Wang et al., 2020).

U.S. Response to COVID-19 and Personal Protective Equipment Shortages

After the COVID-19 epidemic began, there were no clear ideas for countries to proceed and control the spread of the virus. The spread of the virus is not only through air by coughs and sneezes, but also through contact surfaces as well as community transmission (CDC, 2020; WebMD, n.d.). Different countries responded differently to this situation; preparing PPE for the healthcare frontline workers and general people was a critical and immediate job in every country.

The U.S. government, like other countries, needed proactive responses to ensure the safety of citizens. Travel restriction was one of the first U.S. responses to COVID-19. It was mandatory to confirm that the virus was not carried through travelers from other countries. Screening at the port of entry and mandatory home quarantine was assured for travelers (Department of Homeland Security, 2020).The United States' second action was stay-at-home orders instructing citizens to restrict their commuting outside their homes if not required

(Mervosh et al., 2020). Practicing social distancing and wearing masks in public areas were also recommended by the U.S. government and the CDC (Barthel, 2020; Kopecki, 2020). Many felt that despite these actions, the U.S. government made a delayed response to the COVID-19 crisis (Pew Research Center, 2020).

In addition, the United States faced large PPE shortages during COVID-19. Hospitals and other healthcare providers were faced with extremely high demand for PPE products. The radical increase in demand significantly affected the supply chain network, which caused many shortages of supplies and many disruptions in distribution. Most hospitals had only a few days of PPE on hand when COVID-19 hit their regions, and the national stockpile was exhausted quickly. Most hospitals traditionally relied on third-party organizations for supply chain management, who in turn focused primarily on low-cost supplies primarily manufactured outside of the United States. Many of these offshore suppliers, naturally, had diverted their production to the needs of their own countries. Several U.S. companies that supplied a significant percentage of PPE found that some parts of their supply chains were disrupted (Arangdad et al., 2021).

The United States had not only failed to provide a prompt, transparent response, but it also was not successful in delivering enough PPE to frontline healthcare providers (Zurcher, 2020; Robbins & Garde, 2020). A large number of deaths could had been avoided by timely response (Sebenius & Sebenius, 2020). However, no country was sufficiently prepared for such a pandemic, and the U.S. government is no exception. However, in some ways the U.S. response to COVID-19 has been remarkable. In response to these shortages, many companies, organizations, and universities made great efforts to help with this crisis.

Many apparel companies quickly turned their production lines to making masks and gowns. Nonwoven companies focused the melt blown and spun bond nonwovens facilities and expertise to produce specially designed fabrics that can be delivered to U.S. manufacturers to produce respirators.

Many non-textile manufacturers (e.g., Ford and Honeywell) also responded to contribute to the manufacture of PPE (Ford, 2020; Honeywell, 2021). Many universities also responded to the current critical situation. Even though most universities suspended their normal operations in March, switching to an online format to protect students, professors, staff and some research laboratories remained open to assist with PPE supplies. Institutions around the country have worked hard in the past months to assist with the manufacture of materials required for the construction of PPE, to assemble finished products, or to produce innovative products and technologies as an alternative to the shortages of traditional PPE items.

Management of COVID-19 Data in the United States

The current PPE needs were identified using the data on patients in hospitals, in states, and in regions of the country. In the first few months, PPE needs were further complicated by a lack of testing resources and extremely long turn-around times for test results. This situation created large numbers of “suspected COVID-19 cases” that had to be treated exactly as a known case, creating double or even triple use of PPE. Further complicating the PPE demand estimates was the fact that many PPE were single use, meant to be used once and thrown away. When PPE shortages became critical, many hospitals started reusing these PPE, and many organizations created and published methods for cleaning and reusing what were considered disposable items.

Data helps to evaluate existing situations, predict upcoming crises, and make influential decisions (Lithios, 2020).

The Strategic National Stockpile (SNS) was created in the United States in 2002 through federal funding and other initiatives related to the Public Health Security and Bioterrorism and

Response Act. The expressed purpose was to create a special inventory of items that could provide healthcare workers and other professions with critical supplies during a health crisis. This stockpile included many PPE items such as gloves, masks, and surgical gowns as a result of both real and simulated health emergencies (Yorio et al., 2019). During the recent pandemic, these supplies were quickly exhausted. There was much uncertainty and a lack of unified knowledge in sourcing for specific product components and testing requirements, as well as a lack of connection between product suppliers, healthcare organization needs, and the healthcare systems.

Handfield et al. (2020) have reviewed the current situation of the SNS and made several recommendations for major changes in the SNS. The development of strategic plans was recommended, which can be incorporated in the supply chain. It is essential to use all resources from the federal SNS for PPE manufacturing and exploring the quality of needed products and the required storage space for such products, as well as keeping inventories relevant with the most proper product replacement system (Handfield et al., 2020). The recommendations of Handfield et al. (2020) and others should be a priority for the U.S. Administration and Congress to rethink the SNS.

Prediction Models Based on Personal Protective Equipment Needs

The COVID-19 pandemic has been a fundamental challenge to managing supply and demand on a massive scale. High demand of critical products and shortage in supply had a huge impact on healthcare supply chains. We are just beginning to understand and study what worked and what didn't.

Although the current pandemic has been addressed primarily by the healthcare systems and public health organizations, concepts from operations management shed light on many of the challenges faced by the healthcare industry. It is clear that many of the operational problems during the COVID-19 pandemic stem from the current supply chain management systems that have been primarily designed for cost efficiency. The challenges managing supply chains during a pandemic require flexible systems designed to quickly adapt to a rapidly changing landscape. COVID-19 has taught us that healthcare systems must have more flexible backup systems for supply chains to stymie disease spread and adequately equip providers to care for patients in their time of need (Patrinley et al., 2020).

Three Short Examples

The following three short examples illustrate many of the problems in creating working supply chains for PPE during a pandemic.

One of the first shortages was protective gowns for nurses, patients, and others in healthcare settings. Soon, pictures were appearing in the national press of nurses wearing garbage bags in an effort to be at least partially protected. Although a high percentage of hospital gowns were usually made by a major U.S. company, it was soon discovered that their major supplier of fabric in Central America had been closed, and existing fabric supplies were soon exhausted. Two of the largest cotton yarn manufacturers in the world are located in the United States, and there is actually a large weaving capacity. A number of companies soon created a coalition to respond to the crisis, providing millions of yards of high-quality fabric for the production of gowns. But as soon as low-cost supplies were available from other countries, the gown manufacturers sourced the cheapest fabrics, and the U.S. manufacturers were left holding millions of yards of fabric in unsold inventories.

Face shields were also in extremely short supply during the early days of the pandemic, and some healthcare providers are still reporting difficulties in sourcing the numbers they need. There were many stories of organizations quickly pivoting to make these shields. Many organizations, especially research labs in universities, have numerous 3D printers. Designs for the face shields were rapidly shared, and soon tens of thousands of face shields were being produced. One university even partnered with a nearby medical school whose students worked in its mechanical engineering labs in shifts, producing shields for all the nearby hospitals. Although there were some problems sourcing the polymers for the shields, the main bottleneck quickly became the elastic for the headbands for the shields. The shield makers, almost all not connected to the apparel and textile manufacturing complex, found themselves competing with the manufacturers of surgical face masks and N95 respirators for short-supply, narrow-width elastic fabrics.

The most critical PPE shortage was the N95 respirators needed in healthcare settings. The N95 respirator is designed to achieve a close facial fit and efficiently filter airborne particles when properly worn. The FDA states that the N95 should be discarded after each patient encounter. They should not be shared or reused. The N95 respirator is evaluated, tested, and approved by the National Institute for Occupational Safety and Health (NIOSH) as per the requirements of 42 C.F.R. Part 84. The CDC did not recommend these for the general public since they are critical supplies that should be reserved for healthcare workers and other medical first responders. The shortages of these critical PPE items were so great that many healthcare workers were asked to wear theirs for several days or even weeks or resorted to wearing the less efficient surgical masks or even homemade cloth masks.

The N95 respirator is typically made of three layers of nonwoven materials, two spun bond and one melt blown, with the spun bond layers providing structural integrity and the melt blown layer providing the filtering capability. Although the United States is still the number one producer of nonwovens in the world, almost all plants were already running at close to full capacity prior to the pandemic, with many making other critical healthcare supplies. It was difficult, but not impossible, for these companies to pivot to add the melt blown fabric capabilities to produce the critical filter layer. But almost all of the high-quality nonwoven machinery is made outside the United States; backlogs for the German equipment that could help companies ramp up to meet the N95 needs were 3 to 6 months, even with expedited deliveries. Other equipment was needed to add automated lines to mold the completed respirators into the critical shapes required, to attach the headbands, and to package the completed product. Although these upgrades to the nonwoven manufacturing plants were available—even after a rather long delay—the potential manufacturers were reluctant to make the multi-million-dollar investments with the uncertain market conditions and the almost certain feeling that the healthcare providers would switch quickly to lower-cost offshore products as soon as they were again available (FDA, 2020).

The previous examples illustrate several major issues. Almost all of the critical PPE could be manufactured in the United States and distributed quickly to the healthcare providers and others with critical needs. No producer of the end items controlled the supply chains end-to-end needed to produce these items. Most producers were concerned that critical elements of the supply chain would revert to the lowest cost producers outside the United States as soon as possible. These concerns were well founded. Most potential producers of PPE were concerned that they would not be able to recoup their investments in new equipment or facilities, as purchasers would revert to sourcing the lowest cost products when they could. Many also feared that critical elements of the supply chain would soon be refocused on higher-end products as retail customers returned or companies returned to normal production.

Other countries countered these concerns in several ways. Taiwan created an adequate supply of face masks quickly by providing loans for new equipment that would be forgiven after a specified number of masks were provided to the government. This strategy worked amazingly well. Not only were the supplies created for use in Taiwan quickly, but these companies soon became exporters of these masks to other countries still facing shortages (Jao, 2020). The United Kingdom, also facing severe shortages, used a different approach. The government guaranteed purchase of a large number of critical PPE items if manufacturers quickly ramped up production to provide what was needed. This is almost the identical approach the United States used for accelerating the creation and production of the vaccines by Pfizer. Some countries just made direct grants for needed equipment to encourage companies to become capable of producing needed PPE.

Compounding the problems addressed previously was the lack of reliable test facilities for assuring that the products created or rapidly sourced from new suppliers and distributors met the quality standards created by the FDA, CDC, and NIOSH. Examples of products that failed even the basic measures of quality were common. Sometimes over 80–90% of masks were found to be substandard in a large shipment. Some healthcare providers enlisted labs at local universities to provide basic testing on shipments of suspect quality. Other organizations created in-house facilities to do at least rudimentary testing. The largest provider of the critical N95 mask in the United States sued five vendors who targeted officials in three states by offering nonexistent N95 respirators. Federal agents in the United States seized more than 10 million fake 3M brand N95 masks in early 2021 (Long, 2021). In the United Kingdom, 50 million masks bought by the government were not used by the National Health Service (NHS) because of safety concerns (Kemp, 2020).

COVID-19 and the U.S. Textile Supply Chain

The U.S. textile supply chain is actually far stronger than most people realize. In 2019, the U.S. textile supply chain accounted for 585,000 jobs and $29 billion in exports, with the total value of shipments of man-made fibers, yarns, fabrics, apparel, and non-apparel sewn products around $76 billion. New investments in 2018 (the last year data are available) were $2.5 billion. The United States is the world's second largest exporter of textiles (Glas, 2021).

The economic crisis of COVID-19 led to sharp drops in demand from retail customers, forcing many textile and apparel companies to reduce capacity. Clothing sales fell by 49% in March 2020, 87% in April, and 63% in May. Some textile companies were running at 10% of capacity at the height of the crisis (Glas, 2021). The capabilities of the U.S. textile and apparel supply chains and the available capacities during the COVID-19 pandemic makes one wonder whether even the most basic coordination efforts and leadership by the U.S. government could have created surpluses of critical PPE rather than the shortages which still exist in 2021. We feel that any of the approaches used to stimulate production mentioned previously could have made major impacts.

There are currently many proposals in the U.S. Congress for stimulating the production of PPE in the United States and providing a more stable manufacturing base before the next crisis. Information on some of these initiatives is available at the website created by the National Council of Textile Organizations, *MakeAmericaPPE.org.* One worry is that we will create nationalistic policies to restrict exports when questions arise as to how much inventory resides in stockpiles (Finkenstadt & Handfield, 2021). These are desperate measures taken by governments under pressure by citizens concerned about their own healthcare system shortages and ignore the bigger issues facing governments in the face of a crisis: the lack of global stewardship for combating a world crisis (Finkenstadt & Handfield, 2021).

Discussion and Conclusion

When the demand for products changes rapidly and dramatically, the supply chain is stressed. Normally, products are manufactured only in sufficient quantities to cover the anticipated normal demand, leaving little ability to increase production if needed (Patel et al., 2017)

There was uncertainty and a lack of unified knowledge in sourcing for specific product components, testing requirements, and a lack of connection between product suppliers, needs, and the healthcare systems.

One of the pervasive challenges throughout this pandemic has been the lack of information. Data needed to understand what was needed, when, and where was often totally lacking or of extremely poor quality. Over 6 months into the pandemic, the federal government was still making changes to who was responsible for the data collection and analysis on even such basic information as number of cases, hospitalization, ICU utilization, and deaths (Davenport et al., 2020).

Blindly ordering materials based on flawed best guesses will create many shortages, overstocks, and even wrong supplies similar to what we have experienced with COVID-19.

The second part of our research was to develop an understanding of the capabilities of U.S.-based companies to produce large quantities of specific PPE for both the current pandemic and future needs. Unlike supply chains in the automotive and aerospace industries with well-defined tiers of suppliers with which we were familiar, we found healthcare systems were heavily reliant on group purchasing organizations that are focused almost entirely on the costs of the end items. These organizations have little knowledge of the critical supply chain elements. Future healthcare materials acquisition programs should be based on much deeper understandings of the entire supply chain. Clear specifications not just for end products but also for intermediate materials must be developed. Quality measurement plans must be implemented for each critical phase of the supply chains. There needs to be a strong feedback from end users back through each stage of the supply chain to quickly address problems. We feel that this work has implications for anyone in the healthcare supply chain space (DoD task forces, DLA, DHA, FEMA, DHHS, DHS, state, local, and healthcare systems). Even beyond the healthcare supply chain space, this work should be of use to anyone concerned about better approaches to prepare for contingency sourcing.

References

Arangdad, S., Godfrey, A. B., Hoque, A., Joshi, P., Prasath Manoharan, A., & Vhanmane, A. (2021). *COVID 19: Situation, action, and tracking of PPE demand in the US* [Manuscript submitted for publication].

Barthel, M. (2020, April 21). *These grocery stores are the best at promoting social distancing, according to local app*. WAMU. https://wamu.org/story/20/04/21/these-grocery-stores-are-the-best-at-promoting-social-distancing-according-to-local-app/

Beier, F. J. (1995). The management of the supply chain for the hospital pharmacies: A focus on inventory management practices. *Journal of Business Logistics*, *16*(2), 153–173.

Centers for Disease Control and Prevention. (2004, June). *Guidance for the selection and use of personal protective equipment (PPE) in healthcare settings*. https://www.cdc.gov/hai/pdfs/ppe/ppeslides6-29-04.pdf

Centers for Disease Control and Prevention. (2020, February 18). *Coronavirus disease 2019*. https://www.cdc.gov/media/releases/2020/p0117-coronavirus-screening.html

Davenport, T. H., Godfrey, A. B., & Redman, T. C. (2020). To fight pandemics, we need better data. *MIT Sloan Management Review*, *62*(1), 1–4.

Department of Homeland Security. (2020, March 13). *Department of Homeland Security outlines new process for Americans returning from certain European countries, China, and Iran*.

Food and Drug Administration. (2020). *N95 respirators, surgical masks, and face masks*. https://www.fda.gov/medical-devices/personal-protective-equipment-infection-control/n95-respirators-surgical-masks-and-face-masks

Finkenstadt, D. J., & Handfield, R. (2021, March 28). Blurry vision: Supply chain visibility for personal protective equipment during COVID-19. *Journal of Purchasing and Supply Management*.

Ford. (2020, April 13). *Ford to produce respirators, masks for COVID-19 protection in Michigan*. https://corporate.ford.com/articles/products/ford-producing-respirators-and-masks-for-covid-19-protection.html

Glas, K. (2021, March 19). An in-depth overview of the current PPE landscape. *Textile World*. https://www.textileworld.com/textile-world/2021/03/an-in-depth-overview-of-the-current-ppe-landscape/

Handfield, R., Finkenstadt, D. J., Schneller, E. S., Godfrey, A. B., & Guinto, P. (2020). A commons for a supply chain in the post-COVID-19 era: The case for a reformed strategic national stockpile. *The Milbank Quarterly*, *98*(4), 1058–1090.

Holshue, M. L., DeBolt, C., Lindquist, S., Lofy, K. H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., Diaz, G., Cohn, A., Fox, L., Patel, A., Gerber, S. I., Kim, L., Tong, S., Lu, X., Lindstrom, S., … Pillai, S. K. (2020). First case of 2019 novel coronavirus in the United States. *New England Journal of Medicine*, *382*(10), 929–936. https://doi.org/10.1056/NEJMoa2001191

Honeywell. (2021, January 11). *Honeywell delivers 225 million N95 and surgical masks in December for health care and government use* [Press release]. https://www.honeywell.com/us/en/press/2021/1/honeywell-delivers-225-million-n95-and-surgical-masks-in-decembe

Iyengar, K. P., Vaishya, R., Bahl, S., & Vaish, A. (2020). Impact of the coronavirus pandemic on the supply chain in healthcare. *British Journal of Healthcare Management*, *26*(6), 1–4.

Jao, N. (2020, April 13). "Mask diplomacy" a boost for Taiwan. *Foreign Policy*. https://foreignpolicy.com/2020/04/13/taiwan-coronavirus-pandemic-mask-soft-power-diplomacy/

Johns Hopkins University & Medicine. (n.d.). *COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Johns Hopkins Coronavirus Resource Center. Retrieved January 27, 2021, from https://coronavirus.jhu.edu/map.html

Joseph, A. (2020, January 13). *Woman with novel pneumonia virus hospitalized in Thailand—The first case outside China*. STAT. https://www.statnews.com/2020/01/13/woman-with-novel-pneumonia-virus-hospitalized-in-thailand-the-first-case-outside-china/

Kemp, P. (2020, August 6). *Coronavirus: Safety concerns halt use of 50 million NHS masks*. BBC News.

Kim, S. H., & Kwon, I. W. G. (2015). The study of healthcare supply chain management in United States: Literature review. *Management Review: An International Journal*, *10*(2), 34.

Kopecki, D. (2020, March 15). *CDC recommends canceling events with 50 or more people for the next eight weeks throughout US*. CNBC. https://www.cnbc.com/2020/03/16/cdc-recommends-the-cancellation-of-events-with-50-or-more-people-for-the-next-eight-weeks-throughout-us.html

Long, C. (2021, February 17). *Feds seize 10 million fake N95 masks in growing COVID-19 investigation*. Associated Press.

Mervosh, S., Lu, D., & Swales, V. (2020, March 24). See which states and cities have told residents to stay at home. *The New York Times*. https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html

Occupational Safety and Health Administration. (2004). *Personal protective equipment*. https://www.osha.gov/Publications/osha3151.pdf

Patel, A., D'Alessandro, M. M., Ireland, K. J., Burel, W. G., Wencil, E. B., & Rasmussen, S. A. (2017). Personal protective equipment supply chain: Lessons learned from recent public health emergency responses. *Health Security, 15*(3), 244–252. https://doi.org/10.1089/hs.2016.0129

Patrinley, J. R., Berkowitz, S. T., Zakria, D., Totten, D. J., Kurtulus, M., & Drolet, B. C. (2020). Lessons from operations management to combat the COVID-19 pandemic. *Journal of Medical Systems*, *44*(7), 1–2.

Pew Research Center. (2020, April 16). *Most Americans say Trump was too slow in initial response to coronavirus threat*. Pew Research Center. https://www.pewresearch.org/politics/2020/04/16/most-americans-say-trump-was-too-slow-in-initial-response-to-coronavirus-threat/

Robbins, R., & Garde, D. (2020, September 18). *An expert's take on what the U.S., UK did wrong in Covid-19 communications—And what others did right*. STAT. https://www.statnews.com/2020/09/18/an-experts-take-on-what-the-u-s-u-k-did-wrong-in-covid-19-communications-and-what-others-did-right/

Rondinone, D., Godfrey, A. B., Anil Khamkar, A., Padmanabhan Anilkumar, A., Jose, A., Rucha Girgaonkar, R., & Arangdad, S. (2021). *The United States supply chain challenges for personal protective equipment (PPE) during the COVID-19 pandemic* [Manuscript submitted for publication].

Scher, I. (2020, March 13). *The first COVID-19 case originated on November 17, according to Chinese officials searching for "patient zero."* Business Insider. https://www.businessinsider.com/coronavirus-patients-zero-contracted-case-november-2020-3

Schnirring, L. (2020, January 16). *Japan has 1st novel coronavirus case; China reports another death*. CIDRAP. https://www.cidrap.umn.edu/news-perspective/2020/01/japan-has-1st-novel-coronavirus-case-china-reports-another-death

Sebenius, I., & Sebenius, J. K. (2020, June 19). *A faster response could have prevented most U.S. Covid-19 deaths*. STAT. https://www.statnews.com/2020/06/19/faster-response-prevented-most-us-covid-19-deaths/

Lithios. (2020, April 23). *The importance of coronavirus data tracking*. Lithios Apps. https://lithiosapps.com/the-importance-of-coronavirus-data-tracking/

Wang, X., Zhang, X., & He, J. (2020). Challenges to the system of reserve medical supplies for public health emergencies: Reflections on the outbreak of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic in China. *Biosci Trends*, *14*(1), 3–8.

Winsor, M., Shapiro, E., & Deliso, M. (2020, April 27). *US reaches 100,000 coronavirus deaths*. ABC News. https://abcnews.go.com/Health/coronavirus-updates-covid-19-cases-us-health-care/story?id=70897864

WorldOMeter. (n.d.). *Coronavirus update (Live): 100,284,461 cases and 2,149,460 deaths from COVID-19 virus pandemic*. Retrieved January 26, 2021, from https://www.worldometers.info/coronavirus/

Yorio, P. L., Rottach, D. R., & Dubaniewicz, M. (2019). Quality assurance sampling plans in US stockpiles for personal protective equipment. *Health Security*, *17*(2), 140–151.

Zurcher, A. (2020, May 13). *Coronavirus response: Things the US has got right—And got wrong*. BBC News. https://www.bbc.com/news/world-us-canada-52579200

PANEL 16. MANAGING TECHNOLOGY ACQUISITION

Wednesday, May 12, 2021	
1:45 p.m. – 3:00 p.m.	**Chair: Lieutenant General Neil Thurgood, USA,** Director of Hypersonics, Directed Energy, Space & Rapid Acquisition ***Exploring Performance in Science and Technology (S&T) Programs*** Eric Plack, Air Force Cost Analysis Agency Edward White, Air Force Institute of Technology Jonathan Ritschel, Air Force Institute of Technology Clay Koschnick, Air Force Institute of Technology Scott Drylie, Air Force Institute of Technology ***Strategies for Addressing Uncertain Missions and Uncertain Technologies*** William Rouse, Professor, Georgetown University Dinesh Verma, SERC/Stevens Institute of Technology Edward Hanawalt, General Motors ***Analyzing Digital Transformation using the Zachman Framework and SysML*** Mark Kassan, USAF AFMC/ENS John Colombi, Air Force Institute of Technology

Lieutenant General Neil Thurgood, USA—is the Director for Hypersonics, Directed Energy, Space and Rapid Acquisition, Office of the Assistant Secretary of the Army (Acquisition, Logistics and Technology), Redstone Arsenal, Alabama. He assumed duties in March 2019.

