

SYM-AM-24-118



PROCEEDINGS
OF THE
TWENTY-FIRST ANNUAL
ACQUISITION RESEARCH SYMPOSIUM

VOLUME III
ACCEPTED PAPERS NOT PRESENTED

**Acquisition Research:
Creating Synergy for Informed Change**

May 8–9, 2024

Published: May 1, 2024

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the federal government.

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 Department of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, please contact:

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

E: arp@nps.edu

www.acquisitionresearch.net

Copies of symposium proceedings and presentations; sponsored faculty and student research reports and posters may be printed from the **NPS Defense Acquisition & Innovation Repository** at <https://dair.nps.edu/>



ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

Table of Contents

How Epideictic Rhetoric Influences Source Evaluation Boards in Oral Presentations..	1
Organization Culture Roadblocks to Innovation.....	13
Obstacles, and Possible Solutions, to the Analysis of Department of Defense Outsourcing for Goods and Services	18
Using Natural-Language Processing and Large Language Models to Restructure DoD Comptroller Budget Materials into Portfolio Views.....	29
Advanced Technologies to Enable Optimized Maintenance Processes in Extreme Conditions: Machine Learning, Additive Manufacturing, and Cloud Technology	45
Resourcing Innovation: Assessment Learning Model – Business Acumen	63
Collaboratorium: A Multi-Stakeholder Approach to Advancing Innovative Defense Acquisition.....	73
Agile Development of Hardware-Reliant Systems	94
Large Language Model (LLM) Comparison Research	100
Qualitative Data Analysis of PPBE Reform Recommendations in the Open Literature	118
Leveraging Digital Transformation for Resourcing Investments.....	135
Leveraging Machine Learning and AI to Identify Alternative Parts to Increase Parts Availability and Improve Fleet Readiness	146
Open Source Software (OSS) Transparency for DoD Acquisition	154
Investing in the Future: Trends in the Defense Department’s Science and Technology Funding.....	166
Capability-Based Software Cost Estimation (CaBSCE): Proposing a New Method to Estimate Software Costs	181
Acquired and Deployed but Not Adopted: Lost Mission Effectiveness Without Resilient Chat Afloat.....	199
A Quantitative Analysis of the Relationship Between DoD Contractors’ Financial Health and Their Merger and Acquisitions Spending Using Panel Data Regression ..	215
Reimagining Defence: Architecting Estonia’s Force Management for the Uneasy Future.....	233
Defense Civilian Training Corps (DCTC) Integrated Curriculum and Development Pilot-0 Implementation.....	247



Design Revision of Walters Probes—High Sensitivity Differential Temperature, CT2, Probes for Atmospheric Optical Turbulence Characterization265

Test and Evaluation of Revised Concept Walters Probes—High Sensitivity Differential Temperature, CT2, Probes for Atmospheric Optical Turbulence Characterization271

Leveraging Digital Transformation Innovating the Acquisition Workforce – Product Support Edition277

The Acquisition Game: Learning Through Play287

Improving Comptroller Benchmarks on Program Spending.....301

Assessing the Impact of Continuing Resolutions (CRs) on DoD Contract Obligation Rates321

Unlock the Hidden Secrets of AI Transformation on the Workforce.....341





PROCEEDINGS
OF THE
TWENTY-FIRST ANNUAL
ACQUISITION RESEARCH SYMPOSIUM

VOLUME III
ACCEPTED PAPERS NOT PRESENTED

**Acquisition Research:
Creating Synergy for Informed Change**

May 8-9, 2024

Published: May 1, 2024

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the federal government.

Approved for public release; distribution is unlimited.

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



THIS PAGE INTENTIONALLY LEFT BLANK



ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL

How Epideictic Rhetoric Influences Source Evaluation Boards in Oral Presentations

Lisa Shea Mundt—is the Co-Founder of The Pulse, a research and advisory firm dedicated to the empowerment of Government Contractors. She received her BA with a double major in Writing, Rhetoric, and Technical Communications (WRTC) and Media Arts and Design as well as her MA in WRTC from James Madison University. She is currently pursuing her PhD in Writing and Rhetoric from George Mason University. She has 13 years of experience in Federal Proposal Development and procurement trend analysis and reporting. She has touched more than 200 proposal efforts totaling billions of dollars in wins across the Federal Government. [lisa.s.mundt@pulsegovcon.com]

Abstract

Oral presentations have officially been part of the FAR since 1994. Initial critics questioned their auditability, but Federal procurement shops employed rigorously controlled physical environments to provide contractors with equal footing. As the process matured, oral presentations have been a popularized method of source selection and allow a federal buyer to streamline the acquisition process.

Post-Covid, the methods to host oral presentations have moved from in-person to a virtual stage. Now more than ever, contractors are able to fabricate ethos by diligently scripting every word to maintain more substantial control over their intended messaging. This paper examines if favor is subconsciously given to presentations that use rhetorical methods - namely elements of epideictic rhetoric.

This paper will use a qualitative research approach based on relevant Federal and Academic texts and personal experiences as an oral presentations coach. It will examine the history of oral presentations and provide an overview of the principles of epideictic rhetoric. It will connect these elements to communication/retention sciences to show rhetorical favorability. The goal is to provide the Government with tools and recommendations to account for style bias and ensure the right contractors are chosen for mission-critical contracts.

Introduction

Within the Government Contracting Bid & Proposal process, there is an entire marketplace of sales methodologies to better inform the incredibly prescriptive world of proposal responses. Best practice requires a Government Contractor to be intimately familiar with their target customer and use that knowledge to put together a narrative that both a) fits the genre requirements (i.e., compliance) and b) *persuades* a Federal Agency to award them contracted work over a set period of performance. Historically, the method for the Government to acquire said products and services has been through submitted proposals – typically written documentation covering elements such as Technical, Management, Past Performance, and Staffing/Key Personnel (to name only a few). However, with increased pressures for procurement flexibility, innovation, and reform, another style of Government/Industry communication has gained steady traction. Oral Presentations as a Source Evaluation Methodology have become widely utilized through the DoD and Civilian markets.

According to the Army Federal Acquisition Regulation Supplement (AFARS), “Oral presentations are ideal for gathering information related to how qualified the offeror is to perform the work, how well the offeror understands the work, and how the offeror will approach the work.” This is a common theme throughout DoD procurement shops. Another example is the *Air Force Materiel Command (AFMC) Guiding Principles for Fair Opportunity Selection under Federal Acquisition Regulation (FAR) 16.505(b)(1)*, which states, “Under appropriate circumstances, Oral Presentations (OPs) can be an effective method to improve the Government’s evaluation of Offerors’ proposed approaches and may reduce acquisition cycle



time. OPs as a substitute for portions of a proposal can be effective in streamlining the selection process, and may occur at any time in the acquisition process.”

The same document references the pros and cons of utilizing Oral Presentations:

- Pro: “Presentations are best from Offerors’ actual team members instead of professional proposal writers.”
- Con: “Evaluators must be careful about ‘form over substance’. A flashy Offeror presentation (e.g., a great presenter) may mask weak proposal content.”

In a similar vein, the video Evaluating Oral Proposals in Major Government Procurements program, an Innovations in American Government Awards finalist, presented before the National Selection Committee in 1996 is a look back in time to when Oral Presentations were in their infancy. FAA representatives took to the podium to explain the benefits seen in their procurement shop and why oral presentations were credited with increased contracting efficiency. The FAA representative states, “...because proposal writing had gotten to be a real art in the government and those companies that already knew how to do it tended to have an edge. When you turn it more into a job interview as opposed to a proposal writing activity, you’re looking more at the qualifications of the people who will actually perform the work; their ability to turn out paper in accordance with regulations loses some of its significance” (sic).

However, in an almost deafening echo from yesteryears, a member of the receiving panel asked a question that exemplifies the AFMC “con” to using oral presentations as listed above: “It would seem to me that just as you have skilled proposal writers you can have skilled oral presenters. ...have you run into any problems that are related to that – that people say things better than they perform well or the reverse?” (sic).

And therein lies the rub, so to speak. In an effort to identify qualified contractors through rigorous demonstration free from outside impact, the DoD contracting community finds itself against the same issues as with the written proposal process: the influence of consulting, technology, and the effect of style. The heart of any application process is to persuade, and in a hyper-saturated market like Government Contracting, industry uses all the weapons in their armory to stand out in the crowd, including ancient persuasive principles such as epideictic rhetoric. This paper will explore the evolution of Federal oral presentations and how Contracting Officers may arm themselves against giving subconscious favor to presentations that serve style over substance.

Background

In 1994, the FAR was amended in response to two major acquisition reform acts – the Federal Acquisition Streamlining Act (FASA) and the Federal Acquisition Reform Act (FARA). The rewrite included an addition to FAR Part 15 – a source selection methodology called **oral presentations**. According to the Federal Acquisition Regulation Part 15.102(a), “Oral presentations by offerors as requested by the Government may substitute for, or augment, written information. Use of oral presentations as a substitute for portions of a proposal can be effective in streamlining the source selection process. Oral presentations may occur at any time in the acquisition process, and are subject to the same restrictions as written information, regarding timing (see 15.208) and content (see 15.306).”

Since their addition to the FAR, oral presentations can appear as a requirement to bid on a contract in a variety of forms:

- **Written Proposal + Oral Presentation:** Perhaps the most common example, many Agencies will leverage oral presentations as part of a phased acquisition process. This



means a standard written narrative will serve as a down select method, thereby limiting the number of offerors who present to the Government.

- **Prepared Slides in lieu of Written Proposal:** In some instances, Agencies forgo a written proposal submission and instead will ask for the oral presentation slides as the proposal response. The offeror will then present the pre-submitted slide deck.
- **Problem Solving/Challenge Exercise:** If a Government Contractor finds itself pursuing an opportunity through a particularly innovative Agency, then they may need to prepare for a Challenge scenario. The Government will provide a scenario to the presenting group during their allotted time. The offeror's group will have time to work together and develop their presentation response live in front of the Government. They will then have additional time to present. Offerors are evaluated on both of these components.
- **Team Q&A and/or Project/Program Manager (PM) Q&A:** In rare cases, Oral Presentations are limited to a Question/Answer function by either the whole presenting team or just the PM proposed for the contract. More commonly, a Q&A session occurs after the team presents.

Anecdotally, there are several reasons why the Federal Government would implement oral presentations as part of their acquisition strategy. They can be used to evaluate your expertise; test group chemistry; measure leadership ability; put you in a pressure cooker environment (to see how you react under stress); and *to determine if they want you as coworkers*.

This concept of personal preference is often overlooked, especially since contract acquisition practices have added safeguards to the process to turn complicated subjective concepts into areas of scorable objectivity. However, establishing ethos is a foundational heuristic within rhetorical study, and an individual's character and how they hold themselves have direct impact on whether we have confidence in their ability to perform a job.

Oral presentations provide a unique opportunity to directly link key personnel and the awarding agency. As stated previously, for years the primary method of bid communication between Industry and Government was strictly in written form. Although there is a time for dialogue between prospect and customer in the Pre-RFP (Request for Proposal) phase, the person selling is rarely the person doing the work on the contract. Likewise, written narrative proposals are often developed by Proposal Managers and Technical Writers, who, once the bid is "won," have nothing to do with contract execution. Oral presentations provide an avenue for Federal Contracting shops to evaluate prospective contractors not only on their qualifications as a company, but also for an individual's personality, likeability, and knowledgeability based on how they present. Oral presentations are an opportunity for the Government to get a sense of whether these individuals match agency culture – just as much as they are able to evaluate technical expertise. This rhetorical underpinning becomes a subversive part of how the Government evaluates a presentation based on their "Confidence Rating" (more on this later).

According to Acquisition.gov, "Oral presentations may be beneficial in a variety of acquisitions. They are most useful when the requirements are clear and complete and are stated in performance or functional terms. Oral presentations are ideal for gathering information related to how qualified the offeror is to perform the work, how well the offeror understands the work, and how the offeror will approach the work." This vague description does not encapsulate how much impact the degree of presentation quality will have on showcasing "how well" an individual may understand a subject area. This level of potential bias has yet to be academically explored, and though oral presentations are praised as an innovative procurement method, there are some criticisms that warrant more detailed review.



In *Oral Presentations in Negotiated Procurements: Panacea or Pandora's Box*, Sean Michael Patrick Hannaway (2000) provides a detailed look at oral presentations as a groundbreaking phenomenon in the new millennium. Although the musings were mostly positive, when the concept was first formalized into the FAR, some naysayers took issue with the addition and "...contended that the practice would (1) create confusion as to the content of the resultant contract and (2) result in a lack of any record on which an unsuccessful offeror could base its protest. As one commenter put it, the change seemed intentionally aimed at 'not leaving an audit trail'" (p. 463). Additionally, "...critics consistently voiced the concern that emphasis on more efficient source selection methods would come at the expense of small business concerns" (p. 466).

Hannaway's essay also points out the overarching themes that "...technical proposals are prepared by 'professional essay writers' and that they frequently misrepresent the offeror's actual understanding of the work required in a procurement" (p. 473), and although oral presentations have the capacity to mitigate these concerns, there is still the underpinning issue that "...the inclusion of subjective criteria such as past performance may act as an invitation to make a sales pitch and 'dazzle the Government with top-notch speakers'" (p. 470).

Current Day Landscape

Over the course of the 24 years after that article was published, the Government has imposed standards and instructions to mitigate these concerns. Pre-Covid, it became standard practice to videotape the in-person presentations, and the physical requirements were rigorous to ensure fairness. In a post-Covid world, oral presentations were moved into a virtual environment, drastically changing the needs for how industry would respond to these requirements. There are a great many benefits to virtual oral presentations for the Government – it is easy to record a virtual conference for audit purposes, it requires no additional training, and automated systems minimize the threshold for user error.

No longer do individuals need to pay for travel, and administrative burden is removed from the Government by going the route of MS Teams/Zoom. NIH Information Technology Acquisition and Assessment Center (NITAAC)'s Acquisition page echoes these sentiments and states, "Oral presentations not only benefit the federal government but can also streamline administrative burden for responders."

In fact, even pre-Covid, leveraging oral presentations has been lauded as innovative, and in FY17 the GSA's Procurement Innovation Lab Annual Report has use cases on how oral presentation acquisition methodologies saved the Government time and money:

- Expedited \$58M award in 42 days for cloud migration (Procurement Innovation Lab Annual Report, FY 2017, p. 15).
- Increased government's understanding and confidence levels of each offeror's approach for agency enterprise Automated Biometric Identification System (Procurement Innovation Lab Annual Report, FY 2017, p. 16).
- Saved time and increased the technical evaluators' understanding of offerors' capabilities and key personnel for USCIS' agile development and maintenance. (Procurement Innovation Lab Annual Report, FY 2017, p. 18).
- \$63M award in 109 days using innovative acquisition techniques, including oral presentations with scenario-based and on-the-spot questions for TSA agile services (Procurement Innovation Lab Annual Report, FY 2017, p. 20).

As recently as 2022, there are Government-created articles aimed at Federal Program offices to describe the value of implementing oral presentations. An assisted acquisition



contracting officer at NITAAC produced the following table to exemplify reasons for replacing written proposals with oral presentations:

Instead of the written approach which includes...	Consider the oral response...
farming out parts and pieces to subject matter experts and then laboring to create one voice	the team who will be performing the proposed work will be preparing together and rehearsing their unified approach
company standardized formats with aesthetic graphics and charts	the actual subject matter experts telling you what they assessed from your requirements and how they would solve your problem. This is often backed up on slide decks or websites with visuals of projects they've completed to verify what they are saying
a (seeming) regurgitation of the government's requirement, leading to a compliant check list in narrative form and a price	is a tailored oral walk through that clearly demonstrates the value the government can expect to receive. These solutions are also illustrated with compelling and relevant evidence

Note: Table reprinted from NITACC [Oral Presentation](#)

As seen in the table, Federal procurement workers have less-than-positive feelings towards written proposal responses, and the sell to use oral presentations leans heavily on innately rhetorical principles such as teamwork, visuals, and compelling evidence. An intangible concept like *teamwork* is indicative of how the way in which an offeror presents will impact their evaluation and how there is more to an oral presentation than just the information – the *optics* add strength.

Now that Government Contracting is operating in a post-Covid environment, there are additional considerations that make this topic even more critical to examine. In the pre-Covid world, operating under the assumptions of the 1994 FAR amendment, oral presentations were held in person. In those times, it was a rare feat to memorize highly technical presentations, especially in the context of Federal procurement where the outcome of said presentation can make or break a company's bottom line. The difference now is that the move to online presentations allows contractors to diligently script their every word and easily read transcriptions to a virtual audience. Additionally, virtual oral presentations allow for easier ability to virtually caucus between teammates (text/Slack/Teams). The level of presentation authenticity has severely diminished, and the quality of language has increased. Oral presentations are an effective and popularized way to evaluate an offeror's abilities, but what if those abilities are skewed? What I argue today is that the shift to virtual oral presentations increases the need for offerors to employ principles of **epideictic rhetoric** to effectively present in a meaningful, impactful manner to win federal contracts, and that Federal buyers must understand these methods and mitigate potential hindrances that come from them.

Epideictic Rhetoric

Rhetoric is commonly referred to as the "art of persuasion," the origins of which can be traced back to Ancient Greece. The composition and definitions of the practice have changed over the years, and various scholars have added their own assessment and spin to what is fundamentally a practice on how to influence others, primarily through the art of speech. Aristotle, one of the primary fathers of rhetoric, alleged there were three divisions of oratory – political, forensic, and the ceremonial oratory of display, also known as epideictic rhetoric. For



the purposes of this paper, it's important to know that epideictic rhetoric is thought of as the least substantial type of speech, as it focuses on the present state (instead of past or future) and often uses methods of praise and blame to make a point. For this article, I use the academic concept of epideictic rhetoric as a lens to provide insight for Government Officials to understand how long-taught persuasive practices may be impacting the appropriate allocation of taxpayer dollars away from the most deserving contractors. For all intents and purposes, this is an analysis of the dangers of procurement methodologies based on style over substance.

"The Public Value of Epideictic Rhetoric" is a 1996 article written by C. M. Sheard and is the principle modern day text about epideictic rhetoric. Thought of as "mere showpieces," it was Aristotle who "...limited epideictic to the praise or censure of a person or thing (tangible or intangible) and described it primarily as ceremonial speech whose audience serves as spectator rather than judge (as in deliberative and forensic oratory) and whose temporal focus is the present" (p. 768). On the surface, it seems epideictic rhetoric, often found in settings such as eulogies, offers little value, especially when compared to the other two branches (deliberative and judicial). Sheard goes on to explain, "Epideictic discourse fell into disfavor in antiquity as it evolved into a highly figurative, even fictive, mode of discourse that seemed primarily to advertise its speaker's skill" (p. 767).

Of Aristotle's three branches of rhetoric, epideictic has had an alarmingly poor reputation, being conveyed as empty spells cast on passive audiences. A word often used when discussing epideictic speech is poetry (or poetic). Many authors over the years feel that epideictic rhetoric is more concerned with poetic style than with actual substance. In *The Platonic Functions of Epideictic Rhetoric*, Bernard K. Duffy (1983) writes, "...the language of epideictic is prone to be least referential and most poetic because the facts themselves are not really at issue. Rather, the affirmation of ethical standards of judgment and behavior serves as the creative use of language While the forensic and deliberative orator are essentially bound to the facts at hand, the epideictic orator need be less concerned with material realities than with the abstract propositions he aims to affirm" (pp. 90–91). The substantiation for such claims tends to revolve around a lack of civic value due to audience members being seen as passive observers. Some believe there is no consequence garnered from epideictic rhetoric, as the speaker has no explicitly persuasive goal.

Sheard counters, explaining that epideictic rhetoric "...leads to a vision that the audience is not only invited to share but exhorted to help actualize" (p. 780). Epideictic rhetoric is not purely consumed by an audience without inspiring output – simply pretty words for the sake of showing off – because there are always underlying motives behind speech.

Sheard, Perelman, and Olbrechts-Tyteca focus on the innate civic function and how epideictic rhetoric can put thoughts into action – namely how poetic language can move you to perform. In *The New Rhetoric: A Treatise on Argumentation*, Perelman and Olbrechts-Tyteca explain, "Epideictic oratory has significance and importance for argumentation because it strengthens the disposition towards action by increasing adherence to the values it lauds" (p. 50). Who amongst us are not moved by passion, and even today, credibility of leaders may be achieved through poised public speech.

So what does epideictic rhetoric have to do with oral presentations to the Federal Government? Epideictic rhetoric, according to Aristotle, deals with orations that cover praise or blame and are firmly situated in the present. One of its many critiques is that epideictic oration does not invite audience participation. While some might assess that oral presentations lean more towards Aristotle's deliberative rhetoric, the messaging behind these Federal solutions is never up for debate – they are declamatory and impassioned proclamations—a forum to praise the offeror's capabilities without garnering immediate response from the Government. In fact,



the “pitch” component of oral presentations is notably not meant to constitute discussions, and solicitations will often instruct offerors not to speak with the source evaluation board except to ask logistical questions. This is because the concept of “discussions” has contractual subtext in procurement. Discussions occur when a contractor reaches the competitive range and the Government allows the bidder to respond to weakness, deficiencies, and clarification. Discussions in procurement means an offeror may amend their bid, which is why Contracting officials need to be so careful when entering oral presentations for fear the wrong thing gets said.

As an example, a component of instructions pulled from a 2023 National Science Foundation (NSF) solicitation states, “Oral presentations do not constitute discussions... Oral presentations are distinct from the Government’s reserved right to conduct exchanges.” From a legal precedent perspective, according to law firm *Crowell* (previously Crowell & Moring), “GAO ruled that an offeror’s oral presentation and the ensuing question and answer session did not become ‘discussions’ that would trigger other offerors’ rights to revise their proposals, upholding a huge military health services contract award in Sierra Military Health Services (Dec. 5, 2003). Tackling a difficult issue with a fact-bound decision, GAO held that an offeror’s presentation and the Q&A session constitute ‘discussions’ only if agency personnel gave that offeror a chance to revise its proposal in, for example, the answers to the evaluators’ questions.” The purpose of oral presentations is thereby to persuade without conversation, with the offeror acting as storyteller and the Government as an active listener.

Storytelling

Epideictic rhetoric is an unfamiliar concept unless you happen to be a writing, rhetoric, or communications scholar. A more familiar and digestible example of epideictic rhetoric is the concept of storytelling. Within the realm of Federal proposal development, there are common tropes about how to develop written narrative. A cliché amongst Industry is to make text compliant and *compelling*. The compelling notion harkens to the idea of using the rhetorical heuristics of logos (logics), ethos (credibility), and pathos (emotions) to *tell a story* about their solution ... and there’s a scientific reason for that. In *Advertising as Epideictic Rhetoric and Its Implications for Ethical Communication* by Cem Zeytinoglu, the author aptly states that, “...storytelling is epideictic through the metaphors of coherence and fidelity” (p. 35).

According to an article in *The Scientific American*, personal stories make up 65% of our conversations. Further, storytelling reveals a bias for how individuals respond to a call to action. Psychologist Melanie C. Green published a study in 2004 entitled *Transportation into Narrative Worlds: The Role of Prior Knowledge and Perceived Realism*. Green’s research examines the effects of narratives or stories on individuals’ beliefs. Her work focuses on how an audience is heavily influenced by stories that connect to our experiences, and being swept away in a compelling story may impact a listener’s real-world beliefs. Psychologists have examined the genre of psychological realism, which focuses on humanistic motive, to explain how stories with recognizable and believable emotions most often “transport” a listener to a figurative other world – the feeling of being immersed in a story.

According to *The Science of Effective Presentations* by Prezi, “Stories are two times more persuasive than raw data...” and “...storytelling is one of the best ways to persuade people to take action” (p. 6). In this presentation, Prezi cites a study by Deborah A. Small entitled “Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims.” This study examines how when it comes to charitable giving, “people often become entranced by specific, identifiable, victims” (p. 143) as opposed to causes that lean on statistics. Small provides the following examples: “In 1987, one child, ‘Baby Jessica,’ received over \$700,000 in donations from the public, when she fell in a well near her



home in Texas. Similarly, the plight of a wounded Iraqi boy, Ali Abbas, captivated the news media in Europe during the Iraq conflict and £275,000 was quickly raised for his medical care. More than \$48,000 was contributed to save a dog stranded on a ship adrift on the Pacific Ocean near Hawaii” (p. 143). As the Prezi article explains, “People tend to have stronger emotional reactions to vivid information—stories with details that make them come to life—than sterile information—facts, figures, and charts” (p. 7). The Prezi article provides further context: “Stories engage our entire brain. Researchers have discovered that our brains react very differently to stories versus straight information. Numerous studies have found that when presented with metaphors and descriptive words or phrases—things like ‘perfume’ and ‘she had a velvety voice’—trigger the sensory cortex in our brains, which is responsible for perceiving things like smell and touch. That is, the way that our brain handles reading and hearing about sensory experiences is identical to the way it handles actually experiencing them. On the other hand, when presented with non-descriptive information—for example, ‘The marketing team reached all of its revenue goals in Q1,’—the only parts of our brain that are activated are the ones responsible for understanding language. Instead of experiencing the content with which we are being presented, we are simply processing it. Stories are such powerful engagement tools because they engage the whole brain. Vivid imagery brings your content to life—quite literally—in the minds of your audience. Next time you want to hold the attention of a room, tell a story” (p. 15).

Walter Fisher’s idea of Narrative as Human Communication further connects storytelling to epideictic rhetoric. According to Zeytinoglu, “Fisher’s narrative paradigm has two important connections to epideictic discourse. First, storytelling is itself intrinsically ceremonial. In stories humans create a dramatic realism in which the heroes and villains of a society (in their perfect or exaggerated forms) act in certain situations where ethical decisions are made, and social norms and values are reflected in order to demonstrate the desired way of life and character of disposition for the people. Secondly, stories are not documentary and factual in the perfect sense. In fact, they are fictional in form, by which the maker of the story also demonstrates the ability to use language in an attractive and a beautiful way” (p. 35).

Zeytinoglu further suggests that for Aristotle, “...epideictic discourse is the occasion where wisdom merges with style the most. It can be argued that this merger between wisdom and eloquence shows itself mostly in Aristotle’s inartistic *pistis* of *éthos* where the rhetor has to demonstrate a good moral character through his speech” (p. 17). In Federal contracting, one could contend that credibility in oral presentations is gleaned, at least in part, from eloquence of speech.

Impact on Government Acquisition

As put in the decision to GAO protest [Leidos Innovation Corporation, B-415514; B-415514.2; B-415514.3, January 18, 2018](#), “...the responsibility for providing a thorough, **persuasive** response to agency questions as part of an oral presentation falls on the offeror.” Whether or not the FAR or individual solicitations explicitly refer to style, it can be inferred that when offerors present the pitched, non-dialogue component of an oral presentation, then rhetorical appeals are instinctively leveraged given the epideictic nature of the presentations and the evaluator’s penchant for stories.

This presumption can be directly connected to how the Federal Government leverages *confidence ratings* when evaluating oral presentations. FAR Subpart 8.4, Part 13, Subpart 15.3, and 16.505 all cover regulations for the Government to use confidence ratings when evaluating Federal bids. Confidence ratings are considered more innovative than the previously used adjectival ratings (colors coordinated with the terms outstanding, good, acceptable, marginal, unsatisfactory). Per the DHS PIL Bootcamp Workbook, “Confidence ratings provide evaluators



the ability to look more holistically at the strong points and weak points of an offer. Confidence ratings, supported by rationale, are often more helpful to a selecting official.”

“Confidence” does not have a hard and fast definition in the acquisition world, but one might infer it equates to a feeling of trust built on an evaluator’s interpretation of an offeror’s capabilities and benefits resulting in the lowest risk solution. The term is ambiguous and rhetorically driven, for what is trust if not a presenter’s ethos through established credibility, authority, and even likeability? To further exemplify the vagueness of evaluation criteria related to confidence ratings, I’ve pulled snippets from real solicitations from the last 5 years and italicized sentences that are rhetorically charged:

- **Program Executive Office Defense Healthcare Management Systems Workforce 3.0 (WF3) Challenge Scenarios:** “The Government will also evaluate the quality of the solution and solution presentation in a manner that breaks down the fundamental issues being addressed, maximizes understanding of risks and benefits, and successfully *persuades* Government stakeholders on the viability of the plan.”
- “The extent to which the Offeror facilitates a productive yet collaborative team environment to maximize the capabilities of the full team in developing a responsive solution. Collaborative cultures where innovative ideas are scrutinized for their merit rather than by the source of the idea would be of benefit to the Government. The Government will also evaluate the extent to which the Offeror is capable of identifying and mobilizing the best available talent aligned to the challenge including the ability to rapidly brief and onboard subject matter experts in disparate domains. Approaches that *highlight a positive interaction between contractor team members*, provide meaningful contribution from every team member involved, and augment the conversation using subject matter expertise and additional resources. Approaches where a single member or party dominates the facilitation, ideation, and presentation of the solution without inclusive participation amongst other participants and where ideas are valued based on source, rather than value, may present a performance risk.”
- **Department of Homeland Security (DHS), Management Directorate (MD), Federal Protective Service (FPS) Personnel Security Support Services Contract:** “Confidence ratings will be supported by a narrative inclusive of the *noteworthy observations* that assesses the Government’s level of confidence that the offering contractor will successfully perform all requirements.
- **VA Enterprise Data Infrastructure Services EDNS:** The written proposal and oral presentation will each be evaluated to determine the extent to which the proposed approach is workable and the end results achievable. The written proposal and oral presentation will each be evaluated to determine the *level of confidence provided the Government* with respect to the Offeror’s methods and approach in successfully meeting and/or exceeding the requirements in a timely manner.
- **UNITED STATES SECURITIES AND EXCHANGE COMMISSION (SEC) for SEC Enterprise-wide Database Administration (DBA) and Support Services:** “The SEC will evaluate the written responses, oral presentations, and price submissions to select the best-value Offeror. The Offeror that provides the *greatest overall benefit* in their response to the requirement will be selected for the task order award. The approach is based on comparative evaluation and not a tradeoff. Following receipt of responses (including oral presentations) to this TO RFP, the SEC will first check an offeror’s response for conformity with the solicitation. A proposal is acceptable if it conforms to all material requirements of the RFP, otherwise, it is deemed unacceptable. Only acceptable responses will be evaluated. Subsequently, the SEC will perform a



comparative analysis (comparing response to the requirements of the SOW and comparing offerors to each other) to select the contractor that is best qualified to fulfill the requirements, based on the Offeror’s responses to the factors outlined in this TO RFP.”

Recommendations

In my research examining oral presentation solicitation references, there is only one instance where I have read an explicit mention of style, as seen in the **Centers for Medicare and Medicaid Services (CMS) CMS Cloud Information Technology Operations and Maintenance (ITOPS) solicitation**: “The Government is interested in the informational content of the respondent’s answers rather than the particular style or technique used in delivering them.”

This clause acknowledges that the way in which an offeror presents is separate from the material they cover. By including such language, the contracting shop attempts to formally remove the evaluation bias given towards presenters who employ elements of the more stylistic branch of epideictic rhetoric. It begs questions of how rhetorical prowess may impact how a Source Evaluation Board scores presentations, since rational human beings are scientifically swayed by stories.

Based on the science and history of demonstrative oratory, it can be determined that the use of poetic language and storytelling is more persuasive than objective facts, and contractors who leverage such presentation mechanisms are subconsciously held in higher favor since people have a stronger emotional response to vivid information. This means the Government evaluators and their Contractors need to be more attuned to the role rhetoric plays in oral presentations.

Although there will always be bias, especially when leveraging a confidence rating style of evaluation, when the mission is on the line, it is incumbent on the Government to include contractual safeguards from this pre-Aristotelian magic. Key recommendations for the Government to consider are as follows:

Suggestion	Rationale
Leverage a challenge scenario style of oral presentation wherever possible.	Challenge scenarios are the pinnacle type of oral presentations to avoid overly prepared rhetorically charged presentations. By limiting preparation time, you get an authentic look at the individual and how they interact with other members of their team. Hannaway’s 2000 article recommends, “The use of sample tasks, pop quizzes, and questions directly related to the contract requirements has the greatest potential to improve the quality of source selections decisions” (p. 509).
Require only key personnel to be oral presenters. Remove the option to add additional presenters or the request for executive leadership.	There is often an allowance for executive management to attend an oral presentation. This typically occurs in the form of a small business CEO participating in orals to showcase their corporate commitment. Executive leaders who have no impact on project/program operations should not join the presentations since they are well-versed in selling their personal organization.
Where possible, get back to having oral presentations in person.	The fastest way to ensure prospective contractors are knowledgeable in the subject matter is to move oral



	presentations back in person. Without the ability to have a script in front of them and access to reference material, you can better see the natural proclivities of the contract key personnel.
Focus energy on developing more meaningful questions for Q&A and have questions specifically tailored for each of the key personnel positions.	If the Government continues to use pre-scripted oral presentations, then there should be more emphasis on the question and answer portion of the oral presentation period. This provides more opportunity for authentic, ad-libbed material and ensures all members of the team have a unique point of view.
Add caveats to the solicitation to limit the importance of style.	The Government should consider adding clauses like the one referenced from CMS in their evaluation criteria: “The Government is interested in the informational content of the respondent’s answers rather than the particular style or technique used in delivering them.”

The federal government’s ability to succeed in its mission and ensure our national security hinges on its ability to effectively and efficiently engage with industry and gain access to critical technologies and innovation. By acknowledging the impact of scholarly concepts such as epideictic rhetoric, we provide a long-studied subject as a lens to view our constantly evolving Federal procurement institution to improve processes and evaluation procedures.

References

- Air Force Materiel Command (AFMC) guiding principles for fair opportunity selection under Federal Acquisition Regulation (FAR) 16.505(b)(1)*. (n.d.).
- Crowell & Moring. (2004). *When does an oral presentation become a “Discussion”?*
<https://www.crowell.com/en/insights/client-alerts/when-does-an-oral-presentation-become-a-discussion>
- Department of Homeland Security. (2019). *PIL boot camp* (Ver. 2.7.2).
https://www.dhs.gov/sites/default/files/publications/pil_boot_camp_workbook_oct_2019.pdf
- Evaluating oral proposals in major government procurements: Finalist presentation*. (2011).
<https://www.youtube.com/watch?v=hKTMelhBM70>
- Fisher, W. R. (1987). *Human communication as narration: Toward a philosophy of reason, value and action*. University of South Carolina Press. <https://www-jstor-org.mutex.gmu.edu/stable/j.ctv1nwbqtk>
- Green, M. C. (2004). Transportation into narrative worlds: The role of prior knowledge and perceived realism. *Discourse Processes*, 38(2), 247–266. https://doi.org/10.1207/s15326950dp3802_5
- Hannaway, S. M. P. (2000). Oral presentations in negotiated procurements: Panacea or Pandora’s box. *Public Contract Law Journal*, 29(3), 455–509. <http://www.jstor.org/stable/25754407>
- Hsu, J. (2008). The secrets of storytelling. *SA Mind*, 19(4), 46.
<https://doi.org/10.1038/scientificamericanmind0808-46>
- Lael, K. (2022). *Oral presentations*. National Institutes of Health. <https://nitaac.nih.gov/resources/tools-and-templates/oral-presentations>
- Prezi. (n.d.). *The science of effective presentations*. Wilmington University.
<https://www.wilmu.edu/edtech/documents/the-science-of-effective-presentations---prezi-vs-powerpoint.pdf>



Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, 102(2), 143–153. <https://doi.org/10.1016/j.obhdp.2006.01.005>

Zeytinoglu, C. (2007). *Advertising as epideictic rhetoric and its implications for ethical communication* [Doctoral dissertation, Duquesne University]. <https://dsc.duq.edu/etd/1406>



Organization Culture Roadblocks to Innovation

Eric Ferraro—has more than 40 years of experience as a Navy Supply Corps Officer, Senior Consultant with PricewaterhouseCoopers and IBM, almost 20 years as a federal civilian with the Department of the Navy and the General Services Administration, and is currently on the faculty at DAU. He has served in program management roles or executive positions implementing shipboard information systems and defense (and federal) business systems across the Navy and federal government. At DAU he teaches in all executive courses with a focus on tools and skills for program managers as well as executive leadership and organization culture. [Eric.ferraro@dau.edu]

Abstract

The Department of Defense Acquisition workforce and industry partners are “under the gun” to be more innovative and “do things differently.” Although the “Big A” Venn diagram of resources, requirements, and acquisition system can present inherent roadblocks to innovation, the biggest roadblock to overcome is the entrenched culture. This paper seeks to highlight changes needed to create a culture that supports innovation and unleashes the acquisition workforce to meet the needs of the warfighter. The author lays out six major roadblocks to create a high performing organization and a culture that aligns with providing capabilities at the speed of relevance and creating value by doing things differently.

“Culture eats strategy for breakfast (or lunch)” is an often-quoted business maxim attributed to management consultant Peter Drucker although the citation tracing is not definitive. Regardless, the adage is approaching 100% in accuracy as there are countless examples of businesses and government organizations with a great strategy for success only to be overcome by the in-place organization culture which resists change or a culture that is not aligned with the business strategy. There is ample research on organization culture and the impacts on people and getting results. My goal is not to reinforce or challenge those results, but rather to discuss what acquisition leaders need to do to align and create a culture that supports innovation rather than hinders it. I used the research study conducted by professor John Graham and others at Duke University’s Fuqua School of Business as a starting point for what drives culture (Graham et al., 2022). Their research shows how critical culture is to an organization’s success and business results and is driven by formal management practices and governance and more importantly, informal values and beliefs: the day-to-day norms and behaviors. Using this research as a jumping off point and looking through the lens of more than 40 years of experience in the military, consulting industry, and Department of Defense acquisition programs, I’ve identified six roadblocks that are preventing the acquisition workforce from “doing things differently to create value” which is how I’m defining innovation.

Roadblock #1: A Lack of Psychological Safety and Trust

There is a myriad of research on the topic, but from working with students and leaders at the DAU, it is obvious we have many program offices and organizations across the DoD where the workforce does not speak up and does not trust leadership. Many things contribute to this, and it is safe to say it is not intentional but the actions and reactions and lack of leadership skills throughout the organization create a culture of mistrust and hesitancy to try something new or make changes. As described by Dr. Edmonson (Edmonson, 1999) and others, there are four levels of psychological safety: contributor safety; inclusion safety; learner safety; and challenger safety. To successfully innovate, all four levels are needed to move the needle on doing things differently. Trust, although not called out specifically as a level of safety, is inherent throughout and is cited in numerous research efforts and models for high performing teams. From my observations and discussions with organizations, there is a real, or just as critical, perceived lack of trust in leadership, peers, and the larger enterprise (look no further than the number of continuing resolutions over the last decade) which contribute to this. To innovate, you must



experiment, try novel approaches which may or may not work, fail, learn, try again, fail, and learn again before succeeding. If the culture of the organization does not accept this risk, or employees feel the supervisors or team “doesn’t have their back” you have hit a dead end on innovation.

Building trust and psychological safety in teams and organizations requires leadership to be open, transparent, and humble. Admitting mistakes, ensuring all voices are heard and acknowledged (active listening), and empowering/delegating decisions would be a great starting point. Be trusting first, assume positive intent, and collaborate, do not direct as the initial action you take. Be comfortable with being uncomfortable in giving “commander’s intent” and letting your teams move out without the leader directing all the actions or solving all the problems. Embrace and celebrate failure if something new was attempted or as part of an iterative process to produce something new to solve users’ problems (design thinking mindset). The bottom line is without psychological safety and trust as the foundation, not only will you not have an innovative organization, but you will also struggle to deliver results.

Roadblock #2: Losing Sight of the Mission and Purpose of the Work

After President John F. Kennedy set the aspirational goal of “before this decade is out, landing a man on the Moon and returning him safely to the Earth” NASA created a culture of innovation which was best reflected by the answer given by the janitorial staff when asked by President Kennedy what his role was at NASA and the reply was “help put a man on the moon.” This legendary tale exemplifies what is required to create a culture of innovation. Everyone in the organization never takes their eye off the ball and understands their role in meeting the mission. For acquisition program offices the mission is both noble and obvious, but in the day-to-day trenches of dealing with continuing resolutions each year, funding cuts, technical problems, changing requirements, stakeholder and industry concerns, etc., the workforce get wrapped up in today’s “fire” to extinguish and loses sight of the bigger picture and the overall mission of the organization. It is critical for leaders to remind their teams of this bigger picture and to connect everyone throughout the organization to this mission. How often do financial managers, for example, get to tour a ship or factory or meet with the war fighter? If you walk around acquisition program offices, do you see reminders of the mission (photos of what you are providing to the warfighter, scale models, legendary figures, etc.) or sterile walls and a few tacked up mission statements? Both Daniel Pink (2018) and Simon Sinek (2022) addressed purpose in their books and research efforts to drive employee engagement so I don’t consider it a stretch to posit the more engaged your workforce is and focused on the mission, the greater likelihood they will find new ways of doing things differently to be more efficient or solve problems. Reinforcing the mission despite the tactical issues of the day-to-day program execution needs to be a central part of your culture to foster innovation.

Roadblock #3: Not Feeling Appreciated by Leadership or the Organization

Depending upon which poll or research study you look at, anywhere from 50% to 70% of employees cite “lack of appreciation” for leaving a job or for a lack of engagement. In keeping with the NASA references Jack Swigert, the Command Module pilot on Apollo 13, famously said, “Okay Houston . . . we’ve had a problem here.” If looked at another way, a whopping 81% would work harder if their efforts were recognized by the organization. I distinguish between recognition and appreciation primarily since we associate recognition with employee cash awards or various all hands events where employees are publicly recognized. This is indeed a way to show appreciation, or not, depending upon how the individual being recognized wants to feel valued and appreciated. Research by Gary Chapman and Paul White (2019) found five separate ways employees feel appreciated: words of affirmation; quality time; acts of service; gifts; physical touch (high five, fist bump, tap on the shoulder). Cash awards, the go to method



for many program offices, is the least of the five “languages” of appreciation in the workplace according to their research. There’s ample research as well to show that intrinsic motivation (“I feel appreciated for my efforts”) is a much greater method of engagement and innovation than extrinsic motivation such as cash awards for producing a new idea. If your organization is struggling to do things differently, leadership should pulse the workforce and ask how they feel appreciated and devote time and resources (does NOT have to be money) to remove this roadblock. A simple thank you note or email with words of appreciation, brown-bag lunch, or morning coffee with the boss, help with a project, or a fist bump while walking to lunch can go a long way to removing this barrier and improve employee engagement and productivity.

Roadblock #4: Not Feeling “Connected” to Team Members and Leaders

A former colleague pointed out to DAU students that technical problems can be solved, but it requires people to solve and implement the solutions, hence our focus at the Defense Systems Management College on leadership and soft (or more appropriately, essential) skills. Compounded by the isolation and work from home mandates during the 2020–2022 pandemic, more workers feel removed from their teams and “on their own.” Research is ongoing on the long term effects of telework and work from home, and many DoD organizations are struggling with recruiting and retaining personnel if telework is more restrictive. Even with personnel who are routinely together, innovative organizations need leaders to create a sense of team and focus on building relationships. I often use a quotation attributed to Theodore Roosevelt to illustrate this point: “Nobody cares how much you know, until they know how much you care.” By nature, we are social mammals who want to feel part of the pack so considering a substantial percentage of our waking hours are at work, it is incumbent upon leaders and fellow team members to create a sense of one team and belonging. It is beyond just the inclusion aspect of psychology safety and focused on caring for each other with a common sense of purpose. Leaders have a multitude of ways they can do this, but I recommend at a bare minimum meeting one on one with employees or in small groups for large organizations to begin the process of team building and connectedness. As a leader, understanding and learning about your team’s “outside of work” life is critical to building the relationships you will need to get the most from your workforce. Family life, interests, and common bonds between the leader and team members drive these connections and demonstrate caring.

Roadblock #5: Not Feeling Empowered

If asked, most leaders will quickly answer “Of course I empower my workforce” but do they really empower them or just delegate and what is needed to be truly empowered? To drive innovation in the workplace, empowering employees to take charge and move out requires three critical ingredients: the **capabilities** they will need (training, tools, access to expertise if not internal); **resources** (money, people); and **authority** (ability to make the decisions and not have them questioned or altered). I use this empowerment triangle to highlight what it means to truly empower your team and not just pay it lip service. Leaders typically have the hardest time relinquishing authority as many view that as their role (make decisions, give direction and orders). For innovation to take hold it is critical that leaders push decisions and this authority down to the lowest level and more importantly not punish mistakes or “bad” decisions that attempted to do things differently or to learn/experiment. If the boss rolls his eyes, questions decisions, or fails to recognize in a positive way trying something new, the culture will become one of NOT trying out innovative ideas or not feeling empowered, which becomes a downward spiral for both morale and delivering results.



Roadblock #6: A Lack of Actionable Feedback Both Individually and as a Team/Organization

“Feedback is a gift” is part of the Defense Systems Management College’s culture and a robust feedback loop is a critical part of high performing organization’s culture as well. To drive innovation, you must be willing to self-examine as an organization and as individuals and adjust/change/pivot based upon that feedback. The Navy’s “Get Real, Get Better” campaign is an example of this. In our program manager’s course, we include self-reflection as part of our curriculum as another example. Industry is moving away from annual performance reviews and more towards ongoing or more frequent employee feedback, which is also a preference for Gen-Z workers entering the workforce. In his book, *The Culture Code*, Daniel Coyne (2019) points out the effective use of after action reports (ARR) by high performing cultures. Typically, in the DoD, we do AARs when something goes wrong, but Coyne posits successful organizations do these on a routine basis. For example, when the request for proposal (RFP) is released, the program office and contracting team get together to ask these five questions:

1. What results were we expecting?
2. What results did we get?
3. Why did we get these results?
4. What will we continue to do?
5. What will we do differently?

Asking these questions does not need to wait for a major event such as an RFP release but can be embedded in the team/organization’s routine to become part of the culture. Bi-weekly or monthly time set aside to reflect on progress for key program office objectives/goals/milestones that were intended to be achieved will help the organization learn and get better. Apply this to employees as well and set aside time on a routine basis for a check in/reflection with supervisors or as a team. This will also help remove the other roadblocks to creating a culture of innovation (connectedness and psychological safety). Admitting mistakes, demonstrating humility, and acting on feedback is a critical ingredient in leadership. If the boss and team seek feedback to improve, the individual will come to see this as part of the culture and will do the same.

Summary

After experiencing more than 30 organizational cultures (both good and bad) during my career prior to my arrival at DAU, I took on the challenge of trying to identify the good from the bad and what drives a successful culture. At the time, the 809 Commission identified program office culture as an indicator of successful acquisition programs which further drove my efforts. More recently, the focus on innovation and acquisition reform always seems to come down to the “culture” preventing it. Compiling the research from others and building on our existing DSMC curriculum, I pulled together “prerequisites” for a high performing culture, presented in this paper as roadblocks to creating a culture that fosters innovation. Further research efforts could focus on why DoD acquisition programs do not foster psychological safety which might identify specific areas for improvement. Another area that could be investigated is for organizations that are innovative, what aspects of their culture drive their success beyond just the nature of what they are acquiring, and their organization structure (typically very flat and not functionally aligned). I welcome your feedback and further discussion at the contact information provided.

References

Chapman, G. D., & White, P. E. (2019). *The five languages of appreciation in the workplace: Empowering organizations by encouraging people*. Northfield Publishing.



- Coyle, D. (2019). *The culture code*. Random House UK.
- Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, 44(2), 350–383.
- Graham, J. R., Grennan, J., Harvey, C. R., & Rajgopal, S. (2022). Corporate culture: Evidence from the field. *Journal of Financial Economics*, 146(2), 552–593.
<https://doi.org/10.1016/j.jfineco.2022.07.008>
- Pink, D. H. (2018). *Drive: The surprising truth about what motivates us*. Canongate Books.
- Sinek, S. (2022). *Start with why*. Manjul Publishing House.