In this position, LTG Thurgood is responsible for the rapid fielding of select capabilities to deter and defeat rapidly modernizing adversaries, including overseeing development of an Army Long Range Hypersonic Weapon. He leads the Army Rapid Capabilities and Critical Technologies Office mission to rapidly and efficiently research, develop, prototype, test, evaluate, procure and field critical enabling technologies and capabilities that address immediate, near-term, and mid-term threats, consistent with the Army's modernization priorities.

LTG Thurgood most recently served as the Director for Test, Missile Defense Agency, Redstone Arsenal, Alabama. LTG Thurgood most recently deployed from 2017-2018, when he served as Deputy Commander, Combined Security Transition Command-Afghanistan, Operation Resolute Support/Operation Freedom's Sentinel. Prior to his deployment, he served as the Deputy for Acquisition and Systems Management, Office of the Assistant Secretary of the Army (Acquisition, Logistics, and Technology), in Washington, D.C.

LTG Thurgood enlisted in the U.S. Army in 1983. Following his commissioning in 1986 as an Aviation Branch Officer, he served in multiple company grade and battalion aviation positions in both the U.S. and overseas, including multiple combat deployments. LTG Thurgood was then selected and served in the 160th Special Operations Aviation Regiment (Airborne) as a Platoon Leader, Operations Officer and Company Commander.

After transitioning into the Army Acquisition Corps in 1995, he served in various program offices for conventional and special programs. As a Project Manager, LTG Thurgood served in the

Utility Helicopters Office, and as a Program Executive Officer, LTG Thurgood led the PEO for Missiles and Space, at Redstone Arsenal, Alabama. LTG Thurgood participated in operations supporting Operation Enduring Freedom in Afghanistan and Operation Iraqi Freedom in Iraq.

LTG Thurgood holds undergraduate degrees in Business from the University of Utah; a master's degree in Systems Acquisition Management from the Naval Postgraduate School; a master's degree in Strategic Studies from the Air University, Air War College; and a doctorate in Strategic Planning and Organizational Leadership from the University of Sarasota, as well as several professional certifications.

Exploring Performance in Science and Technology Programs

Johnathan Ritschel—Air Force Institute of Technology

Scott Drylie—Air Force Institute of Technology

Clay Koschnick—Air Force Institute of Technology

Eric Plack—Air Force Cost Analysis Agency

Edward White—Air Force Institute of Technology

Abstract

Science and Technology (S&T) programs serve an important function in the defense acquisition process as the initial phase leading to discovery and development of warfighting technology. The results of these programs impact the larger Major Defense Acquisition Programs that integrate the technologies in subsequent phases of the lifecycle. Despite this important role, little prior research has examined the performance of S&T programs. Therefore, we investigate the impact of technological maturation as a critical success factor in S&T programs. The results suggest that S&T programs with mature technologies are more likely to experience above average cost growth and larger contract values while being less likely to experience schedule growth. Additionally, we find the partnership method between the government and contractor matters for both technological maturation and schedule growth. Lastly, the nature of the S&T program is important, with aerospace programs more likely to technologically mature than human systems programs.

An Analysis of Science and Technology Program Performance

Program management focuses on cost, schedule, and performance as the three key measures of success (Meredith & Mantel, 2003; Pinto & Slevin, 1998). A large body of literature identifies critical factors that lead to program success in both private industry (Nasir & Sahibuddin, 2011; Pinto & Slevin, 1987; Zwikael & Globerson, 2006) and the public sector (Rendon, 2012; Rodriguez-Segura et al., 2016; Tishler et al., 1996). Prior analyses of program performance in defense programs, however, have focused almost exclusively on larger, more mature programs that have reached the Engineering Manufacturing Development (EMD) phase of the lifecycle or beyond. An abundance of studies exploring cost growth or schedule growth can be found for these Major Defense Acquisition Programs (MDAPs; Bolten et al., 2008; Cancian, 2010; Smirnoff & Hicks, 2008). Missing from the literature is an exploration of smaller programs that feed basic science and technologies to subsequent acquisition programs or that develop new systems and technologies on a smaller scale. These are the Science and Technology (S&T) programs that are undertaken in defense research labs. This article seeks to bridge that gap through an exploratory analysis of program performance in Air Force S&T programs.

Importance of Science and Technology

The vision to implement science and technology as a centerpiece of our nation's airpower strategy has been around since 1945 (Duffner, 2000). General H. H. "Hap" Arnold, commanding general of the Army Air Forces, enlisted the aid of leading aeronautics scientist Dr. Theodore von Karman to lead the first of these efforts, recommending the creation of an agency devoted exclusively to aeronautical research and development (Gorn, 1988). Over time, that agency has evolved to what is known today as the Air Force Research Laboratories (AFRL; Duffner, 2000).

S&T's enduring importance is demonstrated in the 2019 publication of the Air Force Science and Technology Strategy for 2030. The 2030 S&T strategy aligns with the National

Defense Strategy to empower S&T programs to develop and deliver warfighting capabilities rapidly and effectively (U.S. Air Force, 2019). How does S&T fulfill this need? S&T functions as the initial phase of the acquisition process by which technologies are matured and, where appropriate, transitioned for acquisition by the Air Force (Office of the Chief Scientist of the U.S. Air Force, 2010). Continual advancement in these cutting-edge technologies is crucial, as the Air Force faces ever-changing threats and adversarial advancements in technology.

The Anatomy of Air Force Research Labs

The S&T data analyzed in this paper are from AFRL programs. A brief organizational description is provided for those unfamiliar with the laboratories. AFRL is headquartered at Wright-Patterson Air Force Base (AFB) in Ohio. It is comprised of nine technology directorates in the continental United States and four locations overseas in Hawaii, the United Kingdom, Chile, and Japan, as shown in Figure 1.

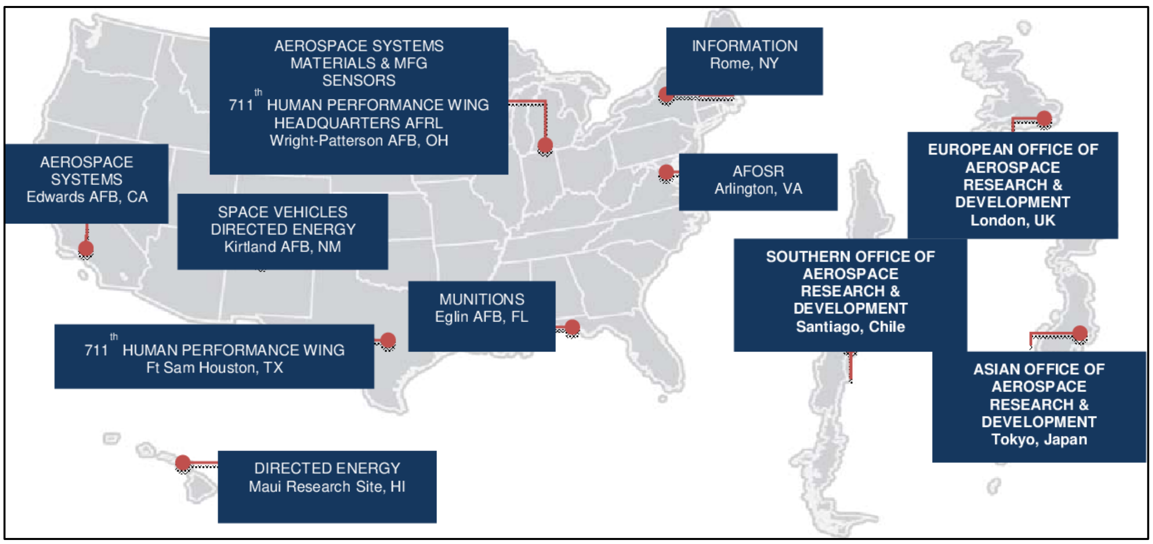


Figure 1: AFRL Locations and Major Offices

Each technology directorate focuses on the development and innovation of leading-edge technologies and is separated by technological capabilities. A list of AFRL's technology directorates, office symbols, and program descriptions are provided in Table 1. The analysis of individual technical directorates will be one of the ways this research segments the data.

Table 1: AFRL Technology Directorates

Technology Directorate	**Symbol**	**Program Descriptions**
Air Force Office of Scientific	AFOSR	Basic Research Manager for AFRL
711th Human Performance Wing	RH	Aerospace Medicine, Human Systems Integration
Directed Energy Directorate	RD	Laser, Electromagnetics, Electro-Optics
Information Directorate	RI	Information Fusion, Exploitation, Networking
Aerospace Systems Directorate	RQ	Aerodynamics, Flight Control, Engines, Propulsion
Space Vehicle Directorate	RV	Space-Based Surveillance, Capability Protection
Munitions Directorate	RW	Air-Launched Munitions
Materials & Manufacturing	RX	Aircraft, Spacecraft, Missiles, Rockets
Sensors Directorate	RY	Sensors for Reconnaissance, Surveillance

Measures of Success: The Role of Technology Readiness Levels

The General Accounting Office (GAO) has identified technology maturation as a critical success factor in product development (GAO, 1999). The Department of Defense's (DoD's) approach to incorporate this critical success factor has been to emphasize Technology Readiness Levels (TRLs) as a measure for selecting mature technologies for inclusion in a program (DoD, 2011). The TRL concept was developed by NASA (Sadin et al., 1989) and has subsequently been adopted by AFRL. A TRL is a tool to measure the technology maturity of a system or subsystem using a nine-level ordinal scale (DoD, 2011). Detailed TRL definitions and descriptions can be found in Appendix A.

It is believed that "programs that enter the Engineering and Manufacturing Development (EMD) phase of the Defense Acquisition System and have immature technologies will incur cost growth and schedule slippage" (DoD, 2009). In an effort to reduce the risk associated with entering the EMD phase of the acquisition lifecycle at Milestone B, DoD Instruction 5000.02 requires technologies to be demonstrated in a relevant environment (i.e., obtain a TRL of at least 6; DoD, 2011). AFRL, through the S&T programs it oversees, serves a key role in the creation and maturation of these technologies to reach those thresholds.

Despite TRLs being identified as a critical success factor, the literature is sparse with empirical examinations. The dearth of analysis is particularly acute for S&T type programs, but even MDAPs have relatively few studies examining TRLs. Dubos et al. (2008) analyzed the relationship between technology uncertainty and schedule slippage in the space industry. Their research resulted in the creation of TRL-schedule-risk curves that are intended to assist program managers to make informed decisions regarding the appropriate TRL to consider when confronted with schedule constraints. The research of Dubos et al. (2008) suggested a close relationship between technology uncertainty and schedule risk, where the more mature a technology is (the higher the TRL), the less potential schedule slippage.

Katz et al. (2015) specifically studied the relationship of TRLs to cost and schedule changes during the EMD phase. They found that weapon systems that achieved a TRL of 7 or greater at Milestone B had a lower probability of schedule slippage during the EMD phase than weapons systems that had a TRL of less than 7. While Katz et al. (2015) found evidence to suggest that technology maturity is related to schedule change, they did not find a relationship with cost changes.

Smoker and Smith (2007), however, found evidence that suggests costs vary exponentially across time as the system's technology progresses through each TRL. Similarly, Linick (2017) found that as the TRL increased throughout the development phase, the percentage of the development cost increased at an increasing rate. As shown by the literature, the extant TRL studies are primarily focused on programs once they reach the EMD stage. To the best of our knowledge, there are no studies that focus solely on S&T programs—a gap this paper is designed to fill.

Data

The data for this research was obtained from the AFRL cost and economics division. S&T programs typically fall below the dollar threshold for traditional standardized reporting such as Contract Performance Reports (CPRs). Instead, the S&T programs receive Funds and Man-Hour Expenditure Reports (FMERs). These FMERs provide the procuring activity visibility into the contractor's expenditures for labor, materials and parts, travel, subcontractors, and other charges. Like CPRs, these reports are required on a periodic basis from the contractor, usually monthly. Unlike CPRs, FMERs do not report standardized cost elements like the ones found in MIL-STD-881D. The initial AFRL dataset consisted of 165 S&T programs with contract start dates spanning from 2009 to 2017.

Research Summary Reports were also collected for these programs. These reports are generated at the start of the program (Initial), during the program (Periodic), and at the end of the program (Final). Research Summary Reports include general information such as the program title, lead technical directorate, and start/end dates. They also include DoD-required information such as performance type, joint capability area, Air Force technical capabilities, and TRLs. An example of a Research Summary Report can be found in Appendix B.

Of the 165 programs obtained from AFRL, 43 are included in the final dataset. Table 2 provides the exclusion criteria and associated number of programs remaining in the analysis.

Table 2. Dataset Exclusions

Category	Number Removed	Remaining Programs
Programs Obtained from AFRL		165
Missing Elements	64	101
Inadequate TD Sample Size	10	91
Less Than 92.5% Complete	48	43
Final Dataset for Analysis		43

As shown in Table 2, programs which had missing elements are excluded. These 64 programs had their costs reported on the FMER in unique ways to include cost burn rates, earned value management graphs, total costs in phases, or simply an overall total cost or labor hours spent. These reporting methods lack the specific elements needed in this analysis to compute percentages of total cost which are used to observe the program's behavior. Of the 101 remaining programs, 10 programs fall under four different technical directorates (RD, RI, RX, and RY). Each technical directorate represents unique programs with different characteristics which precluded aggregation above the technical directorate level. Therefore, the small sample size in these directorates would likely skew the analysis results, especially when observing how these programs behave at the technical directorate level. For these reasons, the programs are excluded from the analysis. Finally, programs with a completion percentage of less than 92.5% are excluded from the dataset. A program's completion percentage is computed using the total cost from the last available FMER to the program's contract value at that time. Previous research determined that a program with a completion percentage of 92.5% or greater accurately predicts the final cost of the program (Tracy & White, 2011). The final number of programs in the dataset is 43, which is sufficient to conduct a robust analysis.

Methods: Contingency Table Analysis

The dataset consists largely of qualitative variables. Therefore, the methodological approach employed is a two-way contingency table analysis. This type of analysis is used to summarize the relationship between two categorical variables based on the data observed. The contingency table analysis uses a 2 × 2 table to test for independence. For each test, the same type of hypothesis test will be implemented, as shown in Equation 1:

$$H_o\text{: The two classifications are independent}$$
$$H_a\text{: The two classifications are dependent} \quad (1)$$

The chi-square distribution is the test statistic used for considering inferences about the category probabilities. If there is a failure to reject the null, the two variables are independent and are not statistically related to one another. If the null is rejected, then the variables are dependent, and a statistical relationship exists between them. The two-way contingency analysis examines the categorical variables (see Table 3) with subsequent discussion on the rationale behind variable selection and categorization.

When highly significant results are found, one of the benefits of a contingency table is that odds ratios and their associated confidence intervals can be produced. An odds ratio is a measure of association for a two-way contingency table. The ratio is the odds of an event occurring in one group to the odds of the same event occurring in another group. In other words, the odds ratio is the ratio of the probability of a property being present compared to the probability of it being absent. If the odds ratio is 1, the two events are independent.

Table 3. Categorical Variables Used in Contingency Table Analysis

Categorical Variables	
Technical Directorate	Cost Growth > 0%
Performance Type	Cost Growth > 33.7%
TRL Increase	Cost Growth > 44.1%
Last Known TRL ≥ 6	Cost Growth > 56.5%
Final TRL ≥ 6	Cost Growth > 60.5%
TRL 1 – 3	Cost Growth > 68%
TRL 4 – 5	Contract Value > $1M
TRL 6 – 7	Contract Value > $3M
TRL 8 – 9	
Schedule Growth > 0%	
Schedule Growth > 33%	
Schedule Growth > 63%	

Categorical variables for the Technical Directorate (TD), Performance Type, and TRLs are obtained from the Research Summary Reports. The TD variable denotes which AFRL directorate is the lead on the program. Such a variable may capture organizational/managerial/technological differences. For this dataset, the TD variable is either RH or RQ. (This limitation is due to the sample size of the other TDs as previously discussed.) The performance type represents the partnership method between AFRL and the contractor. This variable consists of Research, Development, Test & Evaluation (RDT&E) and Small Business Innovative Research (SBIR) relationships. This type of variable may capture differences due to the size, skills, or knowledge of the company types (e.g., small versus large companies). TRL data for the S&T programs are used in seven different categorical variables. TRL Increase indicates if the TRL increases at any point during the program's lifecycle. Last Known TRL ≥ 6 denotes the last reported TRL of the program, while Final TRL ≥ 6 only analyzes programs that have a Final Research Summary Report. The decision to categorize based on TRL level 6 is due to the role this TRL level fulfills in the defense acquisition process. Specifically, a TRL of 6 is equivalent to demonstration in a relevant environment which is needed for a program to enter Milestone B (DoD, 2011). Four variables were created by grouping TRLs based on the maturity of the technology and the product's requirements, as determined in the literature (GAO, 1999). See Figure 2.

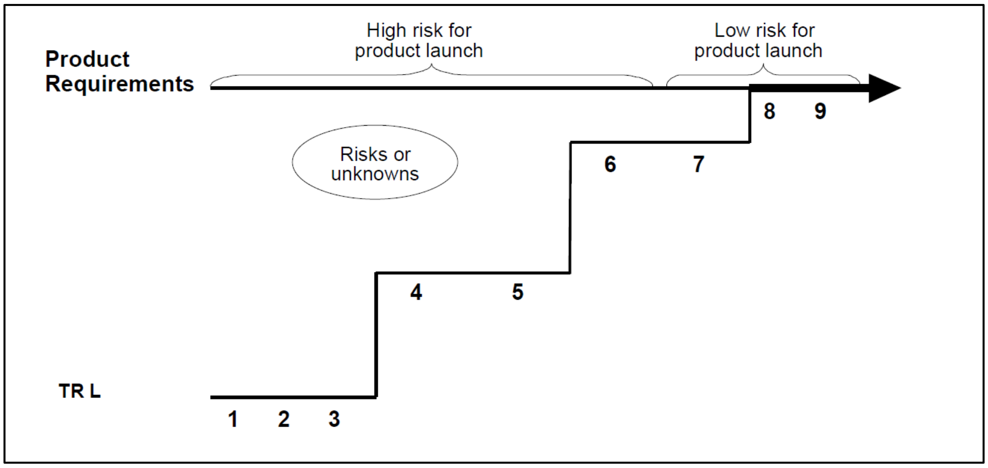


Figure 2. Using TRLs to Match Technology With Requirements
(GAO, 1999)

Additional variables of interest created from the Research Summary Report contract information include schedule growth, cost growth, and contract value. These attributes are commonly studied for acquisition programs at all phases of their lifecycles.

The variables for cost growth, schedule growth, and contract value have been converted from continuous variables to categorical variables in order to be included in the contingency table analysis. Binary (or dummy) variables with methodical break points were created in order to test the relationships at different locations. These breakpoints were derived either from the literature review or from descriptive statistics of the variable itself in the dataset with its mean and/or median. For example, the mean cost growth of the dataset was 68%, which led to the creation of a dummy variable (Cost Growth > 68%), separating programs that are above and below the sample mean. Likewise, Bolten et al. (2008) distinguished mean and median percentages of total DoD and Air Force acquisition program development cost percentages. These thresholds from Bolten et al. (2008) are also examined. A summary of the break points can be seen in Table 4.

Table 4. Break Point Summary

Category	Break Point	Reason	Source
Schedule Growth	0%	Any growth	Dataset
	33%	Median	Dataset
	63%	Mean	Dataset
Cost Growth	0%	Any growth	Dataset
	33.7%	DoD Development – Median	Bolten et al. (2008)
	44.1%	Air Force Development - Median	Bolten et al. (2008)
	56.5%	DoD Development – Mean	Bolten et al. (2008)
	60.5%	Air Force Development - Mean	Bolten et al. (2008)
	68%	Mean	Dataset
Contract Value	$1M	Median	Dataset
	$3M	Mean	Dataset

Results and Discussion

The contingency table results are organized into four sections: technical directorate, performance type, TRL, and growth relationships. Using the chi-square distribution as the test statistic, relationships are identified when Pearson's chi-squared test is significant at a p-value of less than 0.10. For highly significant results (p-value < 0.01), the odds ratio and its associated confidence interval are analyzed. It is important to note the possibility of spurious relationships. Spurious relationships occur when the two variables are associated but not causally related, possibly due to an unknown mediating variable. With the sheer number of 2 × 2 tables generated in this analysis, spurious relationships are possible. Therefore, only highly statistically significant results (p-value < 0.01) will be studied in detail (i.e., full contingency table shown), while the other significant variables are observed solely as potential findings.

Technical Directorate

The TD categorical variable denotes which AFRL directorate is the lead on the respective program: either RH (Airman Systems) or RQ (Aerospace Systems). Analyzing the TD variable resulted in 19 contingency tables to be tested for significance. Two variables were significant at an alpha of 0.10, and two were significant at an alpha of 0.05. The full set of test results is provided in Table 5.

Table 5. Contingency Table Results for Technical Directorate

Variable	TD
Performance Type	
TRL Increase	**
Last Known TRL ≥ 6	
Final TRL ≥ 6	
TRL 1-3	
TRL 4-5	
TRL 6-7	
TRL 8-9	
Schedule Growth > 0%	
Schedule Growth > 33% (Median)	**
Schedule Growth > 63% (Mean)	*
Contract Value > $1.0M (Median)	
Contract Value > $3.0M (Mean)	
Cost Growth > 0%	*
Cost Growth > 33.7% (DoD Dev - Median)	
Cost Growth > 44.1% (AF Dev - Median)	
Cost Growth > 56.5% (DoD Dev - Mean)	
Cost Growth > 60.5% (AF Dev - Mean)	
Cost Growth > 68% (Mean)	
Total Significant Contingency Tables:	**4**

Table Legend:
* **p-value < 0.10**
** **p-value < 0.05**
*** **p-value < 0.01**

TRL Increase is the only TRL variable with a statistically significant relationship to TD. This test suggests that it is more probable to have a program's TRL increase with RQ (Aerospace Systems) programs than with RH (Airman/Human Systems) programs. The RQ programs are comprised primarily of engine and propulsion (hardware) system technologies. The ability to transition RQ through TRL levels may be due to the relationship of hardware versus software (human systems interactions). It is likely easier to make advancements in hardware technologies as the testing, failures, and efficiencies may be more conclusive.

Similarly, the contingency table results suggest that RQ programs are more probable to have cost growth as well as schedule growth that is greater than 33% (the dataset's median) and 63% (the dataset's mean). This could be related to the maturing technology (increasing the TRL) of RQ programs. If the technology is maturing, a program office may be more likely to increase funding and schedule to keep the maturation on track. If the technologies do not mature, it could be that the agile nature of S&T programs allows for early decisions to cancel programs. In summary, the TD results suggest that RQ programs are more likely to technologically mature, have cost growth, and have schedule growth (greater than the dataset mean and median) when compared to RH programs.

Performance Type

The performance type variable represents the partnership method between AFRL and the contractor: either Research, Development, Test & Evaluation (RDT&E) or Small Business Innovative Research (SBIR) relationships. This variable formed 19 contingency tables to be tested for significance. One variable was significant at an alpha of 0.10, two variables were significant at an alpha of 0.05, and two variables were significant at an alpha of 0.01. The full set of test results is provided in Table 6.

Table 6. Contingency Table Results for Performance Type

Variable	Performance Type
TD	
TRL Increase	
Last Known TRL ≥ 6	**
Final TRL ≥ 6	**
TRL 1-3	
TRL 4-5	
TRL 6-7	
TRL 8-9	
Schedule Growth > 0%	*
Schedule Growth > 33% (Median)	
Schedule Growth > 63% (Mean)	
Contract Value > $1.0M (Median)	***
Contract Value > $3.0M (Mean)	***
Cost Growth > 0%	
Cost Growth > 33.7% (DoD Dev - Median)	
Cost Growth > 44.1% (AF Dev - Median)	
Cost Growth > 56.5% (DoD Dev - Mean)	
Cost Growth > 60.5% (AF Dev - Mean)	
Cost Growth > 68% (Mean)	
Total Significant Contingency Tables:	**5**

Table Legend:
* p-value < 0.10
** p-value < 0.05
*** p-value < 0.01

Table 6 test results suggest that an S&T program with an RDT&E performance type is more likely to have or end with a TRL of at least 6 than an SBIR type program is. SBIR programs are developed by small domestic businesses, which potentially provides an agile way to stimulate high-tech innovation. But RDT&E programs are dominated by the larger, more experienced defense contractors. These results suggest that the larger defense contractors may obtain contracts with more mature technologies due to their capacity and ability to develop these technologies when compared to SBIR businesses.

Furthermore, as a potential indication of RDT&E and SBIR working different kinds of programs from the start, one can observe that it is more probable to have contract values greater than $1 million (the dataset's median) with RDT&E performance types, as seen in Figure 3. Testing significance when the contract value is greater than $3 million (the dataset's mean) produces similar results to Figure 3, with an even smaller p-value. Again, this could be due to the differences in the types of contractors involved in RDT&E and SBIR programs. Larger defense contractors possibly obtain larger programs because they have more breadth of experience or capacity, while the small businesses obtain smaller contracts with a more constrained objective; the acquisition community often sees a similar relationship when the large defense contractors are prime on a large system and smaller vendors are subcontractors for a particular subsystem. Additionally, SBIR programs may target uncertain and risky technologies that small businesses research so that AFRL can evaluate which programs have the potential to develop into mature technologies. The scale of these uncertain programs may contribute to lower contract values. In fact, the odds ratio indicates that given the program has an SBIR performance type, the odds of the contract value being less than $1 million is 9.7 times higher than when the program has an RDT&E performance type.

Contingency Table

Last Contract Value>$1M (Med)

Performance Type (Count / Expected)	0	1	Total
RDTE	3	12	15
	7.69231	7.30769	
SBIR	17	7	24
	12.3077	11.6923	
Total	20	19	39

Tests

N	DF	-LogLike	RSquare (U)
39	1	5.0265877	0.1860

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	10.053	0.0015*
Pearson	9.547	0.0020*

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
0.102941	0.022042	0.480764

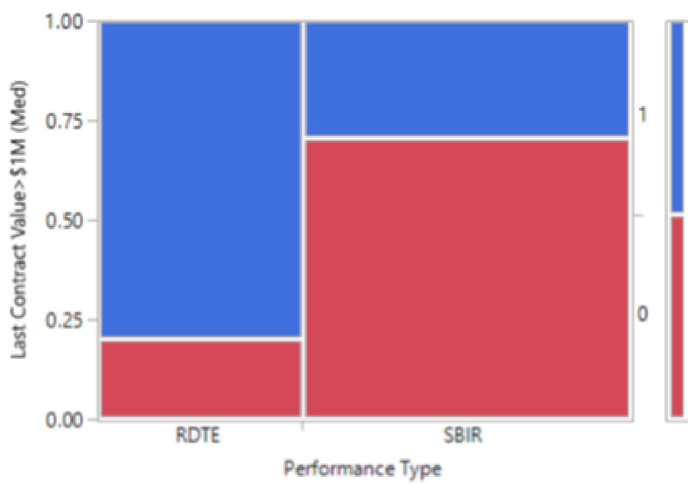


Figure 3. Contingency Table of Performance Type by Contract Value > $1 Million

The Table 6 contingency test results also suggest that a program with an SBIR performance type is more likely to have schedule growth. While test results indicate that RDT&E programs are more likely to have higher TRL levels, the opposite could be said in that SBIR programs are more likely to have lower TRL levels. Less is known about these immature technologies, which could lead these small businesses to spend more time developing them, leading to schedule slippage. This result is consistent with the literature findings of Dubos et al. (2008).

In summary, the results suggest that a program that has a performance type of RDT&E is more likely to have a TRL of 6. Furthermore, highly significant results point to evidence that a program that has a performance type of RDT&E is more likely to have a contract value greater than $1 million. Lastly, the results suggest that SBIR programs are more likely to experience schedule growth.

Technology Readiness Level

TRL data was used to create seven different binary variables, as previously discussed. These seven TRL variables were tested for significance against the 11 performance variables to produce 77 contingency tables. Seven variables were significant at an alpha of 0.10, four variables were significant at an alpha of 0.05, and one variable was significant at an alpha of 0.01. Despite registering significant Pearson p-values, the contingency table results for the seven significant variables at an alpha of 0.10 were found to be invalid. For all seven tests, the expected counts of two of the four cells were less than five. This violates an assumption for a valid chi-squared contingency table test, which states the sample size should be large enough so that the estimated expected count will be equal to five or more. As a further check, Fisher's Exact Test—which is a non-parametric test for small samples—found all seven tests to be non-significant. This result was largely due to the small number of programs with a TRL of 6–7 (5) and a Final TRL of ≥ 6 (4). The full set of test results is provided in Table 7 with special subscript designators on those test results deemed invalid.

Table 7. Significant Contingency Tables for Technology Readiness Level

Variable	TRL Increase	Last Known TRL ≥ 6	Final TRL ≥ 6	TRL 1-3	TRL 4-5	TRL 6-7	TRL 8-9
Schedule Growth > 0%		**	$*_1$			$*_1$	
Schedule Growth > 33% (Median)							
Schedule Growth > 63% (Mean)							
Contract Value > $1.0M (Median)				**			
Contract Value > $3.0M (Mean)		**				***	
Cost Growth > 0%						$*_1$	
Cost Growth > 33.7% (DoD Dev - Median)						$*_1$	
Cost Growth > 44.1% (AF Dev - Median)						$*_1$	
Cost Growth > 56.5% (DoD Dev - Mean)						$*_1$	
Cost Growth > 60.5% (AF Dev - Mean)						$*_1$	
Cost Growth > 68% (Mean)						**	
Total Significant Contingency Tables:	**0**	**2**	**1**	**1**	**0**	**8**	**0**

Table Legend:
- $*_1$ p-value < 0.10, 50% of Expected Counts < 5, Non-significant Fisher's Exact Test
- * p-value < 0.10
- ** p-value < 0.05
- *** p-value < 0.01

The contingency table results suggest that an S&T program is *more* likely to have cost growth greater than 68% (the dataset's mean) with a TRL of 6 or 7 but *less* likely to have schedule growth with a TRL ≥ 6. Such a finding, perhaps unusual for a development program, is both intuitive and precedent in an S&T context. With an early TRL (1–5), there is little knowledge of how the technology will mature. This poses a problem to program managers and cost estimators. As technologies mature, investments are made, which allow costs to grow over their initial estimates. As the technology integrates into a demonstration effort (TRL 6–8), the program is often met with new and unexpected challenges, which tends to increase costs.

These results support previous literature conducted on Air Force programs which concluded that estimated costs vary exponentially across time with the progression through the various TRLs (Smoker & Smith, 2007). However, for more mature technologies, there is a broader knowledge base available for the technology's development due to more completed research. With a higher TRL, and thus more knowledge of the technology available, the better the chance of meeting schedule requirements (Dubos et al., 2008). This literature finding is also consistent with the results found here.

Table 7 results also suggest that an S&T program is *more* likely to have contract values greater than $3 million (the dataset's mean) with a TRL of 6 or greater and *less* likely to have contract values greater than $1 million (the dataset's median) with a TRL of 1–3. The explanation is consistent with the aforementioned cost growth finding. As the program's technology matures, additional investments are made, as shown in the contingency analysis results in Figure 4. In fact, the odds ratio indicates that given the program has a TRL of 6 or 7, the odds of the contract value being greater than $3 million is 14.5 times higher than a program with a TRL other than 6 or 7.

Contingency Table

Last Contract Value>$3M (Mean)

TRL 6-7	Count / Expected	0	1	Total
0	Count	29	8	37
	Expected	26.4286	10.5714	
1	Count	1	4	5
	Expected	3.57143	1.42857	
Total		30	12	42

Tests

N	DF	-LogLike	RSquare (U)
42	1	3.3084592	0.1317

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	6.617	0.0101*
Pearson	7.356	0.0067*

Odds Ratio

Odds Ratio	Lower 95%	Upper 95%
14.5	1.415167	148.5691

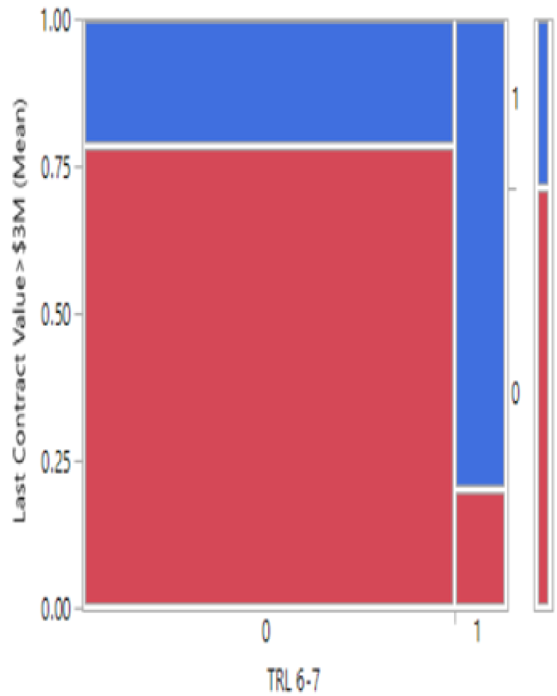


Figure 4. Contingency Table of TRL 6–7 by Contract Value > $3 Million

In summary, the TRL results suggest that programs with mature technologies are more likely to experience larger than average cost growth and larger contract values. Additionally, these programs are less likely to experience schedule growth. Furthermore, the results suggest that programs with immature technologies are less likely to have larger contract values.

Growth Relationships

As previously shown, variables for TD, performance type, and TRL were tested for their relationships with cost growth, schedule growth, and contract value. An analysis was also conducted among these latter variables to analyze their relationships to each other; a total of 63 relationships were tested for significance. Eight tests were significant at an alpha of 0.10, 11 tests were significant at an alpha of 0.05, and 22 tests were significant at an alpha of 0.01. The full set of test results is provided in Table 8.

Table 8. Significant Contingency Tables for Growth Relationships

Variable	Schedule Growth > 0%	Schedule Growth > 33% (Med)	Schedule Growth > 63% (Mean)	Contract Value > $0.9M	Contract Value > $1.0M (Med)	Contract Value > $3.0M (Mean)	Contract Value > $4.0M	Contract Value > $5.0M	Total Significant Cont. Tables
Contract Value > $0.9M		**	**						**2**
Contract Value > $1.0M (Median)									**0**
Contract Value > $3.0M (Mean)									**0**
Contract Value > $4.0M									**0**
Contract Value > $5.0M									**0**
Cost Growth > 0%	**	***	***	***	***	**	**	*	**8**
Cost Growth > 33.7% (DoD Dev - Median)	*	*	***		***	***	***	**	**7**
Cost Growth > 44.1% (AF Dev - Median)	*	*	***		***	***	***	**	**7**
Cost Growth > 56.5% (DoD Dev - Mean)	*	**	***			***	***	**	**6**
Cost Growth > 60.5% (AF Dev - Mean)	*	**	***			***	***	**	**6**
Cost Growth > 68% (Mean)		*	***			***	***	***	**5**
Total Significant Contingency Tables:	**5**	**7**	**7**	**1**	**3**	**6**	**6**	**6**	**41**

Table Legend:
***	p-value < 0.10**
****	p-value < 0.05**
*****	p-value < 0.01**

The contingency table results suggest that it is more probable for S&T programs with larger contract values to experience cost growth. Observing cost growth relationships with the original two contract value variables (using the mean and median of the dataset) provided highly significant results. To explore the sensitivity of these relationships relative to the threshold used to define the binary variables, additional contract value variables were created with lower and higher breakpoints. This additional analysis found contract values greater than $0.9 million to be the lowest threshold for which a statistically significant relationship could be found with amount of cost growth (i.e., cost growth > 0%). As the contract value threshold increased, additional cost growth variables displayed statistical significance until all were significant at a contract value of $3 million. This suggests that cost growth and contract value have a positive correlation with each other.

Table 8 results also suggest that it is more probable for S&T programs with contract values greater than $0.9 million to experience schedule growth above the median and mean (i.e., greater than 33% and 63%, respectively). This was the only contract value variable to result in significant p-values when tested with schedule growth variables. These results imply that programs with contract values less than $0.9 million are less likely to experience schedule growth.

Finally, the results suggest that if S&T programs are experiencing schedule growth, then it is more likely that they're also experiencing cost growth. This seems to contradict the findings that programs with mature technologies are more likely to experience cost growth while being less likely to experience schedule growth. But further analysis of these results suggests that programs with large schedule growth percentages are even more likely to experience cost growth at all amounts. This is because it is the immature technology programs that are experiencing both the schedule and cost growth.

In summary, the results suggest that S&T programs with larger contract values experience cost growth, while programs with smaller contract values are less likely to experience schedule growth. Finally, analyzing the relationship between cost and schedule growth suggests that programs with schedule growth are more likely to have cost growth as well. Deeper analysis revealed that this schedule growth/cost growth relationship is found in those programs with immature technologies.

Conclusion

S&T programs serve an important role in the defense acquisition process. They constitute the initial phase of the acquisition process through discovery and development of warfighting technology. The results of these programs impact the larger MDAPs that integrate the technologies in subsequent phases of the lifecycle. Despite this important role, little prior research has examined the performance of S&T programs. Thus, the overarching goal of this paper was to discern new insights from an analysis of S&T program characteristics in relation to their program's performance.

The literature review identified technological maturity as a critical success factor in product development (GAO, 1999). One measure defense programs use for technological maturity is TRL levels. TRLs, therefore, were an integral component under investigation in this analysis. The objective was to understand how TRLs affect S&T program performance. There are several key findings.

First, the results suggest that aerospace programs are more likely to technologically mature when compared to human system programs. In other words, the AFRL aerospace programs are more likely to increase the TRLs in their programs. To the extent that technological maturity is a measure of success, the aerospace programs outperform. However, this technical performance comes at a cost, as the aerospace programs were also more likely to experience cost and schedule growth. Intuitively, these results are compatible; with proven success in technology maturation, increases in funding and schedule are likely to keep the maturation on track.

Second, the partnership method between the government and contractor matters. The partnerships for S&T programs consist of SBIR and RDT&E relationships. The RDT&E programs are more likely to have and end with a TRL of 6 or more in comparison to SBIR programs. The result is not entirely surprising because, by definition, the larger defense companies comprise the RDT&E category. These larger companies have the capacity and resources to mature technology that the smaller SBIR companies may not possess.

Third, TRLs and program performance are linked. The relationships with TRLs suggest that programs with mature technologies are more likely to experience above-average cost growth and larger contract values while less likely to experience schedule growth. Additionally, the results suggest that programs with immature technologies are less likely to have larger contract values. As technologies mature, additional funds for investments are made, which increases costs over their initial contract values. This is likely to happen when the program is met with new and unexpected challenges as the technology integrates into a demonstration effort (TRL 6–8). Linick (2017) found that as the TRL increased throughout the development phase, the percentage of the development cost increased at an increasing rate. This literature finding is in agreement with these results. Conversely, as these technologies mature, there is a broader knowledge base for their development, which increases the chance of meeting schedule requirements.

Lastly, the analysis of "growth" variables (cost growth, schedule growth, and contract value) provides additional insights on S&T programs. Specifically, the analysis suggests that S&T programs with larger contract values experience larger cost growth at the same time programs with smaller contract values are less likely to experience schedule growth. Further analyzing the relationship between cost and schedule growth, the results suggest that if programs have larger schedule growth, then they are more likely to have larger cost growth as well. Deeper analysis revealed that this schedule growth/cost growth relationship is found in those programs with immature technologies.

Prior examinations of S&T programs are scarce. Thus, the possibilities for future research are vast. The exploratory analysis conducted here focused solely on AFRL programs. S&T programs in the other military services warrant examination. Additionally, one of the more surprising aspects of the data obtained from S&T programs was the reported TRL at various stages of the program's lifecycle. In order for a program to advance past Milestone B into the EMD phase, a program must have a TRL of 6 or greater. Further research into those S&T programs whose technology matured (TRL increased) could identify common characteristics which indicate a higher probability of technological maturation. The exploratory analysis provided here was just the first step of the journey. Through future research and discoveries, the knowledge needed to increase the odds for successful S&T programs is possible.

Disclaimer: The views expressed in this article are those of the authors and do not reflect the official policy or position of the U.S. Air Force, Department of Defense, or the U.S. government.

Appendix A. TRL Definitions, Descriptions, and Supporting Information

TRL	Definition	Description	Supporting Information
1	Basic principles observed and reported.	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development (R&D). Examples might include paper studies of a technology's basic properties.	Published research that identifies the principles that underlie this technology. References to who, where, when.
2	Technology concept and/or	Invention begins. Once basic principles are observed, practical	Publications or other references that outline the application

TRL	Definition	Description	Supporting Information
	application formulated.	applications can be invented. Applications are speculative, and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.	being considered and that provide analysis to support the concept.
3	Analytical and experimental critical function and/or characteristic proof of concept.	Active R&D is initiated. This includes analytical studies and laboratory studies to physically validate the analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.	Results of laboratory tests performed to measure parameters of interest and comparison to analytical predictions for critical subsystems. References to who, where, and when these tests and comparisons were performed.
4	Component and/or breadboard validation in a laboratory environment.	Basic technological components are integrated to establish that they will work together. This is relatively “low fidelity” compared with the eventual system. Examples include integration of “ad hoc” hardware in the laboratory.	System concepts that have been considered and results from testing laboratory scale breadboard(s). References to who did this work and when. Provide an estimate of how breadboard hardware and test results differ from the expected system goals.
5	Component and/or breadboard validation in a relevant environment.	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so they can be tested in a simulated environment. Examples include “high-fidelity” laboratory integration of components.	Results from testing laboratory breadboard system are integrated with other supporting elements in a simulated operational environment. How does the “relevant environment” differ from the expected operational environment? How do the test results compare with expectations? What problems, if any, were encountered? Was the breadboard system refined to more nearly match the expected system goals?
6	System/subsystem model or prototype demonstration in a relevant environment.	Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. Represents a major step up in a technology’s demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in a simulated operational environment.	Results from laboratory testing of a prototype system that is near the desired configuration in terms of performance, weight, and volume. How did the test environment differ from the operational environment? Who performed the tests? How did the test compare with

TRL	Definition	Description	Supporting Information
			expectations? What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before moving to the next level?
7	System prototype demonstration in an operational environment.	Prototype near or at planned operational system. Represents a major step up from TRL 6 by requiring demonstration of an actual system prototype in an operational environment (e.g., in an aircraft, in a vehicle, or in space).	Results from testing a prototype system in an operational environment. Who performed the tests? How did the test compare with expectations? What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before moving to the next level?
8	Actual system completed and qualified through test and demonstration.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation (DT&E) of the system in its intended weapon system to determine if it meets design specifications.	Results of testing the system in its final configuration under the expected range of environmental conditions in which it will be expected to operate. Assessment of whether it will meet its operational requirements. What problems, if any, were encountered? What are/were the plans, options, or actions to resolve problems before finalizing the design?
9	Actual system proven through successful mission operations.	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation (OT&E). Examples include using the system under operational mission conditions.	OT&E reports.

Appendix B. Sample Research Summary Report

Research Summary Report

General Information

Work Unit Title:	Program Title		
Work Unit (WU) #:		**Accession #:**	XXXXX
Start Date:	MM/DD/YYYY	**End Date:**	MM/DD/YYYY
Lead TD:	RQ	**Effort Security:**	Unclassified
Literature Search Order #:	XXXXX	**Lit Search Date:**	MM/DD/YYYY
Work Phase Code:		**WU Category:**	
Responsible Organization:	Aerospace Systems Directorate (AFRL/RQ) WRIGHT-PATTERSON AFB, OH		
Parent Program:	Parent Program Title		

Work Unit Manager

Name:	XXXX XXXX	**Phone:**	(XXX) XXX-XXXX
Office Symbol:	RQTE	**Email:**	xxxx.xxxx@us.af.mil

DoD Required Information

Performance Method:	PROCUREMENT/ACQUISITION AWARD - Contract,	**Lab Core Technical Competency:**	Power, Propulsion, Energy & Alternative Fuels
Performance Type:	RDTE - Research, Development, Test, & Evaluation Work Unit	**Technology Readiness Level:**	2 Technology Concept
Fields of Science & Engineering:		**Data Management Plan Exist:**	No
Joint Capability Area:	Force Application		
Communities of Interest:	1.1 Aircraft Propulsion, Power and Thermal		
Technology Transition Opportunities:	None at this time. Follow on contract will progress TRL of key components to enable future transition.		

Principal Investigator

Name:	XXXX XXXXXX	**Phone:**	XXX-XXX-XXXX
Office Symbol:	XX	**Email:**	xxxx.xxxxx@xx.com

Performing Mechanism

Contract #:	FAXXXX-XX-X-XXXX-XXXX	**Contract Status:**	Complete
Contract Face Value	$XX,XXX,XXX	**TR Due DTIC Date:**	MM/DD/YYYY
Award Date:	MM/DD/YYYY	**TR Draft Due Date:**	MM/DD/YYYY
Tech End Date:	MM/DD/YYYY	**Terminate Date:**	
End Date:	MM/DD/YYYY	**WAWF /DD250Z Date:**	MM/DD/YYYY

Performing Organization

Performer:	NAME OF COMPANY	**Address:**
Division:	XX	
CAGE Code:	XXXXXX	
Performing Org Type:	Industrial Firm	

WU Funding ($K)

PE / BPAC	Source	Funding	Prior	FY18	FY19	FY20	FY21	FY22	FY23	To Complete	Total
		Appropriated									
		Obligations									
		Appropriated									
		Obligations									

Descriptive and Progress/Status

Objective/Restricted

MMM DD, YYYY
12:00:00 AM

Distribution C US Government Agencies and their Contractors only - C-CL Critical Technology (Export Controlled)
Page 1

References

Bolten, J. G., Leonard, R. S., Arena, M. V., Younossi, O., & Sollinger, J. M. (2008). *Sources of weapon system cost growth* (Report No. MG-670). RAND.

Cancian, M. F. (2010). Cost growth: Perception and reality. *Defense Acquisition Research Journal*, *17*(3), 389–404.

DoD. (2009, July). *Technology readiness assessment (TRA) deskbook*. Office of the Director, Defense Research and Engineering.

DoD. (2011, May 13). *Technology readiness assessment (TRA) guidance.* Assistant Secretary of Defense for Research and Engineering.

Duffner, R. (2000). *Science and technology: The making of the Air Force research laboratory*. Air University Press.

Dubos, G., Saleh, J., & Braun, R. (2008). Technology readiness level, schedule risk and slippage in spacecraft design: Data analysis and modeling. *Journal of Spacecraft and Rockets*, *45*(4), 836–842. https://doi.org/10.2514/6.2007-6020

Gorn, M. H. (1988). *Harnessing the genie: Science and technological forecasting and the Air Force 1944–1986*. Office of Air Force History.

GAO. (1999). *Best practices: Better management of technology development can improve weapon systems outcomes* (GAO/NSIAD-99-162).

Katz, D. R., Sarkani, S., Mazzuchi, T., & Conrow, E. H. (2015). The relationship of technology and design maturity to DoD weapon system cost change and schedule change during engineering and manufacturing development. *Systems Engineering*, *18*(1), 1–15. https://doi.org/10.1111/sys.21281

Linick, J. (2017, June 6–9). *Technology readiness level (TRL) vs. percent development cost* [Presentation]. ICEAA Professional Development & Training Workshop, Portland, OR, United States.

Meredith, J., & Mantel, S. (2003). *Project management: A managerial approach*. John Wiley & Sons.