Obstacles, and Possible Solutions, to the Analysis of Department of Defense Outsourcing for Goods and Services

Thomas Bruneau— is a Distinguished Professor Emeritus of National Security Affairs at the Naval Postgraduate School. He joined the department in 1987 and was Chairman from 1989–1995. He was the Director of the Center for Civil Military Relations from 2000–2004 and left U.S. government service in 2013. He has published 15 books, the most recent are a co-edited book, *The Routledge Handbook of Civil-Military Relations*, published by Routledge in 2020, and *Outsourcing National Defense: Why and How Private Contractors Are Providing Public Services*, which was published by Lynne Rienner Publishers in 2023. Between 1998 and 2001, he served as rapporteur of the Defense Policy Board. [tbruneau@globalacademicprofessionals.com]

Abstract

This paper is an attempt by the author to understand why research and analysis of the topic of outsourcing for goods and services by the U.S. Department of Defense (DoD) is extremely difficult. I have researched, written, and published on at least six different topics since receiving my PhD from the University of California at Berkeley in 1970, and this topic on DoD outsourcing is by far the most difficult. In the paper, I seek to explain this difficulty in terms of seven obstacles. The most important obstacles are the following: lack of an established literature on the topic; neglect of the importance of distinguishing classified and proprietary information; generally partial and thus unreliable data; complexity of the gigantic bureaucracy that is the DoD; lack of analysis of innovation, should there be any; unwillingness by those involved in outsourcing to share information; and a stultifying attitude towards innovation by large sectors of the acquisition workforce.

Background

Since obtaining a PhD in political science at the University of California at Berkeley in 1970, I have researched and published on six distinct topics. All six required foreign travel, fluency in other languages, research in documents and personal interviews, and resulted in publications in both blind refereed scholarly journals and books produced by reputable university and commercial presses. In all cases, I sought, through the research and publications, to answer a *why* question.

Beginning in 2009, I began to focus my research on outsourcing, also called contracting out, by the U.S. Department of Defense (DoD). This topic is extremely important both in terms of U.S. national security and the gigantic sums of money involved. The implications for national security are due to the current competition between the United States and its near-peer competitor, China. The amount of money the DoD outsourced in Fiscal Year (FY) 2022, \$414.3 billion, was 57% of that year's total DoD budget of \$728.4 billion (Sanders et al., 2023).

This research is by far the most difficult I have undertaken. This may be due to the fact that in this research, as in the previous six research projects, I am attempting to answer a *why* question based on empirical data, including interview data, as evidence. In this case, the *why* pertains to the unwillingness of the acquisition workforce, totaling some 180,000 military and civilian personnel at the end of FY 2021, to utilize the so-called alternative acquisition authorities in awarding the outsourced \$414.3 billion and to instead rely overwhelmingly on the Federal Acquisition Authority (FAR), despite its limits and liabilities, which are abundantly documented in both government and privately issued reports and studies.¹

¹ I deal extensively with the impediments arising from reliance on the FAR and the option of using Other Transaction Authority (OTAs) in Bruneau (2023). For a list of the alternative acquisition authorities, see GAO (2020, pp. 4–5).



My purposes in this paper are to identify and then provide possible solutions to what I have found to be major impediments to research on the topic of DoD contracting out, focusing on the acquisition workforce and, specifically, on answering the *why* question regarding their reticence to deviate from reliance on the FAR.² As the research involves a wide variety of what I define as obstacles, it is based not only on extensive documentation from government and other sources and interviews with individuals familiar with the topics covered in the research, but also upon my experience as an employee of the DoD between 1987 and my retirement in 2013. During one-half of those 26 years, I had administrative responsibilities for personnel, budgets, and employing contractors, and for all of those 26 years, and continuing until today, I had close contact with officer students from all of the U.S. armed services. Many of these students have been promoted into leadership roles in the Pentagon, combatant commands, their own services, and defense agencies. I have not hesitated to reach out to them for information on the topic of acquisitions. For more than 20 years, I held top secret/SCI clearances, which gave me access to both information and facilities in the Pentagon, intelligence agencies, other defense establishments and U.S. embassies abroad. After retirement in 2013, I worked as a contractor for 3 years and founded a contracting firm, Global Academic Professionals, duly registered in the SAM and with appropriate NAICS numbers. In sum, this paper is based upon written material, personal interviews, and some 30 years of experience as a direct or indirect employee of the DoD.

In conducting my research, which has so far resulted in two books, several journal articles, and a chapter in a book, I have identified the following obstacles to research on the topic of DoD outsourcing. Each obstacle is followed by possible solutions to the obstacles or speculation on how solutions might be discovered.

Absence of an Established Literature on the Topic

Virtually all topics in the social sciences can acknowledge an established literature in books and/or articles in scholarly journals. While Nagle's (2005) book is an excellent history of government contracting and Fox's (2021) *Defense Acquisition Reform, 1960-2009: An Elusive Goal* deals with defense acquisition issues between the 1960s and the early 21st century, neither can be considered fundamental for analysis of contemporary outsourcing by the DoD, let alone on the acquisition workforce. In order to research and write a book on outsourcing by the DoD, I had to construct my own framework, drawing heavily on my and others' conceptual work in civil-military relations.

Due to the absence of an established literature, I also have to draw heavily on two government publications. Congressional Research Service reports and Government Accountability Office (GAO) audits helped fill the gaps left by the absence of an established literature. While timely, both are, however, researched and written at the request of members and/or staff of the U.S. Congress, not by academics. They are better than nothing, but the former are spotty in coverage, and the latter are essentially financial audits.

Possible Solutions to the Obstacle

My 2023 book could serve as a first approximation to establishing a baseline for a founding literature on the topic of acquisitions. The foci, and resulting publications, will depend on how other obstacles identified here are dealt with.

² In addition to the acquisition workforce, the requirements process Joint Capabilities, Integration, and Development System (JCIDS) and budgeting process Planning, Programming, Budgeting, and Execution Process (PPBE) are often identified as major challenges in acquisition reform. Both are currently receiving great attention and scrutiny but are not included in this paper, as I have no unique insights to offer.



The Absence of Empirical Data Due to Two Aspects of What is Crucial in Outsourcing

I was privileged to personally interview the late and iconic Dr. Jacques Gansler on February 23, 2009, at his office at the University of Maryland at College Park. He stated at that time what he had previously published in the Report of the Commission on Army Expeditionary Contracting. I quoted him at that time as saying, “The contract is the nexus between the warfighter and the contracting community” (J. Gansler, interview with author, February 23, 2009). This formulation is found early in the report (Commission on Army Acquisition and Program Management in Expeditionary Operations, 2007, p. 3). I heard essentially the same formulation in my personal interview with a representative of the contracting “industry,” John Gastright, vice president for government affairs, communications, and marketing for Dyn Corp International, on January 9, 2009. He stated, “the contract is the mission statement” (J. Gastright, interview with author, January 9, 2009). In order to be awarded a contract, firms submit proposals to some entity in the DoD, the armed services, defense agencies, and other parts of the DoD. The bids include information that is often both classified and proprietary. According to the late Secretary of Defense Ash Carter (2019), who termed himself “acquisition czar,” there are “about ten million ... separate contracts awarded every year. More than ten thousand ‘contracting officers,’ trained and authorized by law to spend the taxpayer’s money, sign these contracts” (p. 9). A researcher without access to this classified and/or proprietary information can have absolutely no idea why one proposal won out over another. The researcher lacks, in short, empirical data to make a comparison or assessment. One may, however, examine the acquisition authority used in awarding contracts and get some sense of how flexible or nimble they are in bringing in emerging technologies, which appears to be a high priority in the current competition with China.

Possible Solutions to the Obstacle

I can envision no solution for research on what is proposed and what is contracted by the DoD when confronted by issues of secrecy and proprietary information. However, as there is more than one acquisition authority, a researcher may attempt to determine the impact of using different authorities. That was the purpose of my 2023 book, and there is currently a great deal of interest in the DoD in using the authorities listed in the GAO document cited in footnote 1.

Failure to Consider Classified vs. Unclassified Contracts

A further impediment to analysis due to information being classified is exemplified by the following experience, which demonstrates that something to be acquired can be unclassified, or, if necessary, classified and thus hidden from the general public.³ In late 2018, there was a tremendous amount of media-reported agony over personnel at Google who were unwilling to work on Project Maven, which involved the use of AI to identify potential targets for armed U.S. drones. More recently, in early December 2023, it became known that the project was classified and awarded to five firms, including Google. To paraphrase an article in *The Washington Post* in January 2023, 80% of Project Maven—including the drone video analysis work—was transferred to the control of the National Geospatial-Intelligence Agency (NGA; Dou, 2023). Although the NGA is part of the DoD, its procurement is done through a cloud contract (Cornillie, 2020). Orders placed through that contract, known as C2E, which stands for Commercial Cloud Enterprise, do not have to be reported publicly, and the CIA did not announce a vendor list when it awarded the contract. According to the media, Amazon, Google, IBM, Microsoft, and Oracle are the vendors. Also according to the media, Google declined to say whether there has been any change in its employees’ stance. Again according to the media, the DoD’s Chief Digital and

³ This issue is dealt with in a case study contrasting acquisition authorities used by the CIA versus those used by the DoD. See Michaels (2011).



Artificial Intelligence Office, which operates the other 20% of Project Maven, declined to comment. In short, by focusing only on unclassified material, the analysis of acquisitions, such as it is, can miss hugely important contracts.⁴

Possible Solution to the Obstacle

This focus on unclassified-only documents is not really a problem within the DoD, as everyone who deals with classified information, which is everybody, is “read into” various classifications as part of the normal pattern of work. This is, however, a problem with some of the media, which either doesn’t know, or maybe doesn’t care, about security classifications. The solution would be, as it is practiced now in the intelligence community (IC), to state in every interaction with contractors and the media that this or that issue is classified, and at what level, and if the contractors don’t have sufficient clearances, then there is nothing to say.

The Department of Defense is so large and diverse that no single person can begin to understand how it works, let alone implement acquisition reform.

Between 1998 and 2001, I served as *rapporteur* of the Defense Policy Board that provides the secretary of defense and his staff with independent advice on questions of national security and defense policy. I began to understand at that time that the DoD, extending throughout much of the world, and with the giant Pentagon at the center, is huge, cumbersome, and unwieldy. My experience with the DoD since that time, which involved monthly visits to build support for programs at the Naval Postgraduate School (NPS), reinforced this perception again and again. Indeed, a recent Secretary of Defense, Robert M. Gates (2014), states,

The Department of Defense is the largest, most complex organization on the planet: three million people, civilian and military, with a budget, the last year I was there, of over \$700 billion. Nearly everyone there is a career professional, with considerable job security. Every major part of the organization – budget, acquisition programs, and policy - has a constituency both inside and outside the Pentagon. Local and state officials, members of Congress, lobbyists, industry, retired senior officers – everyone has an oar in the water, many of them pulling in different directions. (p. 577)

It is worth noting that Gates (2014) calls specific attention to the trio of acquisition programs, lobbyists, and industry.

Possible Solution to the Obstacle

There have been numerous proposals to modernize and update the DoD. The most recent and, in my view, most feasible, is the Project on National Security Reform (PNSR), spearheaded by one of the two main authors of the Goldwater–Nichols Defense Reorganization Act of 1986 and the first Assistant Secretary of Defense for Special Operations and Low-Intensity Conflict (SOLIC), James Locher. I deal with the PNSR, as well as other proposals for reform of the DoD in Chapter 4 of my 2011 book, *Patriots for Profit: Contractors and the Military in U. S. National Security*.

Problematic Quantitative Data

Empirical data is the *sine qua non* in research. It is telling that the GAO classifies DoD contract management as *high risk*, which means, in the terms of the GAO, that DoD contract

⁴ See Chapter 4 of my book and the sources cited there for the different acquisition authorities available to the intelligence community versus the DoD.



management is not fully accountable in fiscal terms.⁵ Further, the DoD has failed its financial audit 6 years in a row. In addition to the problems of accountability and fiscal mismanagement, the Federal Procurement Data System (FPDS), while the most comprehensive source of information about federal spending on contracts, lacks complete data, and the way data is reported makes it difficult to summarize spending on contracts.⁶ Indicative of the serious problems with FPDS data, a CRS report on the theme of *Department of Defense Contractors and Troop Levels in Afghanistan and Iraq* (Peters, 2021) explicitly discusses the limitations of the data that is available through an update and modification of the FPDS. The report concludes as follows: “Nevertheless, some observers say that despite their shortcomings, the data available through the beta.SAM.gov Data Bank are substantially more comprehensive than what is available on government procurement activities in most other countries in the world” (Peters, 2021, p. 20). It is good to know that U.S. data is more reliable than that of Angola, for example.

Available data is incomplete. For example, other transaction (OT) data is not generally tracked in FPDS-NG. For the DoD, only 10 U.S.C. § 4022 information is tracked, but it does not differentiate between prototype and follow-on contracts. Other agencies do not report their OT data in FPDS-NG because it is optional per the statute or guidance. According to some researchers, FPDS-NG has given those agencies that want to report OT data pushback in allowing them to report the information. <http://go.usa.gov/3cAtG>. Further, there is no accountability for the data being reported. Contracting officers (COs) and acquisition officers (AOs) report this data but generally have little time to ensure accuracy. During the first month of reporting, it is the contractors’ responsibility to confirm their information. Per their instructions, they cannot report until FPDS-NG is correct. This can take 3–4 months for a CO to correct, causing contractor reporting delays since the correction is not prioritized.

Since the credibility and completeness of quantitative data is crucial for analysis by social scientists, it is not surprising that social scientists pay only negligible attention to the DoD despite the DoD’s gigantic budget. Thanks go to Gregory Sanders and his colleagues at CSIS for helping me with some quantitative data to begin to understand what the DoD is outsourcing and with what form of acquisition authority.

Possible Solution to the Obstacle

It seems absurd that the DoD is allowed to get away with tremendous callousness in the use of U.S. taxpayers’ money. The U.S. Congress should not allow the DoD to be so sloppy in the use of data, which, in fact, may play into strategies of major firms in the industry. It is encouraging that the Defense Innovation Board (DIB) has recommended to the DoD major initiatives and reform in data concerning virtually all aspects of DoD outsourcing. The plan is to include the initiatives and reform of data in an upcoming National Defense Authorization Act (NDAA).⁷

⁵ “Over the years since we added this area to our High-Risk List, we have made numerous recommendations related to this high-risk issue, 18 of which were made since the last high-risk update in February 2017. As of November 2018, 41 recommendations related to this high-risk area are open” (GAO, #####). See also GAO (2019).

⁶ See Federal Procurement Data System–Next Generation (FPDS-NG; n.d.). FPDS data are available at www.USASpending.gov.

For the background of the FPDS and discussion of data accuracy issues, see Sargent Jr. et al. (2018).

⁷ For an introductory article on this initiative of the DIB, see Graham (2024).



Acquisition Reform is a Frequent Focus of Attention at the DoD, but Scant Evidence Is Available to Support Any Conclusion

In line with the extreme complexity of the DoD, the theme of acquisition reform is a constant, but there is little evidence of reform. The title of the 2021 book by one of the few experts on this topic, the late J. Ronald Fox is, *Defense Acquisition Reform, 1960-2009: An Elusive Goal*. Another acquisition expert, the late Secretary of Defense, Ash Carter (2019), states in his book, *Inside the Five-Sided Box*, “most of the periodic paroxysms of ‘acquisition reform’ that sweep through government have been amateurish and counterproductive. I know, because I’ve had a front-row seat for several of them” (p. 22).

I personally had much experience with a national level program promising great things for us in managerial positions in the DoD. During the Clinton–Gore administrations, from 1993–2001, the chairman of the other largest department at NPS, Defense Management (which included acquisitions, personnel, budgets, etc.) and I very actively worked to have NPS declared the first U.S. Navy reinvention laboratory under the National Performance Review initiative led by Vice President Gore. We sought flexibility in personnel issues pertaining to GS-level staff. Despite our achieving the designation, no reforms were initiated during our tenure as chairmen, and to the best of my knowledge, no reforms were ever initiated, let alone implemented, at NPS.

The most recent effort in acquisition reform is the Section 809 Panel. An interim report of its recommendations was issued in May 2017, and the final report followed in January 2019. According to the most recent analysis of the reports’ result, dated March 7, 2022, 26 recommendations had been implemented in part or in full. Implementation means, however, in all but two or three of the 93 recommendations, inclusion in various NDAs.

It must be noted, however, that inclusion in an NDA does not mean that the DoD acts on mandates or guidelines in an NDA. For example, in Section 867 of the NDA for Fiscal Year (FY) 2018, Congress directed that

the Secretary of Defense shall establish a preference, to be applied in circumstances determined appropriate by the Secretary, for using transactions other than contracts, cooperative agreements, and grants entered into pursuant to sections 2371 and 2371b of title 10, United States Code, and authority for procurement for experimental purposes pursuant to section 2373 of title 10, United States Code. Sections 2371, 2371b and 2373 are currently codified as sections 4021, 4022 and 4023 of title 10. To effectively make use of these authorities Congress has mandated as part of section 4021:

The Secretary of Defense shall-(1) ensure that management, technical, and contracting personnel of the Department of Defense involved in the award or administration of transactions under this section or other innovative forms of contracting are afforded opportunities for adequate education and training. (NDA, 2017)

These provisions were never implemented if the “adequate education and training” (NDA, 2017) means anything more than online material at the Defense Acquisition University and George Washington University Law school offering one course on innovative contracting. Despite extreme due diligence, I was unable to ascertain if there is anything approaching the general meaning of “adequate” (NDA, 2017). It is telling that in the GAO report on *alternative agreements*, there is no mention of educational institutions in “Figure 1: Key Army Organizations Related to Alternative Agreements and Approaches Supporting Army Modernization” (GAO,



2020, p. 7) nor mention of education/training in “Figure 2: Leading Practices of a Lessons Learned Process” (GAO, 2020, p. 9).⁸

I believe the problem at NPS, continuing with the 809 Panel, and in general acquisition reform, is essentially what J. Ronald Fox (2021) concludes in his book: “The underlying stumbling block [to acquisition reform] has been and continues to be one of implementing and institutionalizing the recommendations required to bring about more professional management” (p. 206).

Possible Solution to the Obstacle

Since I began research on the general topic of DoD outsourcing, first with private security contractors in the Global War on Terror (GWOT) and more recently in the effort to obtain emerging technologies in the Great Power Competition, I have met probably 20 individuals who understand the challenge of acquisition reform and have ideas on how to resolve the problem. If the U.S. Congress would appropriate the funding to hire them for 3 months to develop a reform initiative and then in the NDAA for FY 2025, require its implementation, the problem of acquisition reform could be resolved. Already, the DIB seems to be headed in the right direction, and the realignment within the DoD and huge increase in its budget for the Defense Innovation Unit (DIU) is very encouraging.

Information Obtained Through Interviews is Important in the Social Sciences.

However, a large percentage of those who know anything about acquisitions have good reasons to keep their knowledge to themselves.

According to a CRS report with the most up-to-date data on personnel, the 2020 DoD budget was \$738 billion, of which \$420 billion was obligated in contracts, funding 464,500 full-time contract employees (CRS in Focus, 2021).⁹ The DoD and contracting firms provide support to think tanks, journalists, and Federally Funded Research and Development Centers (FFRDCs). The interpenetration of the DoD and the armed services with the contracting industry is extensive. The biggest obstacle I have encountered in conducting research on acquisitions is the lack of cooperation of people who know anything about acquisitions. I never encountered a similar obstacle to research in any of the six topics I worked and published on prior to working on this one.

Possible Solution to the Obstacle

A simple requirement by the U.S. Congress, and monitored by the GAO, that all entities and individuals that receive money from outsourcing have to publicly declare so would solve the biggest part of the problem. Then, if an office in the DoD could identify benefactors of the U.S. government money, individual scholars such as myself could start with that information, acknowledging awareness of this public money and much more easily obtain information.

Acquisition Professionals are Not Encouraged to Innovate: To Think

The main acquisition authority is undoubtedly the FAR. It encourages training (learning—what to think) versus education (how to think). Consequently, much of the literature on acquisition is in the genre of a check-the-box mentality. Depending on font size, the FAR is at least 1,600 pages long and divided into 53 parts, each dealing with a separate aspect of the procurement process. However, generating FAR-based contracts is routine, as contracting

⁸ On this same issue of lack of education in the use of OTAs see Dunn (2023, p. 1). Pages 1 and 3 from 2017 contain the most recent data for number of employees.¹⁰ I personally have no relationship with either RAND or the DMDC. For insights into the work of the two, see S. M. Gates et al. (2022).



officers have several different contract writing systems available for their work. In addition, adherence to the FAR is a security blanket in that contracting officers bear a tremendous personal responsibility for awarding contracts of the American people's money, and there are simply no incentives to take a risk. There is, in addition, a great distance between contracting officers and operators in the military, currently termed *war fighters*, resulting in a lack of awareness of specific needs of the war fighters, which might possibly result in pressure on contracting officers to innovate or take a risk. I have spoken with numerous senior military officers, active duty and retired, who had responsibility for spending large sums of money. One, with whom I spent a great deal of time over 2 decades and who was responsible for both the B-1 and B-2, informed me that he has absolutely no contact with any contracting officers either working for the U.S. Air Force or, later, in the Joint Staff.

To give a sense of this latter point, I roughly quote written comments I received from one of my graduates, who was a war fighter as the CO of a U.S. Navy warship and later had an important position in the Joint Chiefs of Staff as a U.S. Navy captain:

I would define the relationship [between operators and contracting officers] as one where, like most communities in the military, each wants the others to stay out of their lane, and in fact, has a disdain for the others. As an example, think of aviators—they can think of nothing worse than integrating with the crew when they're embarked on a ship. "We're the aviators." Anything they can do to separate themselves from the SWOs [surface warfare officers] is something they proudly and usually obnoxiously do. Think about SEALs—they see themselves as the only real operators; they talk down to everyone else. The only reason SEALs talk to the other communities is because they need to be taken somewhere, need gear, or need something to support their mission.

And in the case of Supply Officers, theirs is a community of "acquisition professionals" and "contracting officers," and etc., and if you are not one of them, then you are not worthy of discussing details because 1) you wouldn't understand; and 2) you might be trying to analyze or criticize, or worse, audit them, which would get them in trouble. Best to stay far from anyone but their own kind. Their interest is not in improvements to the process or coordination, cooperation, or integration, it's protecting their own community and continuing to do the things the way they've always been done—which is failing the military.

For example, when the submarine world needs a new weapon system based upon tactics of the enemy, the consultation between warfighters and acquisition professionals isn't one (usually) where the acquisition professional understands the TTPs [tactics, techniques, and procedures], but only what widget was asked for by some engineering duty officer, who may or may not be five or ten years past his/her war fighting days—and the result is not a true understanding by the acquisition professional of what is being sought, but merely the what. And what this leads to is that the acquisition professional doesn't ask the right questions which might even lead to a synergy of what the warfighter "wants" aligned with other possibilities that the acquisition professional could bring to the table, but merely a "get what I want" scenario. Imagine if I was a race-car driver and I told an acquisition professional that I want the fastest car, buy it; and instead of asking, "What do you want the car for, what type of racing, will it go around in a circle or need more maneuverability, etc.?", the acquisition professional simply buys me a Ferrari. Turns out, the Ferrari I get doesn't do well against NASCAR contenders, so I lose every race. I contend that this is exactly what we did with LCS [littoral combat ship] as one example. But it occurs everywhere.



Lack of common understanding of needs, requirements, TTPs, applicability, ends, means, ways, and risk—and no real communication to establish ground truths because everyone (communities) has become so siloed in their capabilities or willingness to find mutual understanding.

Possible Solution to the Obstacle

The acquisition workforce must be incentivized to use alternative (to the FAR) acquisition authorities and to be trained in the use of these alternative authorities. The training for other transaction authorities (OTAs) was stipulated in the NDAA for FY 2018 but never implemented. (any other training/education for authority implemented?)

I believe these obstacles help explain the paucity of credible analysis for a phenomenon that is both extremely important for national defense as well as an intelligent use of U.S. taxpayers' money. In a system as dysfunctional as described previously, it only makes sense for a CO to take the safest way out, even if it costs too much and does little to assure U.S. national security. The *why* must be understood in the overall context described previously. There are potential solutions to these obstacles, and it will require more political will than normal to implement them.

Currently, there are several encouraging signs in DoD acquisitions. They are probably caused by an increased and general awareness of the possible threat of China and the use of emerging technologies in the war in Ukraine. These signs include the creation of the high-level DIB in the Pentagon, the much-increased budget and direct reporting authority to the secretary of defense by the DIU, progress in reform of the PPBE process, and experiments and studies of recruitment and training of a new breed of COs. There are also more conferences and workshops put on by a growing number of interested universities and think tanks dealing with innovations in acquisitions, proposals to reach beyond the FAR, and lobbying efforts by non-traditional vendors with the DoD. In my opinion the most important university-based initiatives include the multi-university Acquisition Innovation Research Center, the Boroni Center for Government Contracting at George Washington University, and the Acquisition Research Program at NPS. I have, in addition, identified at least 10 experts (due to background and/or education) who have helped me understand impediments to and possible solutions for acquisition reform. For me, however, at the end of the day, the acquisition system is based on the actions of the COs using, or not, alternative acquisition authorities. Just as the National Performance Review process of Vice President Gore and the transformation process of Secretary of Defense Rumsfeld didn't lead to better performance or transformation, my concern is that the innovation won't lead to innovation. To ascertain progress in this regard, a sample survey of COs would be necessary. The RAND Corporation would be an appropriate entity to conduct such a survey. They have experience in analysis of the acquisition workforce and an ongoing relationship with the Defense Management Data Center (DMDC) in Monterey, CA.¹⁰

¹⁰ I personally have no relationship with either RAND or the DMDC. For insights into the work of the two, see S. M. Gates et al. (2022).



References

- Bruneau, T. C. (2011). *Patriots for profit: Contractors and the military in U. S. national security*. Stanford University Press.
- Bruneau, T. C. (2023). *Outsourcing national defense: Why and how private contractors are providing public services*. Lynne Rienner Publishers.
- Carter, A. (2019). *Inside the five-sided box: Lessons from a lifetime of leadership in the Pentagon*. Dutton.
- Commission on Army Acquisition and Program Management in Expeditionary Operations. (2007). *Urgent reform required: Army expeditionary contracting*. [http://www.army.mil/docs/Gansler Commission Report Final 071031.pdf](http://www.army.mil/docs/Gansler_Commission_Report_Final_071031.pdf)
- Cornillie, C. (2020, November 20). *CIA awards secretive multibillion-dollar cloud computing deal*. Bloomberg Government. <https://about.bgov.com/news/cia-awards-secretive-multibillion-dollar-cloud-computing-deal/>
- Dou, E. (2023, December 21). Pentagon cloud tie-up with Silicon Valley off to a slow start. *The Washington Post*. <https://www.washingtonpost.com/technology/2023/12/21/pentagon-cloud-computing-contracts/>
- Dunn, R. (2023, August 2). Feature comment: New DOD other transaction guide: Retreating from innovation. *The Government Contractor*, 65(28), P.1.
- Federal Procurement Data System–Next Generation. (n.d.). FPDS-NG FAQ. Retrieved April 9, 2024 from https://www.fpds.gov/wiki/index.php/FPDS-NG_FAQ
- Fox, J. R. (2021). *Defense acquisition reform, 1960-2009: An elusive goal*. U.S. Army Center of Military History.
- GAO. (20##). https://www.gao.gov/mobile/high-risk/dod_contract_management/why_did_study
- GAO. (2019, March). *High-risk series: Substantial efforts needed to achieve greater progress on high-risk areas* (GAO-19-157SP). <https://www.gao.gov/products/gao-19-157sp>
- GAO. (2020, October). *Army modernization: Army should improve use of alternative agreements and approaches by enhancing oversight and communication of lessons learned* (GAO-21-8). <https://www.gao.gov/products/gao-21-8>
- Gates, R. M. (2014). *Duty: Memoirs of a secretary at war*. Alfred A. Knopf.
- Gates, S. M., Roth, E., & Kempf, J. (2022). *Department of Defense acquisition workforce analyses: Update through fiscal year 2021*. RAND.
- Graham, E. (2024, February 6). *All Pentagon contracts should secure data rights, says advisory board*. Defense One. https://www.defenseone.com/policy/2024/02/defense-innovation-board-looks-lock-data-access-all-vendor-agreements/393930/?utm_campaign=dfn-ebb&utm_medium=email&utm_source=sailthru&SToverlay=2002c2d9-c344-4bbb-8610-e5794efcfa7d
- Michaels, J. D. (2011, June). The (willingly) fettered executive: Presidential spinoffs in national security domains and beyond. *Virginia Law Review*, 97(4), 801–898.
- Nagle, J. F. (2005). *A history of government contracting*. George Washington University Press.
- National Defense Authorization Act for Fiscal Year 2018, Pub. L. No. 115-91, 131 Stat. 1283 (2017). <https://www.congress.gov/bill/115th-congress/house-bill/2810/text>



Peters, H. M. (2021, February 22). *Department of Defense contractor and troop levels in Afghanistan and Iraq: 2007-2020* (CRS Report No. R44116). Congressional Research Service.

CRS in Focus. (2017, December 17). *Defense primer: Department of Defense contractors*.

Sanders, G., Velazquez, N., Hardesty, E., & Aldisert, A. (2023, December). *Defense acquisition trends 2023: A preliminary look*. Center for Security and International Studies.

Sargent, J. F., Jr., et al. (2018, July 2). *Defense acquisitions: How and where DOD spends its contracting dollars* (CRS Report No. R44010). Congressional Research Service.



Using Natural-Language Processing and Large Language Models to Restructure DoD Comptroller Budget Materials into Portfolio Views

Jose Ramirez-Marquez—is an Associate Professor and Division Director of Enterprise Science and Engineering in the School of Systems and Enterprises at Stevens Institute of Technology. His research focuses on the development of mathematical models for the analysis and computation of system operational effectiveness research and explores the interplay between data visualization and analytical decision-making. He conducted funded research for industry and government and published more than 100 refereed manuscripts in technical journals, book chapters, and industry reports. He holds degrees from Rutgers University in industrial engineering (PhD and MSc) and statistics (MSc) and from Universidad Nacional Autonoma de Mexico in actuarial science. [jmarquez@stevens.edu, 201-216-8003]

Doug Buettner—is the Acquisition Innovation Research Center (AIRC) Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu, 201-216-8300]

Joshua Gorman—is an undergraduate at Stevens Institute of Technology pursuing a Bachelor of Science in computer science. Starting this summer, he will begin his career as a Software Engineer Associate, drawing upon his strong background in modern web development to upgrade pivotal in-house tools. Previously, Gorman garnered valuable experience as a research assistant, actively contributing to diverse application development projects ranging from meteor strike modeling to nuclear fallout predictions. He aims to continue his education by pursuing an accelerated master's program in computer science later this year, to which he is looking to concentrate in cloud computing and computer networking. [jgorman4@stevens.edu, 201-216-8003]

Aashita Patel—is a sophomore at Stevens Institute of Technology. She is majoring in business and technology with concentrations in data analytics and management, and a minor in entrepreneurship. At Stevens, she is participating in an Inclusive Leadership Certification Program to fuel her understanding of diversity, equality, and inclusiveness. She has past professional experience researching Women's Leadership, working as a Sales Associate, and Tutor. After graduation, Aashita plans to work in Business Management. [apatel50@stevens.edu, 201-216-8003]

J. Matthew Mercado—is a graduate student at George Mason University's Master's in Business Administration program. He has past professional experience working as a full stack software engineer, SQL database administrator, network operations technician, IT help desk specialist, as well as a pizza shop manager, a volunteer English teacher, and a coding instructor for elementary school children! Mercado plans to join the Navy Reserve and work towards a PhD someday focusing on data science and/or emerging technology usage within the realm of federal contracting. [jmercad8@gmu.edu, 703-993-6954]

Hoong Yan See Tao—is a Research Project Manager at the Systems Engineering Research Center (SERC). She conducts and manages research in mission engineering, digital engineering, human capital development, systems engineering, and systems management transformation. She was on the Stevens Sustainability Committee and part of the student-led organizing committee for the inaugural Student Sustainability Symposium in 2016. In 2013, she was the president of the INCOSE Student Division at Stevens. She earned her BS and MS in electrical engineering from the University of Texas at El Paso and her PhD in systems engineering from Stevens Institute of Technology. [hseetao@stevens.edu, 201-216-8300]

Victoria Cuff—is an AIRC research fellow aiding OSD in research, focusing on PPBE Commission, workforce initiatives, and improving acquisition management policies. Previously, she was the Chief



Growth Officer at Sagely, a learning and development start-up focused on individualized learning for digital upskilling and reskilling. As the Senior Advisor for Agile Acquisition, she led the BA-8 Pilot for digital technology and created the Digital DNA course for acquisition professionals. Earlier, she contributed significantly to USAF's Kessel Run as the Agile Acquisition lead. Cuff holds a bachelor's in actuarial science, financial mathematics, and master's in applied mathematics from Clemson University. [vcuff@stevens.edu, 201-216-8300]

Philip Antón—is the Chief Scientist at the Acquisition Innovation Research Center (AIRC). He conducts applied research, modeling, piloting, education, and transition of innovative ideas and approaches to improve acquisition outcomes. From 1998 to 2021, Antón conducted policy research at the RAND Corporation. From 2011 to 2016, Antón directed the Acquisition Policy Analysis Center for the Under Secretary of Defense for Acquisition, Technology, and Logistics, analyzing acquisition performance and policies while crafting affordability policy. Antón's MS and PhD are from University of California, Irvine, specializing in computational neuroscience and artificial intelligence. His BS is in computer engineering from UCLA. [panton@stevens.edu, 201-216-8300]

Michael McGrath—is an AIRC Research Fellow and independent consultant with broad government and industry experience in technology, management, logistics, and acquisition. His prior senior DoD positions include DASN(RDT&E), ADUSD(Dual Use and Commercial Programs), a DARPA Program Manager, and OSD Director of the DoD Computer-aided Acquisition and Logistics Support program. He has served on multiple Defense Science Board and National Academies studies. His research interests are in manufacturing, cybersecurity, and data science. McGrath holds a doctorate in operations research from George Washington University. [mmcgrat1@stevens.edu, 703-407-3946]

Abstract

When the yearly President's Budget (PB) is submitted to Congress by the Department of Defense (DoD) Comptroller Budget Documents is complemented by "Justification Books" or "J-Books." These detailed documents provide budgetary details by individual programs, projects, and activities within individual military departments or defense agencies rather than from an integrated portfolio or mission perspective. This disjointed structure makes it difficult for non-DoD insiders (and likely congressional staff for whom these materials are intended) to understand the net operational effect of the requested investments let alone their constituent program elements. Given that final reports Section 809 Panel and subsequent statutes provide that the DoD should use data-driven portfolio management for acquisition and capability investments, we asked how the existing J-Book documents could be restructured to facilitate a portfolio view. Our paper first provides the results from our exploratory use of natural language processing (NLP) techniques to perform a key word search across multiple J-Books to extract and subsequently process the content associated with a key word. For the purposes of demonstration, we focused on using these techniques to identify disparate elements of Joint All Domain Command and Control (JADC2) in these J-Books. JADC2 was chosen as this DoD strategy spans multiple service's Research, Development, Test & Evaluation (RDT&E) J-Book volumes. This research also explored whether emerging large language models (LLMs) could be used to answer different types of portfolio or other questions about DoD spending without changing the existing layout and document delivery approach. We provide the results for our implementation of a dashboard proof-of-concept with an LLM interface from refactoring these budget materials including a temporal analysis of the J-Books content spanning multiple years. The final demonstration's use case is from the perspective of a new congressional staffer trying to understand the differences between these budget materials across the years.

Introduction

The Stevens Institute of Technology's Systems Engineering Research Center (SERC)/Acquisition Innovation Research Center (AIRC) was tasked by the PPBE commissioners to research the following areas,

1. Conduct case studies of technology transition.



2. Provide PPBE process research and analysis with recommendations to determine the following:
 - a. if the process should be the same for programs that breach their Major Defense Acquisition (MDAP) threshold (10 USC 4201), their Major Systems threshold (10 USC 2302d) and non-major systems,
 - b. examine how the Department of Defense (DoD) uses acquisition pathways,
 - c. the legal foundations that drive PPBE and develop a matrix outlining how PPBE components are directed (i.e., statute, regulation, policy, or practice).
3. Explore ways to restructure budgets and artifacts (Portfolio Budgeting, J-Books, and Selected Acquisition Reports [SARs]) around threats, missions, operations, and portfolio levels rather than the level of acquisition programs.
4. Explore potential reform areas that rely on coordinated changes in all three of the PPBE, Requirements, and Acquisition communities.
5. Explore options for restructuring the DoD's President's Budget (PB) Proposal including potential groupings or combinations of budget activities (BAs) paying particular attention to how authorizers and appropriators can still understand the program's phase within potential new structures providing historical insight into how and why the current budget structure was created and evolved.
6. Explore alternative Obligation (Obs) and Expenditure target curves/profiles instead of linear targets. (This task was minimally expanded to investigate the impact of Continuing Resolutions [CRs] with the same analysis.)

Our third and fifth tasks wanted us to explore ways of restructuring budget artifacts by restructuring the PB proposal and overlapped in one significant manner, which is reviewing new streamlined methods for how they are submitted. Hence, one of our early goals was to review The DoD budget documents, also called "Justification Books" or "J-Books." These are detailed documents that justify their budgetary requests and resource allocation for specific programs, projects, or activities within the DoD. These documents are typically produced as part of the President's Budget request to Congress in the U.S. federal budgeting process and serve several essential purposes, such as:

Resource Justification: J-Books provide a comprehensive rationale for allocating resources, including funds, personnel, and equipment, for specific defense-related activities. This justification is essential for decision-makers in Congress who assess and approve the DoD budget.

Transparency: J-Books are designed to enhance transparency in the budgeting process. They help the public, Congress, and other stakeholders understand the DoD's financial needs, the objectives of various programs, and how to utilize resources.

Accountability: They hold the DoD accountable for how it spends taxpayer dollars. By detailing each program's expected outcomes and benefits, J-Books allows Congress to assess whether the proposed expenditures align with national defense priorities.

Program Evaluation: These documents with other associated materials (e.g., National Defense Authorization Acts [NDAAs], SARs, etc.) help evaluate the performance and effectiveness of DoD programs. They often include metrics and performance measures, allowing Congress to gauge the success of these programs.



Program Planning: J-Books assist in program planning by outlining goals, objectives, and expected milestones for various DoD initiatives. This information is critical for managing and tracking the progress of these programs.

Congressional Approval: J-Books play a crucial role in the congressional approval process. They are used by Congress to make informed decisions regarding the allocation of defense funds and to ensure that the DoD budget aligns with national defense and security priorities.

The next section provides a problem statement which is followed by our methodology and approach. This section is followed by our results section, and conclusions section. Our paper finishes with suggestions for future research.

Problem Statement and Our Solution Concept

A comprehensive understanding of the complexities and interdependencies of DoD programs, viewed through J-Books, ultimately advances national defense capabilities, and helps in formulating sound budgetary strategies. However, understanding the budgetary implications of various DoD acquisition programs through J-Books is daunting, especially for complex cross-domain strategies like Joint All Domain Command and Control (JADC2). JADC2, which is woven throughout the different branches of the DoD, is challenging. Tools present within the DoD's Advana

¹ environment (e.g., GAMECHANGER²), and potentially specialized tools built external to the DoD can support human-based analysis. While Advana's capabilities are ever expanding and quite useful, we wanted to completely reconsider how analysis of budgets could be achieved via advanced presentation of the data using modern dashboards with a natural language processing (NLP) and generative artificial intelligence (AI) interface. Generative AI or large language models (LLMs) were just becoming available as we began this research and thus, we envisioned an LLM interface to support queries of documents (Star Trek™ here we come! [Paramount Pictures, 1986]).

However, first, we provide our concept of a Portfolio Analysis Dashboard using Tactical Aircraft in Figure 1.³

¹ Advana is the DoD's enterprise-wide, multi-domain data, analytics, and artificial intelligence (AI) platform that provides military and civilian decision makers, analysts, and builders with unprecedented access to enterprise tools and capabilities—all in a scalable, reliable, and secure environment.

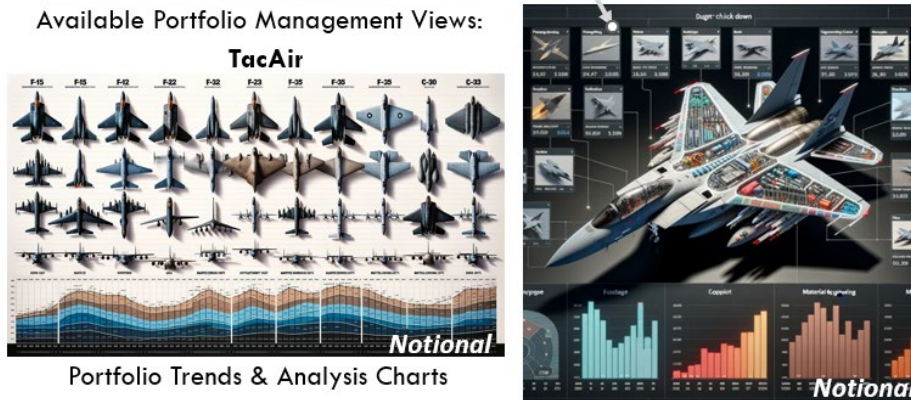
² For additional information on GAMECHANGER, see DIA Public Affairs (2022).

³ Note that OpenAI's ChatGPT and DALL-E terms of service can be found here, <https://openai.com/policies/terms-of-use>, with their publication policy found here, <https://openai.com/policies/sharing-publication-policy>. Stock language provided by OpenAI is amended to be the following: *The authors generated this content in part with GPT-3.5 and GPT-4, OpenAI's large-scale language-generation model. Upon generating draft language and DALL-E concept art, the authors reviewed, edited, and revised the language to their own liking and take ultimate responsibility for the content of this publication.*





Available Portfolio Management Views:



SOURCE: ChatGPT/DALL-E Concept Art

Figure 1. Conceptual Portfolio Analysis Dashboard Using Notional Tactical Aircraft (TacAir)

The primary analyst workstation would allow menu access to various portfolios managed by the DoD (top dual monitor layout in Figure 1 with notional analytical charts and graphs). The arrow depicts a conceptual roller-wheel mouse interface to allow the analyst to select among the various aircraft managed for click-through deep-dives. An alternative concept is the overall Tactical Aircraft (TacAir) chart on the bottom-left ranking the aircraft by funding in the rows and the Fiscal Year (FY) in the columns with a rolled-up portfolio funding view at the bottom.

The following list describes additional monitor-specific analytical concepts:

Left Monitor:

1. Newly Proposed Spending Analysis:

- **Visual Representation:** Charts and graphs displaying the newly proposed spending.
- **Details:** Data shows service system contributions, e.g., Air Force, Army, Navy, and Space Force.
- **Interactivity:** Users can hover over sections of the charts to view exact figures and percentages.

2. Lines of Effort Details:

- **Drill-Down Section:** Interactive selection that allows users to explore various lines of effort.
- **Information Display:** Provides project names, services involved, and allocated budgets.
- **User Interaction:** (L/R) clicking on a graph reveals additional information and historical data.



3. LLM Interface:

- **Summarization:** An LLM interface support for requesting a quick summarization of the observed trends.
- **New Queries:** Interface to the back-end database to create normal language data queries to populate selected charts and figures.

Right Monitor:

1. Past Budget Submissions:

- **Tabular Format:** Information on past budget submissions allowing FY-after-FY comparisons.
- **Graphical Data:** Line graph with trends in budget submissions over time.
- **Accessibility:** Users can sort and filter this data based on different parameters.

2. Latest (Real-Time/Monthly/Weekly) Cost/Schedule Performance Data:

- **Visual Indicators:** Latest data from the EVMS, cost, schedule, and performance risks.
- **Color-Coded Alerts:** Risks are indicated using color-coded alerts, providing an immediate sense of any issues. (Further drill-down with the latest SPO risk management updates for details.)
- **Detailed Metrics:** Information on cost overruns, schedule delays, and other metrics.

3. User Interface and Navigation:

- **Navigation Panel:** Monitors feature intuitive navigation panels for easy access to different sections.
- **Customization:** Users can customize the views and prioritize the information they find relevant.

While we just show the analytical capability concept in this example, one can also envision an LLM enabled J-Books creation capability where the LLM has been trained in the area of application or can simply be used in support of creating more succinct and easier to understand digital data constructs.

Methodology/Approach

Figure 2 shows the overall proof-of-concept flow for our proposed approach. We will discuss in this section the approach taken to implement this analysis flow.



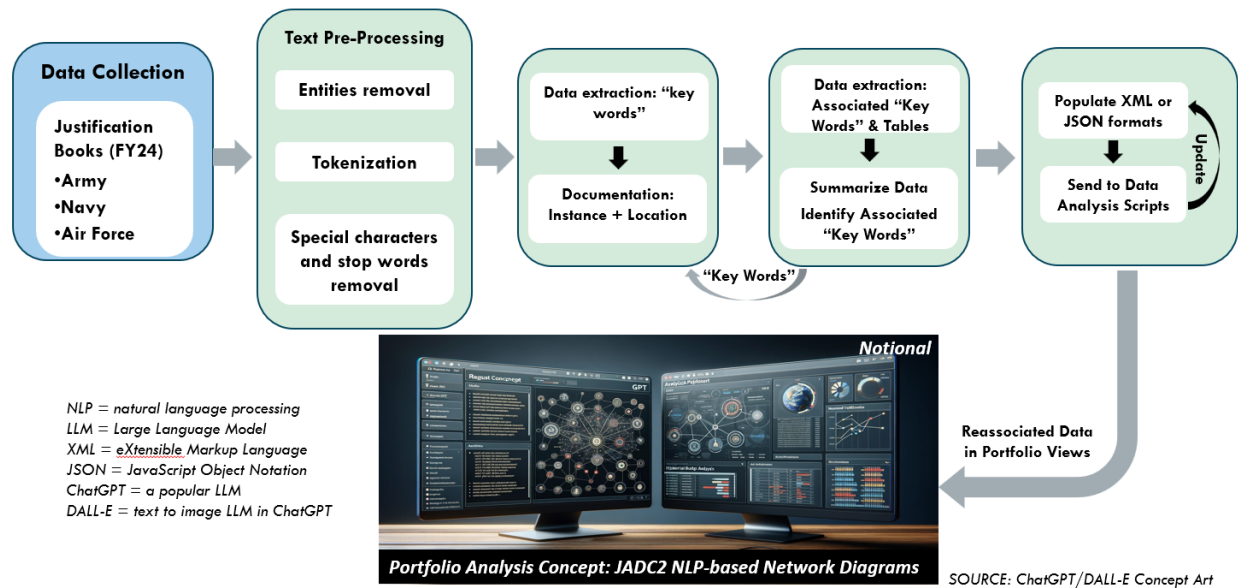


Figure 2. Proof-of-Concept Portfolio Analysis Flow Using JADC2 Keyword Search of FY24 J-Books

Our proposed approach provides an initial proof-of-concept for accumulating programs into “portfolio-like” groupings, leading to a clearer representation of budgetary implications across DoD programs. It is important to note that the information accumulated using this NLP approach on J-Books does not result in a true portfolio (as there is information that will be lacking) but is a step towards restructuring the existing content as elements in a portfolio.⁴ For this reason, the approach using JADC2 demonstrates a “portfolio-like” representation of these budget materials.

Hence, we believe it will help in the following areas:

Inter-Service Synergy: What are the fundamental mechanisms or strategies to enhance synergy and coordination among the Army, Navy, and Air Force regarding related programs to maximize their effectiveness and interoperability?

Technological Integration: Technology integration should be considered a core component of J-Books analysis. It would clarify how emerging technologies are incorporated into different programs and how they intersect across the various branches.

Resource Allocation: How do allocating resources and budgetary decisions within each branch affect the development and implementation of related programs? What are the trade-offs and challenges associated with resource allocation? How have resource allocation and budget requests changed over time? Can we correlate these temporal fluctuations with risk management behaviors of DoD decision makers?

Operational Implications: Within the J-Books framework, it’s essential to understand the functional implications of programs like JADC2 on the specific missions and functions of the Army, Navy, and Air Force. Moreover, it delves into how these branches adapt to accommodate changes to ensure the success of their missions.

This project aims to eventually leverage the promise of NLP techniques to accumulate data associated with programs (or instances) and strategies within J-Books. Moreover, by

⁴ A true portfolio would include numerous other elements that are not present in the publicly available J-Books, such as resource allocation, strategic value, risk assessments, and mitigation costs, among others.

harnessing the power of NLP and generative AI, the project aims to enhance the understanding of budgetary implications across J-Books.

The overall objectives of this project are described in two phases:

Phase 1. Demonstrate NLP J-Books “portfolio-like” accumulation: After identifying J-Book sections associated with keywords (acronyms) through NLP pattern matching, we plan to associate other materials (e.g., NDAAs, SARs, etc.), thus demonstrating “portfolio-like” budgets and displaying their context in a network diagram.⁵

Phase 2. Add analytics: The project then prototyped an analytical front-end of the information for algorithmic approaches to facilitate a deep understanding of the portfolio’s trends. The prototype includes an OpenAI LLM front-end to provide responses to chat-like questions regarding the data and text, with visual displays.

Method Used to Demonstrate J-Books Key Word Extraction

The efficient processing of large volumes of text-based documents presents a significant contemporary challenge in information management. This section introduces an inventive framework designed to simplify content extraction and the identification of content placement within documents. These processes hold immense importance across diverse domains, including academic research and corporate decision-making, as they facilitate rapid and efficient information retrieval while preparing text for decision support.

Prior to following the process as depicted in Figure 2, the researchers utilized Adobe Acrobat in a pre-processing step to apply optical character recognition (OCR) to convert each image in these documents to provide extractable data. OCR can introduce errors into this analytical process (see for example, Lamba and Madhusudhan [2023]). Requiring the Services to not use images in their budget justification documents was our first recommendation to avoid this timely and potentially error prone step.

The framework depicted in Figure 2 contains two key phases: Phase 1, which involves Text Extraction and Summarization, and Phase 2, which focuses on Analytics Extraction.

Phase 1 Text Extraction and Summarization

The framework (see Figure 3) that we implemented in our initial Phase 1 script offered a rudimentary but versatile initial toolset for keyword identification, instance location, and associated content extraction.

⁵ Network diagrams are used as a graphical depiction of the data. See for example Crane et al. for some of the benefits from this kind of analysis.



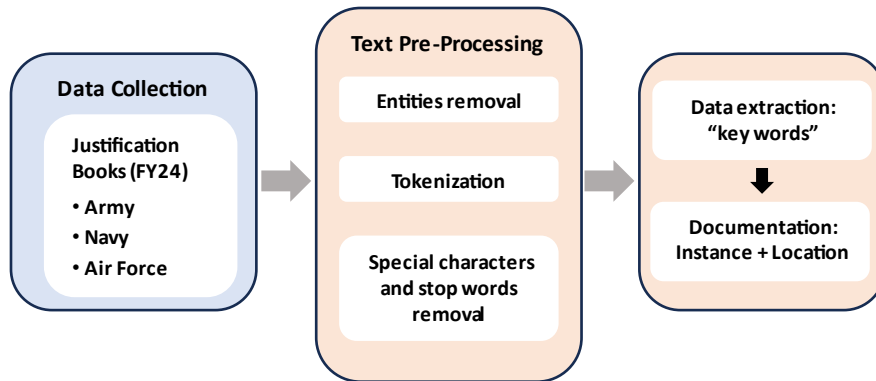


Figure 3. Instance Identification Framework For Phase 1

By automating these processes, the framework enhances the efficiency and effectiveness of information management, allowing users to quickly obtain keyword context and insights from the J-Books.

Phase 2 Prototyped Analytics Front-End

Phase 2 of the framework was intended to provide a prototype of a versatile toolkit for developing analytics, data, text, and visuals. These processes contribute significantly to information organization, retrieval, and insight generation. Figure 4 provides a graphical view of the overall demonstrated capability.

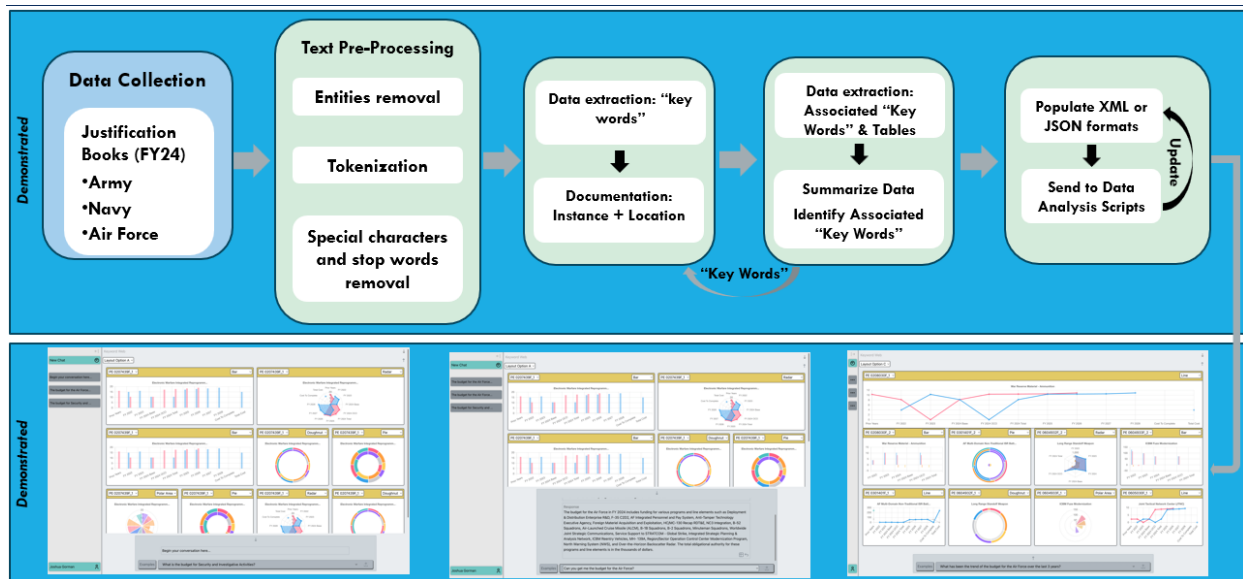


Figure 4. Completed Prototype Work from Phase 1 and Phase 2

Shown in Figure 4 are the completed/demonstrated aspects of the prototype. The top blue box in the figure demonstrates scripts that can extract key words (Phase 1). We were then able to send those results into ChatGPT to extract additional keywords to search for, illustrated as the “Key Words” feedback loop in the top-row of the figure. The output was processed to extract tables from the PDF J-Books files, and successfully associated in the feedback loop of key word search results to create an associated network diagram (Figure 5). We also successfully populated JSON files using the extracted tables from these search results and subsequently used those JSON files in a prototyped front-end (the bottom blue box in Figure 4).



The prototyped front-end (the bottom half of Figure 4) includes a ChatGPT interface to query a back-end database server containing the J-Books extracted JSON information. The current implementation only includes FY24 Air Force RDT&E volumes I, II, IIIa, and IIIb, but could easily be extended to include all the J-Books' contents. Further work is required to fully integrate the NLP processing in the top half of Figure 4 with the front-end and back-end implementation with the ChatGPT/LLM interface shown in the bottom half. Figure 5 is a screengrab of the network diagram results of our JADC2 keyword association. Future iterations of this diagram will include symbols, line type variations and colors to provide visual meta-information regarding the association.

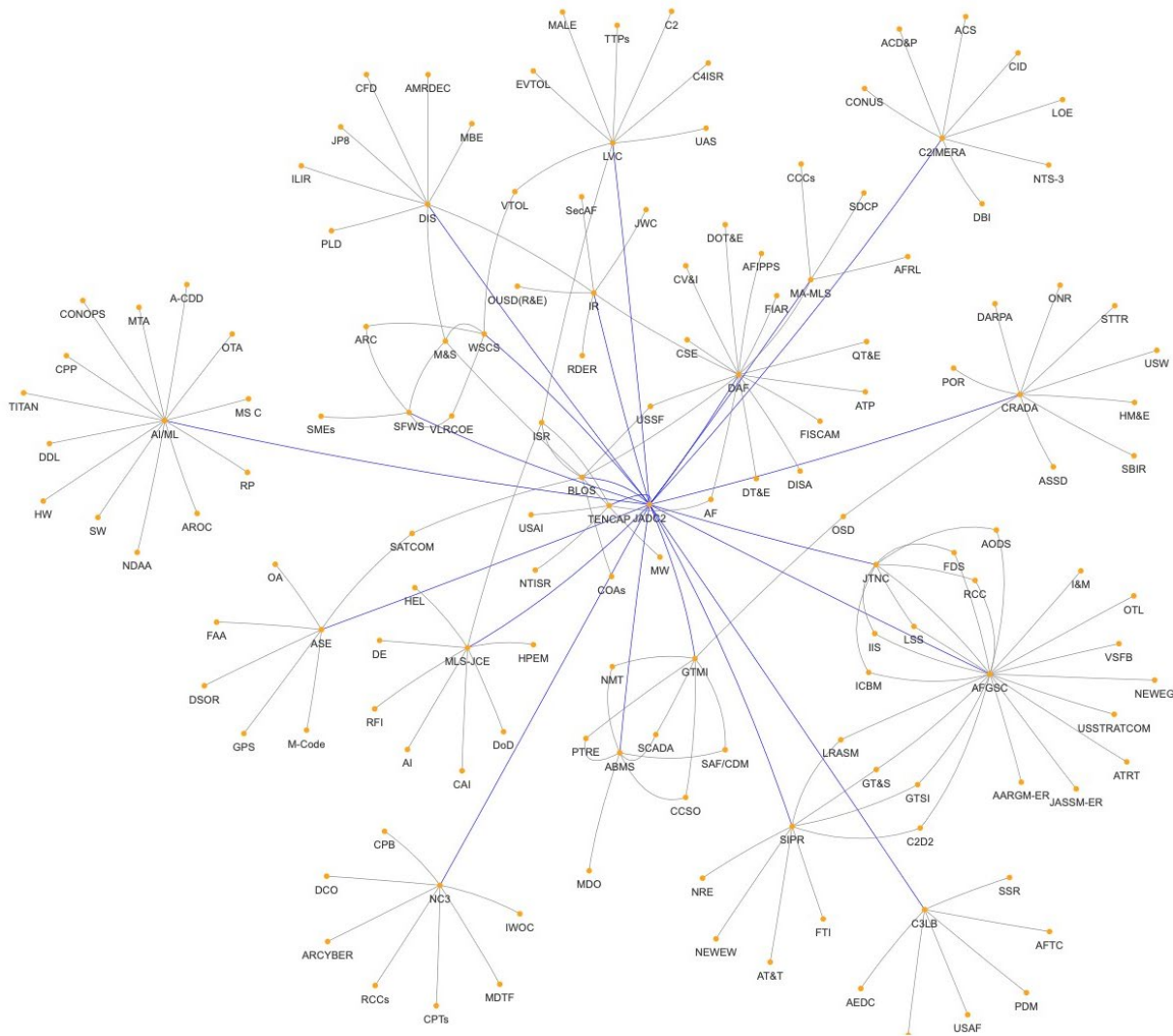


Figure5. Network Diagram Results from Key Word Processing FY24 J-Books for “JADC2” and Associated Acronyms

Note that we made no attempt to identify all the acronyms found by the feedback loop for this proof-of-concept report (i.e., all these acronyms will not be found in the acronym definition section of our final reports [Buettner et al., forthcoming; Ramirez-Marquez et al., forthcoming]). Future functionality should define acronym categories and use other contextual syntax-based “key words” for correlating concepts and terms for association. Figure 6 provides a larger



screengrab demonstrating a changed layout with new tables feeding into the graphics based on the most recent LLM query with its response.

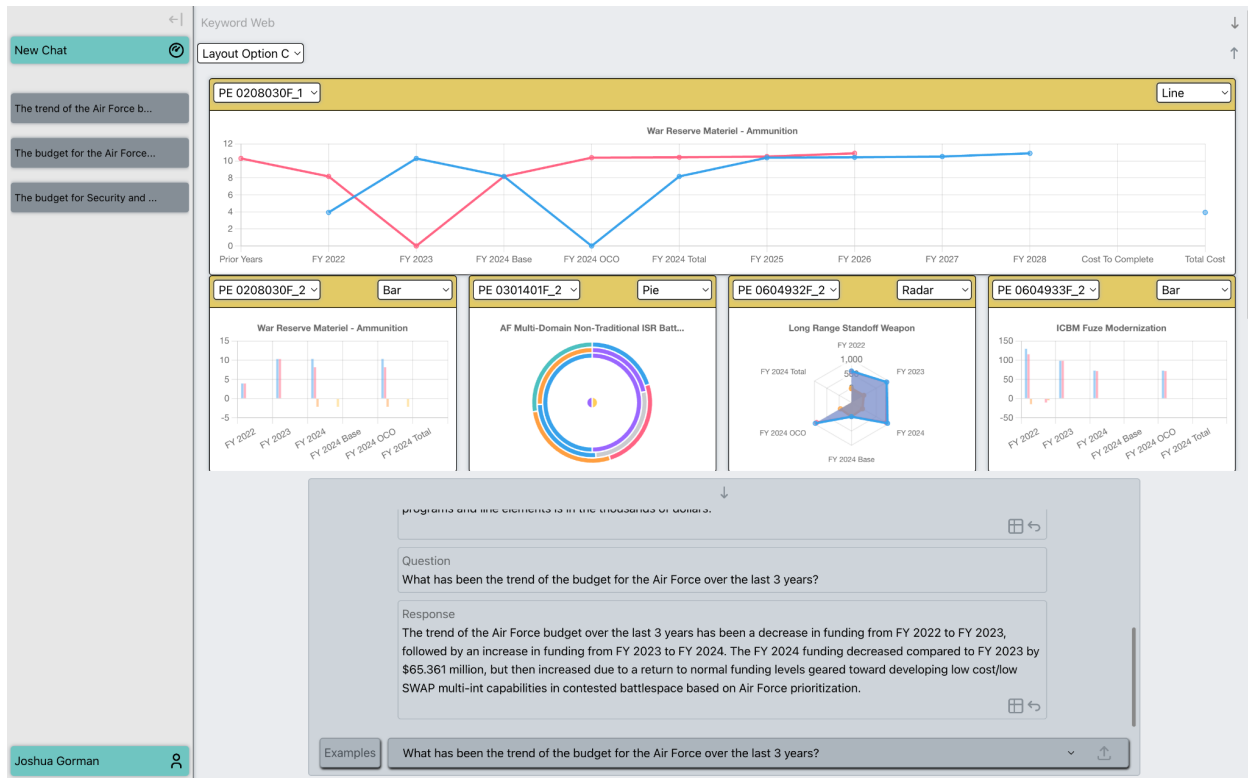


Figure6. Screengrab Demonstrating a Changed Layout with New Tables Based on the Recent LLM Query with the Response

Finally, our Phase 2 implementation was provided directly to the PPBE Commissioners with a video of the primary capabilities implemented by the end of February 2024 and documented in our final reports (Buettner et al., forthcoming; Ramirez-Marquez et al., forthcoming). Table 1 documents the final functionality that was included in our dashboard prototype.

Table 1. Completed Dashboard Prototype Functionality

User Story Name	Story Acceptance
As a User, I want to log in to the dashboard.	The dashboard should have a secure login system. Users should be authenticated against a user database.
As a User, I want to query the database using LLM.	There should be a text input box for users to input their queries. The language model should process the query.
As a User, I want to submit my query and receive relevant data.	Upon submitting the query, the dashboard should send a request to the language model. The language model should interpret the query and generate a query. The database should return relevant data based on the query.
As a User, I want to see the queried data displayed in a table format on the dashboard.	The dashboard should dynamically generate a table to display the queried data. The table should include relevant columns and rows based on the query.



User Story Name	Story Acceptance
As a User, I want to save my queries for future use.	The dashboard should have a feature to save user queries. Users should be able to retrieve and reuse saved queries.
As a User, I want to visualize the queried data with charts and graphs.	The dashboard should provide options to generate visualizations (e.g., bar charts, pie charts) based on the queried data. Users should be able to customize and interact with the visualizations.
As a User, I want to be able to upload new tables to the database.	A portal to upload CSV/XML/JSON files to database. All files if not in JSON need to be converted to JSON.
As a User, I want to be able to use an LLM for my queries and data display.	A User can use an LLM to query the database and plot data.
As a user of the system, I need it to accept an extracted pdf table and populate XML or JSON format.	To achieve this goal, I have written a code to ask the user for a JBOOK and program element. This code then extracts the specific table.
As a User I want to convert tables to XML or JSON format.	The user can convert a PDF table into either XML or JSON.
As a User I want some examples to choose from/general overview.	Create a way to save “templates” which have built in tables and layout. Should be able to choose from a dropdown a “query” sentence which has the desired results. Make sure this is new user friendly, aka intuitive.
As a User I would like to search for additional keywords related to the original search.	
As a User I want to visualize the network of associated keywords.	
Integrate network diagram.	Look for ways to include user options

Conclusions

Trying to use an LLM on a much larger document did not provide satisfactory results, as could be seen in Appendix A of Buettner et al. (forthcoming). Appendix B in that same reference demonstrated the variability of an LLM which could be expected to provide without fine-tuning it to this domain. To support this claim, we provided a quote from an article that claimed this could be overcome.

By fine-tuning the model on specific domains or topics, users can enhance the relevance and uniqueness of the generated responses. Customizing ChatGPT’s behavior by providing explicit instructions or preferences also helps in reducing repetition and tailoring the output to specific requirements.

Despite our overall lack of satisfactory results in the larger analytical context, our results section demonstrated an implementable alternative to obtain satisfactory short summaries of



sub-sections that were extracted from the much larger document.⁶ Hence, we postulate that this proof-of-concept successfully demonstrated a satisfactory LLM summarization of 19 pages from a DoD J-Book, as the existing formatting provided by the USAF easily allows automatic sub-sectioning⁷ of PEs for analysis by an LLM. Hence, we integrated an LLM into our prototype dashboard and successfully had it analyze sections of data from these same AF J-Books using our database implementation.

Importantly, our research can be used as a first step to build a capability to associate programs and their elements into portfolios and demonstrated an implementation of our Phase 2 concept. We felt our concept should provide the capability of reorganizing existing documents into a portfolio budget view. Once that reorganization has occurred, we could have assessed what data is missing for this view to be considered a portfolio budget for PEO management as envisioned by the 809 Panel. This research provided a foundation for refactoring budget materials into a more user-friendly product with links to various parts of the larger portfolio.

Ultimately, this research empowers the DoD and Congress to confidently invest in alternative approaches to understand the complexities and interdependencies of our large DoD programs, how they can begin to associate programs into portfolios, and contribute to a more comprehensive understanding. Such insights enhance the capabilities and effectiveness of military programs and inform more efficient resource allocation, bolstering the overall decision-making processes. This research can significantly influence program development and implementation, ultimately advancing national defense and military strategies.

Initially, we felt a follow-on effort should attempt to characterize the various ChatGPT PDF reader plugins, however, recent updates to ChatGPT-4 have phased these plugins out. Other LLM technologies should be investigated to determine if they can be appropriately trained/tuned to provide summarizations that directly align with congressional staff needs. Further, using a specifically trained LLM in a “common data analytics platform” (PPBE Reform Commission, 2024) could also be a future research step.

Our research team recommended:

- **Funding further research into LLM support of budgetary analysis:** Noting that Advana⁸ has implemented the “GAMECHANGER”⁹ capability, and a blog indicates that at least one tool vendor may have resolved the issues noted in this report¹⁰, further research should assess various LLMs and confirm that an LLM can be trained on the DoD’s corpus of data. The goal is to confirm that a ChatGPT text-based query interface can reliably support and enhance analysts with their tasks.
- **Funding further efforts to reorganize budget documents:** Providing portfolio-like budget views from AI/ML reorganizations without having to make drastic changes to the existing documentation format used by the various services would simply add a processing step to the existing delivery flow. Once completed, these results can support collaborative decisions on what changes and additions should be promulgated to the services to fully enact a portfolio management approach that includes a portfolio budget view during the accumulation of the data.

⁶ Recent changes to ChatGPT4 have made it easier to analyze large documents in addition to phasing out the use of plugins.

⁷ Looking through the larger document, one sees that the PEs are provided in sections that our pattern matching algorithm can use to extract the entire section that the key word search can help identify.

⁸ Advana is the DoD’s enterprise-wide, multi-domain data, analytics, and AI platform that provides military and civilian decision makers, analysts, and builders with unprecedented access to enterprise tools and capabilities—all in a scalable, reliable, and secure environment.

⁹ For additional information on GAMECHANGER, see DIA Public Affairs (2022).

¹⁰ An Associate Professor of Computer Science, from Carnegie Mellon University blog (Neubig, 2023), indicates that many of the issues identified here have been resolved by at least one LLM vendor.



Further Research and Capabilities

To address the questions specified in the Problem Statement section, future work on this research would address the following:

Integrate the NLP Python and Front-End: The proof-of-concept text extraction script leveraged well-formatted J-Books to extract important contextual information, such as cost data, and used other acronyms found in these sections to identify connections. We would demonstrate additional concepts to integrate the NLP scripting approach into the implemented front-end to populate the back-end database for additional analytical capabilities.

Expand Front-End and Back-End Functionality: The existing capabilities would be expanded to provide additional useful analytical processing.

Extend Processed Capabilities: Include processed J-Books across fiscal year boundaries with the addition of other resources, such as including the National Defense Authorization Act (NDAA) in the files processed.

Full-Context Extraction and Portfolio Reformatting: Expand this proof-of-concept text extraction script to leverage these well-formatted J-Books to extract other important contextual information, such as cost data. We would then demonstrate restricting the data into representative portfolios.

Inter-Service Synergy: Implement semantic network analysis to identify critical nodes and their relationships.¹¹ Analyze links between these nodes based on their proximity within the text.

Apply community detection algorithms like the Louvain or Leiden algorithms¹² to identify clusters of related content. This approach can reveal the synergy and coordination among the Army, Navy, and Air Force regarding associated programs. Figure 7 illustrates this type of analysis for a project to identify skills within the acquisition community.

¹¹ Regarding a definition of critical nodes, Lalou et al. (2018) defines them as “those the deletion of which disconnects the network according to some predefined connectivity metrics, such as: maximizing the number of connected components, minimizing pairwise connectivity in the network, minimizing the largest component size, etc.”

¹² The benefits of the Leiden cluster identification algorithm are found in Tragg et al. (2019). The researchers have made their code available on a GitHub site.



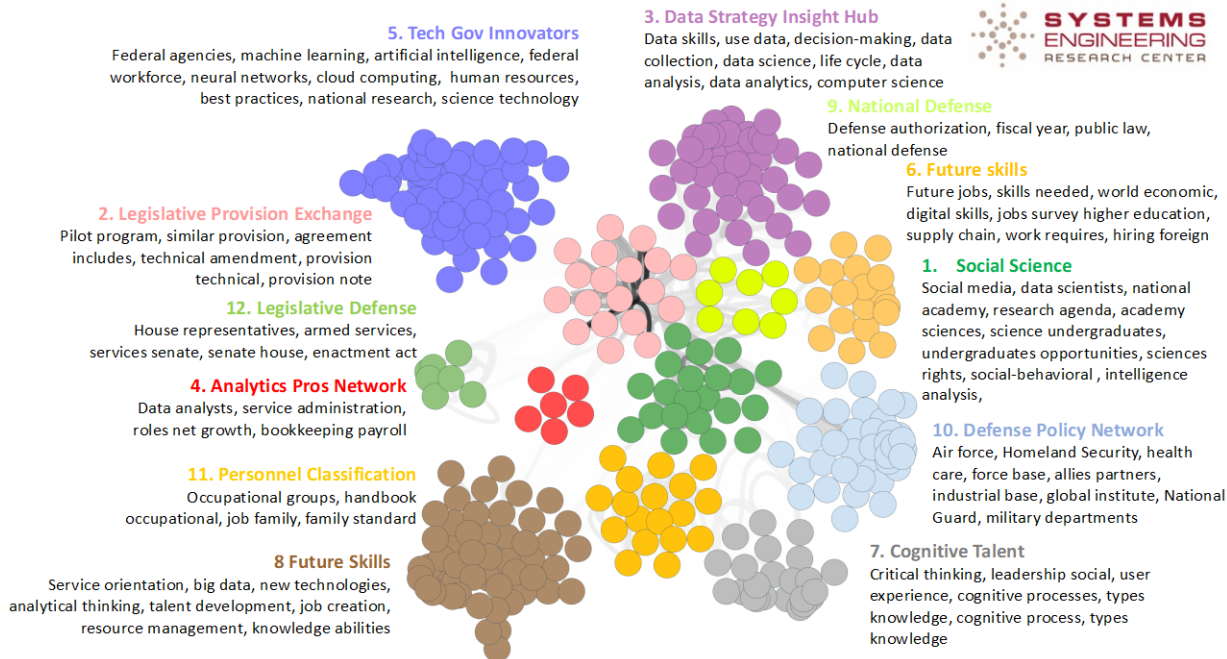


Figure 7. Network Analytics Description Based on Text Extraction

Technological Integration: Analyze text data to identify trends (trend analysis) and patterns in technological integration within the Army, Navy, and Air Force programs through time and programs. Investigate the adoption of emerging technologies across different programs and how these technologies intersect across various branches.

Resource Allocation: Develop models to simulate and analyze resource allocation scenarios within each branch and assess their impact on program development and implementation. Investigate the trade-offs and challenges of different resource allocation strategies.

Operational Implications: Analyze the text data to understand the operational implications of programs like JADC2 on the specific missions and functions of the Army, Navy, and Air Force. Investigate how these branches adapt to accommodate changes and ensure the success of their missions.

Temporal Association and Differencing: Analyze the data across multiple years to identify trends in changes to the budget values and an ability to analyze the specific syntactic meaning of the wording associated with key areas from year to year.

References

- Buettner, D. J., Ramirez-Marquez, J. E., & Anton, P. S. (forthcoming). *ChatGPT budget justification book program element summarization: A proof-of-concept demonstration*. Stevens Institute of Technology, Acquisition Innovation Research Center.
- DIA Public Affairs. (2022, Feb. 7). *Gamechanger: Where policy meets AI*. Defense Intelligence Agency. <https://www.dia.mil/News-Features/Articles/Article-View/Article/2926343/gamechanger-where-policy-meets-ai/>
- Lalou, M., Tahraoui, M. A., & Kheddouci, H. (2018). The critical node detection problem in networks: A survey. *Computer Science Review*, 92–117.



Neubig, G. (2023). *C3 generative AI: Getting the most out of enterprise data*. C3.ai. <https://c3.ai/c3-generative-ai-getting-the-most-out-of-enterprise-data/>

Ramirez-Marquez, J. E., Buettner, D. J., Anton, P. S., Gorman, J., & Patel, A. (forthcoming). *DoD budget justification book keyword search association: A proof-of-concept demonstration*. Stevens Institute of Technology, Acquisition Innovation Research Center.

Tragg, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 1–12.



Advanced Technologies to Enable Optimized Maintenance Processes in Extreme Conditions: Machine Learning, Additive Manufacturing, and Cloud Technology

Dr. Kasey C. Miller— is the Naval Air Systems Command (NAVAIR) Project Manager for Aviation Product Lifecycle Management (AvPLM). Kasey is a retired Marine Major, Military Advisor, and Communications Information Systems Officer. Kasey graduated from Southern Illinois University with a BS in electronic systems technologies and from the Naval Postgraduate School (NPS) with an MS in computer science and a PhD in information sciences. Miller is also a graduate of the Naval War College. Miller has taught in the Computer Science Department at the U.S. Naval Academy. Miller enjoys spending time with his family and unplugging.

Abstract

The way routine maintenance is conducted is not an optimal way to handle maintenance in extreme battlefield conditions. This is a common maintenance problem across various domains, such as repairing battle damage to aircraft or ships without access to a port or depot. The extreme conditions context can also include repairing the Alaska pipeline in the extreme cold, or handling repairs during COVID-19. The researcher examined how modern technology can optimize productivity and reduce the cycle time of the extreme maintenance process. The results of this research found that three emerging technologies: additive manufacturing, cloud in a box, and machine learning (ML), could improve process value, save labor costs, and reduce cycle time. ML had the most significant impact on improving productivity and cycle time. When all technologies were utilized together, productivity and cycle time improvement were more significant and consistent. The research accounted for the riskiness of these technologies, which is necessary to accurately forecast the value added for this extreme maintenance process context. This research is vital because getting correct valued repairs done quickly for the Department of Defense can make the difference between winning and losing a conflict.

Introduction

Parts of this introduction were previously published by Springer Nature in *HCI for Cybersecurity, Privacy, and Trust* (Miller & Mun, 2023). Extreme maintenance conditions, such as during combat operations or personnel shortages as during the COVID pandemic, create many unique repair and maintenance challenges. These challenges include the battle front line availability of technical data or specifications to make the repairs, the lack of parts, and decision support aids to assist with transforming repair data into information and knowledge that lead to making timely decisions. The lack of timely maintenance information leads to uninformed and suboptimal maintenance decisions, especially when edge networks are data-limited, that increases the risk to a given complex repair and the employees making the repairs. For example, the naval enterprise system architecture (ground and aviation) has limited technical data in these edge networks. The communication limitations of the edge networks have led to interaction-based failures, resulting in inefficient information exchanges among maintenance-related systems and repair personnel.

Innovative maintenance approaches with modern IT, can potentially overcome these extreme maintenance case problems. Information systems are needed to provide mechanisms that will enable leadership to make data-driven resource decisions at all levels of the maintenance process by using locally available data derived by leveraging this new IT. With access to the required technical information, and without having to do all the current manual workarounds to get the data needed to make precise repair decisions, the repair team can actually do the repairs in a timely and efficient manner. The problems arise when the



maintenance technicians are forced to do workarounds because they do not have the required technical information available locally.

This extreme maintenance problem requires innovative use of U.S. Naval IT resources to foster the potential of increased process productivity as well as reductions in process cycle times for repair. The three IT artifacts examined in this study (additive manufacturing [AM], machine learning [ML], and cloud in a box [CIB]) should provide the kinds of mechanisms that will enable leadership to make more well-informed IT investment decisions resulting from the innovative leveraging of these new IT technologies.

When deploying new IT solutions in organizations, it is hard for information scientists to gather the required data on IT decisions to determine the impact on employees (Leonard-Barton & Kraus, 2014). If the local maintenance personnel deliver innovative repair ideas, aided by modern IT artifacts that help improve process productivity, these ideas can be embedded in ML, potentially resulting in more optimized processes that add value to their organization. Currently, routine aviation maintenance knowledge (e.g., at the depot level) used to optimize processes is not readily available for potential use in extreme maintenance conditions, and the results of that routine maintenance knowledge are not passed from one generation of maintainers to the next.

Purpose Statement

The purpose of this research is to test the value added of three modern information technology artifacts (i.e., AM, ML, and CIB) to optimize process productivity and cycle time for extreme maintenance conditions. The current research study extends the use of process optimization theory (Castillo, 2011) to include the effect of modern information technology on extreme maintenance process productivity and cycle time. This research is essential because there is a gap in the process optimization literature with regard to optimizing maintenance processes with modern information technology in the context of extreme maintenance. The current research is important because failure to make correct repairs to battle-damaged equipment can make the difference between winning and losing a conflict.

Research Goals

One of the research goals is to make a theoretical contribution to the economics of information technology (EOIT) domain by testing the effects of three new IT artifacts (AM, CIB, and ML) to provide process optimization options that would potentially increase process productivity (i.e., return on investment [ROI]) and reduce process cycle time for extreme maintenance processes. The results of this research should provide greater confidence in decision-makers' IT investment predictions based on information from process optimization model forecasts. The Department of the Navy (DoN) must improve its extreme maintenance processes to maintain readiness in battle conditions. Business process reengineering (BPR) techniques can be used to model the effects of AM, CIB, and ML on productivity and cycle time (Miller & Mun, 2023).

Thus, I propose an information sciences-based investigation of how using modern information technology in extreme maintenance conditions can extend the existing EOIT optimization-focused theories by testing new IT artifacts (AM, CIB, and ML) in a new but pervasive context. For example, AM can provide maintenance technicians with part-generation options that should accelerate the repair cycle. The CIB can house technical information that would feed ML technology and can work in a network disconnected environment (e.g., extreme maintenance at the battle front). The ML IT option under review in this study involves three dimensions: algorithms, systems, and people (Stoica et al., 2017). In this context, ML focuses on accessing technical data (e.g., using the CIB technology), and the ML algorithm learns based on performance feedback from the maintenance personnel.



The types of ML algorithms proposed in the current research are commonly utilized in bioinformatics (Frazier, 2022). These kinds of ML algorithms are used to improve the predictions of the effects of various variables that “repair” biological systems. The results from this domain of research on the use of ML will form the basis for the parameter expectations of the performance of ML to aid repair and maintenance decision-making. This kind of ML should provide extreme maintenance technicians with information to adapt and improve their repair decisions, which include, in particular for the current study, repair evaluation, and parts ordering decisions.

The current research utilizes integrated risk management (IRM) to forecast the effects of using the three IT artifacts to optimize extreme maintenance subprocesses that have been optimized using BPR techniques. By doing so, the current study will expand the scope of EOIT optimization theories through the use of robust forecasting techniques in the context of extreme maintenance decision-making.

Literature Review

This study uses naval aircraft maintenance in particular due to the complexity of the problem. The aircraft battle damage repair (BDR) requires specialized repair and damage analysis, skills, and tools from depot-level maintenance organizations in order to perform complex equipment structure modifications or to perform routine or urgent equipment and system repairs. The baseline model in the current study is derived from the existing depot-level maintenance processes as verified by subject matter experts (SMEs) who perform those depot and extreme maintenance functions during wartime operations. The Forward Deployed Combat Repair (FDCR) teams must be highly mobile and able to operate with very limited communication reach back to the depot resources and repair information. The logistical and maintenance constraints in extreme maintenance conditions (e.g., wartime field theater) will require the U.S. Navy to deploy civilian technicians forward to use new, more timely, and efficient processes by leveraging emerging technologies. This kind of maintenance research has a very high priority, as witnessed by the current efforts that are underway with Navy research teams who are studying the battlefield tactics of the Ukrainian military, including maintaining equipment in extreme battlefield conditions (NPS Information Sciences PhD Seminar Series, Oct 2023).

Baseline process models for extreme maintenance have not been documented previously. BPR optimization techniques require a baseline process model to inform and compare As-Is baseline process performance to To-Be forecasts regarding decisions about how to best utilize IT to optimize core processes (Hammer, 1990; Hammer & Champy, 1993; Housel & Bell 2001). Without such BPR models it is very difficult to justify investment in modern IT options that are designed to optimize processes, especially for extreme maintenance process optimizations that are urgently needed in the U.S. military. For example, if we want to test the potential use of AM, CIB, and ML to optimize the extreme maintenance repair process, we must have an As-Is baseline model to compare to To-Be forecasted improvements. The quantitative methods and models presented in this research will contribute to predicting the impact of modern IT artifacts used as process optimization options in the context of extreme maintenance processes. If the FDCR teams have these three technologies in place, and the IT technologies perform as expected, then the extreme maintenance process cycle time and process productivity performance should show improved optimization.

The current research makes theoretical contributions to information sciences through EOIT by gauging the ability of new IT technology to impact productivity and cycle time in extreme maintenance conditions. The economic theories of EOIT consider the effects of introducing IT on corporate productivity (Goldfarb & Tucker, 2019; Shapiro & Varian, 1999).



Further, in EOIT theory, researchers have hypothesized about the effects of these IT inputs at the process level (summarized in chapter three of Housel and Bell [2001]) on a firm's productivity. The theories ultimately rely on organizational accounting data to test their assertions empirically (Brynjolfsson & MacAfee, 2014; Elliot, 1992; Pavlou et al., 2005). Hitt et al. (1994) framed their research using EOIT and concluded that information technology positively affects an organization's productivity.

This research also seeks to extend process optimization theory (Castillo, 2011) to extreme maintenance processes. In process optimization, value added can be calculated at the subprocess level (Housel & Kanevsky, 1995). In extreme maintenance, the overall core repair process can be decomposed into its subprocesses. Although the outputs of the subprocesses are different, they can be compared by converting them to common units using the knowledge value added (KVA) theory.

These new IT aided models can potentially assist decision-makers by speeding up the data-to-decision (D2D) times and reducing risk (e.g., aircraft downtime). The current research results should be useful in extending EOIT theory by demonstrating how these IT artifacts can potentially be used to speed up the D2D times for repair decisions and how it might lead to overall increases in extreme maintenance process productivity. The results of the current study should help address theoretical gaps in the EOIT research on process optimization by the potential application of process modeling techniques that focus on the use of modern IT artifacts in the context of extreme maintenance requirements.

When applied early in the redesign of processes by modeling the impact of modern IT on process productivity and cycle time, the current study methods can lead to increased IT investment portfolio optimization decision-making within the context of real operations. The IT investment portfolio optimization techniques used in the current study provide a way to generate hypotheses similar to those of a study. Albert and Hayes (2002) found that hypothesis generation efforts should be incorporated early in the acquisition process and tested further with field experiments. Extending prior maritime research by Mun and Housel (2010), the current study will use Monte Carlo simulation with real options. This research addresses the gap in assessing the value of these new IT technologies in process optimization for extreme maintenance conditions.

Analysis

The current study explores several extreme maintenance use cases via modeling and simulation techniques (i.e., the current As-Is approach with the forecasted To-Be approach using new IT). This research is run from the perspective of a Leibnizian (Analytical-Deductive) inquiring system in which the guarantor of the knowledge claims is the self-evidence of the inputs and the deductive soundness of the operations. The validity of this research was established through a clear explanation of the input selection reasoning, a detailed explication of all derived analytical expressions, and a comparison between simulation results and the theoretical predictions of the derived analytical expressions. Complex data analytics packages were used to analyze the data for statistical insight and to process thousands of trial runs on the data and the emerging technologies to provide a complete view of the problem.

A comprehensive view of the problem within extreme maintenance is lacking in that the emerging technology is examined individually and not holistically. There is an absolute necessity to use emerging technology (AM, CIB, and ML) more efficiently within naval aviation maintenance-based decisions. That is why the final model engages all the technologies together (AM + CIB + ML) in the appropriate subprocess. The technical data sets can be challenging to acquire and comprehend. The magnitude of these specialized data sets offers analysis



complexities within an extreme maintenance realm that is large, distributed, and varies from mission to mission.

- **Data Overview**

The data analytics were based on current input from SMEs in extreme maintenance conditions. The analysis was conducted using the data from the surveys discussed in the previous chapter. Field experiments informed the surveys and cost data of the labor by the technicians performing the repairs and managers of those technicians. The surveys were completed by individuals familiar with the current As-Is extreme maintenance process on land and in maritime situations. In either case, the extreme maintenance constraints were applied to the current As-Is process, and forecasts were built into the models. Further, the To-Be models for AM, CIB, and ML were informed by experts in those technologies and the extreme maintenance process. The consistency of different sets of observations that measure the same factors was tested using statistical methods during the data exploration. Further, the correlation between variables of a survey with a pairwise comparison between two variables can be linear or nonlinear and either positive or negative. These linear coefficients are often insufficient and require other tests that check the data points across both the columns and rows for data consistency and reliability (e.g., the intraclass correlation coefficient [ICC] test).

The forecasted events for these models are targeted at three years. Using the technology to forecast further is possible, but future events behaving or occurring in the way expected may have greater volatility. The models can be extended past three years of the study but may require updating of the process maps and data parameters to maintain accuracy and precision. The reliability of the models refers to the repeatability of findings. If the study were repeated, would it yield the same results? If the measurement results are consistent and if the experiment is valid, then the data is considered to be reliable. This section explains the data analysis so another researcher can produce the same stable and consistent results as this study. While the validity of the models refers to how well a test measures what it is purported to measure, validity is more related to how strong the hypothesis outcomes are. It answers the question, are we right? Internal and external validity are tested with multivariate models such as regression and econometrics.

One assumption in this study is the limited data over an extended period. The extreme maintenance conditions for aviation in the modern era are quite new, especially when considering modern weapon systems (fifth-generation aircraft and unmanned aircraft). The processes and technologies in this research are mature but under-documented and mainly untested on a large scale. The data gathered was based on a year of field experiments with an organization that often conducts sea and land repairs. This assumption may affect the generalizability, as not all organizations perform extreme maritime maintenance. Also, the data is collected from military and civilian employees; not all organizations have this blend of employees. Additionally, the U.S. Navy is a private, not-for-profit organization that, of course, may differ slightly from a for-profit public company, yet productivity and cycle time are still driving factors in both non-profit and for-profit organizations. Lastly, it should be noted that more data collected over a more significant period would increase the accuracy and precision of the models.

- **Exploratory Data Analysis**

This subsection provides an exploratory data analysis of the data collected to include statistical tests described in the Analysis section. The variables are reviewed, and insights that will later be used in simulations are generated as parameters and settings for those models. The first variable explored is the subprocess complexity. As discussed earlier, KVA is based on complexity theory and information theory, which is essential to understanding which



subprocesses engage a more significant part of the workforce's time. Further, the learning time for a subprocess is correlated with the complexity of that subprocess. The longer it takes to learn a subprocess, the more complex that subprocess. Table 1 displays the rank order of complexity for the maintenance subprocesses. It shows that the repair subprocess requires the highest learning time and is the most complex sub-process in extreme maintenance and that for most subprocesses, learning time is not as substantial as it is for the repair process. The second most complex subprocess is the depot repair decision, or whether the repair can be completed on-site or needs to be conducted in a higher echelon of maintenance with more access to tools, labor, and infrastructure. Rank Order is a more accurate measure of complexity with a ratio scale than adjusted Rank Order with an ordinal scale.

Table 1. Extreme Maintenance Subprocess Complexity and Learning Time

Sub-Process #	Sub-Process	Rank Order (in complexity)	Rank Order Adjusted	Learning Time (hours)
1	Maintenance Request	2.91	1	5.74
2	Depot Repair Decision	4.45	6	21.35
3	Maintenance Induction	3.73	4	6.34
4	Part Inventory	3.36	2	5.84
5	Repair	5.55	7	41.10
6	Inspection	4.18	5	13.48
7	End Item Delivery	3.45	3	6.14
Total LT				99.98
Correl (RO & LT)				0.95

Learning time is the time someone needs to learn how to perform a particular set of tasks but not the amount of time to actually perform those tasks. The descriptive statistics for the learning time based on the surveys are compiled across the seven subprocesses. The range of learning time is approximately 35 hours, with a mean across the subprocesses of 14 hours. The minimum learning time is around 6 hours, with the maximum learning time being 41 hours. This data review provides parameters for the To-Be models and the four moments. We can also set up a basic statistical test based on the information listed in Table 2 for standard deviation and mean. Data skewness is greater than one, resulting in a positive skew of the distribution. Lastly, the learning time fourth moment or a Kurtosis of 2.9 means that the distribution is more peaked and has fatter tails than normal.



Table 2. Learning Time Descriptive Statistics

Learning Time Descriptive Statistics Summary Statistics	
Sub-Processes	7
Arithmetic Mean	14.28429
Geometric Mean	10.63972
Trimmed Mean	14.28429
SE Arithmetic Mean	4.98427
Lower CI Mean	4.31574
Upper CI Mean	24.25283
Median	6.34
Minimum	5.74
Maximum	41.1
Range	35.36
Stdev (Sample)	13.18715
Stdev (Population)	12.20893
Lower CI Stdev	9.10304
Upper CI Stdev	25.25902
Variance (Sample)	173.90093
Variance (Population)	149.05794
Coef of Variability	0.92319
First Quartile (Q1)	5.99
Third Quartile (Q3)	17.415
Inter-Quartile Range	11.425
Skewness	1.7671
Kurtosis	2.90775

The descriptive statistics just described are visualized with the Box and Whisker Plot shown in Figure 1 to give a spatial visual of the descriptive data. The figure further shows the positive skew of the learning time and average time to complete per subprocess. The learning time and average time to complete is skewed based on the repair and repair decision subprocesses. The X-axis in this chart has no meaning.



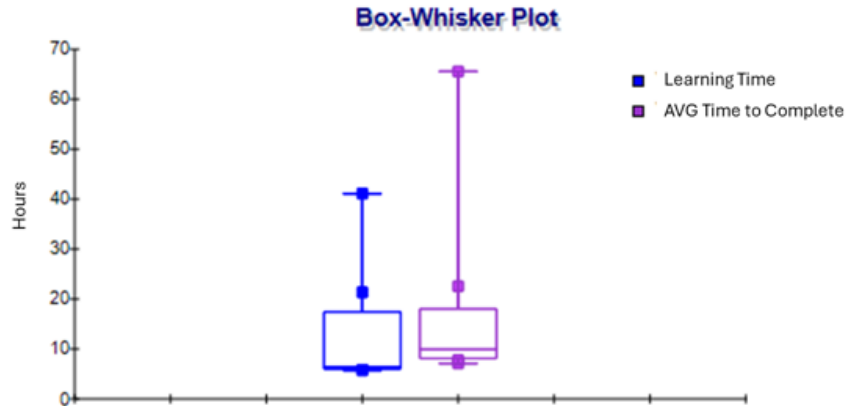


Figure 1. Learning Time and AVG Time to Complete by Sub-Process

The As-Is Expected Project Schedule shown in Figure 2 further shows what subprocesses impact the extreme maintenance most. Furthermore, the Tornado Analysis shows that Repairs, Field Repair Evaluation, and Inspection are the subprocesses that should be targeted for new technologies and process optimization. The repair process and the field evaluation process have the most impact on the overall extreme maintenance process. The delivery of the repaired aircraft and maintenance request subprocesses have the most negligible impact on productivity and cycle time. As with most project management, spending effort on the bottleneck subprocess offers the most room for productivity and cycle time improvement. If time permits, a focus on inspection of the aircraft post repair, maintenance induction, and part inventory subprocesses will be of value because the impact on cycle time and productivity may be minimal in terms of the days it takes to return the item to service.

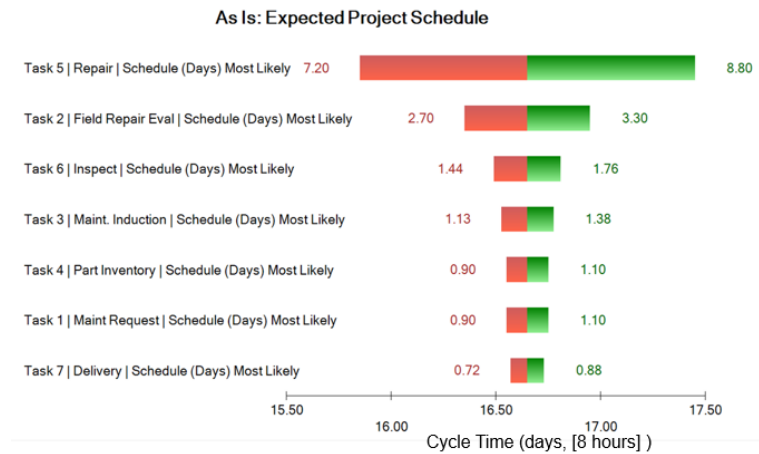


Figure 2. The As Is Project Schedule for Subprocesses

The investment portfolio for the five models (As-Is, AM, CIB, ML, and AM + CIB + ML) is displayed in Figure 3 as a new technology portfolio. This data forecasts the baseline As Is process with the new To-Be processes (i.e., AM, CIB, ML, and AM + CIB + ML). The investment portfolio demonstrates that the new technology reduces cost and schedule, which is vital in project management. The technology that offers the most benefit to the organization is ML. The Y Axis is the number of days expected for the repair, while the X-axis is the cost of the repairs.



So, the technology on the lower left corner of the diagram is beneficial to the organization. For example, ML is completed a day and a half faster and about \$8,000 cheaper over two weeks. In contrast, CIB, followed by AM, also offers gains over the As-Is extreme maintenance process but not to the degree that ML does.