Nasir, M. H. N., & Sahibuddin, S. (2011). Critical success factors for software projects: A comparative study. *Scientific Research and Essays*, *6*(10), 2174–2186. https://doi.org/10.5897/SRE10.1171

Office of the Chief Scientist of the U.S. Air Force. (2010). *Technology horizons: A vision for Air Force science & technology During 2010–2030*. Department of the Air Force.

Pinto, J. K., & Slevin, D. P. (1987). Critical factors in successful project implementation. *IEEE Transactions on Engineering Management*, *34*(1), 22–27.

Pinto, J. K., & Slevin, D. P. (1998, September). Project success: Definitions and measurement techniques. *Project Management Journal*, *19*(3), 67–73.

Rendon, R. (2012). Defense procurement: An empirical analysis of critical success factors. In *Charting a course in public procurement innovation and knowledge sharing* (pp. 174–208).

Rodriguez-Segura, E., Ortiz-Marcos, I., Romero, J. J., & Tafur-Segura, J. (2016). Critical success factors in large projects in the aerospace and defense sectors. *Journal of Business Research*, *69*(11), 5419–5425.

Sadin, S. R., Povinelli, F. P., & Rosen, R. (1989). The NASA technology push towards future space mission systems. *Acta Astronautica*, *20*, 73–79. https://doi.org/10.1016/0094-5765(89)90054-4

Smirnoff, J. P., & Hicks, M. J. (2008). The impact of economic factors and acquisition reforms on the cost of defense weapon systems. *Review of Financial Economics*, *17*(1), 3–13.

Smoker, R. E., & Smith, S. (2007). System cost growth associated with technology-readiness level. *Journal of Parametrics*, *26*(1), 8–38. https://doi.org/10.1080/10157891.2007.10462276

Tishler, A., Dvir, D., Shenhar, A., & Lipovetsky, S. (1996). Identifying critical success factors in defense development projects: A multivariate analysis. *Technological Forecasting and Social Change*, *51*(2), 151–171. https://doi.org/10.1016/0040-1625(95)00197-2

Tracy, S. P., & White, E. D. (2011). Estimating the final cost of a DoD acquisition contract. *Journal of Public Procurement*, *11*(2), 190–205. https://doi.org/10.1108/jopp-11-02-2011-b002

U.S. Air Force. (2019). *Science and technology strategy: Strengthening USAF science and technology for 2030 and beyond*. U.S. Government Printing Office.

Zwikael, O., & Globerson, S. (2006). From critical success factors to critical success processes. *International Journal of Production Research*, *44*(17), 3433–3449.

Strategies for Addressing Uncertain Markets and Uncertain Technologies

William B. Rouse—Georgetown University [wr268@georgetown.edu]

Dinesh Verma—Stevens Institute of Technology [dverma@stevens.edu]

D. Scott Lucero—Office of the Secretary of Defense (Ret.) [dslucero@gmail.com]

Edward S. Hanawalt—General Motors [edward.hanawalt@gm.com]

Abstract

Engineering involves designing solutions to meet the needs of markets or missions. Organizations would like to have the flexibility and agility to address both uncertain needs and uncertain technologies for meeting these needs. This article presents and illustrates a decision framework that enables flexibility and agility and provides guidance on when to pursue optimal, highly integrated solutions. We consider how uncertainties arise, contrasting the automotive and defense domains. We propose an approach to managing uncertainties. We consider how to represent alternative solutions and project the value of each alternative, including how market or mission requirements can be translated into system requirements. Possible use cases for our framework are discussed. A detailed case study of autonomous vehicles for enhancing the mobility of disabled and older adults is presented.

Introduction

Much of engineering involves designing solutions to meet the needs of markets, or perhaps military missions or societal sector needs such as water, power, and transportation. These needs are often uncertain, especially if solutions are intended to operate far into the future.

There is also often uncertainty in how best to meet needs. New technologies may be needed, and their likely performance and cost may be uncertain. Budgets may be insufficient to achieve what is needed. Competitors or adversaries may be creating competing solutions that are similar or superior.

Organizations would like to have the flexibility and agility to address both uncertain needs and uncertain technologies due to performance challenges, organizational experience, supply chains, and so on. This is likely to require ways of thinking and allocating resources that are foreign to many organizations. This article outlines and illustrates these ways of thinking.

To illustrate how companies address uncertainties, consider two experiences at General Motors (GM). Both illustrations involved Ford surprising GM. The first led to a major failure and the second to a substantial success (Hanawalt & Rouse, 2010).

In 1981, General Motors began planning for a complete refresh of its intermediate-size vehicles: the front wheel drive A-cars and the older rear wheel drive G-cars. The GM10 program would yield vehicles badged as Chevrolets, Pontiacs, Oldsmobiles, and Buicks. This program was to be the biggest research and development (R&D) program in automotive history and, with a $5 billion budget, the most ambitious new car program in GM's 79-year history.

The introduction of the Ford Taurus in 1985 was a huge market and business success and a complete surprise to GM. It was one of the first projects in the United States to fully utilize the concept of cross-functional teams and concurrent engineering practices. The car and the process used to develop it were designed and engineered at the same time, ensuring higher quality and more efficient production. The revolutionary design of the Taurus, coupled with its outstanding quality, created a new trend in the U.S. automobile industry, and customers simply loved the car.

The Taurus forced GM to redesign the exterior sheet metal of the GM10 because senior executives thought the vehicles would look too similar. Many additional running changes were incorporated into the design in an attempt to increase customer appeal. The first vehicles reached the market in 1988, approximately $2 billion over budget and 2 years behind schedule.

All of the first GM10 entries were coupes, a GM tradition for the first year of any new platform. However, this market segment had moved overwhelmingly to a four-door sedan style. Two-door midsize family cars were useless to the largest group of customers in the segment; members of the Baby Boomer generation were now well into their child-rearing years and needed four-doors for their children. GM completely missed the target segment of the market. From 1985 to 1995, GM's share of new midsize cars tumbled from 51% to 36%.

The Lincoln Navigator is a full-size luxury SUV marketed and sold by the Lincoln brand of Ford Motor Company since the 1998 model year. Sold primarily in North America, the Navigator is the Lincoln counterpart of the Ford Expedition. While not the longest vehicle ever sold by the brand, it is the heaviest production Lincoln ever built. It is also the Lincoln with the greatest cargo capacity and the first non-limousine Lincoln to offer seating for more than six people.

GM was completely surprised by the Navigator. They had not imagined that customers would want luxurious large SUVs. GM responded with the Cadillac Escalade in 1999, intended to compete with the Navigator and other upscale SUVs. The Escalade went into production only 10 months after it was approved. The 1999 Escalade was nearly identical to the 1999 GMC Yukon Denali, except for the Cadillac badge and leather upholstery. It was redesigned for the 2002 model year to make its appearance and features fall more in line with Cadillac's image.

In 2019, 18,656 Navigators were sold, while 35,244 Escalades were sold. Escalade has outsold Navigator every year since 2002. GM had clearly adapted to the surprise of the Navigator. One can reasonably infer that the company learned from the GM10 debacle. Surprises happen. Be prepared.

We recently studied 12 cars withdrawn from the market in the 1930s, 1960s, and 2000s (Liu et al., 2015). We leveraged hundreds of historical accounts of these decisions, as well as production data for these cars and the market more broadly. We found that only one vehicle was withdrawn because of the nature of the car. People were unwilling to pay Packard prices for Studebaker quality, the two companies having merged in 1954.

The failure of the other 11 cars could be attributed to company decisions, market trends, and economic situations. For example, decisions by the Big Three companies to focus on cost reduction resulted in each manufacturer's car brands looking identical, effectively de-badging them. Mercury, Oldsmobile, Plymouth, and Pontiac were the casualties. Honda and Toyota were the beneficiaries.

This article presents and illustrates a framework for addressing such scenarios. We first consider how uncertainties arise, contrasting the automotive and defense domains. We then propose an approach to managing uncertainties. This leads to consideration of how to represent alternative solutions and to estimate the value of these alternative solutions. We discuss possible use cases for our framework and present a detailed case study of autonomous vehicles to enhance the mobility of disabled and older adults.

Sources of Uncertainties

Table 1 portrays two domains where addressing uncertainties are often central and important aspects of decision-making. The primary domain emphasized in this article is automotive. However, we also want to emphasize the relevance of our line of thinking to the defense domain. The parallels are reasonably self-explanatory, but a few differences are worth elaborating.

In the automotive domain, there are multiple providers of competing vehicles. In defense, there is typically one provider of each platform. Many customers make purchase decisions in the automotive domain while, in defense, there is one (primary) customer making the purchase decision. The lack of competitive forces can lead to requirements being locked in prematurely.

In the automotive domain, vehicles are used frequently. In defense, platforms are used when missions need them, which, beyond training, may never occur. Competitors' relative market positions in the automotive domain change with innovations, for example, in the powertrain. In defense, adversaries' positions change with strategic innovations, for instance, pursuits of asymmetric warfare. As former Defense Secretary James Mattis has said, "The enemy gets a vote on defense planning" (Mattis, 2019).

Automobiles have model year changes, usually 3-year refreshes, and life spans of up to 10 years, typically 6 to 7. The B-52 bomber has been in use for almost 70 years, and the F-15 fighter aircraft has been in use for almost 50 years. There are block upgrades of military aircraft every few years, typically for changes of avionics and weapon systems—rather than body style.

There are similarities that can be seen in Table 1. Uncertainties associated with market needs or mission requirements typically flow down In Table 1. Uncertainties associated with technology typically flow up, for example, when the engineering organization (at the company or vehicle level) is not sure of how to provide a function or whether performance or cost objectives can be met. New technologies enable new military capabilities. The most important weapons transforming warfare in the 20th century, such as airplanes, atomic weapons, the jet engine, and electronic computers, did not emerge as a response to doctrinal requirement of the military (Chambers, 1997, p. 791).

Table 1. Multilevel Comparison of Automotive and Defense Domains

Automotive Domain	Defense Domain
Economy	**Geopolitics**
- Geopolitics (e.g., Regulations, Tariffs, War)	- Military Conflict (i.e., Hot War)
- GDP & Inflation (e.g., Recession)	- Geopolitical Tension (e.g., Grey Zone Conflicts)
- Financial Markets (e.g., Interest Rates)	- Civil Wars (e.g., Migration)
- Energy Markets (e.g., Fuel Prices)	- Soft Power (e.g., Alliances)
Market	**Economics**
- Market Growth/Decline (e.g., Consumers)	- GDP Growth/Decline
- Segment Market Saturation (e.g., Sedans)	- Inflation/Deflation
- External Competitors (Companies)	- Domestic & Allies' Defense Budgets
- Internal Competitors (Brands)	- Congressional Priorities (e.g., Jobs)
Company Priorities	**Defense Priorities**
- Market Strategy (e.g., Positioning, Pricing)	- Engagement Strategies
- Product Management (e.g., Processes)	- Missions Envisioned
- Dealer Management (e.g., Incentives)	- Adversary Capabilities
- Financial Management (e.g., Investments)	- Capabilities Required
- Brand Management (e.g., Rebadging)	- Emerging Technologies
Vehicle	**Platform**
- Price	- Performance
- Design	- Schedule
- Quality	- Cost

Automobile companies are currently wrestling with pursuits of battery electric vehicles and the uncertain rate of market adoption (Liu et al., 2018). Just over the horizon is the opportunity to compete in the driverless car market (Liu et al., 2020), with significant uncertainties about the regulatory environment (Laris, 2020). The case study later in this article addresses this opportunity.

There are also uncertainties associated with where to manufacture vehicles (Hanawalt & Rouse, 2017). Labor costs used to dominate location decisions, but other economic, legal, and political factors are now being considered. Decisions to withdraw from manufacturing in Australia, Canada, and South Korea have resulted.

Product line or program managers in the two domains often have similar questions regarding common uncertainties. A comparison of these questions is shown in Table 2. It is often socially unacceptable to verbalize such questions. Unfortunately, uncertainties not verbalized are seldom well managed (Rouse, 1998).

Table 2. Comparison of Automotive and Defense Domains

Automotive Domain Uncertainties	Defense Domain Uncertainties
Customer future preferences	Mission plans will remain relevant
Customers' future purchases will favor our offerings versus competitors	Mission platforms will remain superior to adversaries' capabilities
Performance of our offerings after development	Performance of mission platforms after development
Affordability over the coming years	Affordability over the coming years
Budgets for our offerings across a range of future needs	Budgets for mission platforms across a range of future needs
Supply chains will be economical, efficient, and secure	Supply chains will be economical, efficient, and secure
Competitors' capabilities will not be perceived to be superior	Adversaries' capabilities will be inferior and certainly not superior
Our enterprise will continue to support our endeavors	Ensuring that sponsors (e.g., Congress) will continue to provide support

Managing Uncertainties

In both the automotive and defense domains there are usually uncertainties about market or mission requirements as well as uncertainties about technologies and abilities needed to meet these requirements. This section outlines an approach to thinking about managing these uncertainties.

Consider a couple of extremes. You are absolutely sure a function will be required, and you are absolutely sure of how to deliver it. In other words, you are not at all uncertain. You should invest to create a solution to meet this need, assuming that you are confident the necessary human and financial resources are available. At the other extreme, you are absolutely sure a function will not be required. Regardless of your ability to deliver this function, you should not invest in creating this solution. Between these two extremes, there are several strategies a company might adopt. The choice depends on enterprises' abilities to predict their futures, as well as their anticipated abilities to respond to these futures. What strategies might enterprise decision makers adopt to address alternative futures? As shown in Figure 1, we have found that there are four basic strategies that decision-makers can use: optimize, adapt, hedge, and accept.

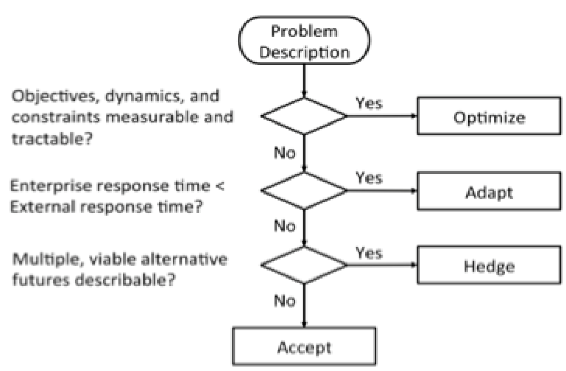


Figure 1. Strategy Framework for Enterprise Decision-Makers (Pennock & Rouse, 2016)

If the phenomena of interest are highly predictable, then there is little chance that the enterprise will be pushed into unanticipated territory. Consequently, it is in the best interest of the enterprise to optimize its products and services to be as efficient as possible. In other words, if the unexpected cannot happen, then there is no reason to expend resources beyond process refinement and improvement.

If the phenomena of interest are not highly predictable, but products and services can be appropriately adapted when necessary, it may be in the best interest for the enterprise to plan to adapt. For example, agile capacities can be designed to enable their use in multiple ways to adapt to changing demands—for example, the way Honda adjusted production capacity but other automakers could not in response to the Great Recession. Their planning was more efficient in the long run; even so, efficiency may have to be traded for the ability to adapt.

For this approach to work, the enterprise must be able to identify and respond to potential issues faster than the ecosystem changes. For example, consider unexpected increased customer demands that tax capacities beyond their designed limits. Design and building of new or expanded facilities can take considerable time. On the other hand, reconfiguration of agile capacities should be much faster, as Honda demonstrated. The value of this approach is widely known in the military. As renown fighter pilot Robert Boyd—inventor of the Observe, Orient, Design, Act (OODA) loop—noted, whoever can handle the quickest rate of change is the one who survives (Boyd, 2004). Similarly, Arie De Gues, head of Strategic Planning for Royal Dutch Shell, stated that the ability to learn faster than your competitors might be the only sustainable advantage (Senge, 1990).

If the phenomena of interest are not very predictable and the enterprise has a limited ability to adapt and respond, it may be in the best interest of the enterprise to hedge its position. In this case, it can explore scenarios where the enterprise may not be able to handle sudden changes without prior investment. For example, an enterprise concerned about potential obsolescence of existing products and services may choose to invest in multiple, potential new offerings. Such investments might be pilot projects that enable the enterprise to learn how to deliver products and services differently or perhaps deliver different products and services.

Over time, it will become clear which of these options makes most sense, and the enterprise can exercise the best option by scaling up these offerings based on what they have learned during the pilot projects. In contrast, if the enterprise were to take a wait-and-see approach, it might not be able to respond quickly enough, and it might lose out to its competitors.

If the phenomena of interest are totally unpredictable and there is no viable way to respond, then the enterprise has no choice but to accept the risk. Accept is not so much a strategy as a default condition. If one is attempting to address a strategic challenge where there is little ability to optimize the efficacy of offerings, limited ability to adapt offerings, and no viable hedges against the uncertainties associated with these offerings, the enterprise must accept the conditions that emerge.

There is another version of acceptance that deserves mention—stay with the status quo. Yu et al. (2011) developed a computational theory of enterprise transformation, elaborating on a qualitative theory developed earlier (Rouse, 2005). They employed this computational theory to assess when investing in change is attractive and unattractive. Investing in change is likely to be attractive when one is currently underperforming and the circumstances are such that investments will likely improve enterprise performance. In contrast, if one is already performing well, investments in change will be difficult to justify. Similarly, if performance cannot be predictably improved—due to noisy markets and/or highly discriminating customers—then investments may not be warranted despite current underperformance.

Lucero (2018) proposed that these four strategies would be differentially relevant for different areas of an uncertainty space with axes involving uncertainties around the requirements and the ability to meet those requirements. We extended his thinking to formulate Figure 2, focusing on uncertainties in developing technologies.

Requirements Uncertainty	Not Feasible	Possibly Feasible	Fully Feasible
Definitely Required	Hedge Via Partnership	Hedge Via Larger R&D Investment	Optimize Technology Capability
Possibly Required	Hedge Via Partnership	Hedge Via Smaller R&D Investment	Adapt If Requirement Emerges
Not Required	Accept Current Situation	Accept Current Situation	License Patents To Others

Technology Uncertainty

Figure 2. Strategies Versus Uncertainties

This figure depicts the space as having nine discrete cells, which makes it easier to explain, but there are unlikely to be crisp borders between areas where the different strategies are applicable.

There are three types of hedges in Figure 2. The upper two cells of the middle column represent company or agency investments in creating technology options to meet possible requirements. The upper two cells of the left column represent licensing, joint development, or other arrangements to buy technology options from partners. The lower cell of the right column represents selling options to others so they can hedge uncertainties.

The criteria on the left of Figure 1 constrain choices of strategies as well as positions in the uncertainty space. If, for example, the objectives, dynamics, and constraints are not measurable and tractable, then optimization may lead to an inappropriate or at least fragile solution (Carlson & Doyle, 2000).

At this point, we have strategies for addressing uncertainties. We now need to address the characteristics of the alternative solutions of interest and then the projected expected utility of each alternative.

Representing Solutions

Whose preferences should guide decisions? While there may be one ultimate decision-maker, success usually depends on understanding all stakeholders. Human-centered design addresses the concerns, values, and perceptions of all stakeholders in designing, developing, manufacturing, buying, and using products and systems. The basic idea is to delight primary stakeholders and gain the support of the secondary stakeholders.

The human-centered design construct and an associated methodology has been elaborated in a book, *Design for Success* (Rouse, 1991). Two other books soon followed (Rouse, 1992, 1993). The human-centered design methodology has been applied many times and continually refined (Rouse, 2007, 2015).

The premise of human-centered design is that the major stakeholders need to perceive products and services to be valid, acceptable, and viable. Valid products and services demonstrably help solve the problems for which they are intended. Acceptable products and services solve problems in ways that stakeholders prefer. Viable products and services provide benefits that are worth the costs of use. Costs here include the efforts needed to learn and use products and services, not just the purchase price.

Figure 3 embodies the principles of human-centered design, built around Set-Based Design (SBD; Sobek et al., 1999), Quality Function Deployment (Hauser & Clausing, 1988), and Design Structure Matrices (Eppinger & Browning, 2012). As later discussed, multi-stakeholder, multi-attribute utility theory (Keeney & Raiffa, 1993) is used to project the value of alternatives. Note that validity, acceptability, and viability in Figure 3 are defined in the above discussion of human-centered design.

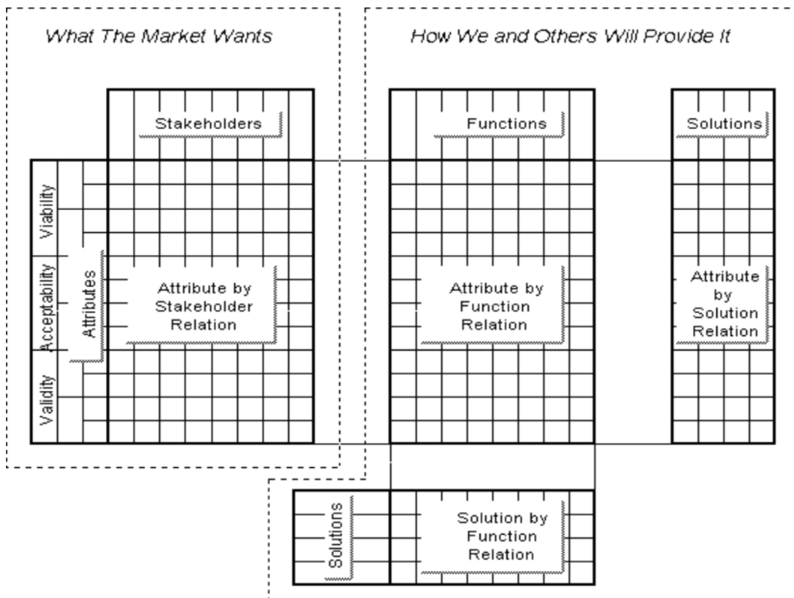


Figure 3. Model Structure for Attributes, Stakeholders, Functions, and Solutions

Sobek et al. (1999) contrast SBD with Point-Based Design. Developed by Toyota, SBD considers a broader range of possible designs and delays certain decisions longer. They argue that, "Taking time up front to explore and document feasible solutions from design and manufacturing perspectives leads to tremendous gains in efficiency and product integration later in the process and for subsequent development cycles." Al-Ashaab et al. (2013) and Singer et al. (2017) report on interesting applications of SBD to helicopter engines and surface combatant ships, respectively.

SBD is reflected in Figure 3 in terms of defining and elaborating multiple solutions, including those of competitors or adversaries. Quality Function Deployment (QFD; Hauser & Clausing, 1988) translates the "voice of the customer" into engineering characteristics. For Figure 3, this translates into "voices of the stakeholders." Design Structure Matrices (DSM; Eppinger & Browning, 2012) are used to model the structure of complex systems or processes. In Figure 3, multiple models are maintained to represent alternative offerings as well as current and anticipated competitors' offerings.

The "What the Market Wants" section of Figure 3 characterizes the stakeholders in the product or service and their utility functions associated with context-specific attributes clustered in terms of validity, acceptability, and viability. The section of Figure 3 labeled “How We and Others Will Provide It” specifies, on the right, the attribute values associated with each solution. The functions associated with each solution are defined on the left of this section. Functions are things like steering, accelerating, and braking, as well as functions that may not be available in all solutions (e.g., backup camera).

Attribute to function relationships in Figure 3 are expressed on a somewhat arbitrary scale from −3 to +3. Positive numbers indicate that improving a function increases the attribute. Negative numbers indicate that improving a function decreases an attribute. For example, a backup camera may increase the price of the vehicle but decrease insurance costs.

Solutions on the bottom of Figure 3 are composed of functions, which are related to attributes of interest to stakeholders. In keeping with the principles of SBD, multiple solutions are pursued in parallel, including potential offerings by competitors. While it is typical for one solution to be selected for major investment, the representations of all solutions are retained, quite often being reused for subsequent opportunities.