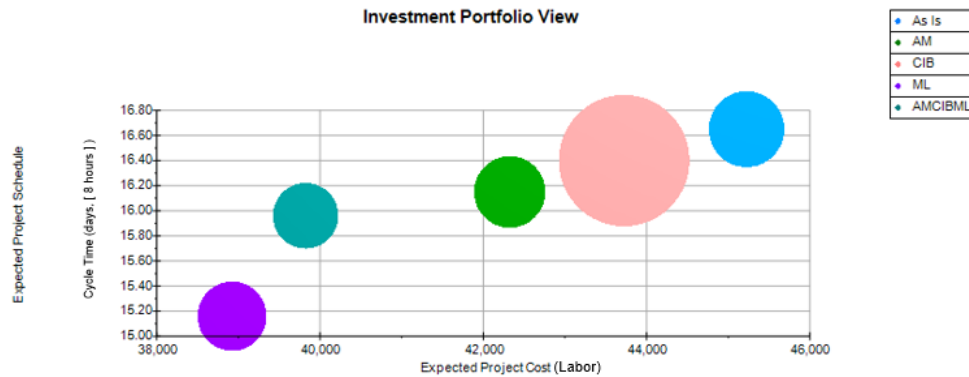


Figure 3. New Technology Investment Portfolio

Based on Monte Carlo simulation, Figure 4 shows the emerging technology’s probability density function (PDFs) and the As-Is process with the expected cost. PDFs are a statistical measure used to gauge the likelihood that an investment will have returns that fall within a range of values and indicate the risks involved. The PDFs in Figure 4 are plotted on a graph that resembles a bell curve, with the data lying below the curve. Also, the skewed angle at either end indicates greater/lesser risk or reward. The wider the curve, the greater the range of possible values. The As-Is process and CIB offer the greatest range and higher risk. In contrast, ML and all the technologies combined represent less variance, less risk, and a higher reward.

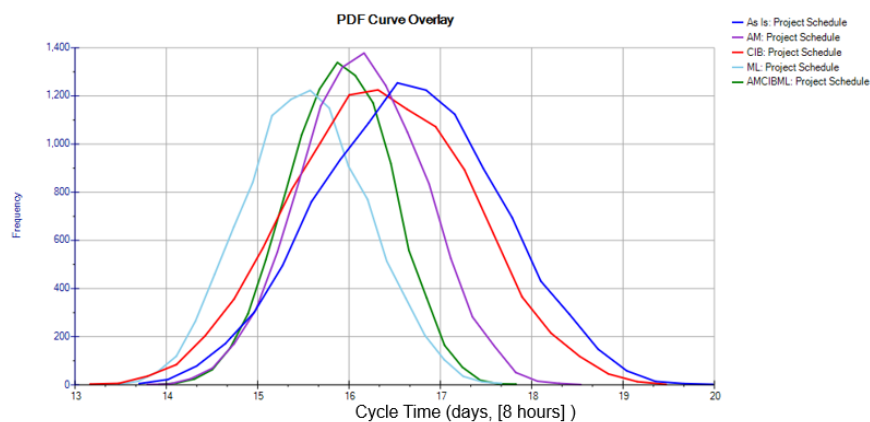


Figure 4. Predicted Schedule Saving of the Emerging Technology

a. HYPOTHESES TESTS

The hypotheses test evaluates the use of new technology against the As-Is case. The eight hypotheses are as follows:



Hypothesis 1: ML-informed repair decisions will lead to improved extreme maintenance process cycle time compared to current extreme maintenance repair prediction decision methods.

Hypothesis 2: ML effects the extreme maintenance process productivity to improve.

Hypothesis 3: Using AM improves extreme maintenance process cycle time compared to traditional supply chain parts acquisition methods.

Hypothesis 4: AM improves extreme maintenance process productivity compared to traditional supply chain parts acquisition methods.

Hypothesis 5: CIB technology improves extreme maintenance process cycle time compared to traditional reach-back methods.

Hypothesis 6: CIB technology improves extreme maintenance process productivity compared to traditional reach-back methods.

Hypothesis 7: AM + CIB + ML technology improves extreme maintenance process cycle time compared to traditional methods.

Hypothesis 8: AM + CIB + ML improves extreme maintenance process productivity compared to traditional methods.

The As-Is, AM, CIB, ML, and AM + CIB + ML cases provide three-point estimates for the minimum, the most likely, and maximum estimates for cycle time. As shown in Figure 4, these point estimates follow a triangular distribution. The cycle time in days is the X-axis, while the rate of change is the Y-axis in Figure 5.

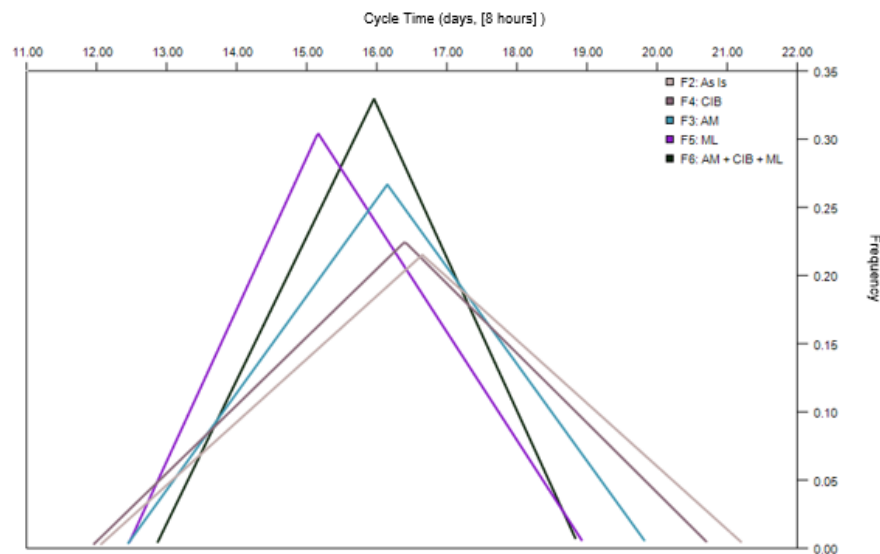


Figure 5. Cycle Time Triangle Distribution

The parameters are then inputted into the Risk Simulator software to generate data based on the As-Is and four To-Be process models and are fully simulated. As the analysis section discusses, these parameters reflect current survey data and SME forecasting input. The hypotheses test data are outputs of simulations run with a thousand data points for each of the



five models using the risk simulator. The data described are shown for the To-Be AM+CIB+ML model in Figure 6, and the simulation results obtained are utilized in the hypotheses tests.

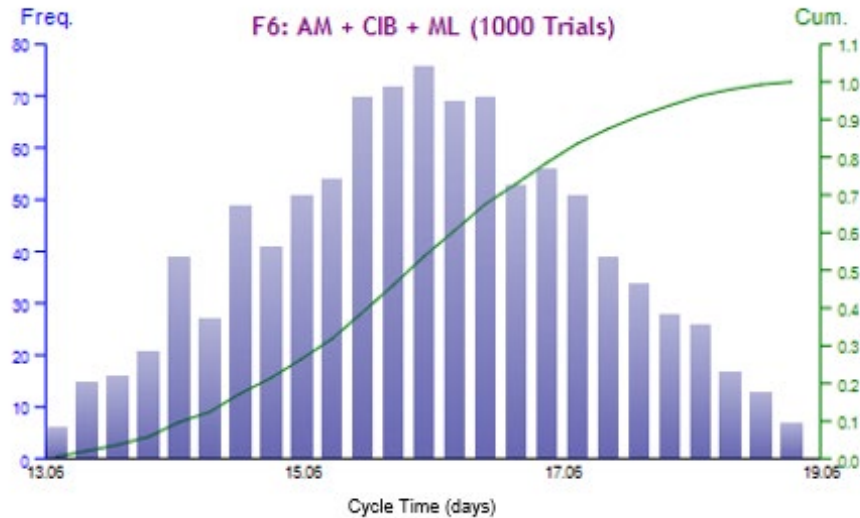


Figure 6. Hypothesis Distribution Simulation Data

The hypotheses tests are parametric two-variable t-tests independent with equal variance. They are not dependent; for example, if the technician fixes an aircraft, runs the test, and then fixes the same aircraft again, that does not fit the conduct of this study. In this study, the technician could be fixing different types, models, or series of aircraft. Since the technicians are repairing different aircraft, the overall aircraft repair process has similar situations. So, similar situations have equal variance, which means the class of aircraft is being repaired by similarly trained individuals. Therefore, the hypotheses test utilizes equal variance. Furthermore, this study is not testing between aircraft, surface vessels, votes, and submarines. This is why we use a parametric two-variable t-test independent with equal variance.

The hypotheses for cycle time are directional hypotheses. Hypothesis one states that ML improves cycle time compared to traditional prediction methods; hypothesis three states AM increases cycle time compared to traditional supply chain parts acquisition methods; and hypothesis five states CIB technology improves cycle time compared to traditional reach-back methods; hypothesis seven states AM + CIB + ML technology improves extreme maintenance process cycle time compared to traditional methods. Table 3 shows a directional main effects hypothesis. The CIB, hypothesis five, is that p-values are less definitive but can still reject the null hypotheses. Finally, hypothesis seven is that AM + CIB + ML technology improves cycle time compared to traditional reach-back methods, which is statistically significant.

The hypotheses tests conducted are multiple T-tests with the As-Is model compared to the To-Be Model and ANOVAs. The simulation data is broken up into groups of hundreds of data points for the AM, ML, CIB, and the three technologies combined. The simulation data was generated with a random seed of one and was analyzed with a pairwise T-test. The data generator allows the simulation of all four To-Be processes. Table 3 shows that AM, ML, and AM+ML+CIB all have an effect on cycle time, while CIB effects are enough to reject the null hypotheses 50% of the time based on the significance level of 0.05. Using AM technology, the null hypothesis can be rejected 70% of the time at the significant level of 0.05. Once ML technology is added, it is 100% of the time at the significant level of 0.05 and 0.01.



Table 3. Cycle Time Hypotheses Tests One Tail

Hypotheses T-Test (Right Tailed, One-Tail) Results (P-Values)				
Sample	AM	CIB	ML	AM+CIB+ML
1-100	0.202949	0.035302	0.000001	0.001389
101-200	0.003927	0.088065	0.000000	0.000096
201-300	0.017506	0.034693	0.000000	0.002298
301-400	0.002766	0.329675	0.000007	0.000085
401-500	0.276834	0.398832	0.000003	0.007049
501-600	0.003153	0.139105	0.000000	0.000107
601-700	0.007370	0.029176	0.000000	0.000009
701-800	0.031494	0.425067	0.000051	0.002155
801-900	0.145179	0.006349	0.000001	0.000153
901-1000	0.004225	0.007521	0.000007	0.000089
Significant ($\alpha=.05$)	70%	50%	100%	100%
Significant ($\alpha=.01$)	50%	20%	100%	100%

An ANOVA was conducted to look across all the independent variables at the same time. A single-factor, multiple-treatment ANOVA was chosen because each factor is applied to the same extreme maintenance repair process. Table 4 demonstrates that one or more technologies have a statistically significant effect at Alpha 1% on at least one of the levels.



Table 4. Cycle Time ANOVA Single Factor Multiple Treatment

Hypotheses Test with ANOVA Results (P-Values)		
Sample	As-Is, AM, CIB	As-Is, AM, CIB, ML, AM+CIB+ML
1-100	0.1770	0.0000
101-200	0.0252	0.0000
201-300	0.0752	0.0000
301-400	0.0172	0.0000
401-500	0.8502	0.0000
501-600	0.0221	0.0000
601-700	0.0396	0.0000
701-800	0.1148	0.0000
801-900	0.0271	0.0000
901-1000	0.0116	0.0000
Significant ($\alpha=.05$)	60%	100%
Significant ($\alpha=.01$)	0%	100%

The power analysis for these tests is post hoc, with two variables, with 10 samples of 100, for the T-test (Figure 7). The Sigma of group one is 16.5346, and the Sigma of group two is 1.634913 with a hundred sample size with two tails and an alpha of 0.05 with minor effects, so the power is only about 12%. Having 1,000 data points does bring the power up to about 74.94%.



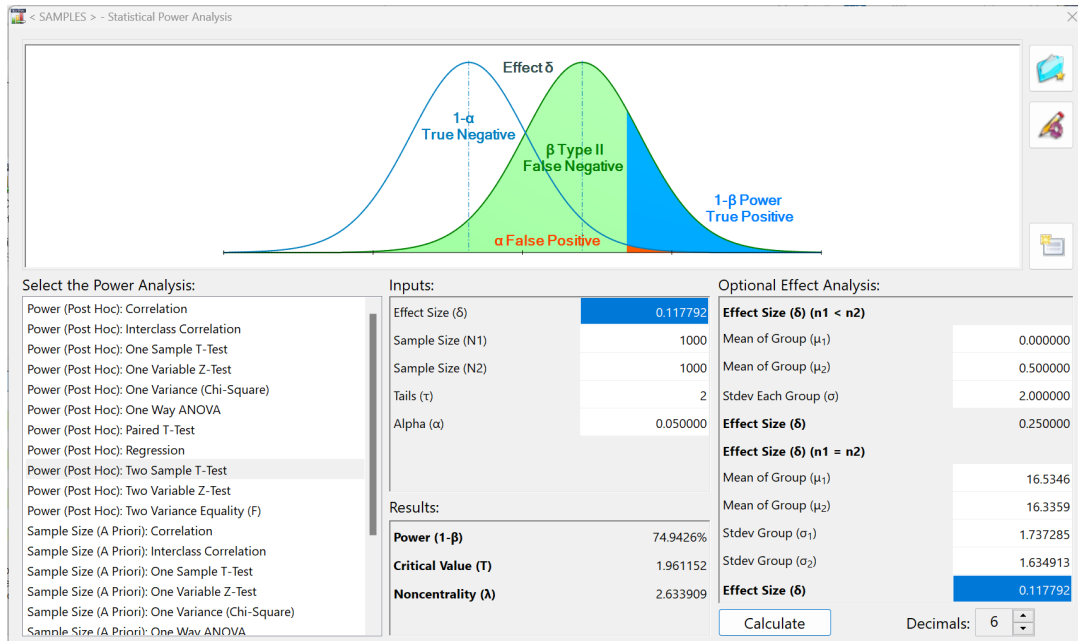


Figure 7. Power Analysis for Hypotheses Tests

At the corporate level, ROI and ROK are productivity ratios in accounting, as seen in Table 5. ROI is based on revenue in the extreme maintenance case for a non-profit organization, like the military, in which there is no revenue. The fact that there is no revenue is not an issue for ROI, as market comparables can substitute for revenue.

Table 5. Productivity As-Is and To-Be ROI

Sub-Process #	Sub-Process	As-Is ROK	As-Is ROI	To-Be ROI AM	To-Be ROI CIB	To-Be ROI ML	To-Be ROI AM, CIB & ML
1	Maintenance Request	77.23%	10.72%	10.72%	10.72%	11.00%	11.00%
2	Depot Repair Decision	95.27%	41.99%	41.99%	41.99%	42.50%	44.00%
3	Maintenance Induction	66.39%	-3.82%	-3.82%	-1.00%	0.00%	1.00%
4	Part Inventory	70.93%	3.75%	5.00%	4.00%	4.00%	4.50%
5	Repair	62.82%	-5.95%	-2.00%	-2.00%	1.50%	3.50%
6	Inspection	100.41%	49.80%	49.80%	50.50%	51.00%	50.50%
7	End Item Delivery	88.57%	29.54%	29.54%	31.00%	30.00%	30.50%
ROI Totals			126.03%	131.23%	135.21%	140.00%	145.00%

Table 6 shows the results of the productivity hypotheses. Hypothesis 2: ML effects process productivity to improve, and hypothesis 4: AM increases productivity compared to traditional supply chain parts acquisition methods. Additionally, hypothesis 6 states that CIB improves productivity compared to traditional reach-back methods. Finally, hypothesis 8 states that AM + CIB + ML improves productivity compared to traditional reach-back methods.



Table 6. Productivity Hypothesis Testing

Hypotheses T-Test (Left Tailed, One-Tail) Results (P-Values)				
Sample	AM	CIB	ML	AM+CIB+ML
1-100	0.000100	0.197900	0.002107	0.000778
101-200	0.033700	0.014390	0.000686	0.001050
201-300	0.004625	0.145225	0.000147	0.000008
301-400	0.014260	0.004070	0.000010	0.001569
401-500	0.000110	0.009500	0.000564	0.000191
501-600	0.056720	0.032137	0.001158	0.006239
601-700	0.011170	0.235800	0.000693	0.056232
701-800	0.004590	0.002750	0.000337	0.002207
801-900	0.000062	0.472000	0.000653	0.021535
901-1000	0.016810	0.387479	0.000003	0.012274
Significant ($\alpha=.05$)	90%	50%	100%	90%
Significant ($\alpha=.01$)	50%	30%	100%	70%

The productivity hypotheses are evaluated using the same methodology as the cycle time hypotheses. The forecasting parameter estimates are derived from the literature review and SME input. The ANOVA results are shown in Table 7.

Table 7. Productivity ANOVA Results

Hypotheses Test with ANOVA Results (P-Values)		
Sample	As-Is, AM, CIB	As-Is, AM, CIB, ML, ALL
1-100	0.0047	0.0069
101-200	0.1032	0.0046
201-300	0.0854	0.0001
301-400	0.0128	0.0003
401-500	0.0035	0.0015
501-600	0.1365	0.0125
601-700	0.1709	0.0204
701-800	0.0071	0.0037
801-900	0.0013	0.0161
901-1000	0.1637	0.0005
Significant ($\alpha=.05$)	50%	100%
Significant ($\alpha=.01$)	40%	70%



Conclusion

In various extreme conditions such as aircraft or ship battle damage repair, extreme cold Alaska pipeline repair, and COVID-19 repair processes, the use of modern information technologies such as ML, AM, and CIB are not being leveraged to optimize productivity and reduce cycle time in these critical maintenance processes. The literature on process optimization does not address the use of modern technology for optimization in extreme maintenance conditions. Therefore, the purpose of this research was to estimate the value added by information technology to optimize process productivity and reduce cycle time for extreme maintenance processes. This research aimed to extend process optimization theory to include the effect of modern information technology in extreme maintenance conditions. It is critical in the DoD context because failure to repair battle-damaged equipment remotely (without access to the depot), correctly, efficiently, and quickly can make the difference between winning and losing a conflict.

Furthermore, extreme maintenance reach-back to the depot for resources or data is problematic, using existing repair processes and systems as the technician must assume they must operate independently. The current research demonstrated that the three technologies (AM, CIB, and ML) technologies potentially offer ways to significantly improve the ROI of the extreme maintenance process and reduce the cycle time of the process. AM alone will potentially decrease cycle time and increase productivity compared to traditional supply-chain parts-acquisition methods. CIB technology will potentially improve cycle time and productivity compared to traditional reach-back methods, in spite of its newness and potential performance volatility. The research clearly demonstrated that ML technology can also be used to improve cycle time and productivity compared to traditional extreme maintenance decision-making prediction methods. The extreme maintenance research findings are summarized in Table 8.

Table 8. Extreme Maintenance Findings Summary

Extreme Maintenance Finding Summary			
Technologies	Cycle Time (Schedule/Cost savings [Labor])	Productivity (Value)	Comments
AM	Moderate improvement	Significant improvement	Technology gains offer an immediate impact with little fielding challenges.
CIB	Slight improvement	Slight improvement	New technology, high volatility, enabler for AM & ML.
ML	Significant improvement	Significant improvement	Highest improvement of all technologies. Implementation might be a challenge due to data availability and extreme maintenance hosting environment.
AM + CIB + ML	Significant improvement	Significant improvement	Recommended option due to improvements in Cycle Time/Productivity, and complementary technologies that reduce risk to increase upside.

The results of this research clearly demonstrated that the three IT technologies have the potential to significantly improve the productivity and cycle time of an extreme maintenance



process. As such, this research extends the current EOIT and process optimization research areas to include this critical context. Further, this research extension can cover the extreme maintenance domain in for-profit (e.g., North Slope Oil extraction operations) and non-profit organizations (e.g., battlefronts without convenient reach back to a maintenance depot).

Research Limitations

One of the issues is that the potential emerging technologies (AM, CIB, ML) have not been tested in extreme maintenance conditions. This study does not test them, and the potential for these emerging technologies is being modeled economically. The dissertation proposes that investment decisions are based on modeling and simulating their value in extreme maintenance. The ML techniques are often subject to the inability to identify flaws and errors, and there are difficulty in identifying scope and reliability models.

Future Work

The real problem facing the U.S. Naval Mission is automating the fleet to include autonomous vessels. By 2045, the U.S. Navy is estimated to have 500 vessels, with at least 150 being autonomous (Tangredi & Goldorski, 2021). If we extrapolate to the aviation fleet, we can expect at least a third to be unmanned. These unmanned aerial systems (UASs) will need maintenance. This is a paradigm shift for extreme maintenance because repairs will not focus on human safety. Maintenance in the future will have more significant gains with new technological improvements, such as AM, ML, and CIB. New technologies that scale will be critical, that is, the AI/ML architecture explored within future Joint Task Force (JTF) extreme maintenance operations. The commanders can shape the battlespace by maintaining combat power and utilizing these system capabilities.

Additional analysis of the warfighting staff and AM, CIB, and ML can transform process optimization and ultimately enable decision-makers to manage extreme maintenance risk based on the data. Also, future work is needed to explore any weaknesses with CIB and address AM cyber vulnerabilities (i.e., data poisoning) in the extreme maintenance use case.

UAS assets' acceptable repair thresholds can change the level of acceptability for parts and repairs in general. As long as a UAS can accomplish its mission, a triumphant return of the asset to friendly territory might not be necessary. The secondary contribution of this research is the use of the methodologies of evaluating emerging EOIT and contextualizing extreme maintenance processes to refracture the existing RO approach to unmanned systems. Future research will take this research and continue to test and refine the model and conduct field experiments where possible. As discussed earlier—the more accurate the data, the better the forecast for the models.

Author Statement

The views expressed in this article are those of the author and do not reflect the official U.S. Navy policy or position of the Naval Air Systems Command, Department of Defense, or the U.S. Government. This is not a product of NAVAIR.

References

- Albert, D., & Hayes, R. (2002). *Code of best practice: Experimentation*. Office of the Assistant Secretary of Defense; Command and Control Research Program (CCRP). <https://apps.dtic.mil/sti/pdfs/ADA457917.pdf>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Co.
- Castillo, E. D. (2011). *Process optimization: A statistical approach*. Springer.



- Frazier, W. (2022). *Predictive maintenance using machine learning and existing data sources*. [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. <https://hdl.handle.net/10945/71062>
- Goldfarb, A., & Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, 57(1), 3–43. <https://doi.org/10.1257/jel.20171452>
- Hammer, M. (1990). Reengineering work: Don't automate, obliterate. *Harvard Business Review*, 1–8.
- Hammer, M., & Champy, J. (1993). *Reengineering the corporation: A manifesto for business revolution*. Zondervan.
- Hitt, L., Brynjolfsson, E., & Walsham, G. (1994). The three faces of IT value: Theory and evidence. *ICIS 1994 Proceedings*, 20. <https://aisel.aisnet.org/icis1994/69>
- Housel, T., & Bell, A. (2001). *Measuring and managing knowledge*. McGraw-Hill/Irwin.
- Housel, T., & Kanevsky, V. (1995). Reengineering business processes: A complexity theory approach to value added. *INFOR: Information Systems and Operational Research*, 33(4), 248–262. <https://doi.org/10.1080/03155986.1995.11732285>
- Leonard-Barton, D., & Kraus, W. (2014, August 1). Implementing new technology. *Harvard Business Review(United States)*, 63(6). <https://hbr.org/1985/11/implementing-new-technology>
- Miller, K., & Mun, J. (2023). Cyber technologies, machine learning, additive manufacturing, and cloud in the box to enable optimized maintenance processes in extreme conditions. In A. Moallem (Ed.), *HCI for cybersecurity, privacy and trust* (pp. 672–684). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35822-7_43
- Mun, J., & Housel, T. (2010). *A primer on applying Monte Carlo simulation, real options analysis, knowledge value added, forecasting, and portfolio optimization* (White Paper). NPS Acquisitions. <https://doi.org/10.21236/ada518628>
- Pavlou, P., Housel, T., Rodgers, W., & Jansen, E. (2005). Measuring the return on information technology: A knowledge-based approach for revenue allocation at the process and firm level. *Journal of Association of Information Systems*, 6(7), 199–266.
- Shapiro, C., & Varian, H. (1999). Information rules: A strategic guide to the network economy. *Journal of Economic Education*. <https://doi.org/10.2307/1183273>
- Stoica, I., Franklin, M., Jordan, A., Fox, A., Joseph, A., Mahoney, M., Katz, R., Patterson, D., & Shenker, S. (2017). *The Berkeley data analysis system (BDAS): An open source platform for big data analytics* (Technical Report AFRL-RI-RS-TR-2017-173).
- Tangredi, S., & Galdorisi, G. (2021). *AI at war: How big data, artificial intelligence, and machine learning are changing naval warfare*. Naval Institute Press.
- United Airlines. (n.d.). *Aircraft on ground (AOG) / Global emergency maintenance (GEM)*. United Technical Operations. Retrieved December 28, 2023, from <https://unitedtechops.com/aog-gem>.



Resourcing Innovation: Assessment Learning Model – Business Acumen

Dr. Bruce Rideout—is currently the Senior Advisor for the Program Management and Acquisition Group (APMG). In his former role as the Director of Continuous Process Improvement (CPI) supporting NAVAIR's Program Executive Office (PEO) for Common Systems (CS), he supported the Organizational Development and Strategic Initiatives for the executive and five program offices. He recently developed the business acumen assessment model to baseline 32 programs across the NAVAIR Command. In addition, he designed and executed the Fleet Readiness Centers (FRC) procurement transformation for the Command, developing and implementing standardized acquisition processes and establishing a common organizational structure at the FRCs. He received the Navy's Meritorious Civilian Service Award for this effort. Other strategic initiatives he assisted: Mission Aligned Organization (MAO) Transformation Core Team, NAVAIR's Knowledge Management System, and NAVAIR University's concept of operations.

Before his current government service, Rideout was a Vice President of Supply Chain Optimization for a private equity firm driving large-scale supply chain transformation and lean operations. He also supported the Federal Aviation Administration's (FAA) strategic programs as the Managing Director of a consulting firm. Rideout has worked in the energy business as a Corporate Vice President leading their business process improvement program and Program Leader for their generation fleet business transformation and infrastructure program leading process improvement in work management, equipment reliability, and enterprise asset management (EAM) systems. When he was at General Electric (GE), he served as a General Manager and chemical operations leader for its plastics business as well as a Six Sigma Master Black Belt driving business-wide improvements.

Rideout is a certified Lean Six Sigma Master Black Belt (MBB), Project Management Professional (PMP) as well as Defense Acquisition University (DAU) certification in Program Management, Science & Technology, and Engineering. He earned his bachelor's degree in mechanical engineering, cum laude, from the University of Florida, a master's degree in public administration from Valdosta State University, and a doctorate from Florida Institute of Technology in business administration. He was named the Graduate Student of the Year in 2021/22. Rideout served in the U. S. Marine Corps as an officer in the Artillery, Air and Naval Gunfire Company (ANGLICO) and commanded a Marine detachment on sea duty. He has completed executive programs at the Office of Personnel Management's (OPM) Federal Executive Institute and GE's Managerial Development Course. His community service roles include the Board of Directors, Chamber of Commerce (Hancock County), Credit Committee (Wyochem Credit Union), and City Commissioner for Planning and Zoning (Green River, WY).

Jennifer Chermansky—Ms. Chermansky's federal career spans over 26 years, which has shaped broad Department of Navy acquisition financial management knowledge. As the Director of the NAVAIR Program & Business Analysis Department (PBAD) since 2020, she leads all the manpower, processes, and tools associated with the Business Financial Managers (BFMs) of NAVAIR headquarters. Her responsibilities include the professional development of BFMs, establishment, sustainment, and improvement of financial management processes within the Planning Programming Budgeting and Execution (PPBE) construct, and refinement of business and financial tools utilized by the NAVAIR enterprise financial management community. She is an enabling leader of 900+ BFM professionals, assigned to programmatically and functionally diverse business units and NAVAIR commands, including headquarters and a nationwide network of field activities.

Before the PBAD Department position, Chermansky served for two years as a Division Head for the Naval Air Warfare Center Aircraft Division (NAWCAD) - Program and Business Analysis Department (PBAD) and lead BFMs who supported major Naval Air Systems Command (NAVAIR) programs across PEO AIR Anti-Submarine Warfare (ASW) (PEO (A)) and PEO Joint Strike Fighter (PEO JSF) Joint Program Office acquisition programs. Chermansky served as a NAVAIR Comptroller Budget Division Director for nine years within two separate divisions: the Program Executive Office for Unmanned Aircraft & Weapons (PEO U&W) and Program Executive Office for Common Systems (PEO CS) division and the PEO AIR Anti-Submarine Warfare (ASW) (PEO (A)) division. She fiscally managed all investment appropriations across numerous Major Defense Acquisition Programs (MDAP) within those divisions. She



was instrumental in the creation of financial management tools, budget documentation, policies, processes, and procedures spanning the entire NAVAIR Enterprise. Over 12 years, Chermansky served as a BFM for several MDAP program offices, including several joint, such as PMA257 – Harrier (AV-8B; lead BFM), PMA265-F/A-18, PMA275 V-22 Osprey (lead BFM), as well as PMA281 (Naval Aviation Mission Planning) and PMA299 (MH-60R).

Chermansky earned a Bachelor of Arts degree in economics from St. Mary's College of Maryland in St. Mary's City, MD, and a Master of Science degree in business management from the Florida Institute of Technology in Melbourne, FL. She is Level III in the DoD Financial Management Certification Program and Advanced Level certified for the Defense Acquisition Workforce Improvement Act in Business-Financial Management. She has been recognized with a DoN Meritorious Civilian Service Award.

Abstract

The Navy's strategy continues to highlight the need for new thinking and innovative acquisition approaches to meet the demands of the Fleet. As stated in Rand's DoD FY2018 National Defense Authorization Act report (Werber et al., 2019), there are gaps in the acquisition workforce's business acumen knowledge that current training approaches have not been able to resolve. The NAVAIR enterprise sought to baseline the current levels of understanding in a program office's business acumen skills while engaging the workforce in an organic learning environment.

This research uses the Capability Maturity Model (CMM) maturity levels overlaid across the Planning, Programming, Budgeting, and Execution (PPBE) system to offer an innocuous way to ascertain a program office's capability in the business acumen domain. The model applies maturity levels as the backdrop across PPBE processes and subprocesses and uses questions to assess how well a program office executes its work practices. Improving maturity levels for work practices has been demonstrated to improve performance (Information Systems Audit and Control Association [ISACA], 2023). This research is grounded in Experiential Learning Theory (Kolb & Kolb, 2008) and uses facilitated sessions to examine leading practices that can be instrumental in fostering a learning culture within an organization. Follow-up surveys and analysis will answer the research question: What is the efficacy of the assessment learning model in improving business acumen/PPBE processes?

The Assessment Learning Model (ALM) provides a common approach for baselining a Defense Acquisition program office's proficiency in a specified domain while deploying a learning element to utilize in perpetuity. The ALM provides a foundation of how well processes and procedures are used to increase performance and improve decision-making across the enterprise. The ALM provides an authoritative learning assessment model that can be tailored to different domains.

Keywords: Acquisition innovation, capability maturity model, PPBE system, experiential learning, enabling leadership, assessments, audits, business acumen.

Assessment Learning Model

Statement of Research Issue

The Navy's strategy continues to highlight the need for new thinking and innovative approaches to meet the demands of the Fleet (Department of the Navy Research Development Test & Evaluation, 2017; Department of the Navy, 2016). The realities of funding shortfalls and leaner workforces combined with increased global threats require more advanced business processes and management cultures to meet the dynamic needs of the fleet (Boyatzis, 2011). The nation's adversaries have increased their capability to acquire and deploy military assets faster. The near-peer adversaries challenge the core of NAVIAR's value proposition to acquire, deliver, and sustain aviation and weapon systems for the Fleet.

To address these challenges, the Assessment Learning Model (ALM) proposes a common approach for baselining a Defense Acquisition program's proficiency in a specified domain. Furthermore, the model can deploy an organizational learning element to be utilized in



perpetuity. The ALM offers a path to reach deeper into a program's structure and engage the agency of mid-tier employees. The ALM provides an authoritative learning model that can be tailored to different domains. The model leverages existing leading (work) practices and can effectively foster a learning culture within an organization. This proposed study will launch an assessment learning model and validate the efficacy of this approach to improve business acumen.

Background

The U. S. Navy has launched the *Get Real, Get Better* effort, described as a new way of thinking and problem-solving. It is designed to improve outcomes in capability, affordability, and availability. The Acquisition Learning Model is an effort to build learning teams and expand critical thinking to deliver warfighting capability to the Fleet at an affordable cost. According to the Virtual Acquisition Office (VAO) Research Institute (2022), business acumen is a skill that needs further development in the acquisition workforce and can save considerable funds when it is further developed. Possible explanations for limited expertise in business acumen include the pressures and tensions that program teams experience in executing their singular duties, which inhibits collective thought or establishment of holistic, adaptive business approaches.

The assessments are not *graded* intrinsically but highlight less mature processes or areas that should be addressed. During the assessment process, team interaction is crucial in establishing an organic learning environment and emphasizing that the output depends on the quality of the input. Although improving working practices is essential, establishing and then applying the learning organization elements is the goal. Experiential learning and enabling leadership play a key role. The effort should yield increased organizational know-how through cognitive work efforts (Rideout, 2023). The assessment is not for comparison with other entities but rather to identify through a collaborative effort the maturity levels of targeted business processes and the management culture while promoting dialogue between organizational boundaries.

Research Objective

The ALM is designed to provide a foundational view or baseline of how well processes and procedures are used to increase performance and improve decision-making across the enterprise. This concept paper examines the Business Acumen use of the Assessment Learning Model across 32 NAVAIR program offices. This research is grounded in Experiential Learning Theory (Kolb & Kolb, 2008). Experiential learning is a cyclic learning process that Kolb and Kolb (2005) posit is more than knowledge delivery; it is the creation of knowledge through grasping and transforming experience from cognitive work effort. Nonaka et al. (2006) suggest that organizational knowledge can be created and provide a competitive advantage.

The focus is to evaluate the business acumen proficiency of a program office's financial management by examining its use of the Planning, Programming, Budgeting, and Execution (PPBE) system. The four sub-processes of the PPBE system examined are the Program Objective Memorandum (POM), Budget Formulation, Spend Plan Management, and Unliquidated Obligation Management (ULO). This ALM also includes key management cultural parameters, such as Decision Rights, Patterns of Interaction, and Distribution of Information.

Technical Concept and Research Design

As stated earlier, the Navy has recognized the need to improve vital operating domains and apply elements of a learning organization to accelerate learning. For NAVAIR, the business acumen domain was initially identified as an immediate area of focus. The NAVAIR enterprise sought to baseline the current levels of understanding in a Program Office's business acumen skills while engaging the workforce in a learning environment. As described by the Capability



Maturity Model (CMM) and modified by the Capability Maturity Model Integration (CMMI) approaches, the maturity levels offer an innocuous way to ascertain a program office’s capability in the business acumen domain. Improving maturity levels for work practices has been demonstrated to improve performance (Information Systems Audit and Control Association [ISACA], 2023).

➤ **Research Description**

This research applies the CMM/CMMI maturity levels as the backdrop across the PPBE system processes to assess how well program offices execute them. The CMMI approach uses an appraisal process (ISACA, 2023) to determine if specific practices are in place, where the ALM examines how well or effectively the team executes a specific process or sub-process. The ALM offers a modular approach that can be used across multiple domains and establishes learning organization elements and infrastructure to increase proficiency across the enterprise.

A process’s maturity level does not equate to *good* or *bad*. Instead, it is a compass for navigating an organization’s gaps and areas that can be further cultivated. Many variables and factors affect the maturity level that a process may fall into, such as workforce experience and skill level, external factors that are out of the organization’s control, and lack of resources within specific areas or as a whole. Figure 1 shows the definition of maturity levels as published in open-source documents.

Characteristics of the Maturity Levels

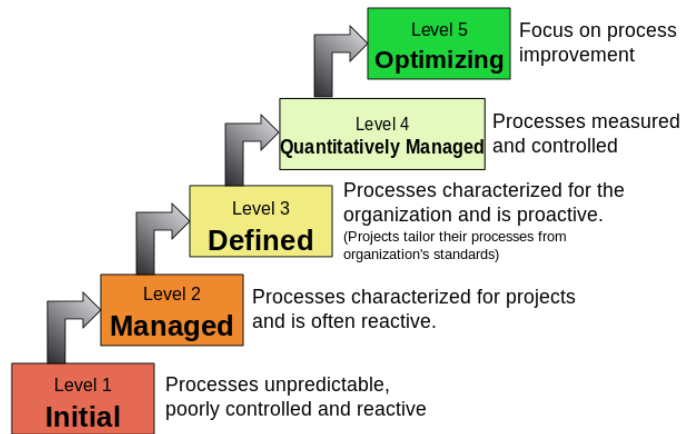


Figure 2. CMM Levels of Maturity
(Godfrey, 2004)

➤ **Research Questions**

This project addresses the following prime research question:

What is the efficacy of the assessment learning model in improving program offices’ business acumen/PPBE processes? The following three clarifying questions provide insight into the prime research question:

1. Was the facilitative approach effective?
2. Is the maturity model design an effective approach?
3. Is the assessment a practical learning approach?

➤ **Research Methodology**

Domain selection. Selecting the domain requires consideration of the right altitude or



level to provide the most effective learning experience. As stated in Rand’s Department of Defense (DoD) FY2018 National Defense Authorization Act report (Werber et al., 2019), there are gaps in the acquisition workforce’s business acumen knowledge that current training approaches still need to be able to resolve. Although there is no standard definition of business acumen for the acquisition workforce (Werber et al., 2019), the Office of Personnel Management (OPM) defines business acumen as consisting of three main elements: financial management, workforce development, and digital transformation (as shown in Figure 2).

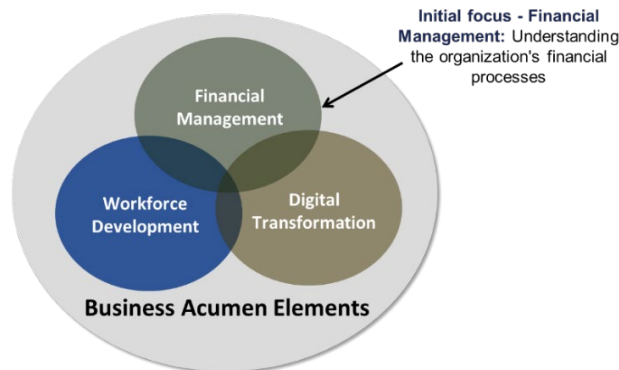


Figure 3. OPM Definition of Business Acumen
(U. S. Office of Personnel Management, 2023)

The financial management element of business acumen for the acquisition workforce is described by the VAO Research Institute (2022) as “Understanding the organization’s financial processes. Prepares, justifies, and administers the program budget. Oversees procurement and contracting to achieve desired results. Monitors expenditures and uses cost-benefit thinking to set priorities” (p. 1). The initial focus was on the financial component of business acumen. It further selected the significant financial management processes for programs: Planning, Programming, Budgeting, and Execution, with greater emphasis on the programming, budgeting, and execution phases. During the assessment, the participants also discussed the relationship of financial management to the workforce development and digital transformation sections of the POM’s definition of business acumen.

After choosing the process and sub-processes to query, the next step was to develop the key criteria to determine the maturity levels in each sub-process. In the business acumen initiative, the PPBE system served as the framework for assessment and is used at all DoD levels. The PPBE documents support the Future Years Defense Program (FYDP). The Congressional Research Service (2022) states, “PPBE is one of DOD’s three main acquisition-related decision support systems” and, therefore, a critical process. The project investigators, assisted by a selected core team, developed a workbook that developed the detailed criteria to be asked during the assessment.

Assessment deployment. A pre-brief to key program leaders was conducted and helped describe the premise of the assessment and how to use the model for best results. It provided an opportunity to describe which team members should attend for a maximum range of participation across functional areas and various skill levels. The key program office leaders introduced a three-hour working session by providing opening remarks for the assessment to level set their team. The core team presented kick-off slides to the program office team, which assessed the team and explained the importance of the effort and the mechanics during the session to include the minimum assessment criteria. Active facilitation from the core team during the session was necessary.



Survey launch and data collection. The core team developed and launched a survey of participants to determine the efficacy of the effort. The survey addressed the research questions as quantitative and qualitative, with numerical scores and open-ended questions.

Data Analysis

➤ **Demographics**

There were 163 participant responses. Table 1 describes in detail the demographics solicited in the survey: functional area, job level, acquisition experience, and highest program level experience. The data shows that the functional areas are evenly split between the program and financial domains. The job level data shows that more than one-third of participants were at a senior level in the program, which connotes significant program financial responsibilities. Similarly, more than half of the respondents have more than 15 years of acquisition experience and overwhelmingly have a large program or ACAT I experience. This demographic data shows the survey participants have deep acquisition experience and knowledge, which provides the survey results a measure of validity.

Table 1. Survey Participant Demographics

Functional Area		Job Level		Acquisition Experience		Program Level	
Program Management	51%	PM/PDPM DPM	22%	21+ years	32%	ACAT I	85%
Financial Management	45%	Level I	14%	16-20 years	26%	ACAT II	5%
Other	4%	Level II	7%	11-15 years	25%	ACAT III	3%
		Level III	1%	6-10 years	9%	ACAT IV	5%
		Support	56%	0-5 years	8%	Other	2%

➤ **Facilitative Effectiveness**

In each of Figures 3 through 5, this study posits that Strongly Agree and Agree can be characterized as having met a positive threshold. Figure 3 shows that 83% of the respondents agree that adequate time was allowed for the assessment. The facilitative approach was 84% effective at creating open dialogue, and 80% felt that their input was considered. Overall, the facilitative approach was considered 81% effective.



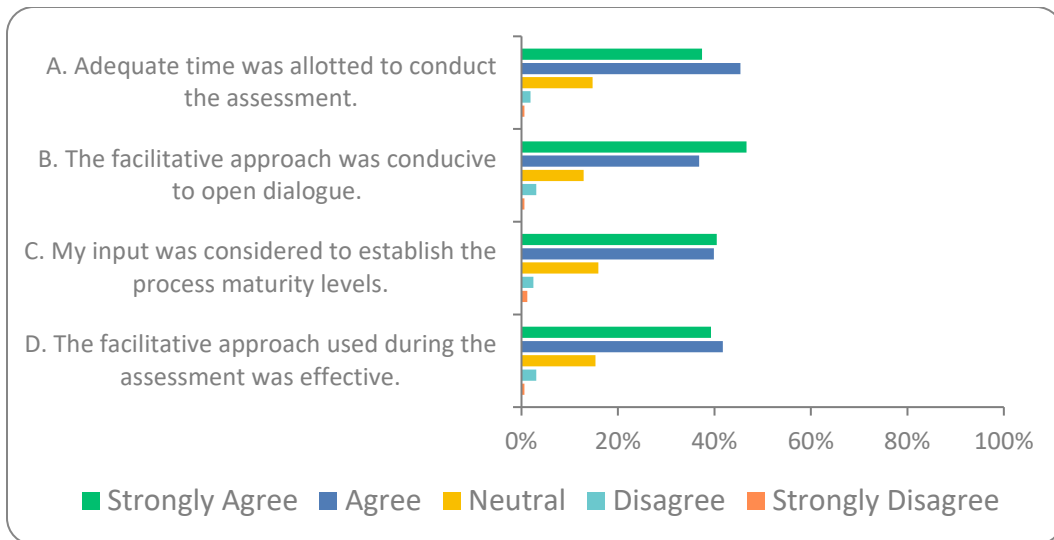


Figure 3. Facilitative Approach Effectiveness

➤ **Model Design Effectiveness**

Figure 4 provided results on the model’s design effectiveness. Seventy-three of the respondents described the tool/learning model as easy to use, while 68% described the maturity levels as clear to use. The survey participants recorded that the assessment tool/learning model was useful and provided a baseline of the program’s business process at 69% and 70%, respectively. These numbers display a clear signal that the assessment learning model answers the research questions.

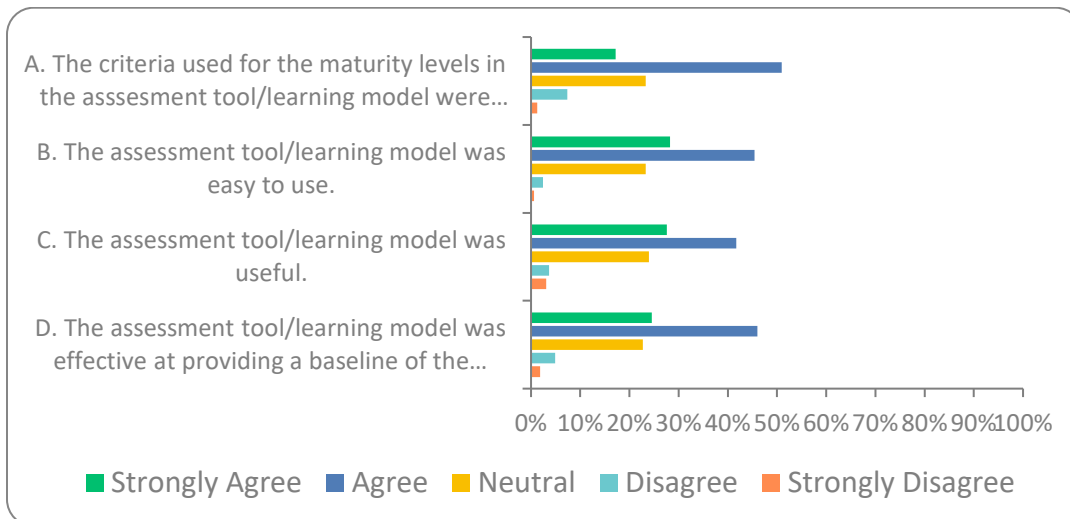


Figure 4. Model Design Effectiveness

➤ **Assessment Effectiveness**

Figure 5 shows that 74% of the survey respondents believe the assessment approach provided a practical learning environment. Sixty-nine percent thought the assessment helped them understand the internal process differences between the organizational elements in the program office. In addition, 64% thought the assessment provided a potential road map on how to mature existing business and management culture processes. Finally, 69% believe the assessment can be effective in baselining other program office processes and domains.



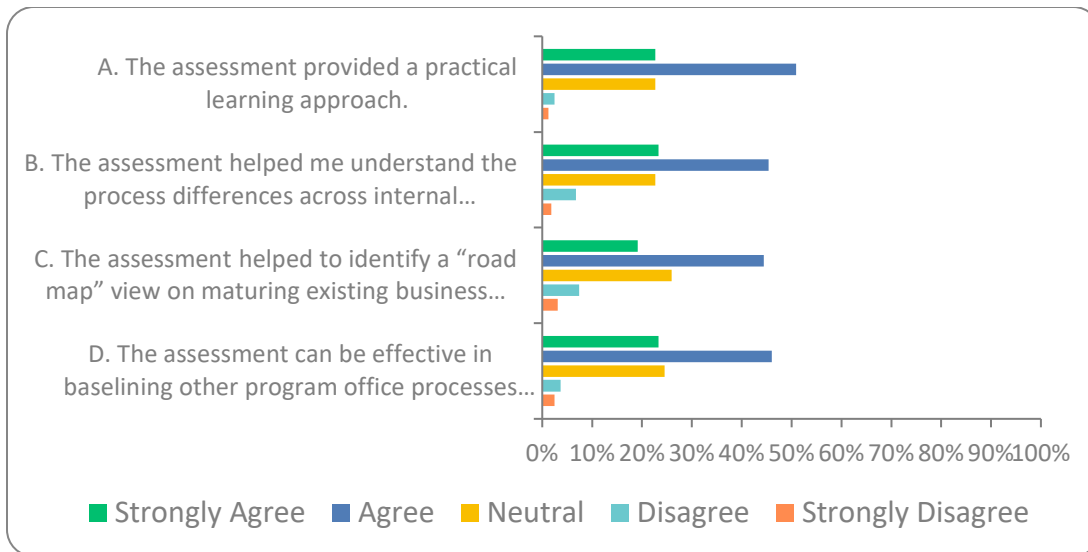


Figure 5. Assessment Effectiveness

Research Relevance and Conclusions

To aid in the continuous development of the acquisition workforce, as noted in the DoD (2015) *Acquisition Workforce Strategic Plan*, this research used qualitative and quantitative methods to explore the efficacy of deploying the assessment learning model for the chosen domain. The overarching concept is to create an interactive and positive learning environment for the program offices that accelerates learning.