There are additional considerations beyond SBD, QFD, and DSM. Uncertain or volatile requirements can be due to evolving performance targets (Ferreira et al., 2009) or surprises by competitors or adversaries (e.g., the Ford Taurus). Both causes tend to result in expensive rework. In the realm of defense, the end of the Cold War ended the need for a 70-ton self-propelled howitzer (Myers, 2001). Advances in anti-ship cruise missiles and a challenging performance envelope doomed the Expeditionary Fighting Vehicle (Feickert, 2009).

Decision-making may involve more than one epoch (Ross & Rhodes, 2008), including both near-term and later decisions. For example, at GM, Epoch 1 involved creating an Escalade as a rebadged GMC in 1999. Epoch 2 involved offering an Escalade as a unique upscale SUV in 2002.

Another issue is the costs of switching from one solution to another (Silver & de Weck, 2007). A surveillance and reconnaissance mission adopted an initial solution of a manned aircraft with an option to replace this solution with an unmanned air vehicle (UAV) several years later (Rouse, 2010). A deterrent to switching was the very expensive manned aircraft, which would no longer be needed. This problem was resolved by negotiating, in advance, the sale of the aircraft to another agency, effectively taking it “off the books.” Thus, there can be significant value in flexibility. “A system is flexible to the extent that it can be cost-effectively modified to meet new needs or to capitalize on new opportunities” (Deshmukh et al., 2010).

Identifying options can be difficult (Mikaelian et al., 2012). What can you do when, and what will it cost? Rouse et al. (2000) discuss case studies from the semiconductor industry. Rouse and Boff (2004) summarize 14 case studies from automotive, computing, defense, materials, and semiconductor industries.

Projecting Value

Using the framework provided by Figure 3 and principles from SBD, QFD, DSM, and so on, one can create multi-attribute models of how alternatives address the concerns, values, and perceptions of all the stakeholders in designing, developing, manufacturing, buying, and using products and systems. The next issue of importance is the likely uncertainties associated with the attributes of the alternatives. These uncertainties involve what the market or mission needs—or will need—and how well solutions, in terms of functions and underlying technologies, will be able to meet these needs.

The expected value of an alternative can be defined as the value of the outcomes a solution provides times the probability that these outcomes will result. The probability may be discrete, or it may be represented as a probability density function. For the former, the calculation involves multiplication and summation; for the latter, the calculation involves integration.

Following Keeney and Raiffa (1993), we approach this problem using multi-stakeholder, multi-attribute utility theory. We can define the utility function of stakeholder i across the N attributes by

$$u_i = u\ (x_{1i}, x_{2i}, \ldots, x_{Ni}) = u(\mathbf{x}_i) \qquad (1)$$

where the bold **x** denotes the vector of attributes. The utility of an alternative across all M stakeholders is given by

$$U = U\ [u(\mathbf{x}_1), u(\mathbf{x}_2), \ldots, u(\mathbf{x}_M)] \qquad (2)$$

The appropriate forms of these functions vary by the assumptions one is willing to make. When there are many attributes, a weighted linear from is usually the most practical. The weights in Equation 1 reflect how much a particular stakeholder cares about the attribute being weighted. It is quite common for most stakeholders to only care about a small subset of the overall set of attributes. Those for which they do not care receive weights of zero.

The weights in Equation 2 reflect the extent to which the overall decision-maker or decision process cares about particular stakeholders. For example, is the customer the most important stakeholder, or do corporate finances drive the decision? These weights are usually subject to considerable sensitivity analyses.

Who are typically the stakeholders? We have found that the concerns, values, and perceptions of the following entities are typically of interest:

- Market/Mission
- Customers/Users/Warfighters
- Operators/Maintainers
- Technologists/R&D
- Finance/Budgets
- Current Competitors
- Possible Competitors
- Investors
- Governments (e.g., Regulatory Authorities)

For the case study presented in a later section, we focus on solely the investor stakeholder. Investors in driverless cars are interested in three primary attributes:

- **Competitive Advantage (CA):** To what extent will the investment of interest enable value-added pricing, reduce production costs, reduce operating costs, and leverage existing capacities?
- **Strategic Fit (SF):** To what extent will the investment of interest leverage technology competencies, exploit current delivery architectures, complement existing value propositions, exploit current partnerships and infrastructure, and provide other opportunities for exploitation?
- **Return on Investment (ROI):** What capital expenditures, technology acquisition costs, and labor expenses will be needed? What revenue and profits will likely result?

We will return to these attributes in the case study.

Use Cases

What types of decisions are amenable to the approach just outlined? We have applied this line of reasoning to 20+ projects involving science and technology investment decisions—in particular, investments in R&D, licensing technologies, and capacity expansion. The case study discussed in the next section is an example of this use case.

Another use case involves exploring tipping points in market/mission analysis, where small investments result in sizable performance gains, either for you or for your competitors or adversaries. A good example is when Motorola found that offering pagers in colors substantially increased sales (Henkoff, 1994). Another example is the aforementioned repurposing of a military aircraft. Getting it “off the books” greatly enhanced the UAV investment value and secured the needed resources (Rouse, 2010).

Another use case involves understanding when disaggregated architectures provide higher value than integrated architectures. A good example involves investments in system infrastructure to support modularity and decrease future switching costs. Tight integration may help the current generation of a technology perform better but may undermine the flexibility of the next generation.

A classic use case involves understanding where key points of uncertainty could be resolved with more information. For example, business intelligence that enables determining competitors’ or adversaries’ actual investments versus advertised intentions can enable avoiding investing in competitions that inherently will not happen. This is an important reason for modeling solutions of competitors or adversaries as indicated in Figure 3.

To address these use cases, we need to be able to predict impacts on outcomes (e.g., attributes):

- Impacts of investments on outcomes (e.g., performance, costs)
- Impacts of particular investments on outcomes (e.g., color on pagers)
- Impacts of architectures on outcomes (e.g., performance, costs)
- Impacts of uncertainties on decisions (e.g., strategies, investments)

Performance can include many things:

- Mission performance (e.g., sorties, targets hit)
- Market performance (e.g., sales, profits, earnings per share, share price)
- Platform performance (e.g., speed, quality)
- Platform acceptance (e.g., consumer ratings)
- Platform availability (reliability and maintainability)
- Time to deployment
- Time to market
- Acquisition and operating costs

Linking alternative investments to these types of metrics require models of how investments translate to capabilities, which then translate to platform, mission, and market performance.

Case Study

Assistive technologies (AT) hold enormous promise for the 100 million disabled and older adults in the United States (Rouse & McBride, 2019). Driverless cars have the potential to greatly enhance the mobility of this population with attractive pricing. Note that the platforms of interest are autonomous vehicles, while the market or mission is to provide enhanced mobility to disabled and older adults.

The Auto Alliance hosted a series of three workshops on “AVs & Increased Accessibility” (Auto Alliance, 2019). We focused on physical, sensory, and cognitive disabilities. Approximately 200 people participated in the three workshops from a wide range of advocacy groups, automobile manufacturers, and federal agencies. Workshop participants suggested a large number of needs as well as approaches to meeting these needs. We clustered these needs into 20 categories. Eight categories covered 70% of the suggestions. Definitions of these categories are as follows:

- ***Displays and controls*** concern information that users can see, hear, touch, and so on, and actions they can take.
- ***Locating and identifying vehicle*** concerns users knowing where their ride is waiting and recognizing the particular vehicle.
- ***Passenger profiles*** include secure access to information about passengers, in particular their specific needs.
- ***Emergencies*** concern events inside and outside the vehicle that may require off-normal operations and user support.
- ***Adaptation to passengers*** involves adjusting the human–machine interface (HMI) to best support particular users with specific needs.
- ***Easy and safe entry and egress*** concerns getting into and out of the vehicle, as well as safety relative to the vehicle’s external environment.
- ***Trip monitoring and progress*** relates to providing information as the trip proceeds, particularly regarding route and schedule disruptions.
- ***Onboard safety*** concerns what happens in the vehicle as the trip proceeds, assuring minimal passenger stress and injury avoidance.

An example mapping from needs to technologies is shown in Table 3. Technologies required include hardware, software, sensing, networks, and especially enhanced HMI. HMIs need to enable requesting vehicle services, locating and accessing vehicles, monitoring trip progress, and egressing at destinations to desired locations.

Table 3. Market Needs Versus Enabling Technologies (Auto Alliance, 2019)

Needs	Technologies				
	Hardware	**Software**	**Sensors**	**Networks**	**HMI**
Displays & Controls	Hardware for Displays & Controls	Tutoring System for HMI Use	Use and Misuse of Displays & Controls	Access to Device Failure Information	Auditory, Braille, Haptic, Tactile, Visual Displays
Locating & Identifying Vehicle	Vehicle-Mounted Sensors	Recognition Software	Integration of Sensed Information	Sensors of External Networks	Portrayal of Vehicle & Location
Passenger Profiles, Privacy	Phone or Smartphones, Tablets	App to Securely Provide Profile Information	Recognition of Passenger	Access to Baseline Info. on Disabilities	Portrayal to Assure Recognition
Emergencies	Controls to Stop Vehicle & Move to Safe Space	Recognition & Prediction of Situation	Surrounding Vehicles, People, Built Environ.	External Services—Police, Fire, Health	Portrayal of Vehicle Situation
Adaptation to Passengers	Adjusting Entry, Egress, Seating	Learning Passenger Preferences	Sensing Reactions to Adaptations	Access to Baseline Info. on Adaptations	Portrayal to Enable Change Confirmations
Easy & Safe Entry & Egress	Sufficient Space to Maneuver	Capturing Data on Space Conflicts	Surrounding Vehicles, People, Built Environ.	Networked Access to (e.g., Bldg. Directions)	Portrayal of Surrounding Objects
Trip Monitoring & Progress	Speedometer, GPS, Maps	Predictions of Progress, Points of Interest	Surrounding Vehicles, People, Built Environ.	Access to Traffic Information (e.g., Accidents)	Portrayal of Trip & Progress
Onboard Safety	Securement of Wheelchairs & Occupants	Capturing Data on Securement Conflicts	Sensing & Recording Safety Risks	Access to Best Practices on Safety Risks	Portrayal of Securement Status

The wealth of assistive technology (AT) and supporting technologies in Table 3 suggest a substantial need for seamless technology integration to avoid overwhelming disabled and older adults, or indeed anybody. We expect that artificial intelligence (AI)–based cognitive assistants may be central to such integration. The question of who might provide which pieces of an overall integrated solution is addressed in this case study.

The question of interest in this case study concerns how an automotive original equipment manufacturer (OEM) should position itself relative to this immense market opportunity. We begin with SBD. The hypothetical OEM wants to consider five alternative solutions, indicated as scenarios in Table 4 because each includes a market strategy as well as a solution.

Predominant uncertainties include competitors' strategies, technologies (particularly software), abilities to execute, and time. The third scenario, ally with advocacy groups, merits elaboration. The key idea is an American Association of Retired Persons (AARP)–branded vehicle—for example, similar to the Eddie Bauer branding of the Ford Explorer—with better paint job, leather seats, heated seats optional, and interior accents. This co-branding alliance with Ford lasted 20 years.

The next step in applying the methodology outlined in this article is characterization of Competitive Advantage (CA), Strategic Fit (SF), and Return on Investment (ROI) for the set of five scenarios. We then want to consider uncertainties associated with each scenario, which for this case study will be characterized using discrete probabilities.

The expected utility of each scenario $E[U_S]$ can then be calculated using

$$E[U_S] = W_{CA} \times P_{CA} \times U_{CA} + W_{SF} \times P_{SF} \times U_{SF} + W_{ROI} \times P_{ROI} \times U_{ROI} \qquad (3)$$

where $W_{CA} + W_{SF} + W_{ROI} = 1$ and P_{CA} , P_{SF} , and P_{ROI} are the probabilities of achieving U_{CA} , U_{SF} , and U_{ROI} , respectively. As noted earlier, in many situations, probability density functions are needed rather than discrete probabilities. The calculation then involves integration rather than multiplication and summation.

Once we have the scenarios ranked by $E[U_S]$ we will return to consideration of the optimize, adapt, hedge, and accept strategies from Figure 1.

Table 4. Set of Solutions Considered

Scenario	Examples	Uncertainties	Confidence in Requirements	Ability to Respond
Provide total vehicle package	OEM itself or acquisition of autonomous vehicle player	Can OEM really compete against the technology companies?	Hardware is high; software has some unknowns	Strength in integration; easier when OEM controls
Provide vehicle platform to host intelligent software	Alliance with Amazon, Apple, Google, Microsoft, or Uber	Why will intelligent platform players source OEM's vehicles?	Basic vehicle platform design is known, but can OEM do this at lowest cost?	Time to integrate software, which will evolve faster than hardware
Provide vehicle platform to host user-centered HMI	Alliance with advocacy groups for disabled & older adults	Why will user-centered HMI players source OEM's vehicles?	How will HMI requirements impact vehicle design?	Time to integrate software, which will evolve faster than hardware
Provide vehicle platform without alliance	OEM will manufacture desired platforms	Why will major players source OEM's vehicles?	Basic vehicle platform design is known; can OEM do this at the lowest cost?	Time to integrate software; design in modularity
Provide integrated mobility services	OEM will provide total mobility experiences	Can OEM competitively manage an end-to-end service?	Auto OEMs do not really understand business model, but does anyone?	Longer time to build out entire ecosystem

Table 5 summarizes assumed probabilities and utilities for the five scenarios. The risk associated with CA is primarily a requirements risk (i.e., the market risk of not having the right offering or best offering). The risk associated with SF is primarily a technology risk (i.e., the risk of not creating, or being able to create, a competitive technology platform). The risk associated with ROI includes both requirements and technology risks.

The reasoning underlying the assumptions in Table 5 is as follows:

- **Competitive Advantage:** U_{CA} is high if providing total solution, moderate if only providing vehicle; P_{CA} is low without strong partners, not just branding partners
- **Strategic Fit:** U_{SF} is high if only providing vehicle, moderate if also providing intelligent software; P_{SF} is high if only providing vehicle, moderate if integrating partners' intelligent software
- **Return on Investment:** U_{ROI} is high if providing total solution, moderate if partnering, low if only providing vehicle; P_{ROI} is low if providing total solution, moderate if partnering or only providing vehicle

The scenarios differ significantly in terms of probabilities of success and utilities if successful. The scenarios also differ significantly in terms of costs of success. Scenarios 1 and 5 represent total up-front commitments and the net present value (NPV) of financial projections would underlie ROI calculations. Scenarios 2 and 3 represent hedges against the risks of not being a player. For these scenarios, net option value (NOV) would be the metric in ROI calculations. Scenario 4 represents an accept strategy, as it exploits existing capabilities and will require the least investment.

Table 5. Assumed Probabilities and Utilities for the Five Scenarios

Scenario	Competitive Advantage		Strategic Fit		Return on Investment	
	P_{CA}	U_{CA}	P_{SF}	U_{SF}	P_{ROI}	U_{ROI}
Provide total vehicle package	Low (P = 0.1)	High (U = 0.9)	Moderate (P = 0.7)	Moderate (U = 0.5)	Low (P = 0.1)	High (U = 0.9)
Provide vehicle platform as host	Moderate (P = 0.3)	High (U = 0.9)	Moderate (P = 0.7)	High (U = 0.9)	Moderate (P = 0.3)	Moderate (U = 0.5)
Provide vehicle platform to host HMI	Low (P = 0.1)	High (U = 0.9)	Moderate (P = 0.7)	High (U = 0.9)	Moderate (P = 0.3)	Moderate (U = 0.5)
Provide vehicle platform only	Low (P = 0.1)	Moderate (U = 0.5)	High (P = 0.9)	High (U = 0.9)	Moderate (P = 0.3)	Low (U = 0.1)
Provide integrated mobility services	Low (P = 0.1)	High (U = 0.9)	Moderate (P = 0.7)	Moderate (U = 0.5)	Low (P = 0.1)	High (U = 0.9)

Boer (2008) suggests how to value a portfolio that includes some investments characterized by NPV and others by NOV. He argues for strategic value (SV), which is given by

SV = NPV + NOV (4)

The NPV component represents the value associated with commitments already made, while the NOV component represents contingent opportunities for further investments should the options be "in the money" at a later time.

Figure 4 provides results for $E[U_S]$ with varying assumptions regarding the relative importance (weighting) of CA, SF, and ROI. The overall results are as follows:

- Scenario 2 has the highest $E[U_S]$ unless SF dominates
- Scenarios 2 and 3 have the highest $E[U_S]$ if ROI and/or CA dominate
- Scenario 4, followed by 2 and 3, has the highest $E[U_S]$ if SF dominates
- Scenarios 1 and 5 have the lowest $E[U_S]$ across all weighting assumptions

Discussion

These results reflect, of course, the assumptions in Table 5. These assumptions could be varied to assess their impact, but given that W × P × U occurs in all the underlying equations, the variations of W in Figure 4 reasonably reflect the range of possibilities.

Scenario 1 embodies a significant technology risk in a very competitive market, while Scenario 5 involves a significant requirements risk in attempting to provide services not typical for an OEM. Both of these risks could be hedged with acquisitions of a software company (Scenario 1) or a service company (Scenario 5). This might be difficult, as the market capitalizations of the automotive OEMs are much lower than the capitalizations of likely and attractive acquisition targets.

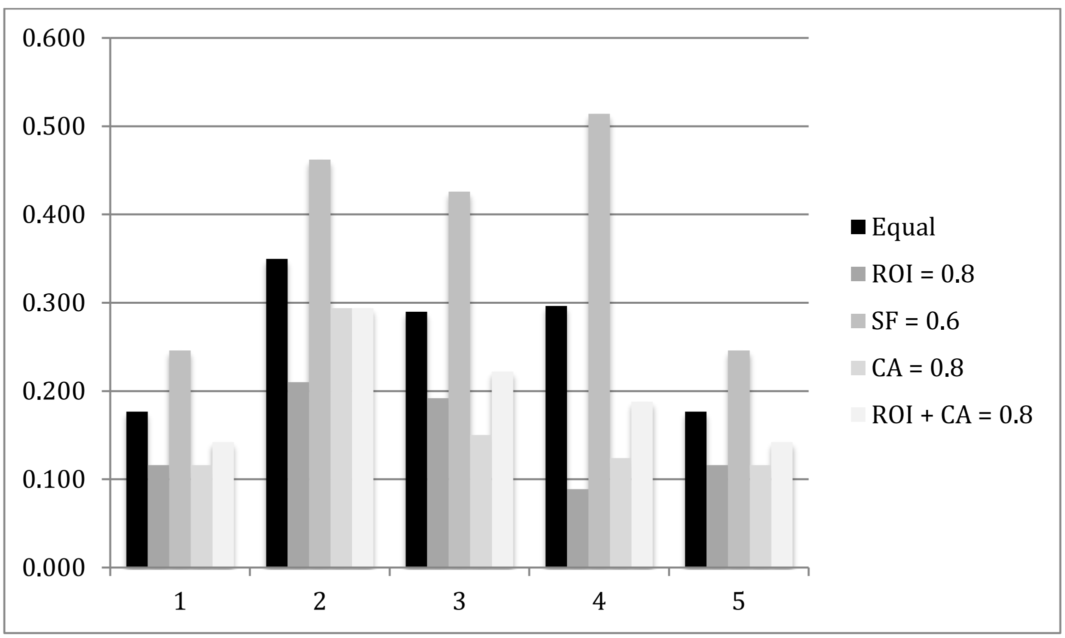


Figure 4. Expected Utilities for the Five Scenarios With Varying Weights

Scenarios 2 and 3 represent hedges against these risks as well but result in dividing the share of the vehicle that the OEM will provide and, hence, its revenues and profits. Nevertheless, they are attractive because they decrease the competition and provide key technologies. These scenarios also allow the freedom to pursue other strategies as uncertainties resolve themselves.

Scenario 4 focuses on leveraging SF. It represents acceptance by the OEM of whatever leverage is provided by its core competencies. This also involves acceptance that they will have to compete with the other automotive OEMs that want to provide the vehicle platform. They are quite familiar with this type of competition.

The resulting overall strategy involves a portfolio of three investments:

- Substantial investment in Scenario 2—a hedge against market and technology risks
- Moderate investment in Scenario 3—a hedge against Scenario 2 not resulting in a partner
- Baseline investment in Scenario 4—acceptance of a traditional role in the automotive marketplace

With the strategies decided, one is ready to apply the QFD and DSM aspects of Figure 3 to the functionality in Table 3. This requires that the set of stakeholders be expanded to include:

- OEM
- Partners
- Suppliers
- Car Service Providers
- Car Service Customers

It also requires characterizing competing offerings, whose likely functions, features, and pricing will have been sleuthed via business intelligence.

This illustrates the multilevel nature of the methodology. The first question is which of the business scenarios make sense and, for those that make sense, determining the appropriate strategy for pursuing each scenario. The idea is to iteratively refine the chosen scenarios and strategies, which will influence the nature of investments—for example, whether one makes a total commitment up front (NPV), hedges uncertainties with smaller investments (NOV), or simply accepts one's current position and waits to see how the market develops.

Conclusions

Engineering involves designing solutions to meet the needs of markets or missions. Organizations would like to have the flexibility and agility to address both uncertain needs and uncertain technologies for meeting these needs. This article has presented and illustrated a framework that provides this flexibility and agility. We considered how uncertainties arise, contrasting the automotive and defense domains. We proposed an approach to managing uncertainties. We considered how to represent alternative solutions and project the value of each alternative. Possible use cases for our framework were discussed. A detailed case study of autonomous vehicles to enhance the mobility of disabled and older adults was presented.

We did not consider but need to acknowledge broader risks. It is quite imaginable that driverless car technologies, once deployed, will lead to inadvertent failures with substantial consequences (Danzig, 2018). It is also possible that sweeping organizational and societal trends will substantially disrupt this seemingly immense market opportunity (Rouse, 2019, 2020). The current pandemic is a case in point. The impacts of climate change are on the horizon.

Understanding and managing uncertainties need to be core competencies in companies, agencies, and institutions. As this article has argued, uncertainties need to be rigorously and systematically addressed. Managing for success must also include forecasting and managing potential failures.

Acknowledgments

This article is based upon research supported by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract W15QKN-18-D-0040. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the U.S. Department of Defense.

References

Al-Ashaab, A., Golob, M., Attia, U. M., Khan, M., Parsons, J., Andino, A., Perez, A., Guzman, P., Onecha, A., Kesavamoorthy, S., Martinez, G., Shehab, E., Berkes, A., Haque, B., Soril, M., & Sopelana, A. (2013). The transformation of the product development process into lean environment using set-based concurrent engineering. *International Journal of Concurrent Engineering: Research and Applications*, *18*(1), 41–53.

Auto Alliance. (2019). *Assessing transportation needs of people with disabilities and older adults: Report of workshop 1*.

Boer, F. P. (2008). *The valuation of technology: Business and financial issues in R&D.* Wiley.

Boyd, R. C. (2004). *The fighter pilot who changed the art of war*. Back Bay Books.

Carlson, J. M., & Doyle, J. (2000). Highly optimized tolerance: Robustness and design in complex systems. *Physical Review Letters*, *84*(11), 2529–2532.

Chambers, J. (1997). *The Oxford companion to American military history*. Oxford University Press.

Danzig, R. (2018, May 30). *Technology roulette: Managing loss of control as many militaries pursue technological superiority*. Center for a New American Security. https://www.cnas.org/publications/reports/technology-roulette

Deshmukh, A., et al. (2010, November 9–10). Valuing flexibility [Paper presentation]. *Proceedings of the 2nd Annual SERC Research Review Conference*, College Park, MD, United States.

Eppinger, S. D., & Browning, T. R. (2012). *Design structure matrix methods and applications*. MIT Press.

Feickert, A. (2009). *The Marines expeditionary fighting vehicle (EFV): Background and issues for Congress* (CRS Report No. RS22947). Congressional Research Service. https://apps.dtic.mil/sti/pdfs/ADA498364.pdf

Ferreira, S., Collofello, J., Shunk, D., & Mackulak, G. (2009). Understanding the effects of requirements volatility in software engineering by using analytical modeling and software process simulation. *Journal of Systems and Software*, *82*(10), 1568–1577.

Hanawalt, E. S., & Rouse, W. B. (2010). Car wars: Factors underlying the success or failure of new car programs. *Journal of Systems Engineering*, *13*(4), 389–404.

Hanawalt, E., & Rouse, W. B. (2017). Assessing location attractiveness for manufacturing automobiles. *Journal of Industrial Engineering and Management*, *10*(3), 73–89.

Hauser, J. R., & Clausing, D. (1988, May–June). The house of quality. *Harvard Business Review*, 63–73.

Henkoff, R. (1994, April 18). Keeping Motorola on a roll. *Fortune*.

Keeney, R. L., & Raiffa, H. (1993). *Decisions with multiple objectives: Preference and value tradeoffs*. Cambridge University Press.

Laris, M. (2020, April 17). Downsides of self-driving cars could swamp benefits if DC region fails to act, study says. *The Washington Post*.

Liu, C., Rouse, W. B., & Belanger, D. (2020). Understanding risks and opportunities of autonomous vehicle technology adoption through systems dynamic scenario modeling: The American insurance industry. *IEEE Systems Journal*, *14*(1), 1365–1374.

Liu, C., Rouse, W. B., & Hanawalt, E. (2018). Adoption of powertrain technologies in automobiles: A system dynamics model of technology diffusion in the American market. *IEEE Transactions on Vehicular Technology*, *67*(7), 5621–5634.

Liu, C., Rouse, W. B., & Yu, X. (2015). When transformation fails: Twelve case studies in the automobile industry. *Journal of Enterprise Transformation*, *5*(2), 71–112.

Lucero, D. S. (2018, July 11). The mash-up rubric: Strategies for integrating emerging technologies to address dynamic requirements. *Proceedings of 28th Annual INCOSE International Symposium*, Washington, DC.

Mattis, J. (2019, October 13). Interview. *Meet the press*.

Mikaelian, T., Rhodes, D. H., Nightingale, D. J., & Hastings, D. E. (2012). A logical approach to real options identification with application to UAV systems. *IEEE Transactions on Systems, Man, and Cybernetics: Part A. Systems and –s*, *42*(1), 32–47.

Myers, S. L. (2001, April 23). Pentagon panel urges scuttling howitzer system. *New York Times*, A1.

Pennock, M. J., & Rouse, W. B. (2016). The epistemology of enterprises. *Journal of Systems Engineering*, *19*(1), 24–43.

Ross, A. M., & Rhodes, D. H. (2008), Using natural value-centric time scales for conceptualizing system timelines through Epoch-Era Analysis. *Proceedings of INCOSE International Symposium*, *18*(1), 1186–1201.

Rouse, W. B. (1991). *Design for success: A human-centered approach to designing successful products and systems*. Wiley.

Rouse, W. B. (1992). *Strategies for innovation: Creating successful products, systems, and organizations*. Wiley.

Rouse, W. B. (1993). *Catalysts for change: Concepts and principles for enabling innovation*. Wiley.

Rouse, W. B. (1998). *Don't jump to solutions: Thirteen delusions that undermine strategic thinking.* Jossey–Bass.

Rouse, W. B. (2005). A theory of enterprise transformation, *Journal of Systems Engineering*, *8*(4), 279–295.

Rouse, W. B. (2007). *People and organizations: Explorations of human-centered design*. Wiley.

Rouse, W. B. (2010). Options for surveillance and reconnaissance. In W. B. Rouse (Ed.), *The economics of human systems integration* (Chap. 15). Wiley.

Rouse, W. B. (2015). *Modeling and visualization of complex systems and enterprises: Explorations of physical, human, economic, and social phenomena*. Wiley.

Rouse, W. B. (2019). *Computing possible futures: Model based explorations of "what if."* Oxford University Press.

Rouse, W. B. (2020). *Failure management: Malfunctions of technologies, organizations, and society.* Oxford University Press.

Rouse, W. B., & Boff, K. R. (2004). Value-centered R&D organizations: Ten principles for characterizing, assessing & managing value. *Journal of Systems Engineering*, *7*(2), 167–185.

Rouse, W. B., Howard, C. W., Carns, W. E., & Prendergast, E. J. (2000). Technology investment advisor: An options-based approach to technology strategy. *Information-Knowledge-Systems Management*, *2*(1), 63–81.

Rouse, W. B., & McBride, D. K. (2019). A systems approach to assistive technologies for disabled and older adults. *The Bridge*, *49*(1), 32–38.

Senge, P. (1990). *The fifth discipline*. Doubleday/Currency.

Silver, M. R., & de Weck, O. L. (2007). Time-expanded decision networks: A framework for designing evolvable complex systems. *Journal of Systems Engineering*, *10*(2), 167–188. http://strategic.mit.edu/docs/2_16_SE_10_2_TDN.pdf

Singer, D., Strickland, J., Doerry, N., McKenney, T., & Whitcomb, C. (2017). Set-based design. *Society of Naval Architects and Marine Engineers, Technical and Research Bulletin*, 7–12.

Sobek, D. K., Ward, A. C., & Lifer, J. K. (1999). Toyota's principles of set-based concurrent engineering. *Sloan Management Review*, *40*(2), 67–83.

Yu, X., Rouse, W. B., & Serban, N. (2011). A computational theory of enterprise transformation. *Journal of Systems Engineering*, *14*(4), 441–454.

Analyzing Digital Transformation using the Zachman Framework and SysML

Mark Kassan—is graduating with a Master of Science degree in Systems Engineering from the Air Force Institute of Technology as a part-time student. He currently supports the leadership of the Air Force Digital Campaign full-time as a plans and programs engineer of the Air Force Materiel Command Engineering and Technical Management Directorate, Systems Engineering Division. [mark.kassan.2@us.af.mil]

John Colombi, PhD—is a professor and curriculum chair of the Systems Engineering program at the Air Force Institute of Technology, Department of Systems Engineering and Management. Before joining the faculty, he served 21 years of active duty in the Air Force as a developmental engineer. He is author or coauthor on over 100 publications. [john.colombi@afit.edu]

Thomas Ford, PhD—is a Principal Systems Engineer with KBR (Government Solutions) assigned as an Assistant Professor of Systems Engineering at the Air Force Institute of Technology, Department of Systems Engineering and Management. Prior to joining KBR, he served for 22 years as an Air Force officer. He has authored or co-authored 14 publications on topics relating to model-based systems engineering and interoperability. [thomas.ford@centauricorp.com]

Major Warren Connell, PhD—is an assistant professor of systems engineering at the Air Force Institute of Technology, Department of Systems Engineering and Management. Prior to joining the faculty, he served for 24 years of active duty in a variety of software development and cybersecurity roles, most recently as the Cyber Test Lead for the F-35 Joint Program Office. [warren.connell@afit.edu]

Abstract

The Air Force Materiel Command is undergoing a digital transformation to increase the speed of delivering new warfighter capabilities. This Digital Campaign consists of six Lines of Effort (LOEs) formed with diverse goals to transform the enterprise. This research investigated using the Zachman Framework and Systems Modeling Language to analyze this transformation. Extensive modeling captured the as-is Preliminary Design Review (PDR) process and mapped LOE goals as primary impacts to Zachman cells. This led to an identification of a to-be digital PDR process. Secondary impacts were also identified and traced using a relationship analysis. Four discoveries were made. (1) Enterprise modeling in Zachman is analogous to a system decomposition under typical systems engineering approaches. (2) As long as the transformation goals do not change, the Zachman cells, and those entities mapped into those cells, will be directly affected by the new digital enterprise. (3) Different from past process transformation efforts, the Digital Campaign has focused on technology upgrades to drive process change. (4) Lastly, model analysis revealed transformation gaps within certain cells that should be covered with new goals. This research provides a formal, model-based methodology for guiding enterprise-wide improvements in pursuit of Air Force digital transformation.

Introduction

Pressure is being put on the United States Air Force to maintain its dominance over potential adversaries as the speed of technology is increasing (Brown, 2020). In addition, the Air Force's time to field its most advanced and complex weapon systems has been increasing over the past 50 years. This is allowing these potential adversaries to develop and field new capabilities faster than the Air Force. For example, back in the 1970s, the F-16's concept to field averaged about 6 years, whereas the latest aircraft developed, the F-35, will exceed 20 years from concept to full operational capability. It is believed that this fielding time will only continue to increase unless the Air Force makes a paradigm shift in the way it acquires new capabilities (Alia-Novobilski, 2020).

This need for change has been realized by senior Air Force leadership including the former Assistant Secretary of the Air Force for Acquisition, Technology, and Logistics, Dr. Will Roper, and the commander of the Air Force Materiel Command (AFMC), General Arnold Bunch. The primary focus of this transformation is the use of digital models and artifacts integrated across the lifecycle. To address this digital transformation, Gen Bunch in March 2020 established a Digital Campaign to drive the whole enterprise to move towards transforming and create an environment to promote change in six lines of efforts. These six lines of effort (LOE) address (1) Information Technology Infrastructure, (2) Models and Tools, (3) Standards, Data, and Architectures, (4) Lifecycle Strategies and Processes, (5) Policy and Guidance, and (6) Workforce and Culture.

Digital Campaign LOE teams are trying to understand and improve a very large and complex enterprise comprised of many distributed organizations, people, and processes that are highly intertwined. The processes have been continuously evolving since the 1960s (Fox, 2011). The Digital Campaign is getting things done by grit, experience, and instinct to overcome complexity in transforming a very large enterprise. As a result, without a rigorous and structured effort to break down the complexity, identify, map, and unravel the interactions, and transcribe individual processes and digital flows, the Digital Campaign is bound to miss critical aspects. This is where an effort to model an Enterprise Architecture (EA) can introduce a formal methodology to provide the insight needed to successfully complete the digital transformation. This paper demonstrates a methodology for modeling the AFMC acquisition enterprise to visualize and gain insight into the digital transformation effort.

This research used a systems engineering approach to build a System Modeling Language (SysML) model within the Zachman Framework for a technical review of the AFMC acquisition enterprise. Once that was completed, the Digital Campaign goals were mapped as requirement changes to the model. The primary effects of these changes on the enterprise's people, products, and processes were studied and documented. The research then used the inherent structure found in the Zachman Framework of the AFMC EA to show how secondary impacts can be identified. The final step of the research identified gaps with the Digital Campaign digital change approach following a systems engineering approach applied to the SysML model of the enterprise.

This exploration addresses the digital transformation of the AFMC acquisition community and directly supports the AFMC Digital Campaign, its goals, and activities. The AFMC acquisition enterprise is large and complex; modeling all of it to the appropriate fidelity would take considerable amount of time beyond the scope of a single effort. Therefore, this research focused on the particular event of a Preliminary Design Review (PDR) within the Technology Maturation and Risk Reduction (TMRR) phase of a defense acquisition. The PDR process involves sufficient personnel, resources, and data artifacts within an acquisition program to provide enough model elements permitting adequate research analysis.

The enterprise is assumed to be acquiring a major capability, either a new program or a major modification to an existing weapon system that would require a PDR within the defense acquisition process. As always, there are shortcuts and tailoring activities available to a program manager and chief engineer.

The next section addresses the Zachman framework used for this research. Then, the paper describes the SysML method used to build the as-is and to-be enterprises. Next, the paper addresses the primary impacts analysis from mapping the Digital Campaign goals. The paper then details the analysis for secondary impacts and gap analysis, and finally concludes with findings and future research opportunities.

Zachman Enterprise Architecture Framework

An architecture framework is a tool for describing an architecture using "conventions, principles and practices established within an application-specific domain and/or stakeholder community" (IEEE, 2011). It presents unique stakeholder perspectives in views that communicate information of concern to that stakeholder about the system.

The earliest beginnings of EAs can be traced to an IBM methodology in the late 1960s called Business Systems Planning (BSP), the purpose of which was to deliberately plan information systems by collecting data through interviewing managers and then developing a top-down plan involving models representing a logical structure that could be implemented. (Kotusev, 2016). There were several improvements of the BSP through the 1980s when John Zachman introduced his framework internally to IBM.

John Zachman published his original framework in 1987 and in 1992 extended his framework into 30 categories in a matrix where there were five perspectives (planner, owner, designer, builder, and subcontractor) in rows and six interrogatives (*what*, *how*, *where*, *who*, *when* and *why*) along the columns. Each of the 30 cells in this matrix is unique, suggesting that it serves as a "periodic table" for entities. This resulted in a diagram representing a different abstraction and perspective of the EA (Sowa & Zachman, 1992). In 2011, Zachman updated his framework matrix to version 3.0 and titled it *The Zachman Framework for Enterprise Architecture*. This version, shown in Figure 1, is a matrix of six perspectives (executive, business management, architect, engineer, technician and the enterprise) and six interrogatives (*what*, *how*, *where*, *who*, *when*, and *why*).

There are many different definitions of an enterprise in the literature. The major theme in many documents is that an enterprise is an organization or activity whose boundary is defined by a common mission and who uses technology, processes, and resources to perform that mission (Bernard, 2012). In the Air Force, as in any complex large organization, there are numerous enterprises. This research focuses on AFMC as an organization who performs an acquisition mission of delivering capability to warfighters. This includes the executing programs and the command and center support organizations that can provide enterprise-level processes, technologies, and resources to program offices where goal achievement is focused.

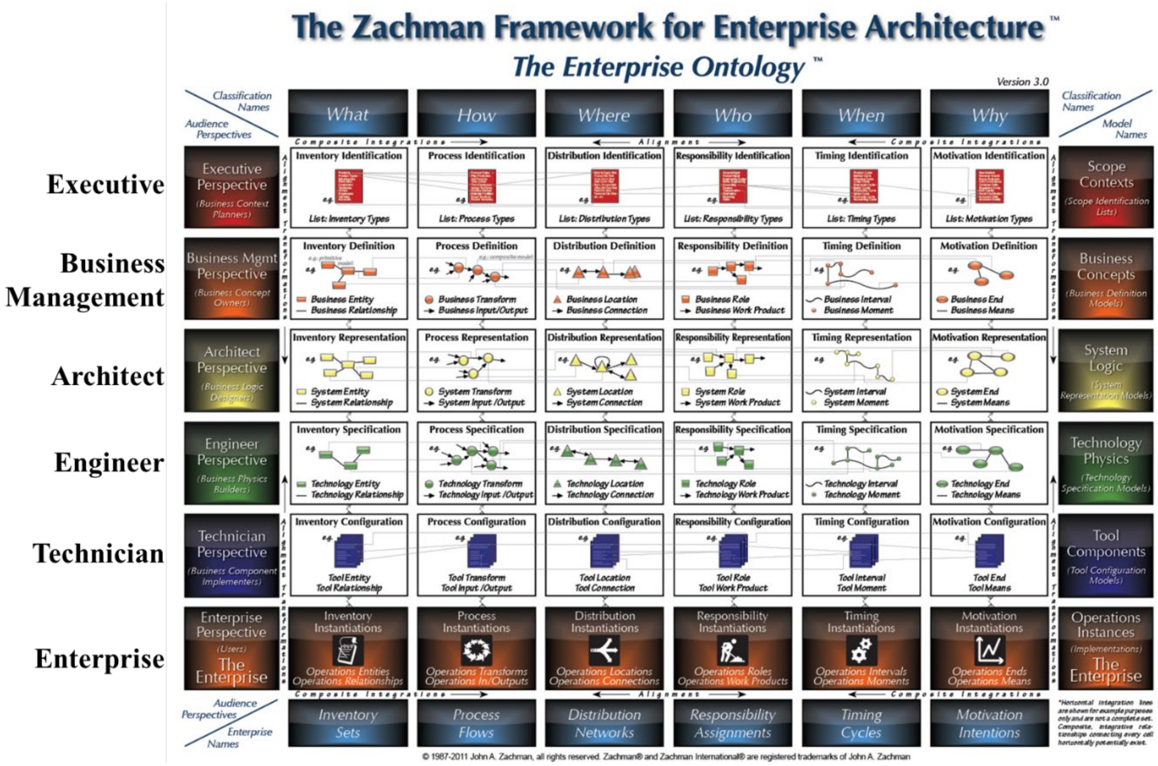


Figure 1: Zachman Framework, published with the permission of John A. Zachman and Zachman International®, Inc. (https://www.zachman.com)

Modeling Methodology

Dassault No Magic CAMEO Systems Modeler v19 was used as the modeling tool because it was available in the Air Force Institute of Technology (AFIT) academic environment. It is also one of the SysML modeling tools used by a large portion of the Air Force acquisition community. There are other tools also used within the Air Force enterprise. These include Sparx Systems Enterprise Architect, IBM Rational Rhapsody, SPEC Innovations Innoslate, and Siemens' Systems Modeling Workbench.

The Zachman Framework is a structure for visualizing a complex enterprise. John Zachman is clear that he does not prescribe a method for his framework, so it is left up to the user to determine the method. This research effort chose to use a systems engineering approach using the SysML language as outlined in Figure 2.

The PDR was picked because it is a prominent technical review within the acquisition process that the majority of Air Force programs must go through. The PDR is also relatively universal in that there are similar documents needed for the review across different programs. The review generally takes place between an external entity (contractor) and a program office, where the contractor takes the time to prove to the program office and other stakeholders that it has met the system requirements in allocating the requirements down to subsystems, software, and components of the needed system. In addition, risk and affordability are looked at during a PDR (IEEE, 2015).

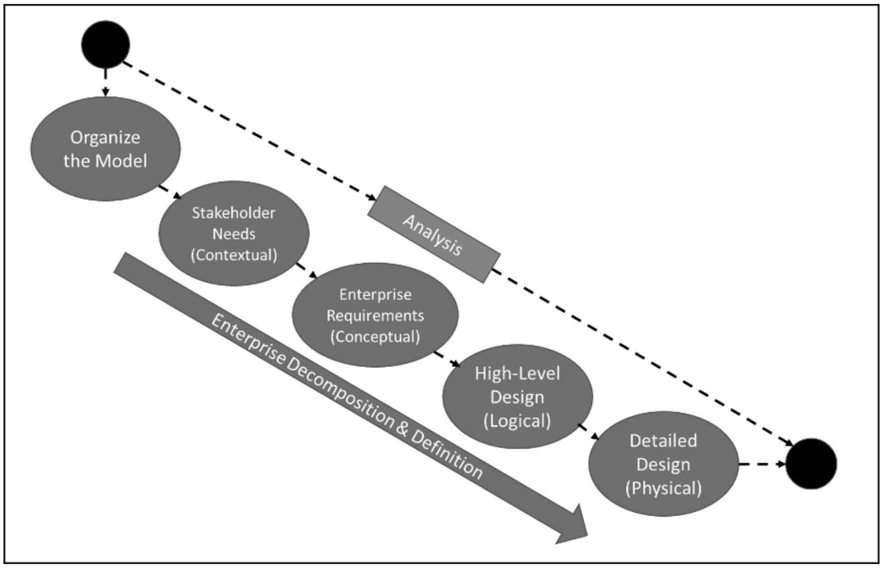


Figure 2: Systems Engineering Method for Enterprise Design

Organizing the Acquisition Enterprise Architecture Model

The first step of this approach is to “organize the model” as shown in Figure 2. To accomplish this step within CAMEO, a package model was built as shown in Figure 3, which represents the Zachman Framework of Figure 1. A top-level package is created for each perspective. A package is a folder that establishes a way to contain and organize related information within a model.

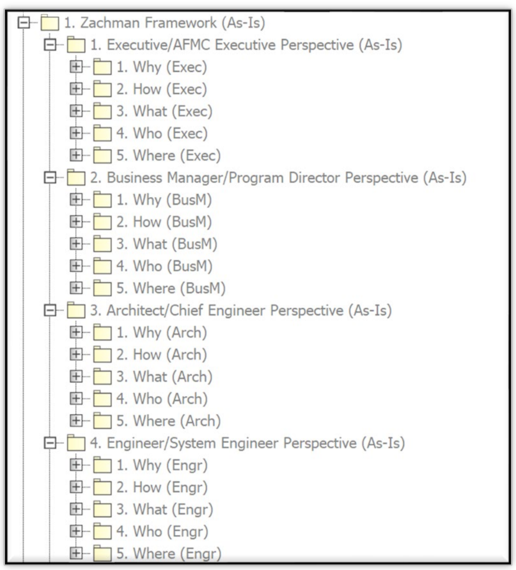


Figure 3: Package Structure Representing Zachman Framework Matrix

The perspectives represent the stakeholders within the enterprise and who participate in the overall enterprise outcome. For this exercise, only four perspectives (Executive, Business Management, Architect, and Engineer) were addressed. The other two perspectives are the Technician (contractor) and the Enterprise (instantiations). Within each of the perspective packages were packages representing the Interrogatives (columns) of the Zachman

Framework. These sub-packages represent the intersection of the perspectives and the interrogatives and will be referred to as a cell. It is within each cell that the modeling artifacts will reside in the form of diagrams, entities, and relationships. To understand what diagram will go in each cell, research was done to define each of the four perspectives.

The Executive perspective is known as a contextual perspective. This is the person who is setting the strategy for the enterprise. This person is concerned with depicting in broad terms, the basic scope of the enterprise (Sowa & Zachman, 1992). This research defines this person as an AFMC executive with duties to understand and provide the overall resources and data needed to meet many customers' materiel requirements within the DoD acquisition process.

The Business Management (or business manager) perspective is known as a conceptual perspective. This is the person who runs the execution organization. This person's perspective is from someone who has to work within the enterprise business and cares about the business products and processes and how they interact (Sowa & Zachman, 1992). This research defines this person as a program office director responsible for producing a product and related data that meets a customer's materiel requirement.

The Architect perspective is known as a logical perspective. This is the person who designs discipline into the organization. This person is concerned with the details of the materiel solution, data products, and the business processes that produce those data products (Sowa & Zachman, 1992). This research defines this person as a chief engineer responsible for the detailed processes that produce the materiel solution, and the data products required to meet a customer's materiel requirement.

The Engineer perspective is known as a physical perspective. This is the person who is responsible for applying specific technologies to solve the problems of the organization. This person is concerned with the constraints of the technology and processes used to produce the data products and must adapt the information technology to meet the enterprise requirements (Sowa & Zachman, 1992). This research defines this person as a systems engineer responsible for applying available information technology and support personnel to the program office business processes that produce the materiel solution and data products.

The perspective definitions are the rows of the Zachman Framework. The five columns for this research are represented as the interrogatives: *why*, *how*, *what*, *who* and *where*. The sixth column, the *when* interrogative, was not considered. Definitions of these columns are described in the following paragraphs.

The *why* column describes the motivation of the enterprise. These are typically described in terms of goals and objectives. Within a model these are best represented as requirements diagrams depicting relationships of goals to sub-goals (or objectives) of the enterprise (Sowa & Zachman, 1992). This research uses requirements diagrams that link the executive's stakeholder requirements (those of the customers' having a materiel solution need) down to the engineer's IT requirements used to meet the requirements flowing back up to the stakeholder's requirements.

The *what* column can be described as the data artifact. Generally speaking, this is the "things" of the enterprise (Zachman, 1987). For this research, the things are the data products being produced, shared, consumed, used, and stored by the enterprise. These are represented by blocks, Block Definition Diagrams (BDD), and their relationships.

The *how* column can be described as the function artifact and is the column where business processes of the enterprise would be described for creating and transforming the enterprise products (Zachman, 1987). Activity Diagrams (ACT) were used to describe the processes of concern for each perspective. These perspectives included the executive's scope of the overall acquisition process to the more detailed processes of the architect and engineer, who are involved in preparing and conducting a PDR event.