The assessment was conducted in a team environment, emphasizing open conversation among the team members to allow for an organic learning environment. As noted earlier, the value of the assessment learning model is in the dialogue and exchange of points of view within the organization. The team environment provides a cross-functional aspect and efficiently reaches multiple team members. For example, in the business acumen domain initiative, 32 program offices were assessed, with an average of approximately 16–18 participants in each assessment for an engagement of 575 people.

As noted in the data analysis section, the quantitative results provide insight into the efficacy of the assessment learning model. The data shows that the three clarifying questions were answered. The facilitative approach is effective, the maturity model is an effective design, and the assessment provides a practical learning approach. These data results provide insight into an accelerated learning tool with the maturity model that provides an innovative approach.

The top qualitative responses were distilled into key themes. The participants described the strengths of the assessment tool/learning model as providing valuable cross-talk and illuminating or providing insight into process areas that need improvement. Potential areas for improvement were mostly in facilitation improvement to refine the content and improve the pre-assessment communications. The majority of respondents stated in the qualitative section of the survey that *no improvements were needed*.

The net promoter score (NPS) model (scale 1–10) was used to ask if the participants would recommend the assessment tool/learning model to a colleague. The resultant distribution was roughly equally stratified in thirds. Promoters (9–10) accounted for 30%, Passives (7–8) 36%, and detractors (1–6) 34% of the respondents.



In conclusion, the qualitative and quantitative data provide sufficient evidence that the assessment and model approach provide a useful and effective method to baseline and improve processes and performance in a given domain.

References

- Boyatzis, R. E. (2011). Managerial and leadership competencies: A behavioral approach to emotional, social and cognitive intelligence. *Vision*, 15(2), 91–100.
- Chermansky, J. (2023). *Overview: Business acumen learning model* [Unpublished Microsoft PowerPoint file]. Naval Air Systems Command.
- Coker, S. (2023). *Business acumen learning model workbook* [Unpublished Microsoft Excel file]. Naval Air Systems Command.
- Congressional Research Service. (2022). *Defense primer: Planning, programming, budgeting, and execution (PPBE) process*. <https://www.everycrsreport.com/reports/IF10429.html>
- Curtis, B., Hefley, W., & Miller, S. (2001). *People capability maturity model (P-CMM) version 2.0* (CMU/SEI-2001-MM-01). Carnegie Mellon, Software Engineering Institute. <https://insights.sei.cmu.edu/library/people-capability-maturity-model-p-cmm-version-20/>
- Department of the Navy. (2016). *Navy civilian workforce framework*. https://www.navy.mil/navydata/people/cno/Richardson/Resource/Navy_Civilian_Framework.pdf.
- Department of the Navy Research Development Test & Evaluation. (2017). *30-year research & development plan*.
- DoD. (2015). *Acquisition workforce strategic plan*. https://asc.army.mil/web/dod_acq_workforce_hscp_fy16-21/
- DoD. (2017). *The planning, programming, budgeting, and execution (PPBE) process* (DoD Directive 7045.14).
- Godfrey, S. (2004). *Using CMMI for improvement at Goddard Space Flight Center (GFSC)* (NASA presentation, systems engineering lecture series). https://ses.gsfc.nasa.gov/ses_data_2004/040601_Godfrey.ppt
- Information Systems Audit and Control Association. (2023). *CMMI Institute*. <https://cmminstitute.com/cmml>.
- Kolb, A., & Kolb, D. (2005). Learning styles and learning spaces. Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4(2), 193–212. <https://doi.org/10.5465/AMLE.2005.17268566>
- Kolb, A., & Kolb, D. (2008). Experiential learning theory: A dynamic approach to management learning, education, and development. *Armstrong: Management, learning, education, and development* (pp. 42–68).
- Nonaka, I., von Krogh, G., & Voepel, S. (2006). Organizational knowledge creation theory: Evolutionary paths and future advances. *Organization Studies*, 27(8), 1179–1208.
- Office of the Secretary of the Navy. (2021). *The planning, programming, budgeting, and execution process* (SECNAV Instruction 7000.30). Department of the Navy.
- Ramanujan, S., & Kesh, S. (2004). Comparison of knowledge management and CMM/CMMI implementation. *Journal of American Academy of Business, Cambridge*, 4(1), 271–277.



- Richardson, J. (2017). *Navy leader development framework*. Chief of Naval Operations. https://www.navy.mil/cno/docs/NLDF_2.pdf.
- Rideout, B. W. (2023). *Complexity leadership theory: The influence of enabling leadership and experiential learning on emergent outcomes and organizational know-how* [Doctoral dissertation, Florida Institute of Technology].
- Rideout, Whetstone, & Guffey. (2018). *Should cost workbook* [Unpublished Microsoft Excel file]. Naval Air Systems Command.
- U. S. Office of Personnel Management. (2023). *Executive core qualifications, OPM*. <https://www.opm.gov/policy-data-oversight/senior-executive-service/executive-core-qualifications/>
- Virtual Acquisition Office Research Institute. (2022). *Taking care of business, shaping smart decisions with business acumen*. www.GoToVAO.com
- Werber, L., Ausink, J., Daugherty, L., Phillips, B., Kuntson, F., & Haberman, R. (2019). *An assessment of gaps in business acumen and knowledge of industry within the defense acquisition workforce. A report prepared for the U.S. Department of Defense in compliance with section 843(c) of the fiscal year 2018 National Defense Authorization Act*. Rand Corporation.



Collaboratorium: A Multi-Stakeholder Approach to Advancing Innovative Defense Acquisition

Eliza Thompson—is a technology and security policy specialist, with experience implementing discovery and matchmaking programming for the defense and information integrity spaces. At Becera, Eliza helps government agencies and stakeholders identify, test, and apply emerging technologies to address operational challenges. She was part of the core team which designed a Department of State-funded platform that brought together government, technologists, and civil society to identify and assess solutions for countering foreign influence operations. Eliza has an MSc in European Policy from the London School of Economics and a BA with honors in Political Science from Stanford University. [eliza@becera.co]

Anna Jorstad—is a graduate student at New York University, set to complete her MS in Global Security, Conflict, and Cyber Crime in May 2024. She has been recognized with the Center for Global Affairs Excellence and Achievement Award in recognition of her academic performance. Anna holds a BA from the University of Virginia, where she double majored in Foreign Affairs and Religious Studies. Her specialties include strategic policy and analysis related to emerging technologies, global security, and cyber issues, aligning with her academic background and research interests. [asj8358@nyu.edu]

Christina Nemr—is the founder and CEO of Becera, a company that bridges the public and private sectors to provide tech match-making that address national security challenges. She has helped develop research, policy, programming related to national security issues in the information environment for over a decade in collaboration with governments, civil society, academia, and the private sector, with increasing focus on understanding and applying emerging tech against complex issues. She has degrees in International Relations and Forensic Psychology. [christina@becera.co]

Bev Corwin—was a small business owner, consultant, and entrepreneur for 20 years in the data center services industry and has experience in information ecosystems, accessibility, cybersecurity, and information security. Bev was a co-founder of the Information Architecture Institute IAI, which later merged with World Information Architecture Association WIAA. Bev currently serves as VP of Professional Development for AFCEA NY Founders Chapter and serves as chapter liaison for AFCEA International Small Business Committee and Innovation Committee. She also runs Technology Transfer Days where she facilitates collaborative learnings between the private sector and the U.S. military, NGA, NASA, and the Department of Energy to amplify innovative technology. She serves on advisory boards, and is a technical and strategy advisor for international, humanitarian, resilience, readiness, and emerging technology businesses. [bev@becera.co]

Abstract

The U.S. Department of Defense (DoD) faces significant challenges in effectively transitioning innovative technologies from research and development to operational use, a phenomenon known as the “valley of death.” This issue has critical implications for national security, as delays in technology acquisition hinder the United States’ ability to keep pace with global competitors. While addressing these challenges requires a multifaceted approach of updating policies, cultures, and ecosystems, one area worth exploring is how the defense innovation ecosystem engages less-traditional stakeholders, such as civil society, academia, independent researchers, and small businesses. Fostering knowledge sharing and collaboration across these diverse communities allows the DoD to tap into a broader range of perspectives and technical expertise, leading to more effective technology discovery and development. This paper first analyzes the policy, process, personnel, and budgeting hurdles hindering defense technology innovation and then examines current successes and future opportunities for less-traditional stakeholder engagement in the defense innovation landscape. Building on lessons learned from an initiative called “Technology Transfer Days” (TTDs), the paper proposes a framework for a technology matchmaking collaboratorium, the Defense Innovation Discovery and Collaboratorium Platform (DID Collaboratorium), which can provide comprehensive resources for a whole-of-community engagement strategy around defense technology development and acquisition needs.



Key words: valley of death, less-traditional stakeholders, emerging technology, innovation, collaboratorium, technology transfer days (TTDs), technology discovery

Introduction

The well-known “valley of death” phenomena—namely the U.S. Department of Defense (DoD) and its broader investment community spending billions annually on development, only to see a lack of sustained scale-up of new technologies—has been explored at great lengths. The issue presents an urgent challenge for the United States and its global allies, as it hinders the United States’ ability to swiftly transition to and leverage the latest technologies crucial for safeguarding national and international security.

Delays in innovation and technology acquisition across both the public and private sectors are worrying, as competitors such as China continue to make breakthroughs in the technology-military nexus. Historically, the United States has always been a global leader in both public and private sector innovation—fostering a synergy between commercial and public interests to drive breakthroughs across industry, academia, and government (Lawrence, 2023). To push back against today’s most imminent threats, the DoD’s innovation ecosystem must find its footing again and ensure the latest mission-critical technologies are not only integrated into military programming but also arise out of such programming directly. Research and innovation must have a seat at the table.

While the DoD has taken meaningful steps to improve the identification and acquisition of innovative technologies, such as implementing numerous research and commercialization efforts to improve innovation, substantial gaps in technology discovery and acquisition remain. These gaps persist despite private sector interest in engaging with the government around technology research and development in efforts to address today’s foremost global and national security challenges (Defense Innovation Board [DIB] Strategic Investment Capital Task Force, 2023).

This paper explores the following two-part research question:

- 1) How does the U.S. defense community currently integrate knowledge sharing and less-traditional stakeholder engagement approaches to help socialize across silos to innovation?
- 2) How could the U.S. defense community more effectively engage less-traditional stakeholders and communities of interest to undo barriers preventing defense innovation from reaching its full potential?

Given the significant attention already given to this topic, it is important to explore the issue of technology innovation and acquisition across the defense community through a different lens. This paper examines how the U.S. defense community resources innovation across silos, including from less-traditional sources, to apply technology to protect national security interests. By examining the ways in which the defense ecosystem successfully engages a community of less-traditional stakeholders, this paper attempts to better the understanding of the widening schism between private and public sector innovation with the aim of bettering technology discovery, knowledge sharing, and investment.

In this paper, we define less-traditional stakeholder engagement as the practice of engaging and incorporating a multi-disciplinary and multi-dimensional set of stakeholders—including civil society, academia, independent researchers, small-scale startups and businesses, and solo inventors and entrepreneurs—into the identification and decision-making process. These stakeholders are frequently underrepresented across traditional defense communities, causing the DoD to miss out on opportunities to incorporate a broader range of



perspectives and technical knowledge. The broad discovery, amplification, and inclusion of these stakeholders can help facilitate otherwise untapped innovations, as well as provide the diversity of ideas and backgrounds needed to ensure technology is not only practical but reflective and ethical.

By exploring current success stories, as well as future opportunities for the DoD and its investment ecosystem to engage with less-traditional stakeholders, a stronger understanding of the importance of knowledge sharing and fostering communities of interest around emerging technologies can be gained. This in turn can be applied to future technology discovery and development efforts across the defense community.

This paper aims to serve as a key resource in better understanding how public and private sector stakeholders can more effectively engage with each other moving forward to ensure the latest technologies are successfully applied in defense efforts. To achieve this, the paper first explores the policy, process, personnel, and budgeting hurdles that the private and public sectors face when looking to collaborate in the defense market. Using “Technology Transfer Days” (TTDs) as a baseline model and case study, the paper then identifies and evaluates existing mechanisms that have been successful in facilitating innovation and collaboration between the public and private sectors across the technology-defense ecosystem.

The paper finally builds on this analysis to provide a framework for a living, dynamic knowledge environment for fostering collaboration and information-sharing that facilitates innovation. A key element of this framework involves the development of a Defense Innovation Discovery and Collaboratorium Platform (DID Collaboratorium), which would provide crucial resources to enable a whole-of-community engagement strategy around defense technology development and acquisition needs.

Research Methodology

The research methodology for this paper followed a three-pronged approach. First, the authors conducted desk-based research to better understand and identify the underlying challenges in this space, including current factors driving barriers to innovation across the defense community ecosystem. Second, the authors conducted a survey across the defense innovation and acquisition community to gain perspectives on the impact of multi-stakeholder community engagement on defense technology identification, assessment, and acquisition. The 10-question research survey was disseminated across relevant networks in the defense technology innovation and acquisition space from February to March 2024 and resulted in 20 responses (see Figure 1). Third, the authors integrated insights from five primary interviews on defense acquisition and matchmaking.



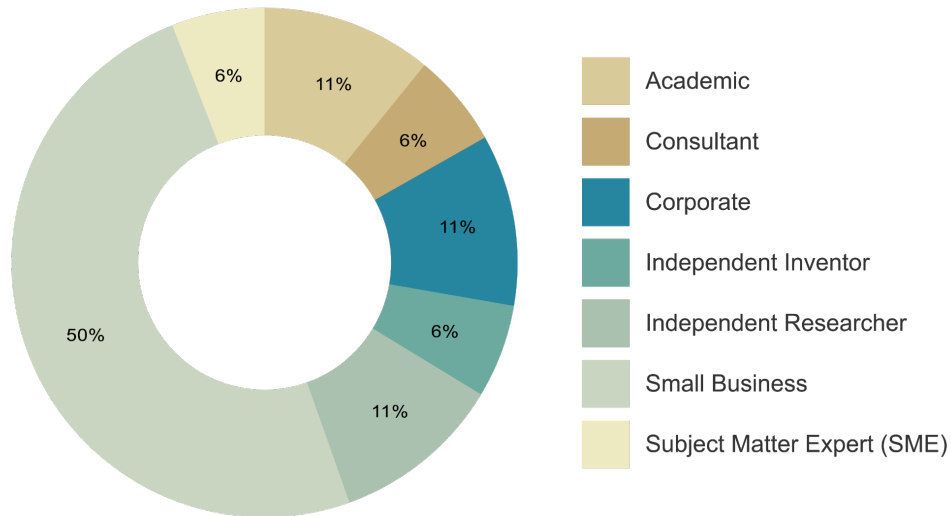


Figure 1. Breakdown of Survey Respondents by Sector

Assessing Internal Barriers to Innovation

What are key aspects of the valley of death?

The United States remains the largest performer of research and development (R&D) globally, with \$806 billion in gross domestic expenditures on R&D in 2021 (National Science Board, 2024). However, while the absolute amount of federally funded R&D increased from 2011 to 2021, the share of the total U.S. R&D funded by the federal government decreased from 30% to 19% during this time period. Moreover, the business sector leads the way in R&D funding for experimental development and applied research, with the business sector funding 87.6% of experimental development compared to the federal government funding 11% (National Science Board, 2024).

The DoD's requirements and acquisition processes were largely designed for a time when the DoD was the largest funder of global research and development. Today, however, many mission-critical technologies are driven by the commercial sector, and the DoD's processes have not adapted to this new reality. In fact, the DoD's industrial base has shrunk by 40% over the past decade (The White House, 2022). The DoD has struggled to effectively leverage new technologies, with the gap between private sector and public sector innovation widening in recent years in areas such as generative AI.

A key aspect of this problem is the lack of engagement with and support for a diverse range of stakeholders, including small businesses and entrepreneurs. The DoD also struggles with engaging other less-traditional stakeholders, such as civil society and independent researchers, two essential voices for ensuring technology development is practical, ethical, and sustainable. Moreover, industry feedback in a survey disseminated by Becera highlighted that the DoD has a tendency to set requirements that are too specific or limited in scope. This can limit the adoption of technologies, even if they feature mission-critical capabilities that are very close to what is needed. It can also lead to wasted time and money developing new technologies when the off-the-shelf solution would have been adequate in meeting the DoD's mission-critical needs from the get-go (DIB Strategic Investment Capital Task Force, 2023).



Barriers in Internal Bureaucratic Processes and Network Access

The unique nature of how the DoD operates and the demands placed on work flows due to internal bureaucratic processes stifles rapid innovation and the ability to keep up with evolving technology. For example, experts highlight the Authority to Operate (ATO) model as a major barrier to accelerating innovation such as AI adoption (Allen, 2023). Under this model, every software system that operates on the DoD's Information Network (DoDIN)¹ and processes government data must receive an official ATO from a certified DoD authorizing official, requiring extensive written documentation demonstrating how the software will comply with various cybersecurity and operational controls.

While having clear controls and standards in place is important for ensuring the integrity of security systems, most DoD components have only a single authorizing official to weigh the benefits and risks of allowing new applications. Moreover, a study conducted by the Center for Strategic and International Studies (CSIS) found that a risk-averse approach has taken root across this ATO process, resulting in long evaluation processes that simply do not keep pace with technology evolution. As outlined by the CSIS, this means the development of, for example, an AI capability cannot begin until the development environment and all the necessary software pieces have received an ATO (Allen, 2023). Contrast this with the commercial and academic sectors, where an AI developer can download widely available open-source AI development frameworks to use with their own datasets.

This reflects a broader trend across the DoD and its bureaucratic processes, which are largely designed and operate under a risk-averse culture. Numerous stakeholders have identified this limitation, including Deputy Assistant Secretary of the Army for Research and Technology William Nelson, who has stated that more experimentation and risk need to be incentivized in order to push forward new technologies and innovation across the DoD (Metzger, 2023). Namely, allowing for more risk would help current processes support innovation and allow for the integration of the latest solutions developed by industry and research.

Additionally, the DoD often imposes excessive security requirements on less-traditional businesses, limiting their ability to showcase their capabilities (DIB, 2024). A DIB study recommends avoiding imposing security requirements on less-traditional businesses in RFPs until the government has a clear understanding of their capabilities, recommending instead that an independent classification system for potential bidders could be created that would allow them to demonstrate their suitability for different types of contracts without revealing sensitive information (DIB, 2024).

Gaps in Data Access and Management

Data-related challenges remain a key barrier to innovation and the deployment of emerging technology across the DoD. While the DoD and wider intelligence community has amassed considerable amounts of data over the last decades, there are key limitations in terms of how this data is organized, stored, and accessed that present obstacles to innovation (Allen, 2023). Often, this data is siloed so that training data is generally application-specific, and there are also issues with the diversity and variety of data due to barriers around acquisition of sufficient data that is diverse enough (Allen, 2023).

There is also the issue of data accessibility, with much of the DoD's data siloed across different levels of classification (Allen, 2023). While classification systems are critical to

¹ According to the U.S. Cyber Command, the DoDIN is a federated environment of 46 combatant commands, services, and DoD agencies and field activities. This includes over 15,000 unclassified, classified networked and cloud environments globally, as well as 23 Cyber Security Service Providers. [Link](#).



information security, this leads to the compartmentalization of information, which in turn hinders the ability to develop, assess, and operationalize new technology. Small businesses, entrepreneurs, and researchers who may lack the knowledge of classification systems or the resources to pursue security credentials can therefore often be left out of the ability to access or use data for technology development or training.

Weaknesses in Talent Acquisition

Innovation requires a workforce equipped with the ability to perform technologically advanced activities and research, with a clear need for capabilities in critical and emerging technologies. From scientific publications to patent activity, leadership in technology development comes directly from the ability to educate, train, and retain talent. However, in recent years, overlapping strategies and entrenched practices have hindered the DoD's ability to recruit and retain technologically savvy employees, creating missed opportunities (Weisner, 2023). Key gaps include lack of mentorship programs and opportunities as well as a culture defined by lack of trust in junior talent (Weisner, 2023). These issues are compounded by bureaucratic processes that slow hiring and make workflows unappealing to those used to operating in fast-paced, dynamic environments characteristic of private sector technology development.

Limitations in Technology Identification and Information Sharing

The DoD also faces issues with its technology identification and matching ecosystem, which hinders its ability to identify and develop technology at the pace necessary to ensure emerging technologies are properly leveraged and deployed. There are issues with the DoD's "commercial technology pipeline" (CTP)² through which innovative commercial technologies are identified. A study by the RAND National Defense Research Institute (NDRI) identified key challenges and gaps in the CTP, including a lack of alignment with stakeholders around shared mission or common goals, objectives, and outcomes (Kotila et al., 2023). The NDRI also found gaps across the incentive structures for CTP stakeholders, as well as a lack of clearly defined metrics or accountability mechanisms to check progress against goals (Kotila et al., 2023).

The NDRI study also found that the CTP lacked both formal mechanisms and requirements for information sharing with stakeholders, including no clear coordination or collaboration across stakeholders (Kotila et al., 2023). This includes the ability to share information around promising emerging technologies, available resources and programs, and ongoing research across stakeholder groups. These limitations reflect a larger issue across the DoD of not adequately engaging less-traditional stakeholders to identify new technologies, which impedes collaboration and contributes to the valley of death problem.

Burdensome Approach to Dual-Use Technology

A DIB study examined the DoD's ability to acquire dual-use technologies and determined it encounters significant self-imposed obstacles, including limiting investment in startup research and development and creating a burdensome acquisition process for less-traditional companies (DIB, 2024). More effectively supporting the identification of and investment in dual-use technologies will help provide more resources and opportunities for innovation. Engaging less-traditional stakeholders is a key component of this, as it removes constraints around who may be a beneficial collaborator for the DoD.

² The RAND National Defense Research Institute (NDRI) defines the CTP as the activities, functions, and processes around the DoD's identification, development, and transitions of innovative commercial technologies from the private sector to DoD for military use. [Link](#).



Assessing Private Sector Barriers to Engagement With the DoD

How does the private sector view opportunities and challenges in the defense innovation ecosystem?

Small businesses and start-ups frequently offer innovative emerging technologies that are potentially mission critical for the U.S. defense community, but they face uphill battles transitioning to sustain DoD operations at scale. Small businesses and start-ups typically lack the resources of more established businesses, making it difficult to overcome resource-intensive government requirements and a disconnected ecosystem (Marinelli, 2023; Mcnamara et al., 2024). Due to uncertainty of the viability of contracting with the DoD, many opt to pursue less risky commercial opportunities instead. Some businesses that achieve commercial success will later seek to contract with the DoD, but others will not. This can result in the delay or full separation of potentially mission-critical technologies from warfighters (Mcnamara et al., 2024; Tucker, 2024).

Limited Opportunities for New Entrants

Industry feedback in the authors' survey highlighted that small businesses face significant hurdles in securing even minimal funding when seeking to contract with the DoD. Acting as the DoD's venture fund, the SBIR and Small Business Technology Transfer (STTR) programs serve as crucial entry points for small businesses (DIB Strategic Investment Capital Task Force, 2023). However, these programs have increasingly been constrained by lock-in, creating a highly competitive landscape with limited funding opportunities for new entrants. The DoD's preference for closed, proprietary systems limits interoperability and the implementation of innovative technologies. These closed-system architectures severely hinder small businesses and new entrants, favoring a small number of repeat private sector partners (Mcnamara et al., 2024).

Additionally, because some businesses receive multiple SBIR/STTR awards each year, a disproportionate share of funding has consistently been allocated to a small number of awardees. Between 2010 and 2019, 90% of Phase I SBIR/STTR funds were awarded to previous contract awardees (Bresler & Bresler, 2020). Between 2012 and 2021, the top 5% of businesses receiving the most Phase I/II SBIR contract awards received 49% of the DoD's Phase I/II SBIR/STTR funding. Additionally, of the mere 16% of Phase I/II awardees that received Phase III contracts, 61% made less in Phase III contract revenue than they did in Phase I/II funding (Bresler & Bresler, 2023).

Resource Demands Make Engagement Non-Viable

Steep resource demands imposed by burdensome requirements compound the negative ROI from Phase I/II to Phase III, representing a significant barrier to private sector business engagement with the DoD (Bresler & Bresler, 2023; Decker & Sheinbaum, 2024). DoD qualifications such as reporting requirements, technical certifications, cybersecurity certifications, licensing, and security clearances can be cost-prohibitive and time consuming for small businesses, delaying awards and taking time away from actual work. Many required licenses and certifications require a contract despite contracts first requiring licenses and certifications, favoring existing awardees.

There is also a large backlog of the DoD's complex individual and facility security clearance applications, hindering new entrants from accessing classified environments and information critical to progressing their work for the DoD. Additionally, larger businesses are prioritized in DoD testing facilities, causing further delay to already burdened small businesses (Mcnamara et al., 2024; Decker & Sheinbaum, 2024). Businesses often wait up to four years for the DoD to finalize funding and contracts, during which time they often are unable to engage in and profit from commercial business due to the DoD's strict intellectual property requirements



(Mcnamara et al., 2024; Decker & Sheinbaum, 2024). These demands ultimately are cost-prohibitive, pricing out many small businesses from working with the DoD due to the risk of going out of business during the acquisition process (Marinelli, 2023; Decker & Sheinbaum, 2024).

Barriers Around Government Requirements and Framework

The bureaucratic DoD acquisition process is “a complex web of entry points and intricate regulations” that small businesses and new entrants struggle to navigate (Tucker, 2024). While there are many entry points to the DoD market, this information is obfuscated by its complexity and lack of centralization. This causes businesses to struggle with a lack of awareness of the value of their tech for the DoD, potential opportunities, and the requirements they must meet (Kotila et al., 2023). Even when aware of entry points, industry feedback in Becera’s survey revealed that many private sector businesses find DoD application portals and opportunity postings confusing and difficult to follow. This confusion can discourage new entrants who forgo the DoD’s bureaucratic barriers in favor of less complex commercial markets. Unprepared businesses may fail to complete the acquisition process due to an inability to complete the requirements or by running out of funding, rather than failing due to their tech lacking value to the DoD’s mission (Kotila et al., 2023). This results in missed opportunities for innovation and collaboration with the DoD and wider defense community.

Disconnect Between the DoD and Private Sector

A disconnect between the DoD and the private sector has created significant barriers to collaboration and led to failed innovative tech acquisitions. Without a clear understanding of what all parties need and a common goal, it becomes challenging to achieve effective and meaningful collaboration (Decker & Sheinbaum, 2024). Compared to repeat contract awardees, small businesses and new entrants are more susceptible to suffer from this disconnect due to a lack of experience and connections (Kotila et al., 2023).

Within the DoD, there are multiple levels of stakeholders across the defense innovation ecosystem who do not share a mission or consistently share information amongst themselves. This fragmentation has led to businesses receiving inconsistent guidance, creating significant confusion and misunderstandings that can delay contracts or cause businesses to fail to complete the acquisition process (Kotila et al., 2023). Businesses may communicate with program managers, but the actual purchasing power lies with other parties who handle contracting. Businesses typically have minimal contact with end-users, the warfighters, and instead must make their product appeal to those with actual buying power who may not understand end-users’ actual needs or the technology itself.

Consequently, businesses may create a product that meets warfighters’ mission-critical needs but fail to be adopted due to their products not meeting the contracting party’s requirements. Or businesses might create a product that appeals to the contracting party’s requirements and may get adopted, but it fails to meet warfighters’ actual mission-critical needs (Ferry, 2024). This lack of clear communication and disconnect harms private sector trust in the DoD, deterring engagement with the DoD and hindering innovative technology adoption (Kotila et al., 2023).

Recognizing these concerning trends, the DoD has made meaningful progress through measures such as its recent empowerment of the Defense Innovation Unit (DIU; Blank, 2024). While the DoD has begun moving in the right direction, change has been slow. To ensure the United States does not fall behind in global power competition, it is critical that the DoD address the challenges that perpetuate the valley of death with expedience. The sections that follow highlight ways to leverage the opportunities that exist.



Power of Innovative Stakeholder Engagement

What are the opportunities for the defense community in leveraging less-traditional stakeholders to enhance innovation?

Enhanced Knowledge Sharing and Communication

A lack of holistic and widespread stakeholder engagement hampers communication, collaboration, and knowledge sharing, which are all essential elements for ensuring innovation can take root. Many agencies have a “wait-and-see approach” before diving into new terrain, adopting a risk-averse approach largely out of necessity. Yet, engaging stakeholders who have already conducted research on security, policy, and ethical considerations can help mitigate the pitfalls of such an approach. Creating more spaces that allow for engagement with civil society, academia, and organizations that have strong knowledge sharing ecosystems already in place can help the defense community enhance their own knowledge sharing capabilities. Engaging less-traditional stakeholders can help encourage cross-functional learning and communication to bridge these gaps and address many of the issues which make a culture of risk-averse investment necessary. Workshops, for example, help to provide a mechanism through which individuals can share information at both a technical and multi-disciplinary level.

Knowledge sharing networks, environments, and ecosystems can also help encourage the sharing of open-source technologies and technical knowledge that could help with infrastructure enhancement. This can build upon existing “open innovation” methodologies, refining them for trust, security, and IP protection to ensure they adequately meet government requirements. Moreover, stakeholder initiatives can help those less familiar with defense requirements and expectations better understand these needs (see Figure 2).

Of respondents, 76% felt that the initiatives helped them understand defense requirements or needs

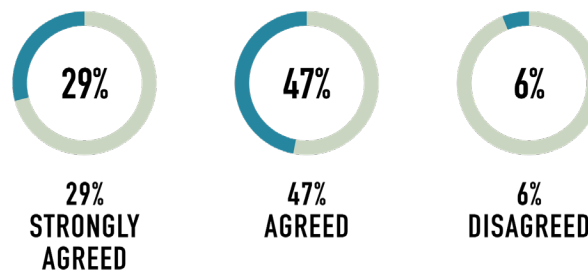


Figure 2. Select Respondent Feedback³

Enhanced Data and Information Accessibility

It is important to have access to quality data at a large scale to develop and operate new technologies such as AI systems that use machine learning. Engaging with stakeholders can help ensure there is adequate access to data that reflects operational environments. It is important to keep in mind data privacy and protection requirements and ensure data is sourced and used ethically. Eliminating unnecessary complications can help bring less-traditional stakeholders into the fold.

³ In Figure 2, the 76% referenced combines both “strongly agreed” and “agreed.” The 6% referenced combines both “strongly disagreed” and “disagreed.”



There are positive lessons from across the DoD that serve as valuable lessons of how data and information sharing restrictions can be decreased to help foster more collaboration and innovation. In 2024, the Pentagon updated its classification policy for space programs to reduce the information-sharing restrictions that make it hard for collaboration with allies, industry partners, and other stakeholders (Albon, 2024). The policy changes focused on updating out-of-date policies around what information could be shared around certain programs to reduce the overclassification of things to the point where collaboration was severely restricted. While this specific example was aimed more at interagency and foreign government engagement, it provides an example of how new policies can help decrease barriers to information sharing, thereby making it easier for stakeholders to engage.

Task Force 59 offers another success story (Vincent, 2023). It is being undertaken under a data-as-a-service model, or a contractor-owned contractor-operated model. This approach means that the data that the system generates is unclassified and can therefore live on a commercial network and move at the speed of commercial development. Due to this unclassified nature, however, all the development is taking place without classified data sources, which is a key gap in development and future deployment and integration.

Incentivizing Private Sector Engagement

Engaging less-traditional stakeholders can also help foster new opportunities for small businesses, civil society, and independent researchers who otherwise may feel a disconnect between their work and government opportunities. This thereby can help incentivize more private sector engagement, breaking down current barriers to public-private partnerships. Less-traditional stakeholders can also help identify and overcome miscommunication and misconceptions.

The DoD has introduced several initiatives that show the positive outcomes of efforts to engage small businesses, including the establishment of the Innovation Pathways gateway for small businesses to engage with the Pentagon on new systems, as well as the Rapid Defense Experimentation Reserve, which offers edge experimentation to new equipment to move prototypes through validation to production (DIB, 2024). The Pentagon's Office of Strategic Capital will also employ financial tools such as loans and guarantees to support startup-built solutions (DIB, 2024).

The DoD and its wider investment community can engage the commercial sector and entrepreneurial community, such as VCs and founders, by leveraging existing networks to gain feedback and identify common interests and goals that can lead to collaboration and cooperation, thereby increasing incentives for the private sector to want to engage the DoD. The DIB study recommends that the DoD develop a flexible and agile approach to acquiring commercial dual-use technologies by ensuring the research and development stage is connected to the acquisition stage to alleviate the pressure from startups involved in dual-use development to first need to demonstrate commercialization of products (DIB, 2024).

Stronger Technology Identification and Matchmaking

Strengthening less-traditional stakeholder engagement can help create an ecosystem for technology identification, matchmaking, assessment, and deployment, which the DoD currently struggles with. The DoD must go beyond its normal community of stakeholders and investment ecosystem to ensure that collaboration is fostered across less-traditional stakeholders as well, such as academia, civil society, startups, small business, and individual technologists and entrepreneurs.

As outlined earlier, the federal government is falling behind the business sector in funding experimental development. Moreover, the private sector is engaging other aspects of



the economy and civil society in a more strategic and holistic way, allowing considerable innovation to emerge from collaboration not necessarily seen 30 or 50 years ago. For example, U.S. universities are frequently leveraging their intellectual property by licensing protected discoveries to outside entities, including startups and small companies; in 2021, U.S. universities executed around 8,000 new technology licenses; 78% were executed with startups or small companies (National Science Board, 2024).

Venture capital also plays a substantial role in innovation and technology matchmaking through the investment in startups, with the U.S. venture capital market investing more than both the European and Chinese venture capital markets as of 2022 (National Science Board, 2024). The federal government can more effectively leverage these invest and collaboration ecosystems to ensure they are identifying the most promising and innovative technologies.

One key area for engagement is facilitating introductions between industry, academia, civil society, small business, and the DoD to create a matchmaking process where new conversations and relationships can be built. Stakeholder engagement initiatives can help both public and private sector actors identify technologists and entrepreneurs to partner with, providing a valuable mechanism through which collaborations can develop (see Figure 3). This can in turn help the DoD to identify technology it otherwise may have missed, as well as provide an ecosystem where all stakeholders are engaged from the beginning to ensure technologies are assessed and deployed to the best of their ability.

Of respondents, 76% felt that the initiatives introduced them to a network of entrepreneurs or partners

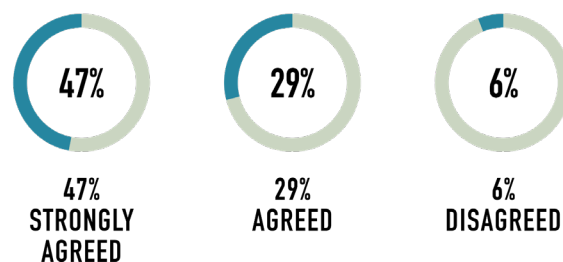


Figure 3. Select Respondent Feedback⁴

Beyond more informal engagement and matchmaking, the defense community can leverage frameworks and platforms, such as the potential DID Collaboratorium, that provide more formalized technology matchmaking. For example, technology matchmaking platforms that bring government users together with private sector, academia, and civil society technologists can help identify, test, and apply emerging technologies to address specific operational and administrative challenges. While current platforms such as Vulcan provide technology scouting and collaboration, more collaboratorium-oriented platforms engaging less-traditional stakeholders can go beyond simply identifying technologies, helping to facilitate data and information sharing, as well as address barriers around technical hindrances and risk-exposure.

By connecting the public and private sectors, these platforms provide a mechanism through which stakeholders can communicate effectively with federal government end-users, as

⁴ In Figure 3, the 76% referenced combines both “strongly agreed” and “agreed.” The 6% referenced combines both “strongly disagreed” and “disagreed.”



well as come to a mutual understanding of the challenges faced across specific areas and the breadth of solutions available to address them. A successful example of one such technology matchmaking collaboratorium was a Department of State–funded platform that brought together government and civil society stakeholders with technology companies and entrepreneurs offering solutions around information integrity.

Opportunities for Iterative Technology Development and End-User Prioritization

End-user feedback is essential for ensuring technology development meets operational needs and goals. Removing barriers to the DoD’s acquisitions of innovative technologies requires a fundamental shift towards prioritizing end-user needs and improving communication among stakeholders. Presently, higher-level management within the DoD serve as decision-makers in the acquisition process. This is the market that private sector businesses must appeal to in order for their technology to be acquired. However, these parties are typically far removed from technology end-users and lack a clear understanding of their actual needs. This disconnect impacts all stakeholders in the wider defense innovation ecosystem, deterring private sector engagement with the DoD and hindering effective innovation adoption (Ferry, 2024).

End-user input should be prioritized in all stages of the acquisition process, as they represent the community that acquired technologies are meant to serve. Unlike higher-level management, end-users have firsthand knowledge of their own needs, preferences, and what technologies are mission-critical. This makes their feedback invaluable to DoD decision-making and technology implementation success. Early involvement in research and development would enable faster identification of technologies of value to the DoD and more effective communication with the private sector businesses that possess them, increasing the chances of timely and impactful innovation adoption (Office of Public Health and Science Health and Human Services Department, 2011; Husted et al., 2021). By testing prototypes and providing other ongoing feedback during research and development, end-users can make businesses aware of the adjustments necessary for products’ successful implementation or enable the DoD to pivot to other options earlier. Ultimately, a shift in market prioritization to end-users would preserve resources, allow for more timely implementation of innovation, and increase the DoD’s strategic military advantage (Metzger, 2023; Leonard-Barton & Kraus, 1985).

In addition to making end-user input central in the acquisition process, the DoD should implement a participatory model. This model would emphasize valuing all stakeholders’ participation in order to leverage their unique skills and expertise to find optimal solutions (Office of Public Health and Science Health and Human Services Department, 2011). Decision-makers should utilize a cyclic iterative process to repeatedly collect insights and analysis from all stakeholders throughout the acquisition process to ensure their needs and priorities are understood (Leonard-Barton & Kraus, 1985). As part of this iterative process and continued collaboration among stakeholders, the DoD should establish formal means of information exchange and consider mechanisms such as establishing focus groups, holding frequent meetings, and conducting surveys (Husted et al., 2021). These changes would foster increased communication, collaboration, trust, and creativity among the defense community and private sector stakeholders and ultimately would improve technology adoption outcomes (Office of Public Health and Science Health and Human Services Department, 2011).

Improved Talent Recruitment and Public-Private Partnerships

The DoD can break down silos that impede growth by fostering a culture of multi-stakeholder collaboration through targeted outreach and training. The DoD can help create an ecosystem informed by and incorporating less-traditional stakeholders by leveraging training and awareness building programming that integrates diverse perspectives, methodologies, and knowledge bases, leading to a more comprehensive understanding of technologies and problem



areas. Such programming has the potential to help improve upon existing paradigms and solutions, foster unconventional ideas and research, and help both public and private sectors to overcome operational ruts.

Public-private partnerships help draw the talent needed to ensure a capable, innovative base for a resilient workforce, especially in areas such as AI where the private sector has adopted a clear leadership role in innovation. As outlined by the Under Secretary of Defense for Acquisition and Sustainment, recruiting a trained and skilled workforce to the defense industrial workforce requires public-private collaboration at all levels to build a robust talent pipeline, starting at the local levels (Clark, 2023). A key element of this is also ensuring that recruiting networks incorporate less-traditional stakeholders who can help foster appealing and welcoming environments.

Moreover, stakeholder initiatives and public-private partnerships can help new entrants across the defense innovation community make valuable contracts with government counterparts (see Figure 4). For example, events that focus on introducing small businesses and individual entrepreneurs to government and military contacts can help provide new opportunities for these stakeholders. Such events also benefit government participants as they are exposed to new businesses and technologies, as well as a new pool of potential talent for future recruitment.

Of respondents, **82%** felt that the initiatives introduced them to valuable contacts within the government or military

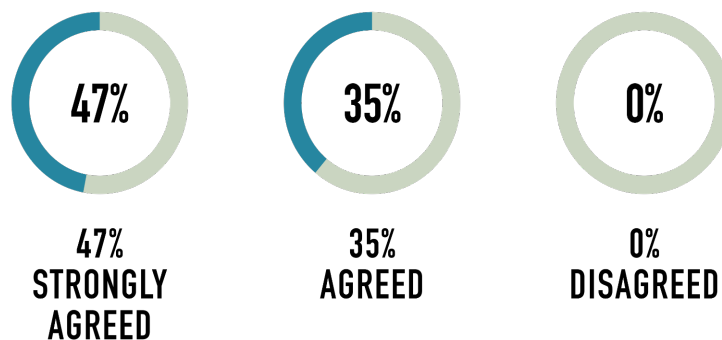


Figure 4. Highlighted Respondent Feedback⁵

Building Towards Cultural Shifts

Implementing changes to defense acquisition frameworks cannot be successful without first addressing the culture issues internal and external to the DoD that would undermine these efforts. Different stakeholders in the commercial acquisition pipeline face unique cultural challenges, but the most pressing and pervasive challenge is a culture of risk avoidance. In order to obtain the best innovative technologies for its warfighters, the DoD must embrace a culture of experimentation and risk-taking (Velte, 2023). This culture would promote continuous

⁵ In Figure 4, the 82% referenced combines both “strongly agreed” and “agreed.” The 0% referenced combines both “strongly disagreed” and “disagreed.”



improvement and creative problem-solving and enable the U.S. military to swiftly adapt to evolving challenges so that it may maintain an edge in the global power competition.

This culture of experimentation and risk-taking should be informed by a wartime-like sense of urgency. It is necessary that the DoD acquires innovative mission-critical technologies in a timely manner. Industry feedback in Becera's survey highlighted a need for greater agility, risk tolerance, and an acceptance of short-term failure within the DoD. It is important that stakeholders within the DoD take on the attitude that risk and short-term failures are a necessary part of a timely technology acquisition process, not something to be feared (DIB Strategic Investment Capital Task Force, 2023). The quicker a technology fails, the quicker improvements can be made to that technology or resources can be shifted to a different product. Long term, this minimizes lost time and funds and allows the DoD to obtain superior technology and resources (Metzger, 2023).

To enable this culture shift, the DoD should move away from excessive oversight, processes, and requirements that undermine timely technology adoption. Rather than the Pentagon, management closer to the field should be granted more flexibility and decision-making power, with safeguards in place to ensure prudence and accountability remain (DIB Strategic Investment Capital Task Force, 2023). DoD budgets and acquisition processes should be simplified and modernized to allow for greater risk-taking and adaptability to evolving end-user needs and technology capacities (Mcnamara et al., 2024).

Defense innovation stakeholders should be united by the shared mission of timely adoption of mission-critical technologies, coordinating and collaborating to ensure success (Kotila et al., 2023). This mission must apply to not just higher-level management, but also middle and lower-level leadership (Tucker, 2024). The DoD should strive to appoint leaders with an understanding of innovative technologies and military applications to avoid missing opportunities due to ignorance. This can be achieved through talent acquisition, training, and engagement with private sector businesses (Mcnamara et al., 2024; Clark, 2023).

Management within the DoD that embraces this culture and prioritizes results should be incentivized with increased budgets and other rewards. Conversely, risk-avoidant management that obstructs progress should face disincentives such as budget cuts and removal if they are unable or unwilling to adapt (DIB Strategic Investment Capital Task Force, 2023). Cultivating strong leadership united by this shared mission is crucial to ensuring the DoD market is conducive to innovation and that the U.S. military maintains its competitive advantage.

Moreover, fostering a collaboratorium bringing together more traditional and less-traditional stakeholders can help break down some of the barriers around communication and accessibility between the defense community and technologists who have never worked with the federal government. Becera survey respondents found that by engaging in TTDs, they were able to become more fluent in terms of what DoD expectations and processes meant. Individuals found such events help them to navigate the defense communities' complex bureaucracies more effectively.



Case Study: Technology Transfer Days

Technology Transfer Days (TTDs) are outreach and engagement events that create converged collaborative learning and discovery environments for stakeholders outside of traditional defense environments. As informal meetups, TTDs have focused on providing accessible, diverse, and inclusive environments for participants to collaborate, experiment, and network around complex defense innovation problems.

TTDs were originally inspired by conversations with stakeholders across the DoD, including from USCENTCOM, JIFX, NPS, USSTRATCOM, USSOCOM, and USINDOPACOM, in recognition of the value add in expanding the ecosystem of defense innovation participants. In addition to engaging the usual defense, government, academic, and industry stakeholders, TTDs have prioritized engaging participation from less-traditional stakeholders. These include small businesses, start-ups, incubators, accelerators, community organizations, subject matter experts, legal experts, veterans, inventors, nonprofits, and public-private partnerships.

TTDs are traditionally free, multi-day events held bi-annually or annually that convene around 150–200 participants in hybrid settings. The participants tend to be a mix from diverse technology sectors who leverage opportunities to mingle with each other and with the local, state, federal government and defense stakeholders also in attendance. A committee of volunteer community advisors serve as meetup leaders, mentors, and provide outreach amplification via their networks. Government and defense participants are invited through email invitations, networking, and relationship development, while other participants are invited through email invitations, extensive networking, and relationship development with local accelerators, incubators, startup groups, universities, research laboratories, and professional associations.

TTDs have been positively reviewed by hundreds of participants who have experienced the benefits of collaborative learning models for sharing experiences, building relationships, and accelerating development. Future iterations aim to build on successes of these models by: 1) scaling the outreach strategy to reach more technology stakeholders outside of traditional engagement circles, 2) incorporating pre- and post-event tech matching to augment in-person collaborative learning, and 3) increasing focus on requirements and needs for more experiments and research. Further information on TTD origins and development can be found in Appendix A.

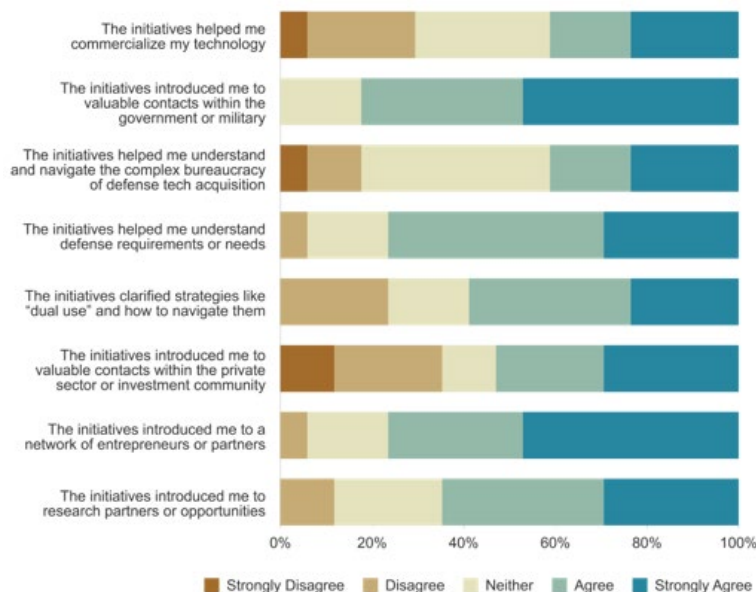


Figure 5. Respondent Feedback



Framework

How can a centralized Defense Innovation Discovery and Collaboratorium Platform help identify and overcome the valley of death?