The *who* column is the people and organization artifact. This column depicts organizational structure as well as the roles of people within the organization. Organizational structure usually shows hierarchal lines of authority or links to who is providing the work product or work service (Sowa & Zachman, 1992). BDDs were used to represent parts of the acquisition enterprise involved in and concerned with the PDR technical event.

The *where* column is the location artifact. This column depicts where the business is occurring or flowing between the enterprise network and sites, depending on the perspective (Sowa & Zachman, 1992; Zachman, 1987). BDDs, blocks and relationships were used to represent locations, where locations are defined as a place for organizations or IT systems within the acquisition enterprise.

This effort used the nomenclature of Figure 4 to refer to each Zachman cell. Each cell is the intersection of an interrogative column and a perspective row. For instance, the *what* interrogative column intersects with the engineer perspective and is referred to as the cell of "What (Engr)," where engineer is abbreviated as "Engr." Other abbreviations include "Exec" for executive, "BusM" for business manager, and "Arch" for architect.

	WHY	HOW	WHAT	WHO	WHERE
EXECUTIVE (AFMC Executive Leader)	Why (Exec)	How (Exec)	What (Exec)	Who (Exec)	Where (Exec)
BUSINESS MANAGER (Program Director)	Why (BusM)	How (BusM)	What (BusM)	Who (BusM)	Where (BusM)
ARCHITECT (Chief Engineer)	Why (Arch)	How (Arch)	What (Arch)	Who (Arch)	Where (Arch)
ENGINEER (Systems Engineer)	Why (Engr)	How (Engr)	What (Engr)	Who (Engr)	Where (Engr)

Figure 4: Zachman Cell Definitions Used in This Research

Developing the As-Is Acquisition Enterprise Architecture Model

Once the Zachman Framework is set up and understood within the CAMEO tool, the next step is to start with modeling the contextual level (executive perspective) representing the stakeholder needs, followed by the conceptual level (business manager) representing the enterprise requirements. Once complete with these two perspectives, the research moved to the next step of high-level design. This step is accomplished by modeling the architect's logical-level cells. And finally, the last step of the systems engineering method is to model the detailed design for the engineer's physical-level cells. Relationship analysis is performed throughout this modeling process. Each of these steps is referenced in the process of Figure 2.

The following paragraphs step through the research activity that resulted in the data artifacts produced for each cell within the as-is EA for the *what* interrogative. The as-is EA is a representation of the actual acquisition enterprise for the PDR and related data, organizations, personnel, and processes as viewed by each perspective.

Executive Perspective

The things that the AFMC Executive is concerned about are represented in the block definition diagram of Figure 5. These are the top-level data needed to manage and provide a materiel solution to the customer stakeholder. These data include the design, performance, cost, risk, requirement, maintenance, operational, logistics, security, test, interface, and schedule data of the materiel solution.

The creation of all diagrams for each cell of the executive perspective sets the context of the AFMC acquisition enterprise and the PDR process that this research continues to break down through the modeling of the other framework perspectives. The next perspective is the business manager or program office director perspective.

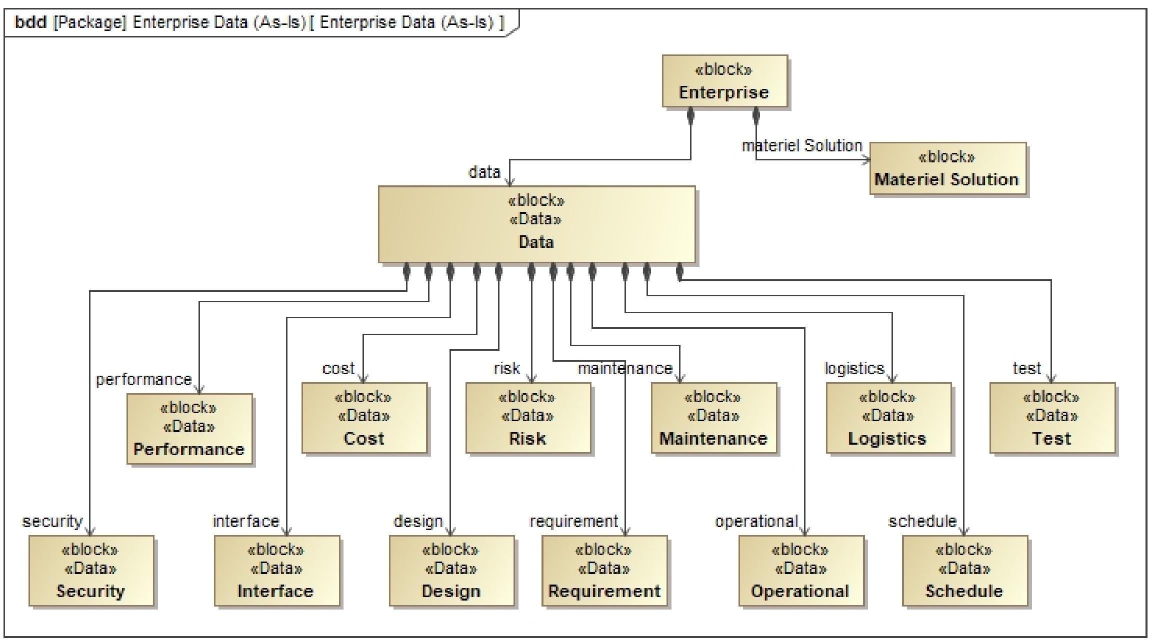


Figure 5: Enterprise Data for the AFMC Executive (Cell: What (Exec))

Business Manager/Program Director Perspective

The program director is concerned with all of the technical and business processes within the AFMC acquisition enterprise. IEEE 15288.2 *Standard for Technical Reviews and Audits on Defense Programs* (IEEE, 2015) was used as the basis for defining the processes and documents. The program director is concerned with the data products that are related to the data that validate the materiel solution. These are represented by the model in Figure 6. As the model shows, the data needed for the materiel solution are contained in the documents developed by other processes of the enterprise. These documents are reviewed and approved in the PDR process.

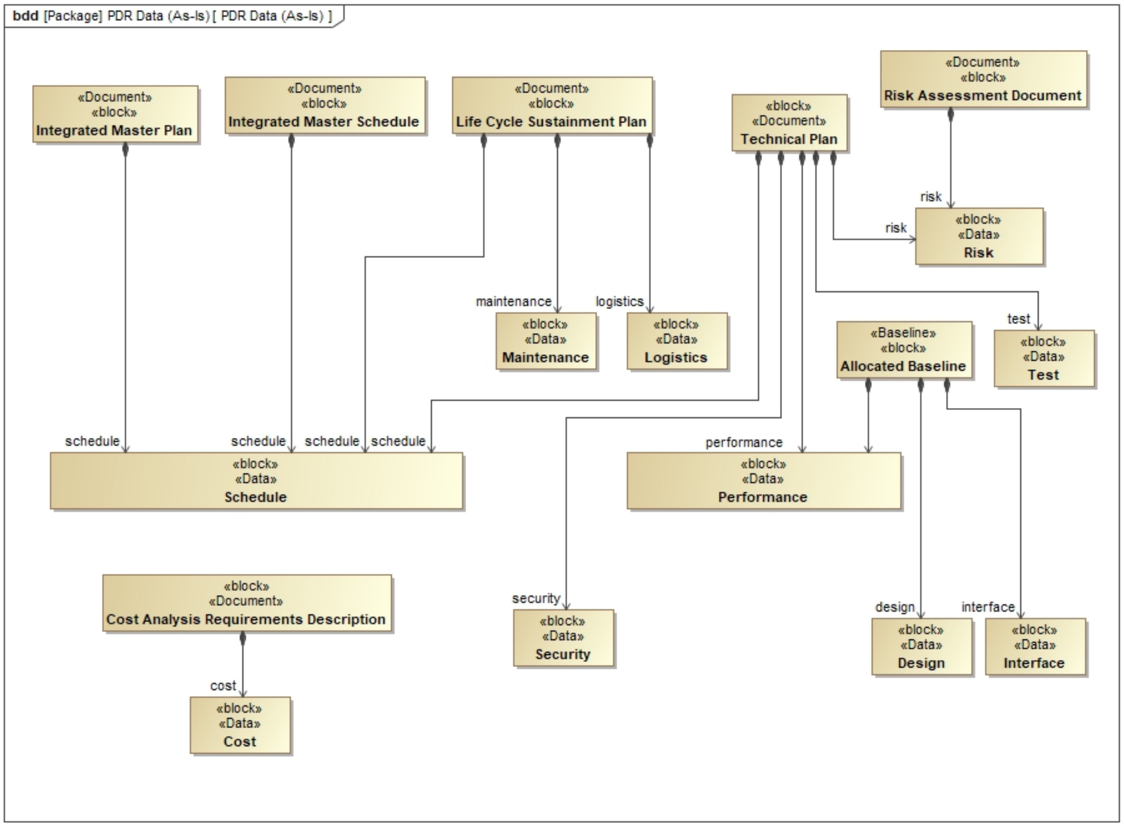


Figure 6: PDR Data for the Program Director (Cell: What (BusM))

Architect/Chief Engineer Perspective

The chief engineer is responsible for setting up, complying with, managing, and executing the PDR process. They have a need to deliver necessary documents for the PDR that represent the system under development. These documents include the Cost Analysis Requirements Document, the Life Cycle Sustainment Plan, the Integrated Master Schedule, the Integrated Master Plan, the Risk Assessment, the documents that represent the Allocated Baseline, and Technical Plans. The Technical Plans include documents such as the Test and Evaluation Plan, the Systems Engineering Plan, several different levels of verification and validation plans, and modeling and simulation plans (IEEE, 2015).

These documents are the entry products that will contain the data required by the materiel solution as shown in Figure 7. In addition, the chief engineer is concerned about other entry documents needed to conduct a PDR which include the presentation document, the PDR membership list, and the PDR agenda. The other consideration is the PDR closure products as shown in Figure 7, which include products such as the PDR minutes, action items, and PDR summary report.

Engineer/Systems Engineer Perspective

The systems engineer cares about the relationship between the documents required for the PDR, and the IT and software tools needed to produce, review, comment on, and approve the documentation, and the IT and software tools needed to conduct the PDR meeting. Software tools include the MS Office products used to create and review documents, SharePoint site, and email used to share documents. The IT tools include the desktop computer, the network, and the server where documents are transmitted and stored. The documents include the PDR presentation, which is normally a PowerPoint-created document projected on a screen in the room and shared during the meeting for all participants to see. There are also several BDDs in this cell such as a BDD to represent the composition of the Allocated Baseline, a BDD to represent the composition of the PDR Entry Products and a BDD to represent the composition of the PDR Closure Products. These are considered the “things” that the system engineer is concerned with.

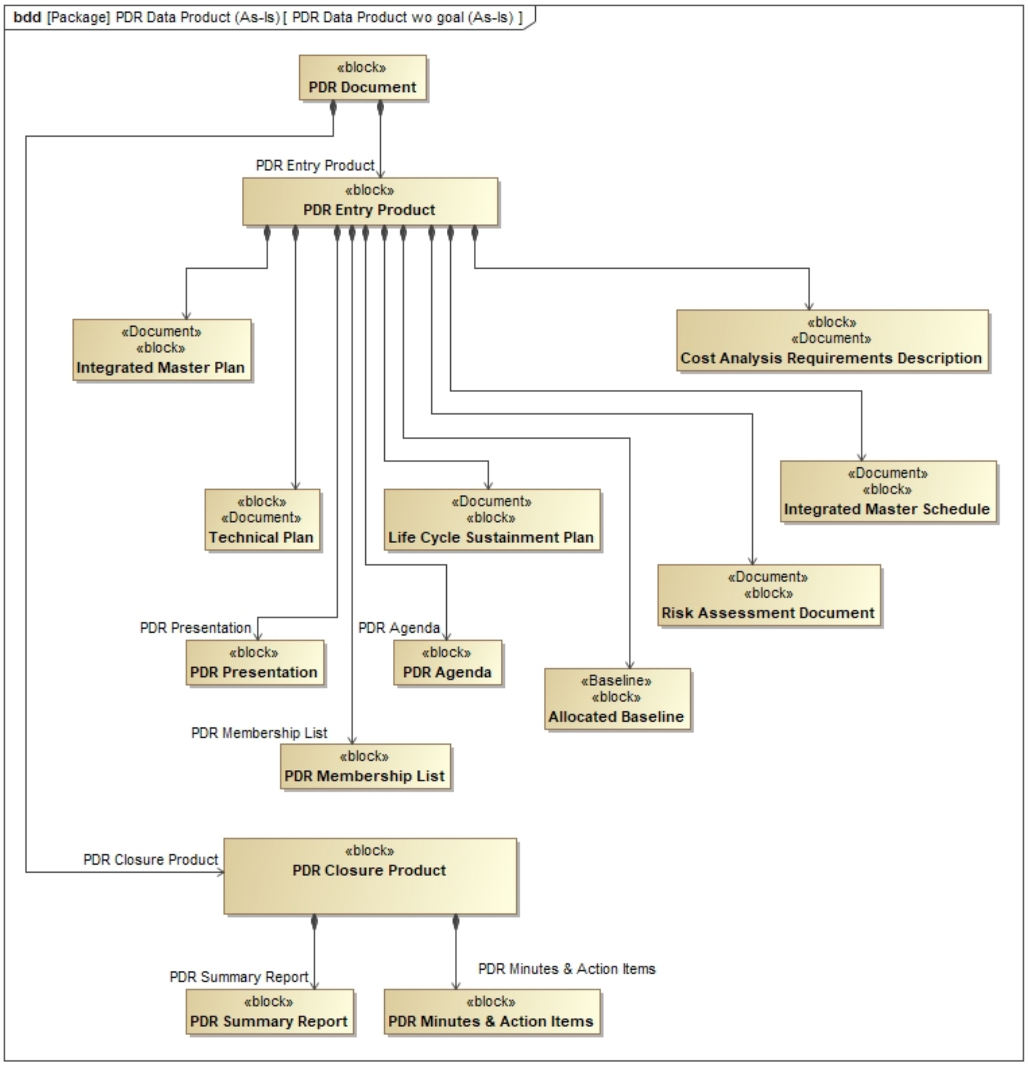


Figure 7: PDR Entry and Closure Products for the Chief Engineer (Cell: What (Arch))

This research modeled the enterprise within a framework that addressed four perspectives. For each perspective, the research addressed all the interrogatives of *why*, *how*, *what*, *who*, and *where* resulting in a picture of the business of conducting PDRs. Only the *what* interrogative was shown here as an example. The next step is to look at what are the primary effects of the Digital Campaign goals on this as-is enterprise model.

Primary Impact Analysis

To address the digital transformation impacts to the as-is AFMC acquisition enterprise model, the next step taken was to map the Digital Campaign’s goals to the Zachman

Framework. The Digital Campaign was established to digitally transform the acquisition enterprise. The Campaign leadership set up six LOE goals to accomplish this transformation. Those goals are shown in Table 1. LOE 4 addresses policy and guidance with its primary objective to review policies outside the AFMC acquisition enterprise and was not considered in this research.

Table 1: Line of Effort Goals of the AFMC Digital Campaign

Line of Effort	Line of Effort Name	Line of Effort Goal
0	Integrated Environment - IT Infrastructure	Provide overarching guidance to influence corporate IT improvement investments to enable a robust, secure infrastructure for the enterprise-wide Digital Campaign
1	Integrated Environment - Tools and Models	Provide an Integrated Digital Environment (IDE) of models and tools for collaboration, analysis, and visualization across the functional domains of AF users
2	Standards, Data, and Architectures	Provide overarching guidance on the use of Government Reference Architectures (GRA) and related standards and datasets for use in an integrated digital environment for application at the enterprise and system levels
3	Lifecycle Strategies and Processes	Develop Life Cycle Strategies and Processes for Technology Transition, System Acquisition and Product Support using an IDE, supporting lifecycle activities from concept development to disposal
4	Policy and Guidance	Assess and define the required policy and guidance updates/changes to enable full implementation of the Digital Transformation
5	Workforce and Culture	Drive culture change across the AFMC enterprise through training and change management, enabling a workforce well versed in Digital Engineering

Figure 8 shows the summary of mapping each LOE goal to its primary cell impacted (green cell). The goal words were reviewed and interpreted to determine what Zachman cells are primarily affected. It was discovered that it is the Zachman perspectives and the interrogatives that are considered when choosing which cells are impacted. This sounds counterintuitive in that one would expect that the entities and/or relationships within the cells need to be considered for impact. It turns out that when a goal mentions influencing, for instance, the infrastructure of the enterprise, it is pretty easy to say that the engineer perspective and the *what* interrogative are primarily impacted. The engineer because this person is concerned with the infrastructure and its constraints in meeting the needs of the enterprise, and the *what* because the infrastructure is a thing that exists within the enterprise. This process therefore is interpretative based on the words of the LOE goal.

LOE 0 and LOE 1 goals impact the *what* interrogative of the engineer perspective because the two goals mention changing the IT infrastructure (LOE 0) and Models and Tools (LOE 1). The LOE 2 goal impacts the *what* interrogative of the architect as the goal mentions using a Government Reference Architecture (GRA) and related standards and datasets to take maximum advantage of an integrated digital environment. This directly impacts the form of the data products (models vice documents) which the chief engineer is most concerned about. The LOE 3 goal impacts the *how* interrogative of the architect. The architect is mostly concerned with the PDR process, which would be impacted under LOE 3 goal achievement. The LOE 5 goal impacts the workforce training and the workforce motivation to change to this new way of business. This primary impact was applied to the *what* and the *who* interrogative of the architect perspective because there would be a change in training (*what*) affecting the skills of the PDR participants (*who*). Another primary impact of LOE 5 was applied to the *why* of all of the perspectives because the goal reads that change needs to occur across the entire enterprise. Therefore, every perspective will be affected by this LOE.

	WHY	HOW	WHAT	WHO	WHERE	WHEN
EXECUTIVE (AFMC Executive Leader)	Change to Stakeholders Requirements (LOE 5)	Conduct Acquisition Process	Enterprise Data	Acquisition Enterprise	Acquisition Enterprise Locations	X
BUSINESS MANAGER (Program Director)	Change Technical Review Requirements (LOE 5)	Conduct Technical Review Process	PDR Data	Program Office	Technology Maturation and Risk Reduction Phase	X
ARCHITECT (Chief Engineer)	Change PDR Requirements (LOE 5)	Change to PDR Process (LOE 3)	Change to PDR Data Products (LOE 2, LOE 5)	Change to PDR Participants (LOE 5)	PDR Location	X
ENGINEER (Systems Engineer)	Change PDR Technology Requirements (LOE 5)	Prepare/Maintain PDR Technology	Change to PDR Technology (LOE 0, LOE 1)	PDR Technology Personnel	PDR Technology Location	X
TECHNICIAN (CONTRACTOR)	X	X	X	X	X	X

Line of Effort (LOE) 0	○Provide overarching guidance to influence corporate IT improvement investments to enable a robust, secure infrastructure for the enterprise-wide Digital Campaign
LOE 1	○Provide an Integrated Digital Environment (IDE) of models and tools for collaboration, analysis, and visualization across the functional domains of AF users
LOE 2	○Provide overarching guidance on the use of Government Reference Architectures (GRA) and related standards and datasets for use in an integrated digital environment for application at the enterprise and system levels
LOE 3	○Develop Life Cycle Strategies and Processes for Technology Transition, System Acquisition and Product Support using an IDE, supporting lifecycle activities from concept development to disposal
LOE 5	○Drive culture change across the AFMC enterprise through training and change management, enabling a workforce well versed in Digital Engineering

Primary Digital Change (green) | (LOE X) = Line of Effort(s) Primary Impact to Cell | PDR - Preliminary Design Review

Figure 8: Summary of Primary Impacts from Digital Campaign Goals Indicated by Green Cells

The LOE goal impacted cells give the modeler a place to start in developing the to-be digital enterprise. Using LOE 2 as an example will provide a deeper understanding. The LOE 2 goal centers on the development of reference models to replace the documents that would ordinarily be produced during an acquisition. In the PDR process, this changes the data products created, used, reviewed, and approved as shown in Figure 9. Instead of an allocated baseline in a series of specification documents, it is documented in a model representing the system (i.e., a system model). Other impacted documents such as the Systems Engineering Plan, the Test and Evaluation Master Plan, and the Integrated Master Plan, as well as other acquisition planning documents have their data show up in a model called an Acquisition Reference Model (current term being used by the Digital Campaign) as shown in Figure 10.

A third model is defined as the Government Reference Model (GRM) and has a relationship to a Government Reference Architecture. DoD defines a GRA "as an authoritative source of information about a specific subject area that guides and constrains the instantiations of multiple architectures and solutions" (DoD CIO, 2010). The GRA is the source of the information that is documented in at least one model or view. The GRM is that set of models and/or views that represents the GRA. Therefore, the GRM contains the data that constrains and guides the solution design contained in a system model including a top-level architecture model, and requirements and rules for a contractor to follow in proposing, creating, and validating the system design.

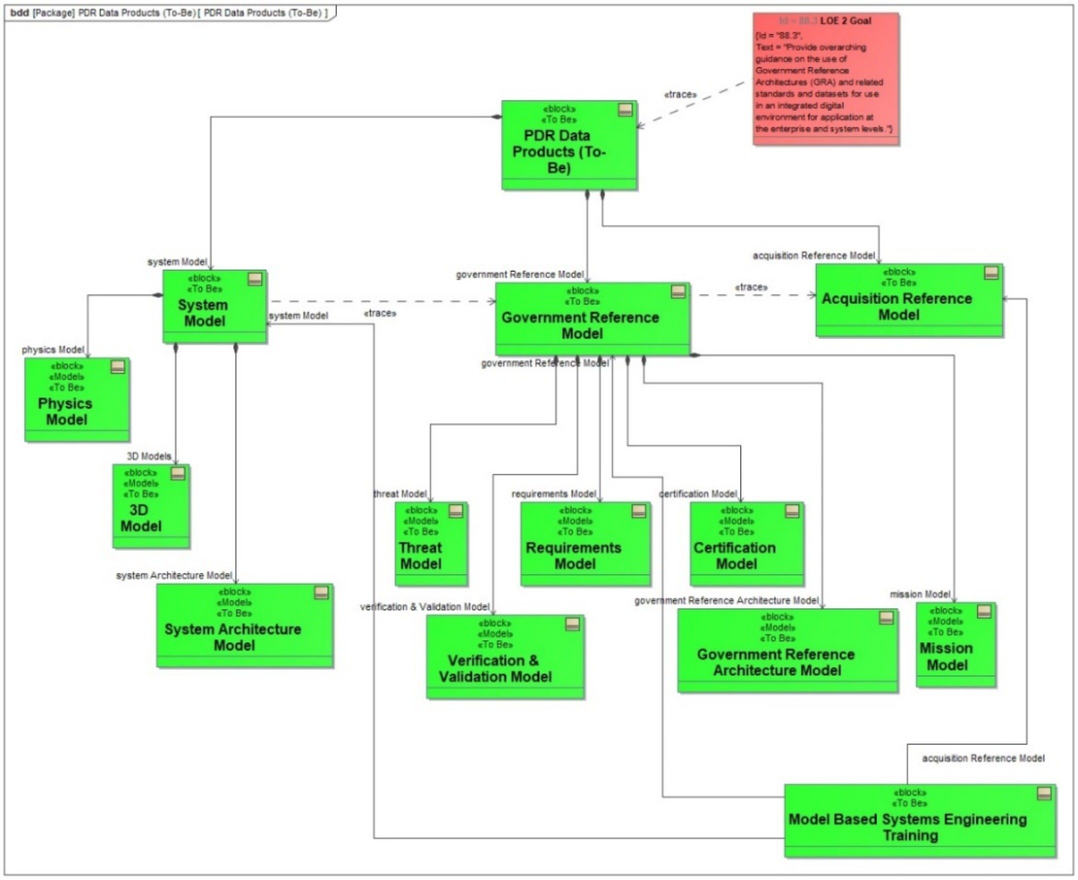


Figure 9: LOE 2 Impact to PDR Data Products (Cell: What (Arch))

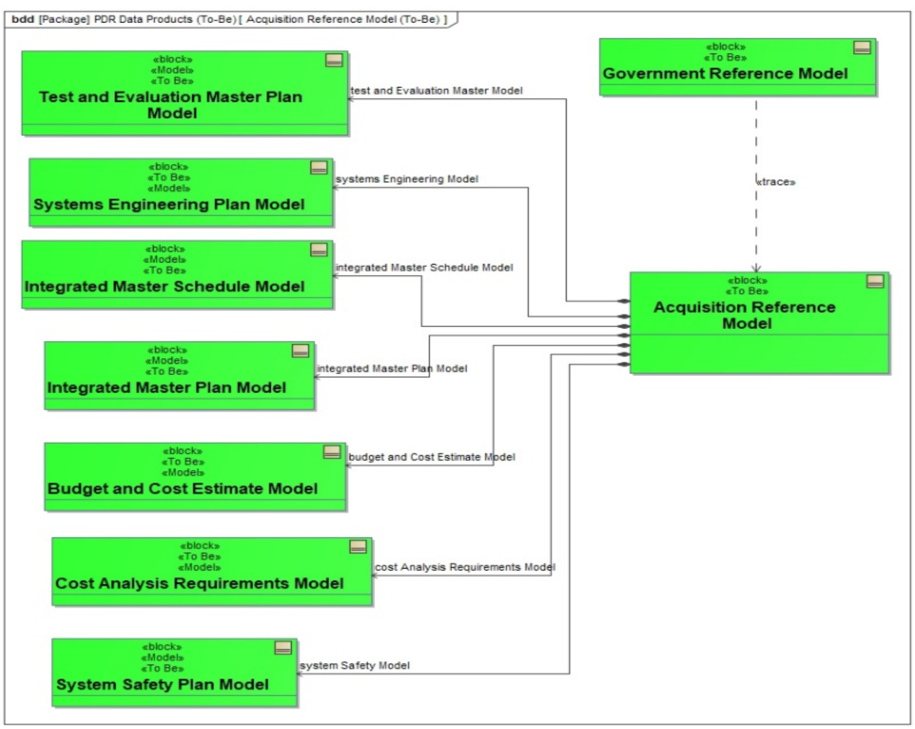


Figure 10: Acquisition Reference Model (Cell: What (Arch))

It is not automatic to discover the LOE goal impacts to the enterprise. SMEs must assess the impacts to their areas of expertise based on the technology state-of-the-art that the organization desires to implement. The modeling using the Zachman Framework does make it easier to see where those impacts are within the framework models, making assessment and assignment much easier.

Secondary Impacts & Gap Analysis

This section explains the analysis behind determining secondary impacts. These secondary impacts are important to visualize because change agents may not understand the cascading effects from their original change requirements. This research also will show how this modeling and analysis effort was used to find gaps within the Digital Campaign's approach.

Tracing Secondary Impacts

LOE 2 is used as an example to show how secondary impacts can be found from the modeling effort. LOE 2 primarily impacts the "What (Arch)" cell as shown in Figure 9. The LOE 2 team is implementing integrated models and data into the acquisition enterprise through architectures and standards as explained in section 4.

A CAMEO tool analysis method is used to display a relation map showing the relationships affected by the LOE 2 goal as shown in Figure 11. This figure shows that there are four secondary impacts based on entity use and relationships to other cells. These are the "Why (Arch)," "How (Arch)," "Who (Arch)," and "What (Engr)" cells. This could be done manually, but the CAMEO relation map capability makes it easy and does the searching automatically.

Changing from Entry and Closure PDR documents of Figure 7 to the PDR models of Figure 12 has an effect on the PDR process occurring in the "How (Arch)" cell as shown in Figure 13. With models implemented and the ability to automate and document model validation as the system model is being designed, a program can now envision conducting continuous PDR reviews. This changes completely the process flow from one major review presentation to a cyclical and agile review flow within the models themselves. Another secondary impact of changing the data products of the "What (Arch)" cell is to the "What (Engr)" cell. So now the models needed for LOE 2 in the "What (Arch)" cell are requiring different tools (such as CAMEO) to build, review, and approve the models instead of the MS Office tools for documents in the as-is enterprise. A last secondary impact of changing the "What (Arch)" cell data products are to the PDR participants of the "Who (Arch)" cell. These individuals will interact with these new tools and models through a desktop system as shown in Figure 14. This desktop is now connected to a new network and cloud environment (a LOE 0 and 1 goal) where the tools reside. The full impact of the LOE 2 goal is shown in Figure 15, where the green-shaded cell is the primary impact of the LOE 2, and the yellow-shaded cells are secondary impacts due to relationships between the cells.

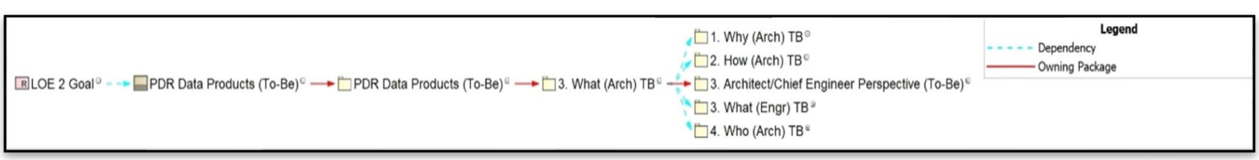


Figure 11: Relation Map for LOE 2 Goal for To Be Enterprise

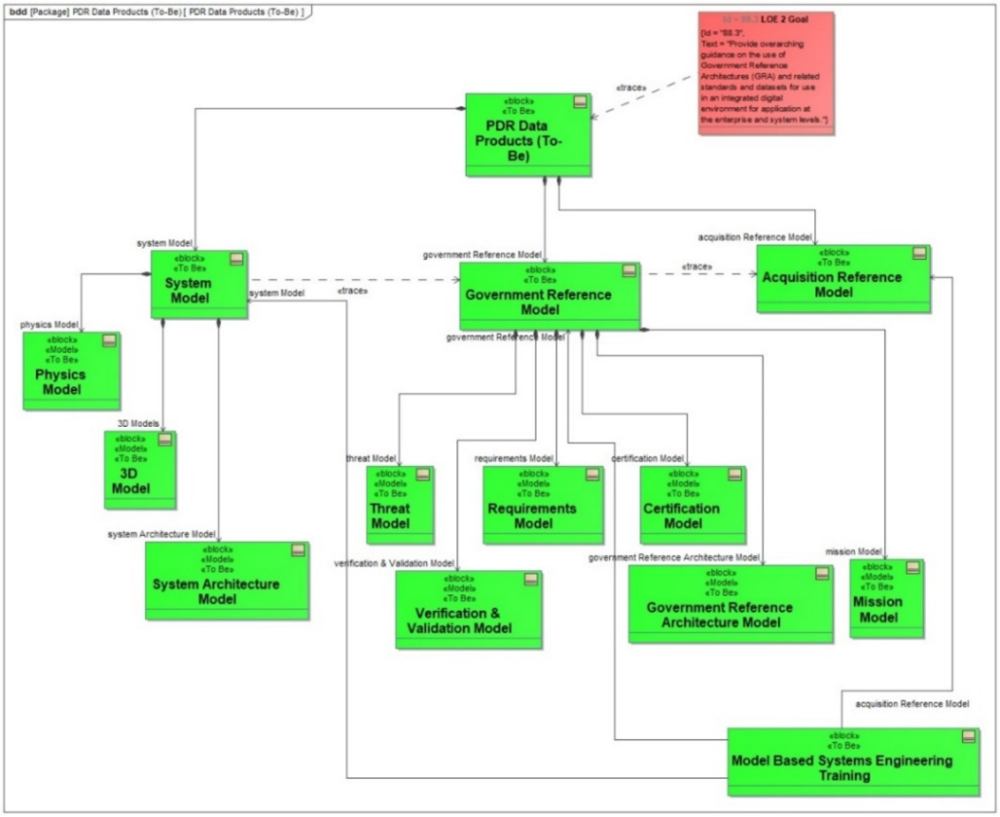


Figure 12: Models Used in To-Be Architecture for PDR (Cell: What (Arch))

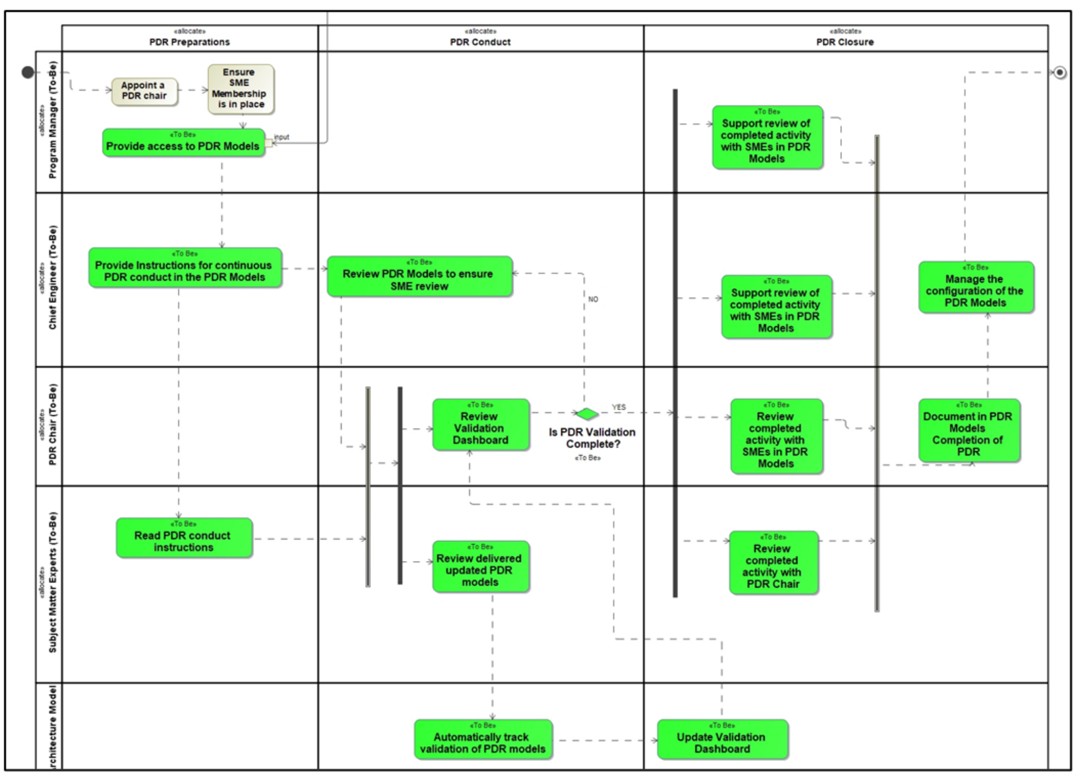


Figure 13: PDR Process for To-Be Architecture (Cell: How (Arch))

Relation maps were used to identify all the cells (packages), which had secondary impacts for each of the LOE goals. Figure 16 shows these secondary impacts for the overall enterprise (yellow cells) that appear outside the primary impacted cells (green cells). As can be seen, many LOE secondary impacts overlap onto primary impacted cells of other Digital Campaign Goals.

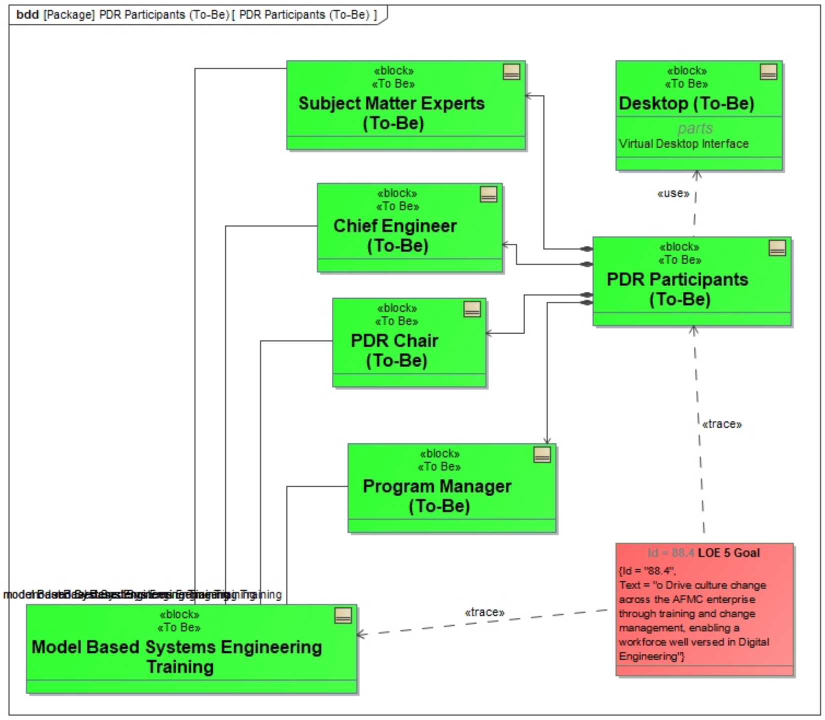


Figure 14: LOE 2 Secondary Impact to PDR Participants (Cell: Who (Arch))

Finding Digital Campaign Gaps

Upon formation of the Digital Campaign, it may have seemed like all six LOEs were aligned based on their goals and that all aspects of the transformation were accounted for. No analysis was done to confirm or deny this conclusion. This research attempted to discover gaps within the LOEs' pursuits using a systems engineering model analysis.

Following a decomposition of requirements into the logical and physical representation of the enterprise from the expected changes of the Digital Campaign, one would expect that the Architect/Chief Engineer and the Engineer cells would be a primary target for the Digital Campaign. The Architect/Chief Engineer perspective represents the logical representation of the enterprise. The chief engineer is concerned with the processes, which are analogous with the functions of the enterprise (system). They ensure the enterprise design complies with the enterprise requirements. In the case of the PDR process, they are the process owner, the data coming in and going out, the participants in the process, and where the processes will take place. In other words, this perspective is an arrangement of related technical concepts and principles that support the logical operation of the enterprise. In order to accomplish a digital transformation of the logical perspective, this research contends that all interrogatives must have primary goals to drive change to the functions of the enterprise, as would similarly be expected by a requirement change to a weapon system.

The Engineer/System Engineer role is concerned with the physical systems that are needed to meet the needs of the digital changes. The physical perspective is where all of the physical systems are for the enterprise. This detailed design must be synchronized with the logical perspective. This perspective is an arrangement of the elements that provide the physical solution to the enterprise change. This research contends for that to happen, all interrogatives of the physical perspective must have a primary goal to drive implementation activities for an efficient transformation.

	WHY	HOW	WHAT	WHO	WHERE	WHEN	
EXECUTIVE (AFMC Executive Leader)	Stakeholder Requirements	Conduct Acquisition Process	Enterprise Data	Acquisition Enterprise	Acquisition Enterprise Locations	X	PDR - Preliminary Design Review
BUSINESS MANAGER (Program Director)	Technical Review Requirements	Conduct Technical Review Process	PDR Data	Program Office	Technology Maturation and Risk Reduction Phase	X	
ARCHITECT (Chief Engineer)	PDR Data Requirements	PDR Process	Change PDR Data Products	PDR Participants	PDR Location	X	Line of Effort (LOE) 2 Provide overarching guidance on the use of Government Reference Architectures (GRA) and related standards and datasets for use in an integrated digital environment for application at the enterprise and system levels
ENGINEER (Systems Engineer)	PDR Technology Requirements	Prepare/Maintain PDR Technology	PDR Technology	PDR Technology Personnel	PDR Technology Location	X	Secondary Digital Change
TECHNICIAN (CONTRACTOR)	X	X	X	X	X	X	Primary Digital Change

Figure 15: Summary of LOE 2 Impacts

	WHY	HOW	WHAT	WHO	WHERE	WHEN	
EXECUTIVE (AFMC Executive Leader)	Change to Stakeholders Requirements (LOE 5)	Conduct Acquisition Process	Enterprise Data [LOE 5]	Acquisition Enterprise	Acquisition Enterprise Locations	X	Line of Effort (LOE) 0 ○Provide overarching guidance to influence corporate IT improvement investments to enable a robust, secure infrastructure for the enterprise-wide Digital Campaign
BUSINESS MANAGER (Program Director)	Change Technical Review Requirements (LOE 5) [LOE 0]	Conduct Technical Review Process [LOE 0]	PDR Data [LOE 0, 1]	Program Office	Technology Maturation and Risk Reduction Phase	X	LOE 1 ○Provide an Integrated Digital Environment (IDE) of models and tools for collaboration, analysis, and visualization across the functional domains of AF users
ARCHITECT (Chief Engineer)	Change PDR Requirements (LOE 5) [LOE 2]	Change to PDR Process (LOE 3) [LOE 2, 5]	Change to PDR Data Products (LOE 2, 5) [LOE 0, 1, 3]	Change to PDR Participants (LOE 5) [LOE 0, 2, 3]	PDR Location [LOE 0]	X	LOE 2 ○Provide overarching guidance on the use of Government Reference Architectures (GRA) and related standards and datasets for use in an integrated digital environment for application at the enterprise and system levels
ENGINEER (Systems Engineer)	Change PDR Technology Requirements (LOE 5) [LOE 0, 1]	Prepare/Maintain PDR Technology [LOE 0]	Change to PDR Technology (LOE 0, 1) [LOE 2, 5]	PDR Technology Personnel [LOE 0, 1]	PDR Technology Location [LOE 0]	X	LOE 3 ○Develop Life Cycle Strategies and Processes for Technology Transition, System Acquisition and Product Support using an IDE, supporting lifecycle activities from concept development to disposal
TECHNICIAN (CONTRACTOR)	X	X	X	X	X	X	LOE 5 ○Drive culture change across the AFMC enterprise through training and change management, enabling a workforce well versed in Digital Engineering

Secondary Digital Change	Primary Digital Change	(LOE X) = Line of Effort(s) Primary Impact to Cell	[LOE Y] = Line of Effort(s) Secondary Impact to Cell	PDR - Preliminary Design Review

Figure 16: Summary of Overall Impacts of LOE Goals on Zachman Framework

To achieve a more aligned transformation following a systems engineering process, the Digital Campaign should have primary goals addressing all of the interrogatives of the architect (chief engineer) and the engineer (systems engineer). The chief engineer perspective has four of the five interrogatives with primary impacts. If the Campaign was to focus also on the *where* interrogative then the perspective would be completely covered (and perhaps the *when* interrogative not covered in this research). Since a majority of this transformation involves implementing state-of-the art technology, an improvement to the Campaign goals would be to also focus on the *how*, *who* and *where* of the systems engineer perspective. Currently, the Campaign focus is only on the *what* and *why* of the system engineer perspective. An example of a goal that might achieve this is the following: “Provide overarching guidance to influence IT locations for robust and secure infrastructure for business activities; ensure an organization and process is in place for the sustainment of IT infrastructure changes.” This goal mentions the “IT locations” taking care of the *where*, the “business activities” taking care of the *how*, and the “organization” being in place taking care of the *who*.

The executive and business manager perspectives are contextual and conceptual perspectives. Setting up goals that would primarily impact these do provide for a complete picture but would not result in permanent changes to the enterprise. They are like ideas, and ideas need to be fleshed out with the logical and physical perspectives. Therefore, it is not as important to address the executive and business manager perspectives with specific primary Digital Campaign goals.

Conclusions and Recommendations

This research presented a method to build a model that addresses the digital transformation of the AFMC acquisition enterprise. The findings of this research are:

1. It is possible to visualize the enterprise and provide better insight into the intricacies and relationships between the people, processes, and infrastructure. Modeling an enterprise into the Zachman Framework using SysML is analogous to a system decomposition and definition using well-established systems engineering processes.

2. As long as the transformation goals do not change, the Zachman cells impacted by the LOEs will be the same. The goals are interpreted using the definitions of the perspectives and the interrogatives to map the LOE goals within the proper Zachman Framework cell. Any enterprise entity mapped into those cells will be primarily affected in a to-be digital enterprise. The focus of this research was the effect of the digital transformation on the PDR process. The focus could have been on other enterprise areas, such as the Air Force Depots or the Supply chain. The result of modeling these processes would be an impact to the part of the model (views) within the affected Zachman cells by the digital transformation goals.

3. Changing process is typically the way the Air Force has gone about change (i.e., trying to be more efficient and effective by leaning the acquisition processes). Guided by the Office of Secretary of Defense (OSD) Digital Strategy, the AFMC Digital Campaign is instead focusing on addressing the *what* (acquisition data using digital system models) and then expecting that the *how* will follow. As shown in this research, changing to digital models from documents and upgrading IT infrastructure is the focus of the transformation. The LOE 3 goal is not to change process to achieve efficiency, but instead, is written to support the change to the Integrated Digital Environment. The Integrated Digital Environment consists of the models, the data, and the standards, as well as the tools and IT infrastructure. This implies that process will change as a result of the technology change. This is visible with mapping the Digital Campaign goals onto a Zachman Framework and performing relationship analysis showing cascading impacts to processes from the technology changes.

4. Lastly, the Digital Campaign should take a more formal and organized architecture modeling approach to transforming the AFMC acquisition enterprise. The model analysis

revealed gaps in the LOE goals that do not cover all Architect's and Engineer's interrogatives. Building an EA would reveal gaps if used as a tool for documenting progress and help ensure an efficient digital transformation.

There are several areas not addressed within the scope of this research that could advance the promise of modeling enterprises undergoing digital transformation. Many enterprises are transforming to digital to increase acquisition speed. This effort did not address *time* as a parameter. This could be an important factor to research in a future effort. This research demonstrated the method based on a single exemplary process – PDR. It assumed that the Digital Campaign goals always will impact the same Zachman Framework cells regardless of the process mapped. This hypothesis could be refined with an effort to map additional acquisition processes. This would contribute to the concept that the gaps identified within the Digital Campaign goals show themselves as Zachman cells where a primary goal does not exist within the logical and physical perspectives. Lastly, a future effort could expand the enterprise boundary. Air Force Materiel Command (AFMC) is an enterprise within the larger Department of the Air Force (DAF). Looking at the entire DAF Acquisition Enterprise could provide more insight into other areas of how acquisition could be improved through digital transformation.

References

Alia-Novobilski, M. (2020, June 5). *AFMC Digital Campaign aims to modernize, streamline life cycle process*. Air Force Materiel Command. https://www.afmc.af.mil/News/Article-Display/Article/2209560/afmc-digital-campaign-aims-to-modernize-streamline-life-cycle-process/

Bernard, S. A. (2012). *An introduction to enterprise architecture* (3rd ed.). AuthorHouse.

Brown, G. Q. (2020). *Accelerate change or lose.* United States Air Force. https://www.af.mil/Portals/1/documents/csaf/CSAF_22/CSAF_22_Strategic_Approach_Accelerate_Change_or_Lose_31_Aug_2020.pdf

DoD CIO. (2010, June). *DoD reference architecture description*. Assistant Secretary of Defense for Networks and Information Integration.

Fox, J. R. (2011, April 29). *Defense acquisition reform, 1960–2009. An elusive goal.* Center of Military History, United States Army. https://history.defense.gov/Portals/70/Documents/acquisition_pub/CMH_Pub_51-3-1.pdf

IEEE. (2011). *ISO/IEC/IEEE 42010:2011. Systems and software engineering, architecture description*.

IEEE. (2015). *IEEE Std 15288.2-2014. IEEE standard for technical reviews and audits on defense programs.*

Kotusev, S. (2016). The history of enterprise architecture: An evidence-based review. *Journal of Enterprise Architecture, 12*(1), 29–37.

Sowa, J. F., & Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture. *IBM Systems Journal, 31*(3), 590–616.

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal*, *26*(3), 276–292.

Acquisition Research Program
Graduate School of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net