No one solution can address the complex web of challenges in identifying, developing, and integrating cutting-edge defense technologies, which inherently requires discovering and engaging a diverse range of sources. However, guided by the literature, Becera survey responses, and stakeholder interviews, and building on the elements outlined previously, this section offers an operational and actionable framework for addressing stakeholder engagement gaps in a comprehensive manner. The framework specifically focuses on the elements necessary for a cross-functional discovery and collaboration platform, namely a DID Collaboratorium, that brings together traditional and less-traditional stakeholders in a secure, user-friendly environment. Such a platform would enable technology matchmaking, facilitate knowledge sharing, sustain a community of interest, and foster collaborative development of innovative solutions to critical defense challenges. Providing this DID Collaboratorium that reduces barriers to entry and accelerates the adoption of new technologies would be a complex but feasible endeavor.

In addition to the recommendations gleaned from Becera desk research, survey, and primary interviews, the following section incorporates lessons learned from the authors' previous experience building discovery and matchmaking platforms for defense-related technologies.

Creating a Defense Innovation Discovery and Collaboratorium Platform

- 1) Showcase technologies in a centralized location:** As several Becera survey respondents noted, a centralized platform where technologies can be discovered and showcased is essential for facilitating collaboration. Technology providers can fill out a profile page, sharing as much information as they're comfortable with. For those companies who choose to have fully public profiles, the platform can automate the process of keeping their profile pages up-to-date by pulling in updated information from their websites, social media, and news articles. To expand towards a fuller Collaboratorium experience, other stakeholders besides those with market-ready technologies can be offered a profile page. For example, stakeholders at academic or research labs, independent entrepreneurs working on prototypes, and venture capitalists or incubators can leverage the platform to showcase their projects, missions, or portfolios.
- 2) Identify stakeholders outside of traditional defense circles:** To broaden the pool of potential innovators and solutions, it is crucial to proactively identify and engage with stakeholders beyond the traditional defense industry. As Becera survey respondents suggested, this involves advertising opportunities to early career academic researchers, independent researchers, civil society organizations, non-profits, and the wider community. Advertisements should not be limited to defense-related publications, outlets, or forums. They should be posted broadly in non-defense channels to identify stakeholders with valuable solutions not traditionally engaged by the defense community. Proactive engagement also requires initiating more local events that bring together technologists, researchers, and others around specific defense tech topics. Organizing and implementing in-person events are valuable opportunities to build meaningful connection points, but virtual events can be equally beneficial, particularly when bolstered by extending an invitation to join the DID Collaboratorium and maintain new networks. As primary interview feedback revealed, engaging more less-traditional stakeholders and sustaining their participation will require thoughtful and deliberate



outreach, information delivered in an accessible format free of defense jargon, and clear communication of the value proposition for participating in defense innovation efforts.

- 3) **Match technologies to defense requirements:** To ensure that innovative technologies are effectively aligned with defense needs, Becera survey respondents emphasized the need to establish processes for centralizing information on the immediate, mid-term, and long-term needs of various defense programs and then matching solutions to those requirements. The DID Collaboratorium can automate the centralization of this information by pulling information from SBIRs, STTRs, Broad Agency Announcements, and other similar requirements documents. Requirements can then be matched to existing technology capabilities as identified in the platform, alerting relevant stakeholders to the existence of potential matches to explore. As input from primary interviews noted, to initiate a secure sharing of information, the platform can employ a double-blind feature, wherein technology provider information on capabilities is not linked to their name, nor are defense requirements detailed or linked to specific offices until the platform makes the match. In this context, the platform can leverage AI to facilitate this matchmaking process. Defense stakeholders can then create a tiered filtering and assessment system to determine which of the automated matches they want to pursue. For example, for those end-users who have established profiles and indicated which requirements relate to their missions, the first tier of automated matchmaking would send them a list with every technology showcased on the platform that is a 50% or above match. For the second tier of facilitated matchmaking, end-users could be prompted to send a brief questionnaire template to the technology providers they think best fit their parameters. Based on the responses, end-users can then schedule a call for further clarifications with the technology providers deemed as prospective fits. Every initial tier of matchmaking can be automated to reduce burden on both the end-user and technology provider.
- 4) **Link technology providers to program officers and end-users:** Facilitating direct connections between technology providers, program managers, and end-users is crucial for fostering collaboration, gathering feedback, and ensuring that innovations are tailored to operational needs. While this can be achieved through in-person meetups, virtual events, and other networking opportunities that allow small businesses and innovators to engage with defense stakeholders, it can be sustained through a directory feature on the DID Collaboratorium platform. To prevent program managers and end-users from being inundated with questions, these stakeholders can upload an FAQ to their profile pages that shares information like deadlines, processes, and preferred methods of engagement. If users type in questions for specific end-users and program managers that have been asked and answered previously, the platform can surface the relevant information. Questions that still remain can be sent directly to the program manager or end-user, who will then be prompted to update their individual FAQ accordingly or to respond to the inquirer directly. Aspects of this engagement feature can also be gamified to encourage responses. For example, the platform can show how quickly a program manager or end-user typically takes to respond to direct messages or update their FAQs upon receiving questions. The platform can then compare their response rates across their peers, thereby eliciting friendly competition to ensure that information is kept timely and up-to-date.
- 5) **Build a networked community of interest to share information, ideas, and contacts:** Fostering a vibrant, collaborative community of interest is essential for connecting silos, facilitating knowledge sharing, and catalyzing collaboration and innovation. Moreover, all stakeholders require clear, accurate, and up-to-date



information about available programs, requirements, and processes. As highlighted in Becera survey feedback, both goals can be achieved through the DID Collaboratorium platform on a specific information forum that offers guidance, FAQs, and resources on topics such as dual-use technologies, acquisition pathways, and security requirements. Through such a forum, users can ask questions, post lessons learned, and share information on events and opportunities. The forum would also enable direct chats with other users who opt in to having their individual profiles shared publicly. The information forum can then incorporate features like upvoting and downvoting to separate signal from noise and surface the most helpful responses and information. It can also incorporate a feedback mechanism whereby users can indicate what further information would help clarify certain topics. A comprehensive and dynamic forum of this nature provides the sense of continuity between stakeholders, resources, and organizations that one primary interviewee said was essential for defense innovation efforts. It also addresses some of the engagement limitations that exist in current defense innovation platforms like Vulcan, as primary interviews revealed.

- 6) **Facilitate a flexible funding pool:** To support the development and maturation of promising technologies, it is essential to establish a flexible funding pool that can close transition gaps and prevent innovations from stalling in the valley of death (Kotila et al., 2023). Becera survey respondents suggested that funding can be made in small amounts to mitigate risk aversion and should be accessible through a simplified application process, with initial funding separated from classified activities to minimize bureaucratic hurdles. Leveraging AI, the DID Collaboratorium can also provide an aggregate of other funding opportunities such as U.S. federal grants to help stakeholders identify and navigate relevant opportunities that otherwise would take additional resources to find. The platform can also provide resources to provide clarity on some of the nuances of federal grant applications and help break down some of the barriers identified around communication and clarity around funding processes. As Becera survey respondents noted, providing timely, risk-tolerant funding that accepts short-term setbacks as learning opportunities will enable the defense innovation ecosystem to help promising technologies reach their full potential and transition into operational use.
- 7) **Provide professional development opportunities:** Semantic, cultural, and technical differences yield misunderstandings and result in valuable collaborations being lost in translation. A comprehensive platform should offer professional development opportunities that enable policymakers to become more technical and technical stakeholders to become more policy- and business-aware. The collaboratorium would therefore offer a collaborative learning environment where professional associations and academic institutions could provide their various credentialing opportunities in emerging technology learning opportunities. The platform can also offer technology exercises and wargaming scenarios to bring together technologists and the defense community, helping to operationalize any collaboration that arises from these professional development opportunities. Another feature could be a dedicated section for talent acquisition, a key gap identified in research. Separate but related to professional development, a job opportunities and candidate forum can offer a centralized place for individuals from across the defense innovation ecosystem to understand what opportunities exist in multiple sectors.
- 8) **Measure outputs and outcomes:** The benefit of a centralized platform like the DID Collaboratorium is the ability to track progress in real time. Transparent efforts to ensure that defense innovation and engagement efforts are delivering tangible results will



encourage continued use of the platform while building trust in the process. Becera survey respondents suggested that a centralized platform should measure metrics including tracking the success of dual-use companies in finding long-term success in the defense ecosystem, demonstrating the impact of new capabilities on mission needs, and understanding what factors contribute to the successful integration of novel technologies into operational use cases. Rigorously evaluating the performance of defense innovation initiatives and sharing insights with stakeholders allows the ecosystem to continuously improve its effectiveness and adapt to evolving challenges and opportunities.

Conclusion

Research and practice make clear that the defense innovation ecosystem faces significant challenges in effectively identifying, developing, and integrating cutting-edge technologies from a diverse range of sources. In addition to traditional defense acquisition processes struggling to keep pace with rapid technological advancements, methods of stakeholder engagement are excluding those outside of traditional defense circles. This is resulting in missed opportunities to leverage potentially game-changing innovations and leaves the defense sector at risk of falling behind in an increasingly complex and competitive global landscape. Existing innovation practices and the lessons they have yielded offer much hope, however. Building on the realities of continuous and dynamic innovation, a DID Collaboratorium offers a broadly encompassing platform that convenes the people, processes, and procedures necessary to ensure an effective defense technology ecosystem.

References

- Albon, C. (2024, January 17). *Pentagon rewrites space classification policy to improve info-sharing*. C4ISRNet. <https://www.c4isrnet.com/battlefield-tech/space/2024/01/17/pentagon-rewrites-space-classification-policy-to-improve-info-sharing/>
- Allen, G. C. (2023). *Six questions every DOD AI and autonomy program manager needs to be prepared to answer*. Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/2023-05/230515_Allen_Six_Questions.pdf?VersionId=zdOM8qS30hmQRV_4JjbPTMzLukCCm3Bk
- Blank, S. (2024, January 5). *The DoD is getting its innovation act together, but more can be done*. Defense News. <https://www.defensenews.com/opinion/2024/01/05/the-dod-is-getting-its-innovation-act-together-but-more-can-be-done/>
- Bresler, A., & Bresler, A. (2020). *The effect of defense-sponsored innovation programs on the military's industrial base*. The Acquisition Research Program of the Graduate School of Defense Management at the Naval Postgraduate School. NPS Archive: Calhoun. <https://calhoun.nps.edu/bitstream/handle/10945/64763/SYM-AM-20-059.pdf?sequence=1&isAllowed=y>
- Bresler, A., & Bresler, A. (2023). *Assessing the effectiveness of defense-sponsored innovation programs as a means of accelerating the adoption of innovation force wide*. Naval Postgraduate School. <https://dair.nps.edu/bitstream/123456789/4868/1/SYM-AM-23-101.pdf>
- Clark, J. (2023, December 3). *Public-private partnership is key to building defense industrial base and workforce*. DoD. <https://www.defense.gov/News/News-Stories/Article/Article/3604968/public-private-partnership-is-key-to-building-defense-industrial-base-and-workf/>



- Decker, J., & Sheinbaum, N. (2024). Shining a light on the Defense Department's industrial base problems. *The Texas National Security Review: The Strategist*, 7(1), 75–88. The University of Texas ScholarWorks. <https://doi.org/10.26153/tsw/50673>
- Defense Innovation Board. (2024). *Lowering barriers to innovation*. DoD. https://innovation.defense.gov/Portals/63/2_%2020240118%20DIB%20LBI%20Study.pdf?ver=dkaQFfnZxzd2EoKaKfIUDg%3d%3d
- Defense Innovation Board Strategic Investment Capital Task Force. (2023). *Terraforming the valley of death: Making the defense market navigable for startups*. Defense Innovation Board, DoD. https://innovation.defense.gov/Portals/63/DIB_Terraforming%20the%20Valley%20of%20Death_230717_1.pdf
- Ferry, J. (2024, January 18). *How to sell technology to the government*. Rogue. <https://www.userogue.com/how-to-sell-technology-to-the-government>
- Husted, K., Finchum-Mason, E., & Suárez, D. (2021). *Sharing power? The landscape of participatory practices & grantmaking among large US foundations*. Evans School of Public Policy and Governance, University of Washington. <https://evans.uw.edu/wp-content/uploads/2021/07/Participatory-Practices-and-Grantmaking-Report-June-2021.pdf>
- Kotila, B., Drezner, J. A., Bartels, E. M., Hill, D., Hodgson, Q. E., Huilgol, S. S., Manuel, S., Simpson, M., & Wong, J. P. (2023). *Strengthening the defense innovation ecosystem*. RAND Corporation. <https://doi.org/10.7249/rra1352-1>
- Lawrence, S. S. (2023, June 29). *Tackling technology's valley of death*. Signal; AFCEA International. <https://www.afcea.org/signal-media/defense-operations/tackling-technologys-valley-death>
- Leonard-Barton, D., & Kraus, W. A. (1985, November). Implementing new technology. *Harvard Business Review*, 63(6). <https://hbr.org/1985/11/implementing-new-technology>
- Marinelli, B. (2023, May 17). *A better bridge across the "valley of death."* Defense News. <https://www.defensenews.com/opinion/commentary/2023/05/17/a-better-bridge-across-the-valley-of-death/>
- Mcnamara, W., Modigliani, P., Macgregor, M., & Lofgren, E. (2024). *Commission on defense innovation adoption*. Atlantic Council. <https://www.atlanticcouncil.org/wp-content/uploads/2024/01/Commission-on-Defense-Innovation-Adoption-Final-Report.pdf>
- Metzger, M. (2023, March 1). *There isn't just one valley of death: Tackling the DOD transition problem*. Defense Scoop. <https://defensescoop.com/2023/03/01/there-isnt-just-one-valley-of-death-tackling-the-dod-transition-problem/>
- National Science Board. (2024). *The state of U.S. science and engineering 2024*. National Center for Science and Engineering Statistics Directorate for Social, Behavioral and Economic Sciences of the National Science Foundation. <https://nces.nsf.gov/pubs/nsb20243>
- Office of Public Health and Science Health and Human Services Department. (2011). *Principles of community engagement* (2nd ed.). National Institutes of Health. <https://www.govinfo.gov/app/details/GOVPUB-HE20-PURL-gpo15486>
- Tucker, P. (2024, January 26). *Innovation needs to go more than skin-deep at the Pentagon, advisory board says*. Defense One.



<https://www.defenseone.com/technology/2024/01/innovation-needs-go-more-skin-deep-pentagon-advisory-board-says/393649/>

Velte, E. (2023, November 14). Barriers to AI and ML: A roadmap to federal adoption. *Forbes*. <https://www.forbes.com/sites/forbestechcouncil/2023/11/14/barriers-to-ai-and-ml-a-roadmap-to-federal-adoption/?sh=456fd4c858fb>

Vincent, B. (2023, May 18). *"It's about the DOD" — New study explores obstacles hindering Pentagon's AI adoption*. Defense Scoop. <https://defensescoop.com/2023/05/18/its-about-the-dod-new-study-explores-obstacles-hindering-pentagons-ai-adoption/>

Weisner, M. (2023, October 26). *DoD hindering recruitment of tech-savvy workers, warfighters: Report*. Federal Times. <https://www.federaltimes.com/management/career/2023/10/26/pentagon-practices-harm-recruitment-of-tech-savvy-workers-warfighters/>

The White House. (2022, February 15). *Fact sheet: Department of Defense releases new report on safeguarding our national security by promoting competition in the defense industrial base* [Speeches and remarks]. <https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/02/15/fact-sheet-department-of-defense-releases-new-report-on-safeguarding-our-national-security-by-promoting-competition-in-the-defense-industrial-base/>



Agile Development of Hardware-Reliant Systems

Tom McDermott— is the Deputy Director of the Systems Engineering Research Center at Stevens Institute of Technology. [tmcdermo@stevens.edu]

Douglas Buettner—is a Deputy Chief Scientist with Stevens Institute of Technology. [dbuettne@stevens.edu]

Abstract

The Acquisition Innovation Research Center (AIRC) and Software Engineering Institute (SEI) sponsored a workshop on Agile Development of Hardware-Intensive Systems. This paper summarizes the results from that workshop and further discusses concepts of agile and DevSecOps in defense acquisition programs that include hardware-intensive development activities. Every technological revolution triggers changes in how work is done and managed, as well as how people are managed. As a result, traditional approaches become insufficient for addressing new problems. Commercial industry has adopted agile practices in software, hardware/software systems, and services to address rapidly changing threats to and opportunities in their business. The Department of Defense (DoD) now faces similar external drivers and must move to agile practices across all acquisition processes and functions. Workshop speakers and participants identified 10 primary themes that should scope transformation of future defense acquisition for all types of systems, not just software.

Research Issue

Rapidly advancing threats and technologies have increased the need for the U.S. Department of Defense (DoD) to develop, field, and upgrade operational capabilities to ensure mission effectiveness and success more quickly. Agile development along with DevSecOps (development, security, and operations) can accelerate acquisition and improve relevance. Industry has successfully applied agile and DevSecOps to software, hardware, and inter-reliant hardware/software systems. The DoD has embarked on this journey—but primarily only for software systems. However, continuous innovation and deployment are a total system concern and involve hardware components in a system as well as software, business process, funding, and all other human-oriented intangible components.

Agile enterprises recognize that deploying new systems or capabilities cannot wait on the slowest components of the system. Instead, all components need to be deployed when ready, and both systems and organizations need to be structured to support modularity and flow. All organizational processes that might delay these components must also make the shift, including program management, contracting, test & evaluation, sustainment & logistics, and financial management.

The DoD has struggled to make the shift to agile. Over years of employing more sequential approaches, the Department, like other organizations, has created siloed organizations responsible for one part of the process with movement to a different stage (silo) triggered by full completion of the activity, coupled with large testing events at the end of development. Alternatively, the core principle behind agile is “flow.” The flow of work should continue consistently across cycles of product strategy, resources, product development and test, and product support. With digital transformation, the DoD can reduce phase durations and cycle times in all phases of development and acquisition. By integrating agile and DevSecOps initiatives with Digital Engineering, the DoD can improve flow, allowing components to react more quickly to changing end-user demands. Agile practices for hardware-intensive systems in Digital Engineering and Digital Acquisition are a primary driver to agile product development, along with focused adaptations in acquisition practices and policy in support of agile transformation.



Research Results

The Acquisition Innovation Research Center (AIRC; 2023) and Software Engineering Institute (SEI) convened a workshop on April 18–19, 2023, focused on Agile Development of Hardware-Reliant Systems. The workshop was co-sponsored by the Office of the Under Secretary of Defense for Acquisition and Sustainment and the Office of the Under Secretary of Defense for Research and Engineering. Forty-eight experts, including 17 speakers from government, industry, and academia, gathered together for a working discussion and sharing of best practices, lessons learned, challenges, and progress. The stated goal was to develop a set of foundational practices and research vectors relating to hardware-related agile, DevSecOps, and Digital Engineering/Acquisition to provide practical advice to programs in applying these techniques to both hardware and software elements of acquired systems while facilitating workforce training and improvement.

Commercial industry has adopted agile practices in software, hardware/software systems, and services to address rapidly changing threats to and opportunities in their business. The DoD now faces similar external drivers and must move to agile practices across all acquisition processes and functions (Vesonder & Hutchison, 2022). Workshop speakers and participants identified 10 organizational goals that should scope transformation of future defense acquisition for all types of systems, not just software. Organizations and programs undergoing digital transformation of their acquisition and engineering processes should work to adopt these goals to drive their transformations.

- 1. Shift Learning to the Left.** This should be added as the 13th principle to the 12 Agile Principles (Manifesto for Agile Development, n.d.). In hardware-reliant systems, agile practices augmented by digital models, prototypes, and test infrastructures help bring learning forward, reduce integration risks, and create more flexibility in long-term design decision points. The workshop highlighted the value of knowledge and learning in the development process. Results emphasized the need to capture and share knowledge, as well as the importance of gaining insights and feedback at various stages to improve the final product. Digital models are the primary knowledge sharing tools, and organizations must manage data and models as a means to truly experiment and prototype in the digital realm to accelerate learning as far left as possible.
- 2. Design for Change.** Intentionality in the early design stage of hardware-reliant systems to accommodate innovation in later stages of product development is an enabler of agility. Related concepts are design for iteration and flexibility. Too often in the DoD, designing to maximize performance and satisfaction of performance requirements results in systems that are inflexible to iteration and change. Choosing the elements of the system to emphasize in this strategy helps anticipate evolution of components that have the most potential for change late in cycle or those for which innovative change will have the most pronounced performance gain. Modularity must be a primary design driver. The concept of flexibility in design along with digital models promotes a strategy of “don’t decide until you have to” lock in hardware design components.
- 3. Design for Flow.** The current inefficiency of the Defense Acquisition System (DAS) could be improved by broad adoption of a few underlying premises of agile: create direct collaboration between users and developers, encourage simplicity, and create continuous flow of value. This is a DoD enterprise-level shift independent of hardware or software acquisition. Agile is fundamentally an approach that seeks to improve process flow. Currently in the DAS, the flow from warfighter need to capability acquisition passes through many organizations and processes before it becomes an acquisition program (of any type). This changes interpretation of needs and requirements, isolates the real customer from the capability development and interrupts the flow of work from need to capability. A further barrier to flow is the transactional nature of DoD acquisition, which can disrupt consistency and interrupt flow. Modular acquisition



practices help here but are rarely used. The workshop highlighted the importance of directly aligning ambitions and efforts of the users and developers with the actual capabilities required in the system, which in agile principles is called encouraging simplicity. Connecting Combatant Commands and their priorities directly to the acquisition process is one strategy noted at the workshop. The DoD requirements process often runs counter to this, preferring to aggregate needs and requirements into larger acquisition efforts. Design for flow requires that every function in the DAS and the primary contractor processes be aligned to maximize flow rates from warfighter need to capability delivery.

4. Overcome the Single Batch Mindset. The historical Acquisition Category (ACAT) I acquisition process remains ingrained in a waterfall mentality, even though alternative pathways are available. The workshop addressed the challenges of the single batch mindset and the belief that everything must be understood before implementation. Instead, one is encouraged to find ways to overcome these barriers and adopt a more flexible and adaptive approach. This implies an enterprise-level shift to allow more frequent delivery of working systems (or system elements) through reconciliation of development and delivery cycles for best effect. Rather than compounding the effect of slower cycles that drive the pace of system-level delivery, a refactoring of the contributing streams of work can assure flow enabled by smaller batches of work. Milestone completion remains important but must be translated into buying down risk, not just criteria completion. Integrating both a consistent work cadence and milestone-driven goals are critical to agile in hardware-intensive systems. A warning: The fact that the Software Acquisition (SWA) Pathway specifically waives ACAT I designation even for large software programs is a signal that the historical DAS has a single batch mindset to overcome. There is a danger that agencies using Middle-Tier Acquisition (MTA) and Software pathways may just bypass onerous ACAT I milestone approval processes and damage the flexibility granted with these other pathways. The concept of “tailoring-in” instead of “tailoring-out” regulatory acquisition requirements based on need was discussed. A better approach is to make all pathways more efficient using agile principles.

5. Decomposition and Partitioning. The workshop noted the concept of decomposing capabilities and finding clever ways to partition them. Breaking down complex systems into smaller, manageable components allows for faster learning and better understanding of individual elements. Agile practice takes advantage of modularity to architect systems that can be evolved over time. Control of interfaces and application program interfaces (APIs) is fundamental to both defining the work in the system and the team skills needed to do the work. Modular Open Systems Architecture (MOSA) precede agile development in both software and hardware systems, as well as the infrastructure used to develop them.

6. Deliver Working Software Frequently. The Agile Manifesto focuses on delivery of working software as the primary measure of progress. All systems including hardware-intensive systems today are software intensive, so programs should continually deploy and redeploy working software into everything they do. Meaningful movement of prototypes from virtual environments to physical realizations to operational use has tangible benefits when the software is reused from one product to another. Programs should embed deployable software into simulation and training systems, allowing all developers and users to experience the operational use of the product.

Figure 1 depicts this concept.



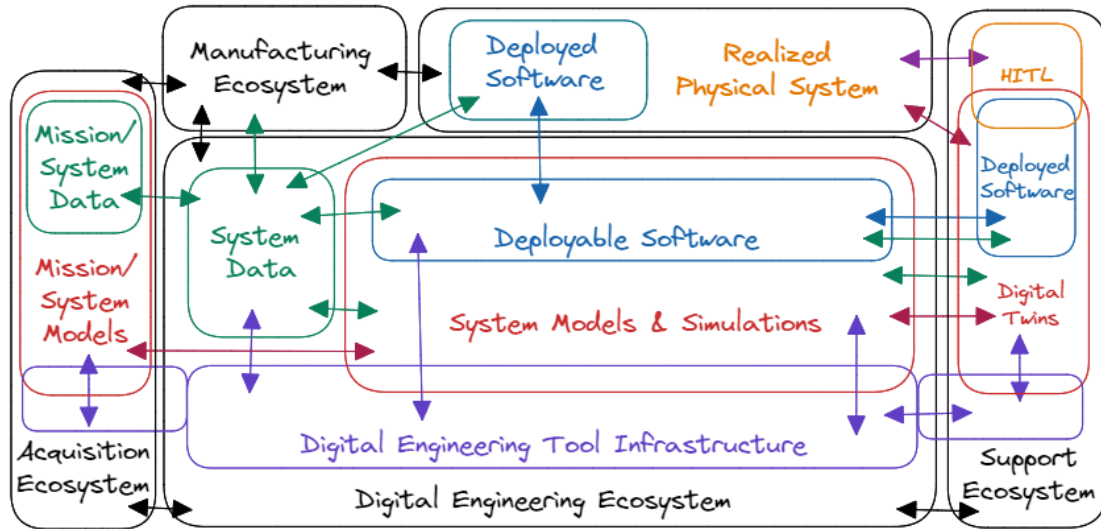


Figure 1. Today's Systems Should Plan for Deploying Working Software Frequently Into All Aspects of Program Planning and Development

7. Hardware-Intensive Agile Requires Front-End Investment. In his paper “Managing the Development of Large Software Systems,” Win Royce (1987) introduced the waterfall model and noted its fundamental flaw: Testing is at the end; therefore, flaws in the design are not identified until the end. Agile in hardware-intensive systems requires front-end investment in test activities and infrastructure to buy down end-item risk. For example, SpaceX’s™ investment in and experience of learning from multiple launch failures is an example of the culture and mindset required for innovation and continuous improvement. The workshop highlighted the value of automating as much as possible while also being mindful of the cost and benefits of automation. Investment in model-based engineering tools, multiple systems-level prototypes, and hardware-in-the-loop environments will be critical for “shifting left” to successful agile implementation in hardware-intensive systems. However, return on investment (ROI) is not easily quantifiable up front.

8. Configuration Management and Branching Strategy. The concept of branching – independent lines of work that stem from a central design – is a practice in both software systems and models. Intentionally integrating branching strategies into simulations, test articles, certification articles, and manufacturing systems is a necessary strategy in hardware-intensive systems. However, branching can lead to configuration management failures and must be implemented appropriately. This is a practice that needs more exploration and lessons learned.

9. Managing the Digital Infrastructure. Organizations need to have dedicated persons or groups to manage integration of their digital tool infrastructures. This is more difficult in hardware-intensive programs because the tools are more diverse and less well-integrated than in today’s software/DevOps environments. Modern tool infrastructures for hardware-intensive systems may also integrate manufacturing systems, 3D printers, robotics, and associated digital engineering tools. Employing a dedicated data analytics team to monitor tool effectiveness and improvements is necessary.

10. Continuous Focus on the Workforce. There is a need to continuously train the entire workforce on agile principles. Much of the DoD organic workforce is familiar with milestone-driven development practices but needs continuous indoctrination into agile methods. The workshop acknowledged the importance of structuring and organizing responsibilities into roles different than those in traditional development. The workshop emphasized the

interconnectedness of individuals, knowledge, and the system being developed, and the need to assemble expertise in specific roles. The value of agile training, independent of selected methodology, is the mindset shift to new roles and ways of doing business.

In summary, these 10 goals can form a checklist for organizations to plan and measure their organizational agile transformation. In DoD acquisition, the government and lead contractor should jointly evaluate these goals. In checklist form, these are:

- () We have integrated digital modeling and simulation with our selected agile methodologies and deployed a test infrastructure to shift learning to the left.
- () We have specified and organized our system architectures for iteration and flexibility as Key Performance Parameters and trained our design teams to consider these qualities at every level of design.
- () We have redesigned all of our organizational processes necessary to deliver capabilities to remove all waste and delay and maximize flow of value to the end user (warfighter).
- () We have redesigned our technical review and certification processes to emphasize flow over batch size and are integrating smaller in-process reviews with larger system/capability reviews to balance capability delivery with appropriate certifications.
- () We have systems engineering processes that embrace modular open systems approaches and work decomposition strategies that emphasize small teams and faster learning, as appropriate to the system.
- () We have created a “software first” mentality in our organizations and are developing and testing critical software capabilities at every stage of design in every deliverable product.
- () We have invested in the digital infrastructure, high-fidelity modeling and simulation tools, and test automation strategies necessary to move testing to the front-end of every component development flow.
- () We have mature digital configuration management approaches allowing design teams to iterate at their level and merge design iterations into the full system design.
- () We manage the digital infrastructure and associated processes and tools as integral to system and product design and employ the necessary data management and analytics teams to maintain flow across our digital tools as well as extract progress.
- () We have deployed appropriate training and incentives to shift the mindset of everyone in the organization from “single batch” to “continuous flow.”

The leading conclusion reached by the participants was that the agile development of hardware-reliant systems is not only possible but is being done today! A number of commercial hardware-intensive product development companies today are “born digital and agile” and are paving the way with lessons learned. It will be more difficult for the DoD and defense industrial base to “become digital and agile,” but this must be a goal of the whole DAS. Agile development of hardware-reliant systems requires a different mindset. Elaborating requirements in periodic demonstrations of new capabilities, with a notable tolerance of early learning failures, is needed. Learning appropriately from short, iterative development cycles that focus on testing with frequent user feedback, can deliver a core set of essential capabilities to warfighters with the rest of the system elements following the initial minimum viable system versions. This does not mean that a critical test and certification of a new hardware element should be shortchanged,



just that the flow of all other components should be organized to continue in digital simulations and prototypes and merge into that test and certification process at the appropriate time.

Fundamentally, the approach “shift learning to the left,” which could be considered a new principle for the agile manifesto, was considered essential for a hardware agile manifesto. Finally, decreasing the “acquisition distance” between the warfighter (combatant commands) and this type of capability acquisition would be an essential enabler towards a more agile practice for hardware-reliant systems.

References:

Acquisition Innovation Research Center. (2023, April 18–19). *Agile development of hardware-reliant systems*. Research Workshop. <https://acqirc.org/events/agile-development-of-hardware-reliant-systems-workshop/>

Manifesto for Agile Development. (n.d.). *The 12 principles behind the agile manifesto*. <https://agilemanifesto.org/principles.html>

Royce, W. W. (1987). Managing the development of large software systems: Concepts and techniques. *Proceedings of the 9th International Conference on Software Engineering (ICSE '87)*, 328–338.

Vesonder, G., & Hutchison, N. (2022, October). *Agile acquisition: History and recommendations*. Acquisition Innovation Research Center. https://document.acqirc.org/technical_reports/reports/1685721563.WRT-1049.3_REPORT_0428.pdf.



Large Language Model (LLM) Comparison Research

William M. Fisher—is a data science fellow in the Cost Analysis and Research Division at the Institute for Defense Analyses (IDA). He has a background in applied math and data analytics and has worked on projects on large language models and space policy research. He has a BS in Computational and Applied Mathematics from the University of Chicago.. [wfisher@ida.org]

Nicholas Wagner—is a research staff member in the Cost Analysis and Research Division at the Institute for Defense Analyses (IDA). He specializes in data analytics and machine learning and has experience with projects in personnel, force planning, maintenance, contracts, health care, and AI policy, along with some IDA data governance on the side. He is currently working remotely out of Arizona. He has a BS in Materials Science and Engineering from Arizona State University and a PhD in Materials Science and Engineering from Northwestern University. [nwagner@ida.org]

Kevin Garrison—is a research staff member in the Information Technology and System Division at the Institute for Defense Analyses (IDA). He specializes in data and analytics for the defense department, acquisition oversight, cybersecurity, AI, and machine learning. He previously worked as the Chief of Analytics for the Office of the Department of Defense Chief Information Officer. He has a BS in Astronautical Engineering from the U.S. Air Force Academy. [kgarriso@ida.org]

Abstract

Over the past few years, large language models (LLMs) have rapidly increased in capability, with OpenAI's GPT-4 being the most prominent example. This case study explores two ways that GPT-4 could be used to assist research tasks: data analysis and writing executive summaries. We chose these tasks because they are common to Institute for Defense Analyses (IDA) projects and because they are often presented as tasks appropriate for LLMs. First, we used GPT-4 to conduct tasks such as data cleaning, exploration, modeling, and visualization. We compared the quality and speed to a human doing the same task. We found analysis quality was insufficient when utilizing AI alone, but improved greatly with a human partner. Using GPT-4 saved about 60% of the time on the data analysis assignment and presents an opportunity for significant cost savings in this area. Then, we used the GPT-4 to generate executive summaries (EXSUMs) for three publicly available IDA publications, and we compared these to the human-generated EXSUMs. We found that the LLM-generated EXSUMs often failed to provide appropriate context for more technical papers, but that given the speed that they are generated and their thoroughness, LLMs still present time- and cost-saving opportunities.

Introduction

As large language models (LLMs) have improved and increased in capability over the past few years, many organizations are examining how they can be employed to enhance research productivity. For IDA researchers, an LLM could be used to automate and/or assist tasks such as data analysis and writing executive summaries (EXSUMs). At the moment, the most prominent LLM developer is OpenAI, who released their most recent model, GPT-4, in March 2023. In early July 2023, OpenAI added a new feature to GPT-4 called Code Interpreter, which was then renamed GPT-4.¹ This feature, which is now automatically enabled for all GPT-4 chats, can write and execute Python code, and allows users to upload files and ask it to analyze data, create charts, and edit the files in place. The limitations include a strict time limit for executing code and an inability to save work beyond the immediate session.

In this analysis, we examined these features of GPT-4 to determine the possible time and cost savings of using an LLM for data analysis and writing EXSUMs. First, we measured how long it took both a human and GPT-4 to conduct various data analysis tasks, such as

¹ ChatGPT Plus subscribers can access GPT-4 at <https://chat.openai.com/?model=gpt-4-code-interpreter>, and they can learn more at <https://www.nytimes.com/2023/07/11/technology/what-to-know-chatgpt-code-interpreter.html>.



exploration, modeling, and visualization, which provided quantitative measures of time to perform non-CUI data analysis tasks. We then compared analysis quality. For writing EXSUMs, we used the GPT-4 API to generate summaries for publicly available IDA publications and compared the output to the existing EXSUMs for the publications. Through this comparison process, we can determine how much of a quality difference there is in the summaries and how much time a researcher can save by using an LLM to summarize a research paper rather than doing it entirely unassisted.

Dataset

We used two datasets for this analysis. The first comes from Kaggle, a public website, and describes Airbnb activity in New York City during 2023 (Kumar, 2023). The data includes information on prices, reviews, locations, and more. We chose this dataset because it has a variety of options for exploratory analysis and modeling, while being simple enough to ensure we would be able to easily understand and critique the decisions made by GPT-4. We also used a dataset from the Department of Defense Office of Local Defense Community Cooperation on state-by-state spending as a defense-specific example (OLDCC, n.d.). This dataset was used to focus more on GPT-4's data visualization capabilities.

Data Analysis Comparison Methods

To determine how much time GPT-4 can save and how it compares in terms of the quality of its work, we performed data analysis operations on the dataset ourselves before turning to the LLM. For each of the following sections, we briefly describe the work we did with that done by GPT-4 and compare the results. The comparison is based on the estimated time taken for data analysis, as well as the time it would have taken to do everything GPT-4 did. This comparison is not perfect, as GPT-4 sometimes went beyond what we were asking or required steps to be run locally, but the work was similar enough to provide estimates of the time GPT-4 can save on these tasks. The code for this section and the EXSUM generation section can be found at https://code.ida.org/users/wfisher/repos/llm_crp_code/browse.

Data Cleaning and Exploratory Analysis

1. Human Task Details

We began data cleaning and exploratory analysis by reviewing the different variables in the dataset to ensure we understood them and agreed with their associated data types. We then performed standard data cleaning operations, such as removing duplicate rows, checking for columns with mixed types, and removing rows with null values. After this, we conducted exploratory analysis by checking the summary statistics and distributions of the numerical variables, as well as unique values of the categorical ones, to better understand the data. We also explored how price varied across different neighborhoods and determined which ones were the most and least expensive.

2. GPT-4 Task Details and Prompts

We began the second half of the data analysis exercise by using GPT-4 to clean the data and perform an exploratory analysis on the dataset, which allowed us to check for data quality issues and understand the nature of the data it was working with before moving on to modeling and visualization.

Our prompt was, "Clean the data and conduct exploratory analysis on the nature of the different variables in the dataset." We used this prompt because these are often the first steps for a human working with a new dataset and to understand how GPT-4 would respond to broad terminology like "exploratory analysis." In response, it looked for three things in particular: columns with missing values, columns with mixed types, and duplicate rows. GPT-4 quickly removed duplicate rows, as well as a column containing license codes that did not appear to be



useful given the amount of missing values and mixed types. However, it inserted zeroes for missing values in the “last_review” and “reviews per month” columns, which is questionable, particularly in regard to using those variables to model something like rental price. For example, more popular, and therefore more expensive, rentals may be reviewed more often than less popular ones, meaning that there could be a relationship between the “last_review” and “reviews per month” variables and rental price that is obscured by inserting zeroes for missing values rather than using something like the mean for the neighborhood.

As far as exploratory analysis, GPT-4 explored descriptive statistics and correlations between numerical variables and frequency distributions for categorical variables. It summarized its main findings in bullet points, which are shown in Figures 1 and 2.

Categorical variables:

- The dataset contains 41406 unique listing names. The most common listing name is 'Water View King Bed Hotel Room'.
- There are 9831 unique hosts. The most common host name is 'Blueground'.
- Listings are spread across five neighbourhood groups, with Manhattan being the most common.
- There are 223 unique neighbourhoods, with Bedford-Stuyvesant being the most common.
- Among the types of rooms, 'Entire home/apt' is the most common.

Figure 4. GPT-4 Findings on Categorical Variables

Numerical variables:

- The average price per night for a listing is about \$200, but the range is vast, going from \$0 to \$99,000. This large range and the high standard deviation suggest that there are some extreme values (outliers) in the price.
- The average minimum number of nights is around 18, but it goes up to 1250, suggesting there are some long-term listings.
- On average, a listing has about 26 reviews. The maximum number of reviews a listing has received is 1842, indicating some very popular properties.
- The 'calculated_host_listings_count' shows that, on average, a host has around 24 listings, but the maximum goes up to 526, suggesting the presence of big players or property management companies.

Figure 5. GPT-4 Findings on Numerical Variables

The bullet points in the two figures provide useful information, such as the skewed distribution of the price, minimum nights, number of reviews, and the most common neighborhood. GPT-4 also makes a significant inference from this exploratory analysis in noting that the skewed number of listings for hosts may indicate the presence of significant real estate figures and property management companies. It also produced histograms for certain



variables, shown in Figure 3 with comments in Figure 4, and calculated correlations between variables, shown in Figure 5 with comments in Figure 6.

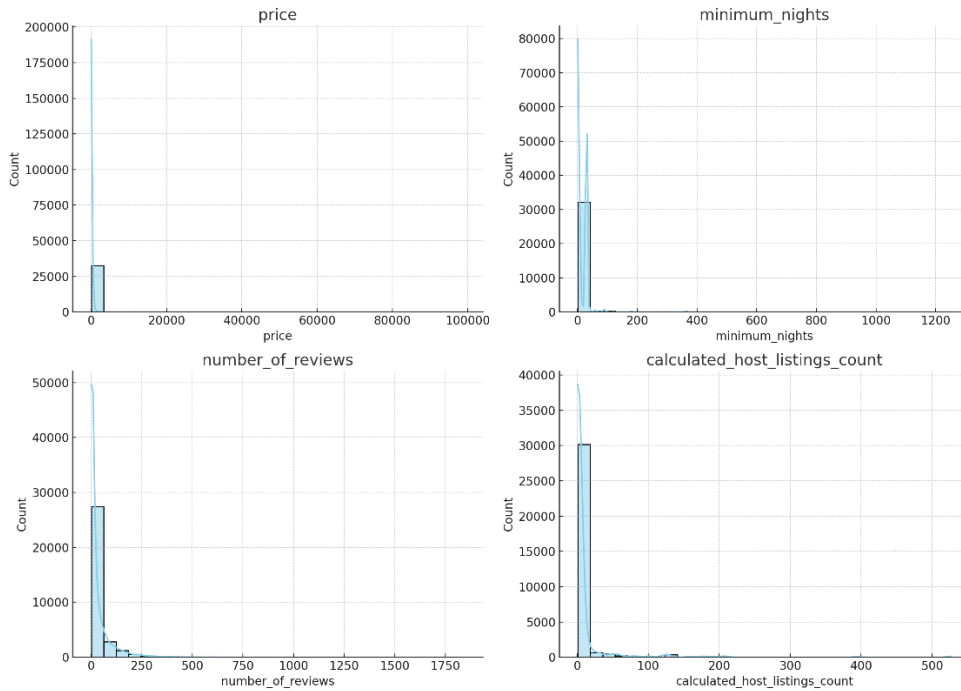


Figure 6. GPT-4 Histograms of Select Numerical Variables

Histograms:

- The `price` histogram shows that the vast majority of prices per night are below \$10,000, with a few extreme outliers above that. This is consistent with the statistics we saw earlier.
- The `minimum_nights` histogram reveals that most listings require a stay of fewer than 200 nights, but there are some that require a much longer minimum stay.
- The `number_of_reviews` histogram indicates that most listings have fewer than 200 reviews, but a few have many more.
- The `calculated_host_listings_count` histogram shows that the vast majority of hosts have fewer than 100 listings, but a few hosts have a significantly higher number.

Figure 7. GPT-4 Comments on Histograms



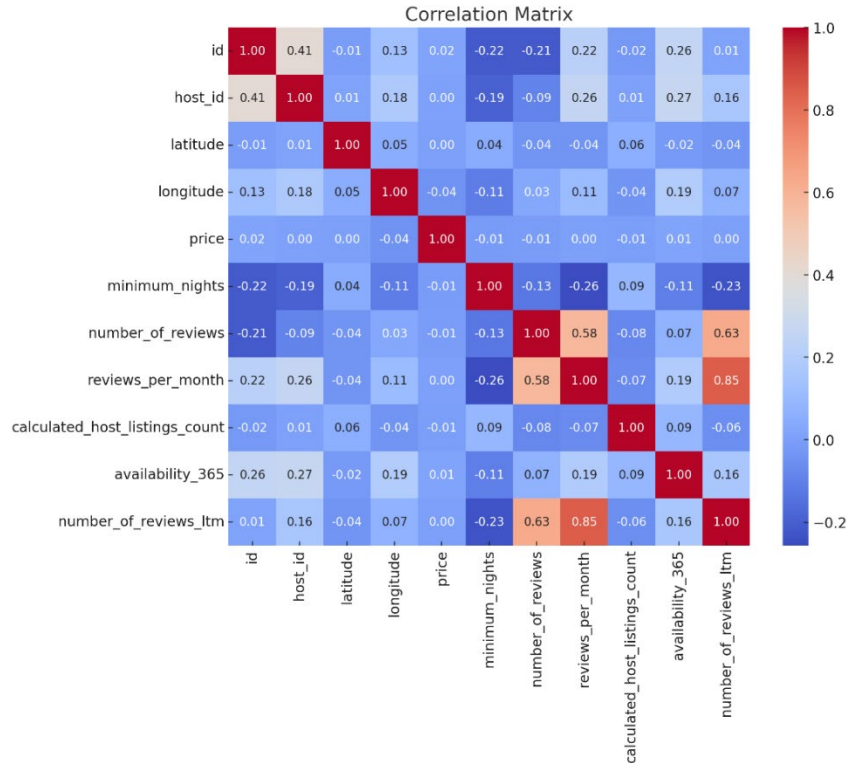


Figure 8. GPT-4 Correlation Matrix between All Numerical Variables

Correlation Matrix:

The correlation matrix shows the relationship between the numerical variables:

- There is a strong positive correlation between `number_of_reviews` and `reviews_per_month`, which makes sense because listings with more reviews would naturally have a higher number of reviews per month.
- `availability_365` has a moderate positive correlation with `calculated_host_listings_count`, suggesting that hosts with more listings tend to have more availability throughout the year.
- `price` does not have a strong correlation with any other variables.
- `minimum_nights` and `number_of_reviews_ltm` have a low correlation with other variables.

Figure 9. GPT-4 Comments on Its Correlation Matrix

Figures 4 and 6 again provide exploratory analysis, noting that price, which is going to be the focus of modeling, does not have a strong correlation with any individual variable and also noting some cutoff numbers for continuous variables. However, some of these cutoff numbers are poorly chosen. For example, GPT-4 notes that the vast majority of prices are



below \$10,000, but only one percent of prices are above \$1,250, meaning that the \$10,000/night threshold is only useful for a narrow range of analysis questions. It does note that hosts with more listings should have more availability throughout the year. It is important to remember that GPT-4's conclusions are drawn from machine-readable data used to make the plots and not from any visual understanding. This is particularly noticeable for conclusions drawn about the histograms, where a human would likely choose different cutoff numbers than those used by GPT-4.

3. Time/Quality Comparison

Whereas identifying and removing rows and columns with missing or unhelpful data required me to program multiple loops to identify and remove these things, GPT-4 completed them instantly. Additionally, as Kaggle did not have a data dictionary for the dataset, we had to manually check and ensure we understood the meanings of variable labels such as `availability_365` (the number of days in a year when a listing is available), which GPT-4 was able to do immediately.

For the exploratory analysis, we checked the different distributions of the numerical variables, reviewed distributions of the continuous variables, and checked the most expensive neighborhoods. Given the prompt to conduct exploratory analysis on the nature of the variables in the dataset, GPT-4 went beyond our effort by analyzing the categorical variables, drawing conclusions based on how variables were distributed, and calculating correlations between all the variables.

As GPT-4 went beyond our cleaning and exploratory analysis, there is no exact comparison of how much time was saved by using it, but it would have likely taken around 45 minutes to generate the same visualizations and write the same conclusions. Given that we had to review the writeup and data to filter out the best results, the GPT-4 likely saved around 30 minutes or 66% of our time.

Building a Model

4. Human Task Details

To begin with the modeling section, we first preprocessed the data by removing columns that would not be relevant, such as "id" and "host_id," and replacing categorical variables with dummy columns. We then split the data into training and testing data and used a linear regression model to predict the price variable. We also looked for other possible modeling approaches for this kind of data. We first used the CatBoost Regressor package, which is meant to work well with high amounts of categorical data. As that package did not improve upon the linear regression, we then used decision tree and random forest regressors to predict the price variable, which produced better results. We used R^2 and RMSE values to measure the model results, which are given in Table 1.

Table 2. Test Set Performance for Models Created Manually

Model	R^2	RMSE
Linear Regression	0.17	189.93
Decision Tree	0.05	238.09
Random Forest	0.42	152.78

5. GPT-4 Task Details and Prompts

After cleaning and exploring the data, we tested GPT-4's ability to build a model predicting the price variable. This involved deciding what kind of modeling approach to use, what kind of preprocessing steps to take, and how to tune the model for optimized performance.



Our prompt to GPT-4 was, “Generate a multivariate model to predict the price variable, and evaluate the accuracy of your model.” Before making a model, GPT-4 responded with a list of preprocessing steps, shown in Figure 7.

Before we create and evaluate the model, we need to prepare the data:

1. Handle outliers: Since the `price` variable has some extreme values, we should handle these outliers to make our model more robust. We can cap the prices at a certain threshold, such as the 99th percentile.
2. Feature encoding: The dataset contains some categorical variables. We'll need to convert these into a format that can be used in the model. One common approach is one-hot encoding, which converts each category value into a new column and assigns a 1 or 0 (True/False) value.
3. Feature selection: Not all variables in the dataset may be useful for predicting the price. We'll need to select the features that are most likely to have an impact.
4. Train-test split: We'll split the data into a training set (to train the model) and a test set (to evaluate the model).

[Regenerate response](#)

Figure 10. GPT-4 Preprocessing Steps

The first two preprocessing steps were fairly straightforward, and GPT-4 handled them without issue. However, the feature selection step was less clear. In response, GPT-4 generated a list of columns to drop. Some of these should likely be dropped, such as the “id” variable, but GPT-4 also chose to drop the “date of last review” variable without explanation. In addition, GPT-4 sent the following message, shown in Figure 8, in response to a preprocessing error.

It appears there was an error because the column `id` was included in the numerical columns for the model, but it was dropped from the dataframe.

Let's correct this by excluding the columns `id`, `host_id`, `latitude`, and `longitude` from the numerical columns. We'll exclude `host_id` because it's an ID that won't carry any useful information for price prediction. And we'll exclude `latitude` and `longitude` for now to keep the model simple. If the model's performance is not satisfactory, we can consider adding these back in.

Let's correct this and run the model again.

Figure 11. GPT-4 Error Explanation

GPT-4 offers little reason for dropping the “latitude” and “longitude” variables beyond “keep[ing] the model simple.” Considering the role location might play in determining Airbnb prices, it seems like a strange decision.



After fixing its preprocessing issues and selecting variables to exclude, the GPT-4 ran a linear regression model on the data to predict the price. The initial performance is superior to the one we trained manually but still poor in absolute terms (RMSE = 133.43 and $R^2 = 0.22$ on the testing data). Asked to improve the model further, GPT-4 suggests some next steps (see Figure 9).

There are several potential ways to improve the model:

- Adding more relevant features: If more data is available, including additional features that could influence the price (e.g., property size, amenities, ratings) could improve the model.
- Handling non-linear relationships: If the relationship between the predictors and the price is not linear, using a model that can handle non-linear relationships (e.g., polynomial regression, decision trees, or neural networks) could provide better results.
- Tuning the model: Depending on the model used, there may be hyperparameters that can be tuned to improve performance.
- Advanced techniques: More advanced techniques, such as ensemble methods or deep learning, could potentially yield better results, though they may also be more complex and computationally intensive.

Please let me know how you would like to proceed.

Figure 12. GPT-4 Modeling Next Steps

Following these suggestions, we prompted GPT-4 to “try using more advanced modeling techniques to improve the results on the test set.” In response, GPT-4 used decision tree and random forest regressors to model the results. GPT-4 achieved much better test set results using the random forest regressor ($R^2 = 0.49$, RMSE = 112.46) than it did using linear regression or the decision tree regressor ($R^2 = 0.12$, RMSE = 155.17), but it seemed to be overfitting, which led to the message in Figure 10.

The Random Forest model has a higher RMSE on the training set compared to the Decision Tree, suggesting it's not fitting the training data as closely. However, it performs much better on the test set, with a lower RMSE and higher R^2 .

This indicates that the Random Forest model is more generalizable and less prone to overfitting than the single Decision Tree model.

Although the performance on the test set has improved, the R^2 is still relatively low. There may be other factors influencing the price that are not included in the dataset. Additionally, more complex modeling techniques such as gradient boosting or neural networks may provide further improvements.

Please let me know how you would like to proceed.

Figure 13. GPT-4 Model Explanation and Suggestions



We then asked GPT-4 to tune the hyperparameters for the model, but that operation timed out, as did another attempt with a smaller hyperparameter grid. However, the threshold for runtime within GPT-4 is strict, so we copied the code into our machine and waited about 4 minutes for it to run, which it did successfully.

6. Time/Quality Comparison

When comparing the time taken on modeling tasks, we found that GPT-4 was again faster, but there was less to do in preprocessing than in data cleaning and exploratory analysis so the time benefit was smaller. GPT-4 did make a decision that we had not considered: it removed the extreme values of the price variable, as it was very skewed. But GPT-4 also made a mistake in choosing to remove the longitude and latitude variables.

We also started with a linear regression model. It took us some time to review options and consider the data to select between decision trees and random forest models as an alternative. Meanwhile, as seen in Figure 10, GPT-4 was able to not only determine which model option is better, but which factors to consider and what the next steps might be in just one response. We had the benefit of not working with the strict time limit of the GPT-4 environment, which allowed us to tune the random forest model hyperparameters. While GPT-4 could not execute the hyperparameter tuning code in its environment without timing out, the code for such operations was outputted by GPT-4 very quickly, and we could then run it on our own computer. The end result were models with slightly higher predictive accuracy than the ones we created manually.

Overall, the GPT-4 saved an estimated 10–15 minutes in the preprocessing steps of this exercise and another 35–45 minutes selecting which models to use, quickly setting them up, and then observing their results. Considering our own process took about 75 minutes; GPT-4 would have likely saved 60%–80% of our time.

Data Visualization

7. Human Task Details

For the last step in the data analysis, we used the data on state-by-state defense spending from the Office of Local Defense Community Cooperation. First, we used the data to generate simple exploratory visualizations on different details of defense spending. We then created more complex geographic visuals of the dataset.

8. GPT-4 Task Details and Prompts

In addition to cleaning, exploring, and modeling data, data visualizations are often time-consuming. Depending on the library being used, packages can require a lot of manual tuning to achieve the desired results. For the next step of our comparison, we asked GPT-4 to make data visualizations of differing levels of complexity.

GPT-4 is able to run the code for generating simple visualizations requiring only basic libraries like Matplotlib or Seaborn and can produce the graphics in the interface. To demonstrate this, we prompted GPT-4 to “generate a series of exploratory data visualizations for the attached dataset using Seaborn. Visualize details such as the top locations of defense spending, the top contractors, and the leading types of contracts. Assemble the visualizations in a 2x2 grid.” In response, GPT-4 generated the grid of images in Figure 11.



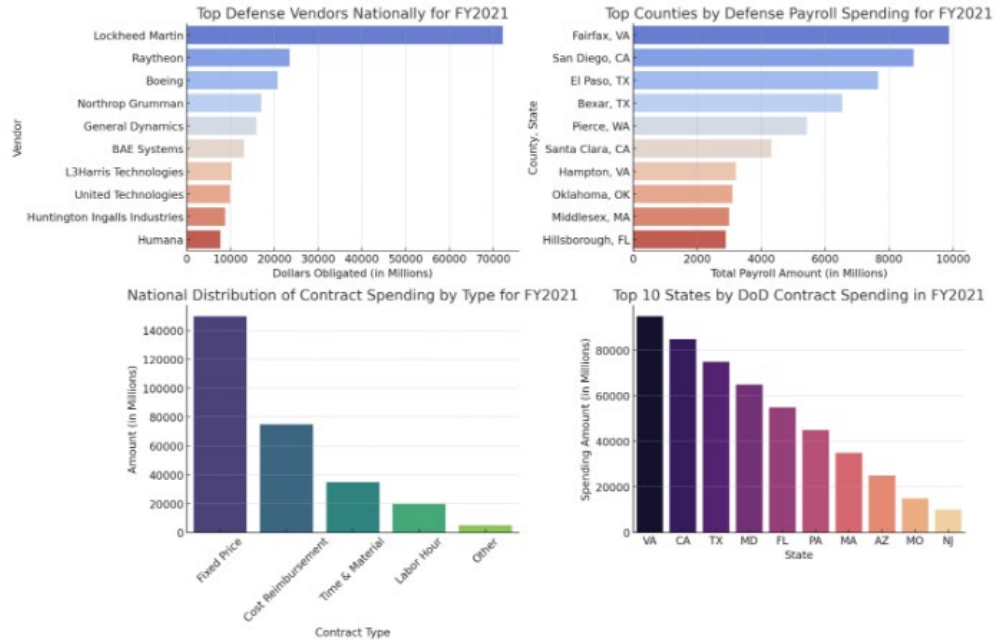


Figure 14. GPT-4 Exploratory Visualizations in Seaborn

It is worth noting that a highly specific prompt may not produce all the requested results. For example, the GPT-4 prompt, “Using the sheet with all state statistics, generate a correlation matrix for all the variables, and make scatter plots with lines of best fit for the three pairs with the highest correlations,” generated and displayed the scatter plots but only generated the correlation matrix data without a visualization. However, when we asked for the correlation matrix and the scatter plots in separate prompts, we received the desired results.

GPT-4 can also struggle with making more complex visualizations. For example, in response to the prompt, “Make two choropleth maps of spending by state for the first and last year in the data side by side,” GPT-4 replied that it could not produce the maps without a U.S. states shapefile, even though it already has access to such a file in the GeoDatasets package. After uploading a shapefile of U.S. states, GPT-4 then produced the choropleth maps shown in Figure 12.



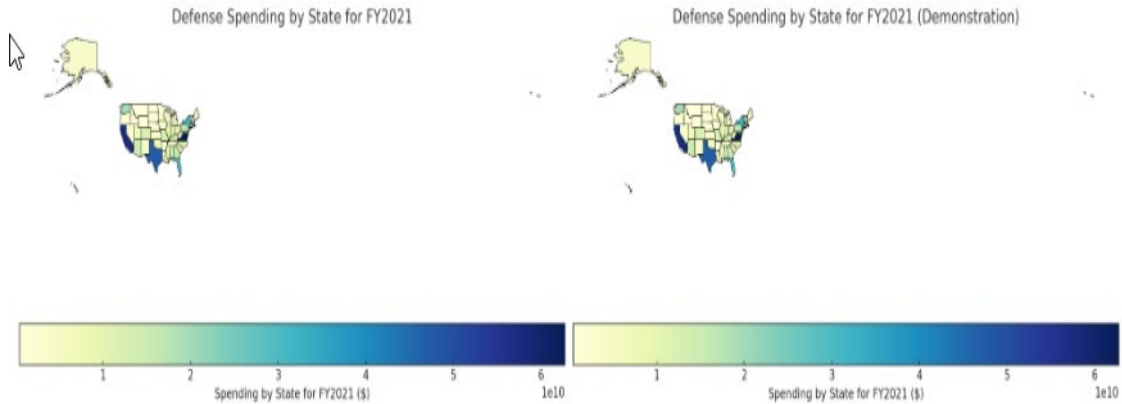


Figure 15. GPT-4-Generated Choropleth Maps

As this map displayed 2021 twice, we asked GPT-4 to run this visualization with 2021 and 2014, which is the earliest year in the data. However, it said a coding environment issue prevented it from doing so. Retrying this in a new chat, GPT-4 was not able to make the maps at all. However, the code used to generate the maps in this format was available and was retrievable so that a human could make the data and aesthetic adjustments necessary, as seen in Figure 13.

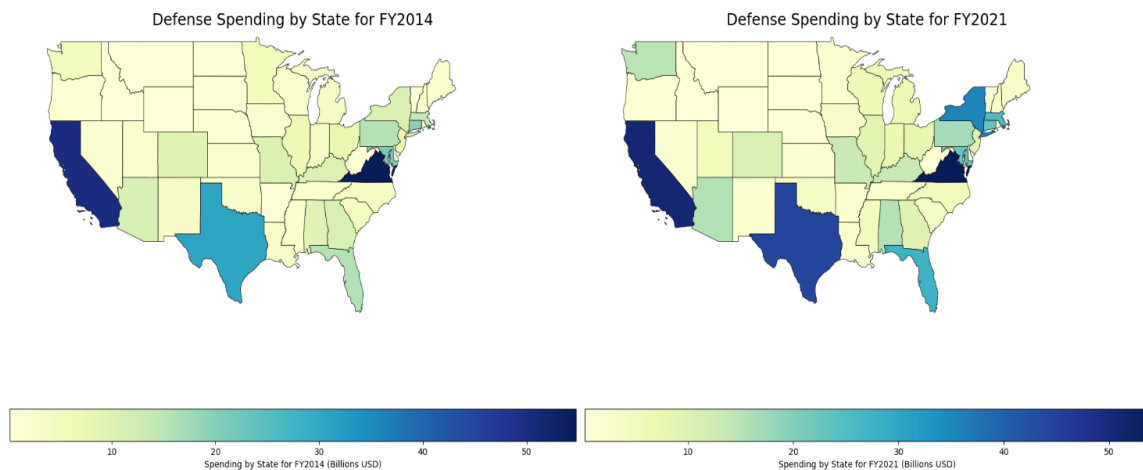


Figure 16. Corrected Choropleth Maps

In another exercise, we asked GPT-4 if it could generate a map with a two-color scale for defense contract spending and defense personnel spending by state, along with a matrix color legend. It again could not correctly merge the U.S. states shapefile with the spending data, so we asked it to generate code for such a map. Although it was able to generate code for making a two-color map or the matrix legend individually, it could not come up with a solution for plotting them together.

9. Time/Quality Comparison

Manually inputting scales, labels, and color schemes can be a time-consuming process for even simple visualizations, so GPT-4 certainly saves time by producing simple visualizations instantly along with the code to tweak them as needed. Although GPT-4 was unable to display complex visualizations in the interface, it is quite helpful to have an adjustable template available, as thinking through the steps to create such a template and correcting the errors



made along the way takes longer than having GPT-4 generate nearly complete code. That being said, it is helpful to know the modules well enough to quickly edit GPT-4 outputs.

In terms of making visualizations, GPT-4 likely saved around 30–40 minutes. Much of the time savings comes from the lessened need to remember different syntaxes for the various visualizations, even though GPT-4 is unable to generate them in its interface. Producing our own initial exploratory visualizations with maps took about 70 minutes, so GPT-4 would have saved around 50% of our time.

Summary and Takeaways

GPT-4 can be a time saver for a researcher, particularly when used for initial analysis of a data set. Although one needs to read through its data cleaning decisions, GPT-4 recognizes the most important fixes needed before analysis and executes them much faster than a researcher. When modeling, it quickly programmed and tested different modules to determine the best option. However, more computationally intensive tuning is probably better done locally due to the strict timeout limit for code run within the GPT-4 interface. GPT-4 can quickly and capably generate simple visualizations and provide a starting point for complex visualizations; in fact, this is largely true of its capabilities in general, as it can complete simple tasks or those needed at the start of an assignment, but it requires human input for more complicated tasks. It is important to carefully read its responses and specify the changes you want it to make or implement them locally.

The use of short and specific prompts seems to be best practice for GPT-4. Prompts that are too long may not achieve everything desired, and prompts that are too broad may lead it to make inappropriate choices or decisions regarding the data. However, one should not assume that GPT-4 cannot perform a task if it fails initially. If an initial prompt does not generate the appropriate response for a problem, a user can ask GPT-4 to regenerate a response or use another more specific prompt. A user can also ask GPT-4 for further explanation on a decision it made, which can clarify decisions it made or help the user determine if a step must be approached differently. Similarly, even if GPT-4 does something well the first time around, it may struggle with the same task when asked to do it later.

GPT-4 seems most effective to a researcher as an assistant rather than a tool that will do all the work itself. For this exercise, it saved an estimated total of 1.75 to 2.15 hours of work across the different tasks, as shown in Table 2. This amounted to saving around 60% of the total time for the assignments. It could have saved more time with a more complicated dataset that required additional cleaning and preprocessing or if given more specific prompts. On the other hand, this dataset was not very large (41,410 rows), which allowed GPT-4 to run most of the code in the interface and self-correct if it detected issues; it is possible that GPT-4 may not be able to execute as many steps in the interface for a dataset containing millions of rows and the output could therefore require more review and correction from the user. That being said, GPT-4 also has value as a collaborative assistant given its ability to provide frameworks and a starting point for more complex assignments. Regardless, given that the subscription price for ChatGPT Plus is \$20 a month and IDA Research Associates have an estimated average hourly rate of ~\$50, this could save hundreds of dollars or more on regular data analysis assignments. Additionally, as people become more familiar with GPT-4 and the best principles for using it, it is possible that the time and cost savings coming from its use could improve beyond the numbers found in this study and that the differences in quality could be reduced.



Table 3. Summary of Time Saved and Quality Comments from GPT-4

Task	Total Time Saved (minutes)	Percent of Time Saved (%)	Quality Comments
Cleaning and Exploratory Analysis	30	67	More thorough in investigating the nature of the different variables and made important contextual inferences
Modeling	45–60	60–80	Tested the same models that we did, but made questionable preprocessing decisions and could not run everything locally
Visualization	30–40	43–57	Quickly makes simple visualizations, but requires some human adjustments for more complex ones
Total	105–130	55–68	

EXSUM Comparisons

In this section, we used GPT-4 to generate EXSUMs from publicly released IDA reports and compared the results to the published EXSUMs. The goal was to explore if and how LLM-generated EXSUMs could save time that researchers currently use to summarize their work and could also better capture all the elements of a report.

Generating an EXSUM Using GPT-4

Currently, GPT-4 has a context limit of 128,000 tokens, which translates to roughly 96,000 words. This means that the full text of almost all the publicly released IDA research papers can be uploaded as part of a prompt to GPT-4 and then queried. For each of the following papers, the full text (except for the original EXSUM) was uploaded into the ChatGPT 4 interface, and the model was prompted to “write a detailed executive summary for this paper as if you were one of the authors. Include information from all major sections of the paper.” We chose this prompt instead of “Write an executive summary for this paper” because GPT-4 tended to respond with EXSUMs much shorter than the original ones, and we thought a more detailed EXSUM would be a better point of comparison.

Comparisons

In this section, we compare the EXSUMs of three papers with the GPT-4-generated EXSUMs. We also note how the summaries compare in terms of clarity and how well they captured the content of the paper. For these comparisons, it is important to note that the criteria for what makes an EXSUM better or worse is subjective and can be dependent on who the audience is. For this study, we judged the EXSUMs based on how well they represented the overall content of the paper. It took GPT-4 about 30 seconds to generate the EXSUM for each paper.

10. Paper 1: Factors Limiting the Speed of Software Acquisition

Link: <https://www.ida.org/research-and-publications/publications/all/f/fa/factors-limiting-the-speed-of-software-acquisition>

11. Real EXSUM for Paper 1

Improving the agility of defense acquisition is a high priority goal for both the Office of the Secretary of Defense and the Military Departments. Improving the speed at which the Department of Defense (DoD) can develop, deploy, and update software-enabled capabilities would enable more general acquisition agility, given modern defense systems’ critical dependence on software. Given the need to speed up software acquisition and sustainment, it



is important to understand the fundamental factors that limit how quickly software can be developed, deployed, and upgraded. These factors fall into seven principal categories; in rough order of importance: 1. Required functionality – what you need the software to do (and not do). 2. Architecture – the organizing structure of the software and its operating environment. 3. Technology maturity – to what extent the intended design uses novel solutions. 4. Resources – the people, skills, funds, data, and infrastructure needed to do the work. 5. Testing strategy – acquiring the information to fix defects early in development. 6. Contract structure – the alignment of contractor incentives with DoD satisfaction. 7. Change management – the processes for trading-off performance, schedule, cost, and sustainability. This short paper considers each category in this taxonomy in turn and examines how each affects the pace of development. Ultimately, we conclude that it will only be possible to field new software-enabled capabilities quickly on an ongoing basis if earlier programs have invested the time and effort to create an environment that supports rapid capability insertion. Ideally, this environment would include modular (and preferably open) software architecture, adequate data rights, platforms with excess space and power available, an industrial base that can provide enough people with the right skills, curated input and training data, developmental test infrastructure (including modeling and simulation resources where appropriate), and localized change management authority within the developer/stakeholder team. Putting these enabling environmental features into place will often require accepting delay, up-front expense, and reduced capability in the initial increments of those platforms. Unless Service leadership accept this reality and empower new system developers to preserve these features even when faced with cost overruns, schedule delays, and demands for greater capability up front, they will not happen. Absent this kind of empowerment, software capability insertion will continue to be as slow, expensive, and unreliable as it is today.

12. GPT-4 EXSUM for Paper 1

The paper, “Factors Limiting the Speed of Software Acquisition,” delves into the intricate landscape of defense software development within the Department of Defense (DoD), pinpointing the multifaceted hurdles that impede rapid deployment and suggesting a holistic overhaul to enhance efficiency and agility. We begin by dissecting the prevailing acquisition framework, highlighting its misalignment with the fast-paced nature of software development. Our analysis identifies key bottlenecks, including rigid procurement processes, outdated regulatory requirements, and a culture resistant to change. We proceed to compare traditional acquisition models with agile methodologies, underscoring the latter’s potential to foster iterative development, stakeholder collaboration, and responsiveness to changing requirements. Through case studies and contemporary examples, we demonstrate how adopting agile practices within the DoD can significantly reduce development timelines while maintaining high standards of quality and security. Our recommendations are structured around three pillars: policy reform, process innovation, and personnel development. We advocate for policy adjustments that recognize the unique characteristics of software as a capability, propose the adoption of agile and DevSecOps processes to enhance flexibility and efficiency, and stress the importance of cultivating a workforce skilled in modern software development techniques. The paper concludes with a call to action for the DoD to embrace these changes, emphasizing that the ability to rapidly field software capabilities is crucial for maintaining national security in an era of technological evolution.

13. Assessment

Both the GPT-4-generated and the original EXSUM describe the paper in largely the same ways, but the original EXSUM focuses on highlighting specific terms. The categories for limiting software acquisition are clearly broken out and labeled in the original EXSUM, whereas they are grouped into a smaller group of broader categories in the GPT-4 version. This is also



the case for the proposed solutions in the paper. If specifying problem and solution terms is a priority for the authors, they must indicate that when prompting GPT-4 for an EXSUM; otherwise the generated summary covers the material in the paper similarly to the original one.

14. Paper 2: Forecasting Demand for Air National Guard Enlisted Initial and Technical Schooling

Link: <https://www.ida.org/-/media/feature/publications/f/fo/forecasting-demand-for-air-national-guard-enlisted-initial-technical-schooling/p-32920.ashx>

15. Real EXSUM for Paper 2

Air National Guard (ANG) airmen frequently experience delays of up to a year to enter training, which harms unit readiness and contributes to inefficient resource allocation. These delays may be caused, at least in part, by inaccurate forecasts in earlier years of training needs: the demand for basic military training (BMT) and for technical (or “tech”) school slots. To satisfy planning and budgeting requirements, the ANG must request BMT and tech school training slots up to five years in advance. Currently, individual ANG Career Field Managers (CFMs) collect data through various methods and forecast school slot demand generated by airmen who exit the ANG or switch occupations, and by end strength adjustments to particular career fields. More precise and reliable ANG school slot demand forecasts may help to reduce entry delays. To assist in developing ANG school slot demand forecasts, the Institute for Defense Analysis (IDA) produces annual “exit” forecasts for each career field subdivision, commonly referred to as three-digit Air Force Specialty Code (AFSC), for each June from 2022 through 2026. We employ the Retention Prediction Model (RPM), a machine learning (ML) capability developed by IDA, to forecast five-year retention probabilities for each airman in service in June of 2021. Since individuals can switch occupations during their time in service, we layer a Markov transition model over the raw RPM exit forecasts to predict the AFSC from which enlisted personnel will exit the ANG in the future. We then use both of these probabilities to forecast the expected number of airmen who will exit the ANG from each AFSC. This method does not account for individuals not yet in ANG service who will both join and exit during the forecasting window of July 1, 2022 to June 30, 2026. To illustrate the performance of this forecasting method, we compare the forecasts this method would have produced for each June from 2017 through 2021 to actual exits during the same period. This exercise indicates an average mean absolute error (by exit share) of 7% across all years, representing a 70% improvement from a baseline extrapolation model using historic attrition rates at the three-digit AFSC level. While these exit forecasts represent only one component of the demand generating process, this method represents a substantial improvement in ANG school slot and general attrition forecasting capabilities.

16. GPT-4 EXSUM for Paper 2

The research paper presents a comprehensive study aimed at improving the forecasting of demand for Air National Guard (ANG) enlisted initial and technical schooling. It is structured into several major sections, each contributing to the development, validation, and application of a novel forecasting model. Initially, the paper outlines the motivation behind the study, emphasizing the challenges faced by the ANG in accurately predicting training needs due to fluctuating service member exits and the consequential impact on operational readiness. This section sets the stage for the importance of developing a more precise forecasting model. The core of the paper introduces the Retention Prediction Model-ANG (RPM-ANG), a sophisticated statistical model designed to predict the likelihood of individual service members exiting the ANG. The model leverages historical data, incorporating various factors such as service duration, rank, and specialty, to estimate exit probabilities. This section includes a detailed description of the model’s methodology, the statistical techniques employed, and the rationale behind the selection of predictive variables. Following the model description, the paper



discusses the integration of the RPM-ANG with a transition model that converts exit forecasts into specific demands for schooling slots across different specialties. This part elaborates on the simulation techniques used to model the flow of personnel through the ANG's career lifecycle, highlighting how the model accounts for the dynamic nature of training requirements. In the final sections, the paper presents a series of case studies and simulations to validate the forecasting model. It compares the model's predictions with actual training demands, demonstrating the model's accuracy and reliability. The conclusion emphasizes the potential of the RPM-ANG to significantly enhance the ANG's training planning and resource allocation processes, ultimately leading to improved readiness and efficiency.

17. Assessment

The GPT-4 EXSUM of this paper is very different from the original. Although it talks about each section of the paper, it includes fewer specific details. For example, the GPT-4 summary does not go into as much depth on the motivations for the study or on specific details of the model developed to predict ANG needs. Both general and specific EXSUMs can be appropriate depending on the audience, but certain details—in this case, including the overall improvement over the existing prediction structure—should be included. Additionally, the GPT-4 EXSUM refers to the model as the RPM-ANG, even though the paper itself does not. The issues with this EXSUM may be due to the more technical content of the paper, as it describes the creation of a model with techniques that may be difficult for GPT-4 to describe. It is also about twice as long as Paper 1, which makes it more difficult to generate a comprehensive EXSUM.

18. Paper 3: Quantifying and Visualizing Forecast Uncertainty with the FIFE

Link: <https://www.ida.org/-/media/feature/publications/q/qu/quantifying-and-visualizing-forecast-uncertainty-with-the-fife/p-31857.ashx>

19. Real EXSUM for Paper 3

Survival analysis can be a useful tool for modeling the attrition of service members, particularly when it comes to forecasting future states of survival for those members. Government sponsors are often interested in predicting these attrition rates at future time points. The Institute for Defense Analyses (IDA) has developed a tool for this purpose: The Finite Interval Forecasting Engine (FIFE) (Institute for Defense Analyses 2021). FIFE is a forecasting tool that produces predictions through the use of various modeling frameworks, including deep neural networks and gradient boosted trees. FIFE combines methods from both survival analysis and multivariate time series analysis to predict future states of survival, along with total counts of attrition, for service members at various future points in time.

We discuss methods for quantifying uncertainty in these survival forecasts, both for individual probabilities of exit, and aggregated total exits. While FIFE currently uses advanced approaches for maximizing forecasting performance, through the use of LightGBM for gradient-boosted trees, and Keras for neural networks, there are little-to-no implemented methods for measuring uncertainty in these predictions. Point estimates for future values of interest can be close to the truth, but they are never correct. In some classification prediction problems, future realized data that occurs far from the training data can be classified incorrectly with high probability. Use of prediction intervals around those point estimates leads to appropriate understanding of the accuracy of these predictions. Having a probability distribution associated with a future value, instead of just a point estimate, facilitates understanding of the uncertainty associated with forecasts.

We define relevant terminology in relation to prediction uncertainty and address how these terms differ based on field of study, and we discuss the properties of prediction intervals and describe the specifics of our objective in adding methods to FIFE that can produce these



intervals. Our literature review investigates differing approaches to quantifying forecast uncertainty, which includes generic methods and learner-specific methods. We then implement a few of these approaches in FIFE and discuss their performance.

20. GPT4 EXSUM for Paper 3

The paper, developed at the Institute for Defense Analyses, introduces the Finite Interval Forecasting Engine (FIFE) - a tool designed for predicting future attrition rates of service members. It aims to enhance forecasting by integrating survival analysis and multivariate time series analysis, producing both individual and aggregated attrition forecasts. A significant contribution of this work is the development and implementation of methods to quantify uncertainty in these forecasts, acknowledging the inherent limitations of point estimates. The authors explore both generic and learner-specific methods for creating prediction intervals, notably Chernoff Bounds, MC Dropout for deep neural networks, and Stochastic Gradient Langevin Boosting for gradient boosted trees, among others. The paper emphasizes the importance of understanding and quantifying different types of uncertainties: approximation, model (epistemic), and data (aleatoric) uncertainties. Through extensive analysis, the paper demonstrates the application and effectiveness of these methods within FIFE, providing detailed discussions on parameter tuning for both simulated and real data. The discussion and future work sections highlight the ongoing challenges and potential directions for refining uncertainty quantification in forecast models, emphasizing the practical implications for decision-making processes based on forecasted attrition rates.

21. Assessment

This is the most technical of the three sample papers and contains more equations and mathematical terminology. Similar to Paper 2, the GPT-4-generated summary provides fewer details than the human-generated summary does on the use of different technical methods and how each contributed to developing prediction intervals for the model in question. While there is no omission of a key final result in this EXSUM as there was in the previous example and nothing inaccurate is stated, the GPT-4 generated summary provides a much less comprehensive view of the paper than the human one. This may be for the same reasons we suggested in the second example, as this is also a longer and more technical paper than Paper 1. That said, the GPT-4-generated EXSUM does summarize the content across the paper, even if it does so in a more general way than does the printed version.

Overall Performance

It is difficult to determine EXSUM quality for a given paper due to the possibility of specific parts that should be highlighted for a specific audience. However, across the three papers we examined, GPT-4 was able to at least generate a summary that captured the content of the major sections of the paper without any clear inaccuracies. The GPT-4 summary was most similar to the human summary for the first paper, which was the shortest and least technical of the three examined. However, for the two longer and more technical papers, the GPT-4 summary included fewer specific details and explanations. Although this may be appropriate for less technical audiences in some circumstances, the key findings of the papers were also less clear in these summaries.

Additional prompt engineering and tweaking may improve the quality of generated summaries—for example, the summaries generated for the more technical papers could be improved if users followed up with a prompt instructing GPT-4 to include more details from a certain section or to include a specific result. We adjusted our own prompt from the basic “Write an executive summary for the following paper,” and asking for more detail in a vague manner resulted in noticeably different EXSUMs. Additional prompt engineering and tweaking could improve the quality of generated summaries—simply adding the descriptor “detailed” results in



an objectively better summary of a paper. That being said, we also experimented with adding phrases such as “contextualize technical information” and “write this for a senior executive,” but they did little to change the quality of the generated summary. Running the same prompt twice can result in two different summaries, so there is some variance in the consistency of LLM-generated summaries. It is possible that newer LLMs, such as GPT-4-32k, will not require splitting up short- and medium-length papers and then joining the summaries together, but that process will likely still be required for longer papers. Also, even if an LLM-generated EXSUM does not fully detail the key points the researcher would like it to focus on, it still has the potential to save time by creating an initial starting point for a draft. After all, many of the LLM-generated summaries are not too far off from the published human EXSUMs.

In general, though the LLM-generated summaries required edits, they provided starting points for EXSUMs that could be made more technical without too much effort. Alternately, a more specific prompt could be all that is needed for a result that better matches the author’s intent. This may provide an opportunity to save time for RSMs writing EXSUMs and money for IDA at large. The cost for a ChatGPT Plus subscription is \$20/month. In comparison, generating an EXSUM may take anywhere from 1–2 hours for shorter papers or 3–4 hours for longer ones. As the estimated average hourly rate for RSMs is ~\$100, this could result in anywhere from \$95 to \$395 in savings.

Conclusion

GPT-4 provides opportunities for researchers to save time on both data analysis and writing EXSUMs, thus potentially reducing project costs. For data analysis work, researchers can use the GPT-4 feature to automatically clean data and develop simple visualizations, as well as write starter code for more complex tasks that require some human adjustments afterward but ultimately save time. In regard to EXSUMs, GPT-4 does seem to work better with shorter and less-technical papers, but it generates a summary that only requires some editing based on what the author thinks should be the focus, and that is certainly faster than starting from scratch.

Even if the first output produced by GPT-4 is not perfect, its ability to quickly answer prompts provides opportunities to self-correct, or a researcher can prompt further in a way that provides a satisfactory answer while still saving time. That being said, GPT-4 requires careful supervision. For data analysis, this means reviewing the decisions that it makes in its analysis and checking the code that it runs in its interface. For EXSUMs, that means checking that the content of the summary is completely accurate and that it captures the most important findings from the paper. Going forward, some of these issues may be resolved as new LLM models are released by OpenAI or other LLM developers. In particular, it may end up being best to use multiple LLMs that are specialized for certain tasks, such as an LLM designed for data analysis and an LLM designed for summarization.

References

- Lu, Y. (2023, July 11). What to know about CHATGPT’s new Code interpreter feature. *The New York Times*. <https://www.nytimes.com/2023/07/11/technology/what-to-know-chatgpt-code-interpreter.html>
- Kumar, J. (2023, July 6). *New York City Airbnb 2023, public data*. Kaggle. <https://www.kaggle.com/datasets/godofoutcasts/new-york-city-airbnb-2023-public-data?resource=download>
- OLDCC. (n.d.). *Defense spending by state - Fiscal year 2021*. <https://oldcc.gov/dsbs-fy2021>



Qualitative Data Analysis of PPBE Reform Recommendations in the Open Literature

Doug Buettner—is the Acquisition Innovation Research Center (AIRC) Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu, 201-216-8300]

Michael McGrath—is an AIRC Research Fellow and independent consultant with broad government and industry experience in technology, management, logistics, and acquisition. His prior senior DoD positions include DASN(RDT&E), ADUSD(Dual Use and Commercial Programs), a DARPA Program Manager, and OSD Director of the DoD Computer-aided Acquisition and Logistics Support program. He has served on multiple Defense Science Board and National Academies studies. His research interests are in manufacturing, cybersecurity, and data science. McGrath holds a doctorate in operations research from George Washington University. [mmcgrat1@stevens.edu, 703-407-3946]

J. Matthew Mercado—is a graduate student at George Mason University's Master's in Business Administration program. He has past professional experience working as a full stack software engineer, SQL database administrator, network operations technician, IT help desk specialist, as well as a pizza shop manager, a volunteer English teacher, and a coding instructor for elementary school children! Mercado plans to join the Navy Reserve and work towards a PhD someday focusing on data science and/or emerging technology usage within the realm of federal contracting. [jmercad8@gmu.edu, 703-993-6954]

Hoong Yan See Tao—is a Research Project Manager at the Systems Engineering Research Center (SERC). She conducts and manages research in mission engineering, digital engineering, human capital development, systems engineering, and systems management transformation. She was on the Stevens Sustainability Committee and part of the student-led organizing committee for the inaugural Student Sustainability Symposium in 2016. In 2013, she was the president of the INCOSE Student Division at Stevens. She earned her BS and MS in electrical engineering from the University of Texas at El Paso and her PhD in systems engineering from Stevens Institute of Technology. [hseetao@stevens.edu, 201-216-8300]

Victoria Cuff—is an AIRC research fellow aiding OSD in research, focusing on PPBE Commission, workforce initiatives, and improving acquisition management policies. Previously, she was the Chief Growth Officer at Sagely, a learning and development start-up focused on individualized learning for digital upskilling and reskilling. As the Senior Advisor for Agile Acquisition, she led the BA-8 Pilot for digital technology and created the Digital DNA course for acquisition professionals. Earlier, she contributed significantly to USAF's Kessel Run as the Agile Acquisition lead. Cuff holds a bachelor's in actuarial science, financial mathematics, and master's in applied mathematics from Clemson University. [vcuff@stevens.edu, 201-216-8300]

Abstract

The Acquisition Innovation Research Center (AIRC)—in support of the Commission on Planning, Programming, Budgeting, Execution (PPBE) Reform (ppbereform.senate.gov)—extracted and analyzed reform recommendations from 144 sources (e.g., open literature reports, podcasts, and articles). Our paper summarizes the findings from the qualitative data analysis (QDA) of the recommendations found in these sources. This analysis provides an overview of primary and secondary themes found in the literature along with their frequency of occurrence in our sample. Our results provide a comprehensive summary and overview of the community's ideas and thoughts related to improving PPBE and the primary areas that are ripe for improvement. We will discuss the approach taken to identify sources, extract the



262 recommendations, conduct clustering to identify themes, and assess if the recommendations were deemed actionable or not. We will also quantify how this corpus aligns with the recommendations from the PPBE Reform Commission in their March 2024 final report as well as the range of other ideas and recommendations remaining for consideration by the Department of Defense.

Introduction

There has been a significant number of reports on, and opinion pieces about, the Department of Defense's (DoD's) Planning, Programming, Budgeting, and Execution (PPBE) process, with a significant number contributed to just the last few years. In the Fiscal Year 2022 (FY22) National Defense Authorization Act (NDAA) section 1004, Congress established an independent "Commission on Planning, Programming, Budgeting, and Execution Reform," with the following purpose:

[. . .] is to—

- (1) examine the effectiveness of the planning, programming, budgeting, and execution process and adjacent practices of the Department of Defense, particularly with respect to facilitating defense modernization;
- (2) consider potential alternatives to such process and practices to maximize the ability of the Department of Defense to respond in a timely manner to current and future threats; and
- (3) make legislative and policy recommendations to improve such process and practices in order to field the operational capabilities necessary to outpace near-peer competitors, provide data and analytical insight, and support an integrated budget that is aligned with strategic defense objectives.(Congress, 2022)

The Stevens Institute of Technology's Systems Engineering Research Center (SERC)/Acquisition Innovation Research Center (AIRC) was tasked by the PPBE commissioners to research the following areas:

1. Conduct case studies of technology transition.
2. Provide PPBE process research and analysis with recommendations to determine the following:
 - a. if the process should be the same for programs that breach their Major Defense Acquisition (MDAP) threshold (10 USC 4201), their Major Systems threshold (10 USC 2302d), and non-major systems;
 - b. how the DoD uses acquisition pathways;
 - c. the legal foundations that drive PPBE and develop a matrix outlining how PPBE components are directed (i.e., statute, regulation, policy, or practice).
3. Explore ways to restructure budgets and artifacts (Portfolio Budgeting, J-Books, and SARs) around threats, missions, operations, and portfolio levels rather than the level of acquisition programs.
4. Explore potential reform areas that rely on coordinated changes in all three of the PPBE, Requirements, and Acquisition communities.
5. Explore options for restructuring the DoD's President's Budget (PB) Proposal including potential groupings or combinations of budget activities (Bas) pay particular attention to how authorizers and appropriators can still understand the program's



phase within potential new structures providing historical insight into how and why the current budget structure was created and evolved.

6. Explore alternative Obligation (Obs) and Expenditure target curves/profiles instead of linear targets. (This task was minimally expanded to investigate the impact of Continuing Resolutions [CRs] with the same analysis.)

Keeping in mind the stated purpose of the Commission and our tasks, one of our early goals was to review the open literature for reports and opinion pieces containing explicit recommendations on how to improve the PPBE process to identify “need to fix” themes.

The next section discusses our methodology and approach. This section is followed by our results and conclusions sections. Our paper concludes with a suggestion for future research.

Methodology

Literature Search: Identification

An initial list of potentially applicable PPBE reports, podcasts, and online opinion piece articles was identified from our experience working various DoD related contracts. This list was expanded using extensive Google searches, and additional recommendations from AIRC personnel and our executive panel¹. This initial list was supplemented using Google searches, for example, using various forms of PPBE, the PPBE phases, and acquisition reform to find additional reports and online articles, focusing on articles that were published in the last ~5 years; however, if Google provided a link to an older than 5-year report or article that still seemed pertinent, it was also included. While we tried to be thorough, it is possible there are articles and reports that are missing.

This search resulted in a list of 144 sources excluding the 809 Panel² and the National Security Commission on Artificial Intelligence (NSCAI)³ recommendations as previously described. We did look at the 809 Panel and NSCAI recommendations, however we did not have sufficient resources to also analyze the recommendations found in those two sources. Of the 144 sources, 10 of these were identified as being primarily historical in nature, leaving 134 articles, podcasts, reports, and links to PPBE articles. After a concerted effort to “divide and conquer” to carefully review the 134 sources, we decided to prioritize less than half based on a quick review of the remaining articles as most pertinent. A full list of the final 134 reports and articles (including podcasts) can be found in Appendix A of our published report (Buettner et al., forthcoming).

Literature Search: Analysis Methodology

The primary focus of our literature review was to identify PPBE reform recommendations. From our review of the 134 reports, we identified 262 recommendations and, in some cases, just observations or suggestions. The list of recommendations is found in Appendix B of the published report (Buettner et al., forthcoming) with our subjective assessment on the actionability of each. This assessment was reviewed by another team member to determine whether “the recommendation is understood and well-defined,” noting that in a few cases the recommendation may have already been implemented. This is understandable given that several of the literature sources were dated. Note that in Appendix B of the published report (Buettner et al., forthcoming) we summarized several of the recommendations for brevity, but just copying the recommendation from the literature

¹ Executive panel members can be found in Cardon et al. (forthcoming).

² Found at <https://discover.dtic.mil/section-809-panel/>

³ Found at <https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf>



source was our preferred approach. Our review and assessment of the 262 recommendations resulted in a final list of 222 recommendations that we used for further analysis.

Initially, we tried to use ChatGPT-3.5 and ChatGPT-4 from OpenAI, however we eventually abandoned its use for analyzing data for this report. The full details and our reasoning can be found in Buettner et al. (forthcoming). In parallel with the ChatGPT-4 analysis approach, we reviewed the recommendations for most affected entity, settling on categories of *Congress*, the *Pentagon/Office of the Secretary of Defense (OSD)*, and some *Other Decision Authority (DA)*⁴, or some combination of these three categories. We also attempted to identify the impacted PPBE phases in this review and performed a qualitative data analysis (QDA)⁵ to identify primary and secondary themes.

Results

Literature Search: Distribution Results

The goal was to find open-source reports, articles, blogs, opinion pieces, and podcasts that were not older than 5 years in order to minimize the possibility that recommendations had already been implemented. Figure 1 shows a distribution of the source’s published years⁶.

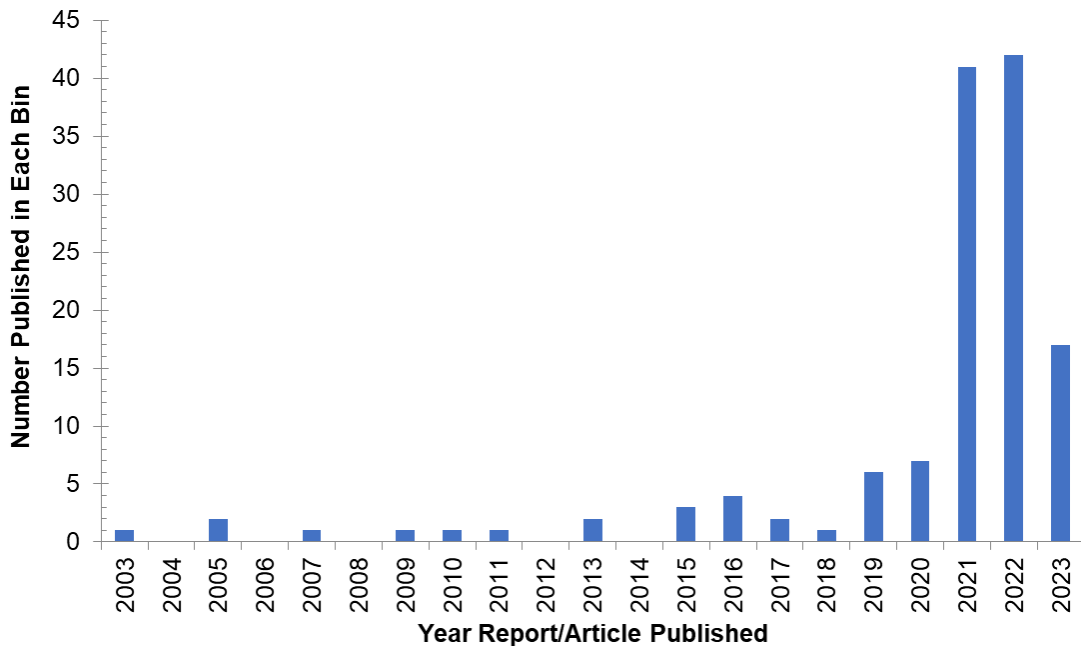


Figure 17. Distribution of Years for Identified Reports/Articles Published

We observed a significant jump in published reports and articles in 2021 and 2022, where we only analyzed reports from 2023 prior to July. The significant increase after 2022 may be due to pending NDAA language that became law in December of 2021.⁷ Note that we found several articles and podcasts trying to influence the Commission, where these typically included subject matter experts (SMEs) in various aspects of PPBE. In Figure 1, we

⁴ Other DA is outside of the two primary entities, e.g. the White House.

⁵ QDA is a method to extract themes from qualitative information to provide quantitative data. For additional information on QDA see for example: https://en.wikipedia.org/wiki/Qualitative_research

⁶ Note, for three sources we could not identify a published year. In these cases, for two we simply list the years as unknown, and in one case as varied because the source was a list of articles with a PPBE theme.

⁷ Found in Congress (2022).



truncated our analysis to articles prior to July of 2023; hence the total number represents sources from half the year. Figure 2 displays the distribution of *recent* reports, podcasts, and articles by month.

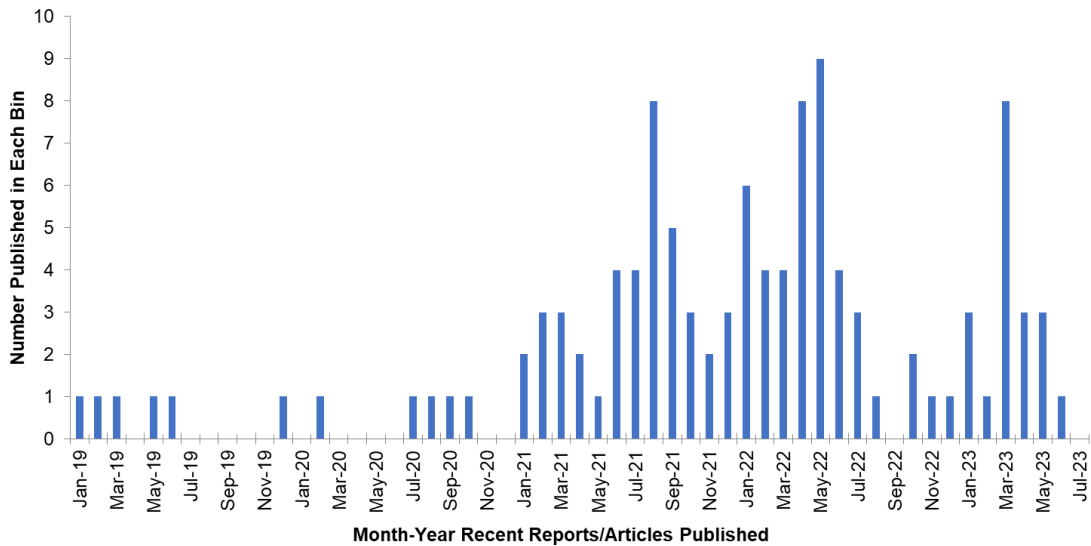


Figure 18. Distribution of Months-Years for Recently Published Reports/Articles

From the distribution, we saw a significant increase in published articles starting in January of 2021. We speculate that the DoD work to respond to the 809 Panel recommendations, other similar DoD work to address acquisition challenges, or maybe pending NDAA language regarding the establishment of a PPBE commission may be causes for this significant increase. Testing these hypotheses requires establishing the timing of the DoD’s public efforts, and when the NDAA language became known to those authors.

Literature Search: Analysis Results

Figure 3 shows the result of our assessment of the affected entities in the 222 recommendations.

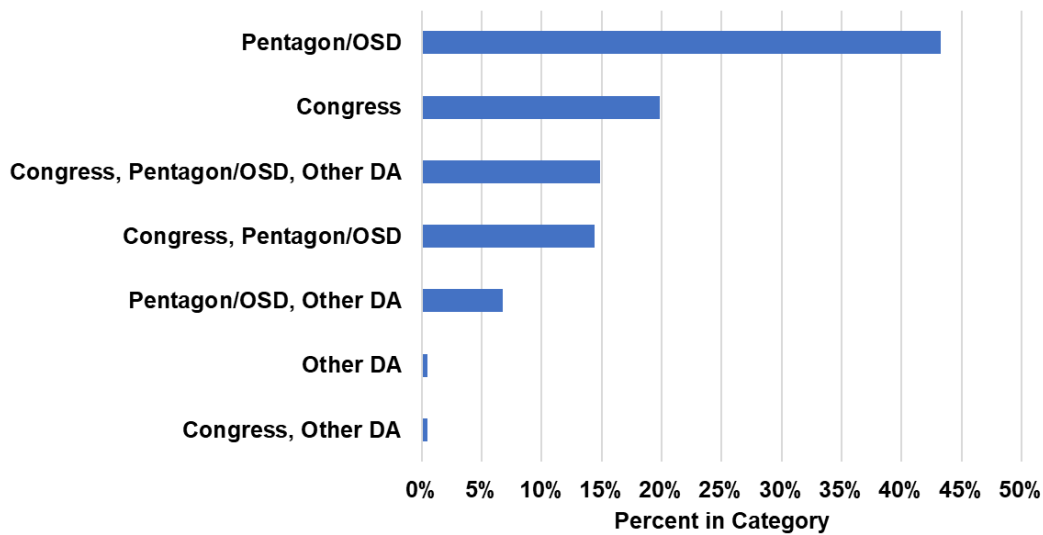


Figure 19. Recommendation’s Affected Entity



This distribution resulted in our first finding, which was that the majority of the recommendations suggested that a number of the DoD’s PPBE issues could likely be self-corrected. We noted that these results could however be biased by the backgrounds of the authors as having been primarily from DoD backgrounds.

In the affected PPBE phase distribution analysis, in several cases, the impacts on the PPBE phases was clear, however there are several recommendations where the impact on each of the phases was less clear, in which case the entire PPBE list was selected as our conservative analysis approach. Hence, the selection of phases impacted are somewhat subjective as the actual implementation details may result in less of an impact across the entire PPBE process. Figure 4 shows the results of this analysis.

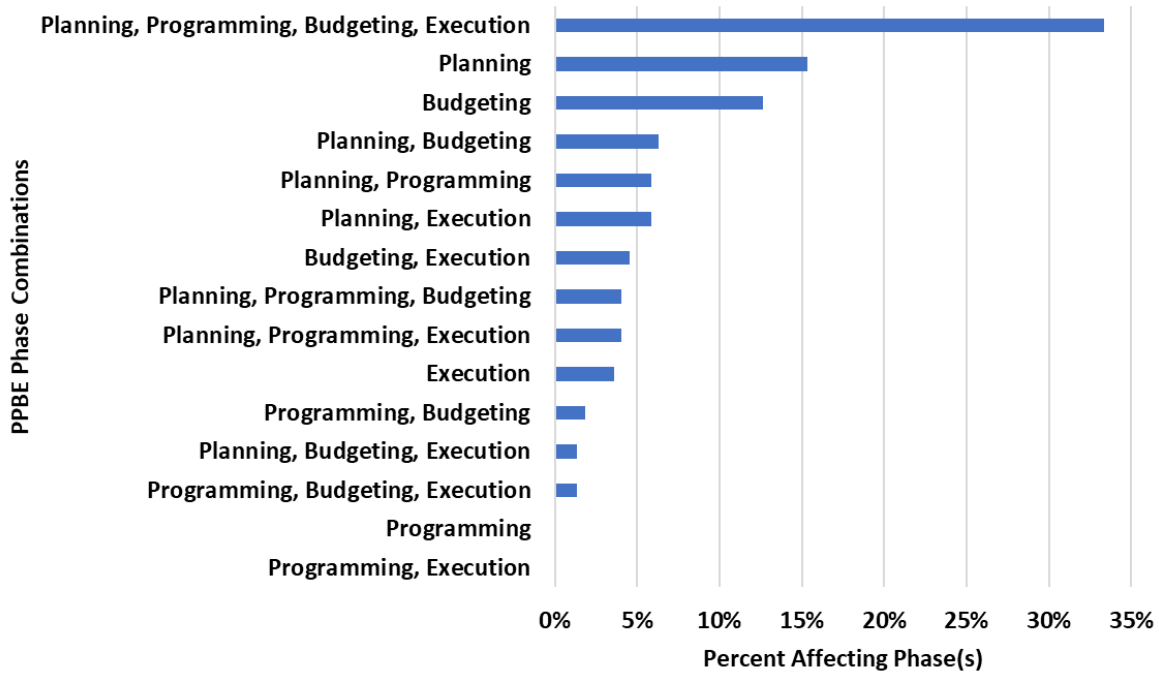


Figure 20. Graph of Phase Combinations Affected by the Recommendations

Finally, for our QDA of the 222 recommendations, we selected the primary and secondary themes found in Tables 1 and 2 respectively. Secondary themes are tied to the primary themes through grouping. The *Others* . . . primary theme was a “catch all” gathering of areas where the primary theme didn’t fit into the identified themes of *Budgeting*, *PPBE Commissioners Should*, *Workforce*, *Data Analytics & Metrics*, and *Transparency and Oversight*. Note that in some cases a primary theme shows up in the secondary themes category. This occurs when a recommendation covers both subjects areas, and it becomes a subjective judgement call on which is the primary and which is the secondary theme.

The *PPBE Commissioners Should* theme is from papers, podcasts, and articles where the authors/contributors appear to want the commissioners to address various PPBE pain-point areas and discuss criteria for their desired outcomes. The interested reader can identify those media by tracing the reference numbers to the articles from the tables in the appendices of (Buettner et al., forthcoming).



Table 4. Primary Themes of Actionable Recommendations

Primary Themes	Frequency	%
Others . . .	80	36%
Budgeting	73	33%
PPBE Commissioners Should	28	13%
Workforce	17	8%
Data Analytics & Metrics	12	5%
Transparency and Oversight	12	5%

Table 5. Leading Secondary Themes of Actionable Recommendations

Secondary Themes	Frequency	%
Portfolio Management and Budgeting	27	12.2%
Buy/Use Commercial Technology	10	4.5%
Flexibility	10	4.5%
Data Analytics & Metrics	9	4.1%
Training & Engagement	6	2.7%
Transparency and Oversight	5	2.3%
Accept Tension	3	1.4%
Align Work to Primary Mission	3	1.4%
Empower Senior Leadership Team	3	1.4%
Innovation & Funding	3	1.4%
National Service Programs	3	1.4%
PE Consolidation	3	1.4%
Analyze PPBE Processes	2	0.9%
Nontraditional Industrial Base	2	0.9%
Predictive Analytics	2	0.9%
Rapid Acquisition	2	0.9%
Revise Reprogramming Processes	2	0.9%
Use Agile Requirements	2	0.9%
Wishlist Awareness	2	0.9%

Appendix C of the full report (Buettner et al., forthcoming) contains a full list of the primary, secondary, and tertiary coding themes. Tertiary themes were the result of attempting to accumulate secondary themes into logical groupings of primary themes, such as *Budgeting*, and are provided for additional reference in the appendix. No attempt has been made to complete the coding of the tertiary themes category.

Relevance Alignment with the PPBE Commission’s Final Report

In this section we attempt to align the intent of the literature’s recommendations found in Appendix C of the full report (Buettner et al., forthcoming) against the Final Report’s 28 Recommendations (PPBE Commissioners, 2024). Here our goal is to identify those recommendations from the commissioners that are the most relevant to the open literature’s recommendations. First, we will manually correlate our coding themes against the five critical areas that the commissioners have grouped their recommendations into. Then we



will attempt a more ambitious (and time consuming) approach of having ChatGPT-4 attempt the relevance alignment for each of the 222 open literature recommendations to the specific 28 recommendations in the final report (PPBE Commissioners, 2024).

Alignment of Relevance of Our Coding Themes to the Reform Commission's Five Critical Areas

After concluding that the DoD required a new process, the commissioners identified five critical areas in their final report with a total of 14 key recommendations in a total of 28 recommendations (PPBE Commissioners, 2024). The five critical areas, with their recommendations, are provided in the following list. Roman numerals in bold identify the critical areas with their associated recommendations in the numbered sub-list. Key recommendations are in the sub-list are in bold.

- I) Improve the alignment of budgets to strategy**
 - 1) Replace the PPBE process with a new defense resourcing system**
 - 2) Strengthen the defense resourcing guidance**
 - 3) Establish continuous planning and analysis**
 - 4) Transform the budget structure**
 - 5) Consolidate RDT&E budget activities
- II) Foster innovation and adaptability**
 - 6) Increase availability of operating funds**
 - 7) Modify internal DoD reprogramming requirements
 - 8) Update values for below threshold reprogrammings**
 - 9) Mitigate problems caused by continuing resolutions**
 - 10) Review and consolidate budget line items**
 - 11) Address challenges with colors of money**
 - 12) Review and update PPBE-related guidance documents
 - 13) Improve awareness of technology resourcing authorities
 - 14) Establish special transfer authority around milestone decision
 - 15) Rebaseline OSD obligation and expenditure benchmarks
 - 16) Encourage use of the defense modernization account
- III) Strengthen relationships between the DoD and Congress**
 - 17) Encourage improved in-person communications**
 - 18) Restructure the Justification Book
 - 19) Establish classified and unclassified communication enclaves**
- IV) Modernize business systems and data analytics**
 - 20) Create a common analytics platform**
 - 21) Strengthen governance for DoD business systems
 - 22) Accelerate progress toward auditable financial statements
 - 23) Continue rationalization of the OSD resourcing systems
 - 24) Modernize the tracking of congressionally directed actions
- V) Strengthen the capability of the resourcing workforce**
 - 25) Continue the focus on recruiting and retention**
 - 26) Streamline processes and improve analytic capabilities
 - 27) Improve training for personnel involved in defense resourcing**
 - 28) Establish an implementation team for commission recommendations

At a high level, we can align these critical areas to our primary coding themes as shown in Table 3.



Table 6. Alignment of the Five Critical Areas to Our Primary Coding Themes

#	Critical Area	Primary Coding Theme	#
1	Improve the Alignment of Budgets to Strategy	Budgeting	73
2	Foster Innovation and Adaptability		
3	Strengthen Relationships Between DoD and Congress	Transparency and Oversight	12
4	Modernize Business Systems and Data Analytics	Data Analytics & Metrics	12
5	Strengthen the Capability of the Resourcing Workforce	Workforce	17
Total Open-Source Literature Recommendations			114

From this alignment we see that the commissioners for the most part agreed with our primary groupings of areas to target. The area that do not align well is the second critical area, *Foster Innovation and Adaptability*, to our primary coding themes; further, the “Others . . .” and “PPBE Commission Should” primary coding themes can align across any of these critical areas. To identify the alignment of these two primary themes to the critical areas we manually aligned the secondary coding themes to these critical areas in Table 4.

Table 7. Alignment of “Others . . .” and “PPBE Commissioners Should” Primary Themes to the Five Reform Commission’s Critical Areas

#	Critical Area	Others ... Theme	PPBE Comm. Should Theme
1	Improve the Alignment of Budgets to Strategy	29	7
2	Foster Innovation and Adaptability	32	3
3	Strengthen Relationships Between the DoD and Congress	0	2
4	Modernize Business Systems and Data Analytics	1	10
5	Strengthen the Capability of the Resourcing Workforce	7	1

This assigns an additional 98 of the open-sourced literature recommendations to these five critical areas. This leaves a total of 16 unaligned recommendations. Of these, we classified six of these recommendations as resulting in the work the PPBE commissioners did. This leaves 10 unaligned secondary themes, which are Establish Task Force Review of Missed Opportunities, Fundamental Scrub of MDAPs, use Micro-services Architectures, Strengthen Capital Market Programs, Acknowledge Pork Barreling, two cases of Use Agile Requirements, and three cases of Accept Tension.

The final alignment by percentages are provided in Table 5.



Table 8. Manual Alignment of the Critical Areas to our Open-Source Literature Coding Themes

#	Critical Area	%
1	Improve the Alignment of Budgets to Strategy	49%
2	Foster Innovation and Adaptability	16%
3	Strengthen Relationships Between the DoD and Congress	6%
4	Modernize Business Systems and Data Analytics	10%
5	Strengthen the Capability of the Resourcing Workforce	11%
6	PPBE Commission’s Work	3%
7	Unaligned recommendations	5%

ChatGPT-4’s Alignment of Relevant Reform Commission Recommendations

Using ChatGPT-4, which allows us to upload the PPBE Reform Commission’s Final Report, we requested that it perform an assessment of the alignment of the intent of the literature’s recommendation against the Final Report’s 28 Recommendations (PPBE Commissioners, 2024). However, due to ChatGPT4 limitations on the number of questions that can be asked of it within a specified period based on system load considerations, we could not complete the entire analysis prior to the deadline for this paper. Hence, the results in Figure 5 are not completely representative, but do provide insight into the fraction assessed in this analysis.

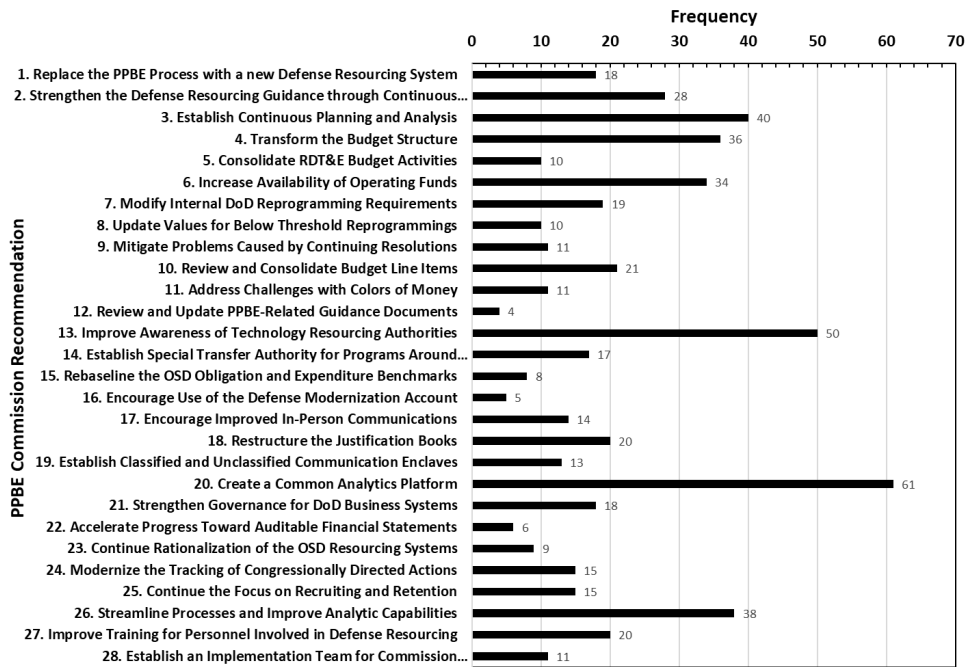


Figure 21. ChatGPT-4 Assessment of the Alignment of Intent From the Open Source Recommendations (Buettner et al., forthcoming) Against the 28 Recommendations from the PPBE Reform Commission’s Final Report

Finally, when doing the analysis for Figure 5, ChatGPT-4 typically returns about two or three of the commissioner’s 28 recommendations depending on the wording in the open literature’s recommendation. In some cases, ChatGPT-4 could return more if the open literature recommendation is in the form of a list, for example. In a spot check of ChatGPT-4’s answers after one or two days of doing the analysis, we found in several cases the recommendation alignment would change. For example, if the LLM indicated that it did not find any of the reform commission’s 28 recommendations that met the intent recommendation we were checking it would say so. However, a few days later it provided



answers that were somewhat reasonable. Hence, we believe for this case, the analysis should only be used as a guide.

Even though ChatGPT-4 has consistency issues and sometimes needs to have areas of alignment pointed out, these results suggest (based on the currently analyzed open literature recommendations) that the most important recommendation for the DoD to implement for this fraction of recommendations is recommendation 20, Create a Common Analytics Platform. With ChatGPT-4 looking for alignment across primary themes, we expected the alignment percentages to differ. This can be seen by comparing the results in Table 8 to those in Table 7.

Table 9. ChatGPT-4 Alignment of the Critical Areas to our Open-Source Literature Coding Themes

#	Critical Area	%
1	Improve the Alignment of Budgets to Strategy	23%
2	Foster Innovation and Adaptability	34%
3	Strengthen Relationships Between the DoD and Congress	8%
4	Modernize Business Systems and Data Analytics	19%
5	Strengthen the Capability of the Resourcing Workforce	15%

Conclusions

The results suggest that a significant fraction (almost half) of the Pentagon’s problems can be “self-corrected,” thus we consider this to be our first finding. There may, however, be some underlying bias based on an inherent clustering of SMEs associated with writing these recommendations having worked the PPBE process from within the OSD and Pentagon; without an analysis of the backgrounds of the authors and their contributors we cannot rule this out. Yet, there were also several recommendations suggesting actions that can be unilaterally taken by Congress or in collaboration with the DoD in support of obtaining a responsive-agile PPBE process.

From the QDA of the recommendations in Tables 1 and 2, we noted that there were several proposed actions to foster trust and transparency through modernized business systems, using, for example, real-time data analytics. As a result, we provided a proposed reference architecture that could digitally enable access to an acquisition program’s data and models. Figure 5 is a notional reference architecture of an integrated modeling environment showing the interacting processes, models and data in a modern digital engineering ecosystem supporting a weapon system’s acquisition. Figure 6 reorientates this notional reference architecture to show how data could flow to various external stakeholders.

These figures can be used as a starting point to discuss the data needs of the various external stakeholders. It is presumed that the realized (actual) cost and schedule data (earned value management [EVM] with performance against the integrated master plan [IMP] and integrated master schedule [IMS]) are the primary information that external stakeholders (e.g., congressional staffers) would need. In a modern digital acquisition, we should only have to identify the type, format, periodicity, and integrity assurances of the data that is required for Congress to fulfill its oversight role. Once identified, a proper mixture of enabling technology with policy and statute can support the visibility needed.

In addition, referring to the recommendations in Tables 1 and 2, we also observed that the 809 Panel’s Portfolio Management and Budgeting recommendation, Buy/Use



Commercial Technology, and Flexibility (under Budgeting) were significantly repeated themes, as were various workforce themes, including training and retention.

A comparative analysis of the open literature’s recommendations identified in Buettner et al. (forthcoming) currently suggests that the most important recommendation for the DoD to implement is PPBE Commission’s Create a Common Analytics Platform from the viewpoint of the literature’s intended areas for improvement. A more comprehensive and detailed manual analysis of these sources against the final report may provide different results, while completing the entire list of 222 recommendations from Buettner et al. (forthcoming) may result in a different prioritization.

Furthermore, no effort has been made to cross-check the recommendations found in the open literature against existing DoD efforts to implement them. For example, we are aware of DoD initiatives to implement 809 panel recommendations, and the Defense Civilian Training Corps (DCTC)⁸ where AIRC staff are supporting other on-going research efforts.

Figure 6 and Figure 7 provide different views of the same SERC/AIRC digital engineering reference architecture.

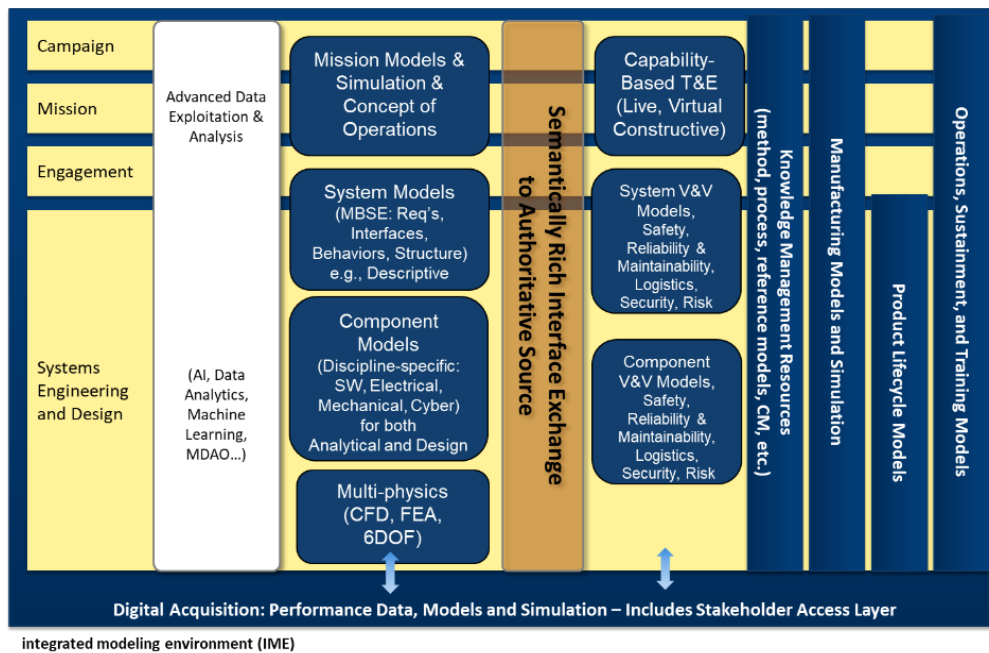


Figure 22. View of a Reference Architecture for an Integrated Modeling Environment (IME) Which Typically Supports the Digital Engineering of Modern Systems

⁸ See <https://dctc.mil/> for more information on the DCTC initiative.

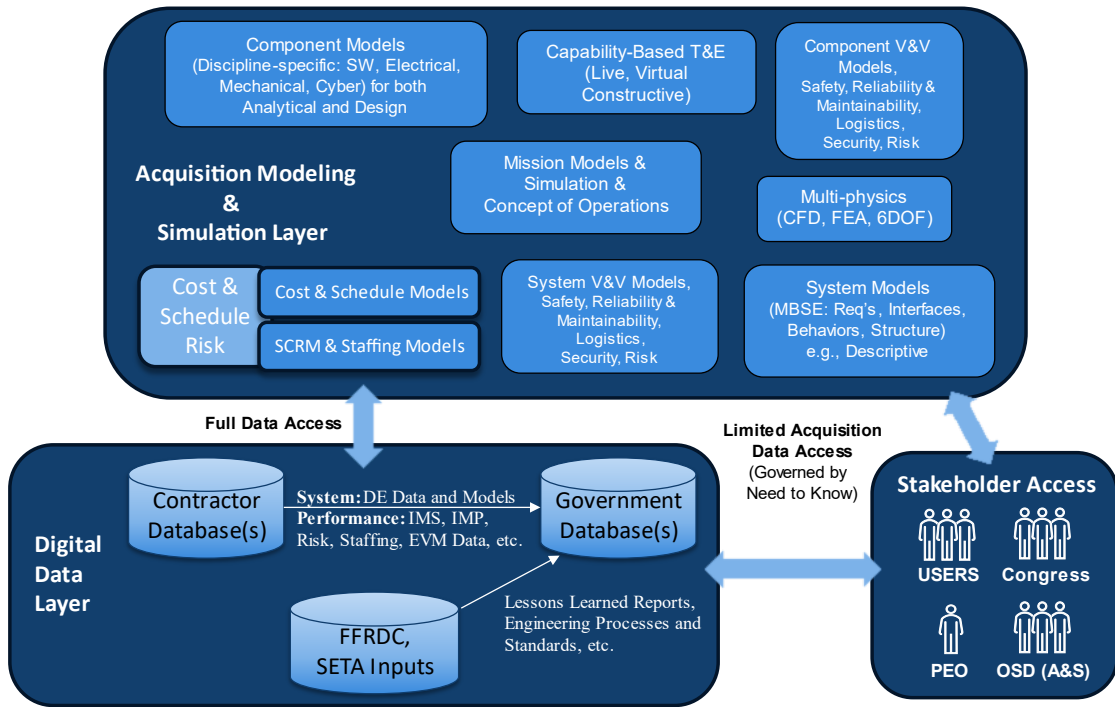


Figure 23. View of the Reference Architecture in Figure 6 Showing the Proposed Data Communication for a Major Defense Acquisition Program (MDAP) that Could Foster Rapid Data Exchange and Support Trust Between All Stakeholders

Figure 6 is intended to show the modeling and data in the various swim lanes of a modern digital acquisition. Figure 7 modifies this view slightly to show a possible digital transformation communication infrastructure that could more rapidly support information flow between success critical stakeholders of the MDAP, the program office, and the contractor.

Finally, in this paper we did a cross-comparison of the recommendations found in the open-literature (Buettner et al., forthcoming) and found a reasonable alignment to the recommendations found in the PPBE Reform Commission’s Final Report in theme (PPBE Commissioners, 2024). In fact, there are cases where the open-literature and the PPBE Reform Commission’s recommendations are identical.

Appendix A: Secondary Themes for “Others...” and “PPBE Commission Should” Primary Themes

Table 10. “Others...” Category: All Secondary Themes

Others...: Secondary Themes	%
Buy/Use Commercial Technology	13%
Empower Senior Leadership Team	4%
Align Work to Primary Mission	3%
Analyze PPBE Processes	3%
Nontraditional Industrial Base	3%
Rapid Acquisition	3%
Revise Reprogramming Processes	3%
Use Agile Requirements	3%
Supply Chain Capacity Focus	1%
Accountability	1%
Acquisition Based Risk Tolerance	1%
Appropriate All-Phase Performance Based Competitions	1%
Clearly Articulated Leadership Role in PPBE Processes	1%
Contingency Planning	1%
Continuous Cybersecurity Verification	1%
Contracting	1%
Create Supply Chain Intelligence Center	1%
Cycle Time Based Contracting	1%
Dedicated Reform Team	1%
Delegate Authority to Emphasize Speed	1%
Digital Engineering	1%
Empower PPBE Process Czar	1%
Establish Informal Strategic Analysis Forum	1%
Establish Long-term Forum	1%
Establish Task Force Review of Missed Opportunities	1%
Evaluate Effectiveness of BA-8 Pilots	1%
Expand SDA Model	1%
Fundamental Scrub of MDAPs	1%
Implement Review Process for Resource Based Needs	1%



Others...: Secondary Themes	%
Industrial Base	1%
Industrial Base Requirements Transparency	1%
Integrated Business Analytics Platforms	1%
Micro-services Architecture	1%
Modernize Defense Research Laboratory	1%
Modular Multiple Award Contracting	1%
Move PEOs and PMs to System Commands	1%
PAF Acquisition Acceleration Tool	1%
Pilot Alternative Allocation Resource Process	1%
Pilot Efficient Streamlined MDA Processes	1%
Prioritize Customers	1%
Proactively Prioritize Joint Needs	1%
Reduce ID/IQ Contract Scopes	1%
Reestablish ADCP as a DMAG	1%
Reform Leadership Structure	1%
Require AI Readiness in MDAPs	1%
Requirements	1%
Research China vs US Processes	1%
Reset Reprogramming Authorities	1%
Resource Allocation	1%
Return Acquisition Oversight to Services	1%
Revise Executive Branch Processes	1%
Revise Requirements Management Processes	1%
Senior Leadership Own & Integrate PPBE Processes	1%
Standardize PPBE Processes	1%
Standardize Printed Parts	1%
Strategy Based Priorities	1%
Streamline Regulations & Implement Authorities	1%
Strengthen Capital Market Programs	1%
Strengthen Concept to Fielding Activities	1%
Strengthen Defense Planning Guidance	1%



Others...: Secondary Themes	%
Strengthen PPBE Execution Processes	1%
Use Customer Focused Cross-Functional Teams Model	1%
Waive Competition for No Cost/Innovation Advantage	1%

Table 11. “PPBE Commissioners Should” Category: All Secondary Themes

PPBE Commissioners Should: All Secondary Themes	%
Accept Tension	11%
Budgeting	11%
Data Analytics & Metrics	11%
Predictive Analytics	7%
Transparency and Oversight	7%
Use Historical View	4%
Acknowledge Pork Barreling	4%
Address Planning Processes Lack of Analytical Framework	4%
Change Timing/Sequencing	4%
Civil-Military Integration	4%
Digitally Transform Business Systems	4%
Equities as a Package	4%
Establish PPBE Lessons Learned Feedback Processes	4%
Identify Business Accountability Processes	4%
Prioritize Expanding Specific Programs	4%
Provide Impactful Recommendations	4%
Rebuild Strategic Analysis	4%
Recommendations Enabling Speed & Agility	4%
Rent Technologies with Marketplace Enabled Sustainment	4%
Simplify Financial Management	4%

References

Buettner, D. J., Cuff, T., Anton, P. S., & Mercado, J. M. (forthcoming). *Planning, programming, budgeting, and execution (PPBE) literature review: Analysis of public*



recommendations. Stevens Institute of Technology, Acquisition Innovation Research Center.

Cardon, E., Drabkin, D., Branch, E., Dyess, R. M., Masiello, W., Smith, M., . . . Anton, P. S. (forthcoming). *Improving defense outcomes through improved interfaces between PPPBE, acquisition, and requirements*. Stevens Institute of Technology Acquisition Innovation Research Center.

Congress. (2022). *FY22 NDAA section 1004*. <https://ppbereform.senate.gov/section1004-fy22-ndaa/> or <https://www.congress.gov/117/plaws/publ81/PLAW-117publ81.pdf>

PPBE Commissioners. (2024). *Commission on planning, programming, budgeting, and execution (PPBE) reform: Defense resourcing for the future—Final report*. Senate.



Leveraging Digital Transformation for Resourcing Investments

Jason Thomas—serves as the Digital Transformation Lead at the Naval Air Warfare Center Aircraft Division (NAWCAD) in Patuxent River, Maryland, where he is responsible for establishing, integrating, and implementing the Command's Digital Transformation strategy. Primary lines of effort include, but not limited to, MBSE, MBE, Sustainment, Data Analytics, Digital Infrastructure, Workforce Development, Test and Evaluation, across NAWCAD and its Echelon IV Commands. He is a Joint and Information Warfare qualified Aerospace Engineering Duty Officer in the Navy Reserve with experience in Space Operations, Surface, Air, and Information Warfare, and is a Systems Engineering PhD student at NPS. [Jason.j.thomas34.civ@us.navy.mil]

Abstract

Resourcing in the DoD is a challenging endeavor. A program must navigate various appropriations that align to both their strategic goals and tactical needs, all the while external factors such as skill set availability and congressional winds may delay or derail well-thought-out plans. A program must also leverage a myriad of subject matter experts to design, develop, field, and sustain its complex weapon systems and when resourcing constraints arise, knowing what products and services provide the best business value is not always clear. Lastly, the lack of feedback mechanism for a specific product or service are not always well understood or even measured. This can lead to questions regarding resourcing needs and impact to secure resources at a future state. This vicious cycle remains unchanged and requires an innovative approach for resourcing priorities in support of program and mission needs. Previous attention in this area has focused on restructuring the PPBE process or purely mathematical models that rarely capture the complex dynamics of the ecosystem. This research supports this thirst for innovation by leveraging digital transformation to identify the relationships in a systems-of-systems approach and provides analytical methods analysis to achieve that end-state.

Introduction

Making resourcing and investment decisions in the Department of Defense (DoD) is challenging. The DoD has many customers, stakeholders, appropriations, and priorities that must be weighed before a budget can be finalized and funding promulgated and initiatives staffed. The procurement process is the backbone for developing and sustaining military capability and is comprised of the Joint Capability Integration Development System (JCIDS), Acquisition Process, and the Planning, Programming, Budgeting, and Execution (PPBE) process (Figure 1; DAU, 2023).

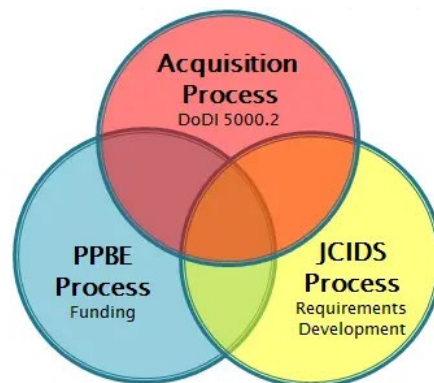


Figure 24. Defense Acquisition System



The JCIDS process is what provides the documentation and baseline of validated capability needs. The PPBE process is used to determine funding requirements across the various defense appropriations and allocate resources to provide the capabilities deemed necessary to accomplish the Department's mission. PPBE is the DoD system for allocating resources among the armed services, defense agencies, and other components, and is the primary mechanism in which program offices or command teams document and request funding to execute their programs (McGarry, 2022). Although well-ingrained, the process has received additional scrutiny over the past few years due to a perception of obsolescence and the need to make adjustments in funding acquisitions given the pace or technology availability and evolving threat baselines (Hale & Lord, 2024). The acquisition process is the overarching approach programs and support teams procure and acquire capabilities directly related to or in support of military weapon systems design and support. The approach, or pathway, to procurement has been modernized in the past few years to allow for different types of systems and needs to be acquired in a manner that more aligns with industry and development timelines (Figure 2).

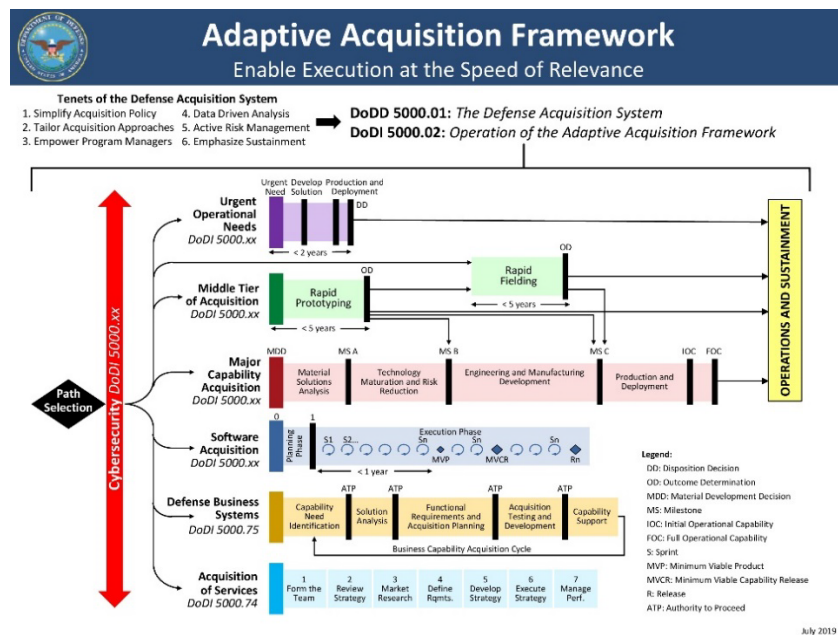


Figure 25. Adaptive Acquisition Framework

It is within each of these independent processes, JCIDS, PPBE, and acquisition, as well as between them, that decisions need to be weighed on what, where, when, why, and how to invest. Fielded weapon systems may need port, facility, or system upgrades to alleviate maintenance, reliability, or production strains. A system command (SYSCOM) or warfare center may want to invest in IT infrastructure for analytical capabilities or to sundown legacy systems. New science and technology opportunities may arise and investments must be planned ahead of time with the right appropriation to incorporate.

With any investment, the key considerations are how much funding is required, when does it need to be paid, and for what? Additionally, any smart investor would want to know who receives the benefit, how is the benefit measured, and when is the benefit realized? Beyond those initial questions, it's important to consider other factors such as what are the short-term and long-term impacts of choosing one path instead of another, and if one path is chosen, are you prevented or limited in your options in the future? This concept of opportunity costs (or opportunity capabilities, readiness, or another measure) must also be part of the decision space

as it aligns to the organizations strategic intent and core values (Fernando, 2024). With precious human capital, financial resources, development time, risk tolerance of global events, insights for investment decisions in the DoD have never been more critically monitored (Beckley & Brands, 2024).

The Funding Process

The yearly PPBE process requires teams to identify their future needs to support their program, the necessary appropriations for the products and services, and the estimates of the funding required. As the PPBE process unfolds, the requested funding becomes authorized and finally appropriated into the year of execution budget. Programs are informed of what they are authorized for and begin the decision process on what to prioritize based on funding received. Planned investments for program offices may include weapon system reliability or maintainability improvements, or training on internal process efficiencies, including analysis and training. Programs must face this challenge each year as part of the budget process and planning for future year's needs. Planned activities, including hiring, may need to be deferred one or more years based on not only funding, but also as a result of events that occur throughout the year. For example, in 2017 the T-45 and F-18 fleets had an increase in physiological episodes that required time, effort, and resources to address what was the number one safety issue in naval aviation (Hudson, 2017). Teams external to program offices, whether it be at the Command level or lower, plan investments for IT capabilities and processing, facility and range improvements, connection nodes between government and industry partner sites, or specific investments to business operations to improve internal process such as new methodologies or training. For non-program teams, it is observed that investments in the DoD tend to be made by aligning to overarching or new strategies, such as a conscious focus to modernize infrastructure or workforce development. The challenges internal to a single program are difficult enough, and are compounded when improvements in one weapon system that are now deferred have a cascading effect to a systems of systems (SoS) capability that will now be delayed. Localized decisions rarely remain contained to a single team, and the result is a constant reactive nature across a program office, capability kill-chain, or a Command.

How a benefit is articulated, conveyed or conceptualized is important. More important is how the needs are defined and understood. A poor understanding of the problem limits the effectiveness or the solution or benefit. Understanding must be the precursor as to whether the solution enabled by the funding request is achievable or dependent on other factors. If a deeper understanding of the needs is not conveyed, the requirement for funding in the POM submission is more likely to be viewed as a "shopping list" of wants. This perception erodes trust between stakeholders at different echelons and further hinders conversations. Furthermore, who asks first, which program is more high profile, or which program is the "flavor of the month" are not sustainable approaches for delivering warfighter advantage at a cost we can afford. Programs have been on a journey to best communicate their needs to those who are not involved in the day-to-day operations of their specific program, receive requests from many programs, have insufficient time to process the requests, and must balance tactical and strategic perspectives.

Digital Transformation

Digital Transformation (DX) is the adoption and integration of digital technologies to improve efficiency and streamline business operations across all sectors of a business, fundamentally altering how value is brought to customers (Gebayew et al., 2018; Hanelt et al., 2021; Libert et al., 2016). Examples of DX technologies are artificial intelligence (AI), cloud computing, autonomy, and advanced model and simulation (M&S; Waugh, 2022). Developing and leveraging advanced M&S, teams at various echelons can run scenarios for what are the best use of their resources, for a given outcome, at a specific point in time. Integrating cloud computing allows for better accessibility, resilience, and processing power for specific



improvements and investments under consideration. AI can highlight indicators of good returns on investment and the environmental factors that strengthen or reduce the benefit.

Despite these well-known axioms, many organizations have continued to struggle with implementing these for resourcing investments in the most optimal way. There remains a lack of understanding in how resources are prioritized and aligned in support of programs. The lack of feedback mechanisms for the funded products and services and their impact on mission outcomes remains unclear in many cases. This paper looks at a novel approach to connect DX with investment decisions with the needed development of robust architectures and heuristics that will provide the right insights when they're needed and to whom. By understanding, planning and embracing the right technologies for the right purpose at the right time, DX offers the opportunity to approach how investment decisions are understood, measured, and made that most optimally achieves an organizations tactical and strategic goals.

Quantifying Investment Impacts

Get Real

The desire to know what initiative or investment is going to have the most impact on affordability, availability, or other metric given limited resources is an ever-present endeavor. With competing priorities, resource volatility, and the POM process requiring confident forecasting of not only funding amount, but also appropriation type, establishing known returns on the investments can feel like an effort in futility or at least a significant handicap when asked to defend budget requests. This reality has only been exacerbated over time with an aging inventory requiring more resources in sustainment accounts. One funding account has seen a four-fold increase in an eight-year span (Figure3).

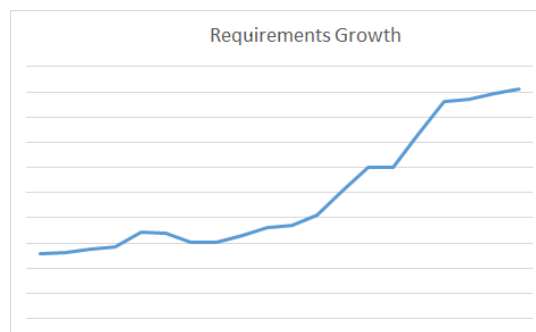


Figure 26. Requirements Increase

A program must be its own best advocate. The saying goes, “you won’t know until you ask,” and requesting additional funding is the responsibility of each program. In fact, additional funding was the reason why weapons systems can achieve their availability targets (Abott, 2019). While no one would dispute that more funding is needed to meet operational demands, it is also fair to ask if the resources were allocated in the most efficient way, and what was sacrificed in the process.

Embracing the Red

Analytical models to assist programs with investment decisions have existed for decades. Within the early acquisition and development communities, models have primarily consisted of requirements development and verification to requirements, mission effectiveness, engineering physics-based models for system performance, and integrated test and evaluation, with only recent developments infusing costs predictions (Byron, 2021; DAU, 2021). Within the logistics, operations and support, and sustainment communities, they have typically focused on



decisions pertaining to maintenance, supply provisioning, funding account adjustments, operational availability, resourcing-to-readiness, logistics footprint, manpower, impacts of reliability improvements, and life cycle costs estimates (Draham, 2017). While conceptually each had value and remains a needed area of better insights, all had significant limitations and all but one was ultimately cancelled (SPA, 2022). Other investment areas need to be considered as well, such as the Central Test and Evaluation Investment Program (CTEIP), as DoD's corporate investment vehicle for modernizing DoD test infrastructure; Capital Investment Program (CIP), which establishes a capability for reinvestment in the infrastructure of business areas in order to facilitate mid and long term cost reductions in order to improve product and service quality, timeliness, and reduce costs and foster comparable and competitive business operations; and the Naval Innovation Science and Engineering (NISE) program, which provides a way to conduct innovative basic and applied research, transition technologies into operational use, develop the workforce, recruit and retain highly skilled scientists and engineers, and purchase state-of-the-art labs and equipment (DoD Financial Mgmt, 1994; Miller, 2022; Paust, 2016). The author has found no such model that was or in development to assess the impact of command-level or team-level investments that tie to mission outcomes. The author also believes that business case analysis (BCA) is a better measure to assess investments since return-on-investment (ROI) may imply a financial return is the primary reason and driver for an investment. Lastly, there is the potential for investment in one account or area to be duplicative to one in another if a broader understanding and traceability is not well understood.

In addition to analytical or system modeling, several initiatives have emerged over the last 20 years to help program teams and resource sponsors define where the next dollar should be spent. These initiatives have been organically generated at a small team level, at a Systems Command (SYSCOM) level, and even at the Echelon 1 level. These efforts focused on resourcing investments to reduce costs or improve readiness of systems as the primary goal. A sampling of such efforts has included the Cost Wise Readiness Integrated Improvement Process (CWRIIP) as an overarching process for all Type/Model/Series (TMS) to prioritize Ready for Tasking (RFT) aircraft and cost degraders at the system and part level; the Predictive Analytics Modul (PAM), which was billed as a "modeling and simulation (M&S) capability to accurately represent and evaluate future maintenance and supply functions for ship- and land-based naval aviation operations"; enterprise PAM (ePAM), which was a scaled version of PAM; the Root Cause Analysis Model (RCAM), which was the next iteration of PAM when PAM and ePAM were unfunded; the Cost and Readiness Impact Model (CRIM), a suite of interactive analytics programs that uses an extensive data set to take into account all aspects of the operational and sustainment environments through advanced data analysis, machine learning and artificial intelligence for the sustainment enabler account; and the Readiness Decision Impact Model (RDIM), which looks more strategically on balancing short term and long term decisions (Draham, 2017; Jenkins, 2007; Myers, 2022; NAVAIR, 2022). In addition to these organically developed solutions, commercial-off-the-shelf tools have been pursued, but tend to come up short due to the "black box" nature of their construction or the dependence on personal heuristics rather than and understanding of how the various elements interplay with one another and drive outcomes.

Always Learning

In aligning to the Navy's Get Real, Get Better initiative, it is instructive to deep-dive one of these analytical models as a way of self-assessing our approaches to the investment decision problem set so that we may self-correct, learn, and get better (Franchetti, 2024). As part of the Re-Imagine OPNAV initiative, CRIM was developed to link investment dollars from the 1A4N (now Air Systems Support) enabler account to mission outcomes, by measuring high-level availability and affordability metrics. This appropriation is used for technical publication and maintenance plan updates, reliability analysis and improvements, mission system software



updates, readiness and condition-based maintenance analysis, policy updates, IT software and license, and many other important products and services to ensure aircraft are available, reliable, safe, and lethal (Figure 4).



Figure 27. Air Systems Support Products

To fully understand the feasibility of such an analytical capability, data from other appropriations and products and services type were used to test the baseline integrated capability of a combination of organic and COTS products. While successful for the initial investigations, and follow on specific and bespoke analysis in support of AirBoards, the underlying model, data capture, and data governance structure were determined to be more limiting. Scalability outside the initial analyses proved to be a challenge and the data captured in various input modules were cumbersome and onerous. Since the focus of CRIM was on the cost and availability metrics, a thorough verification and validation (V&V) effort was required to build credibility with the organization and stakeholders. The V&V was focused on availability with the metric of mission capable aircraft and the affordability calculation. Sample results of two different platforms are shown below for the improvements to mission capable aircraft (Figure 5).

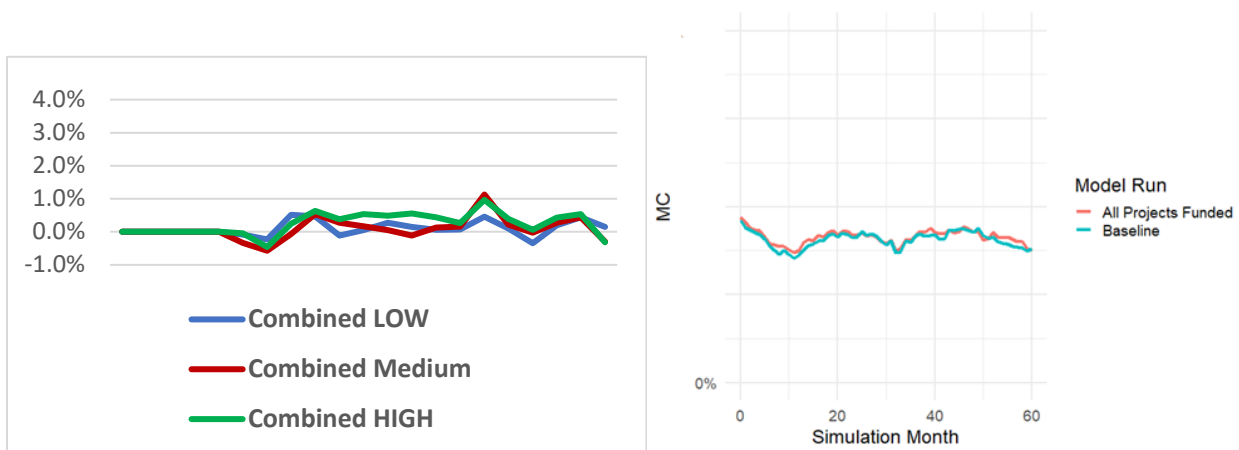


Figure 28. Availability Predictions

The first result shows five initiatives against specific system components with realistic thresholds for low, medium, and high levels of improvements and the simulated change in mission capable aircraft as a result over time. The second platform shows a similar result with



more initiatives on specific components of interest selected by the respective program office. This raises a few questions: 1.) is the SIMLOX discrete event simulation accurate, 2.) were the models built right, 3.) why were these system components selected by the programs, and 4.) how comprehensive was the understanding of the impact to the improvements? These results call into question how well the COTS simulation product captures the sustainment ecosystem and actions, products, and services therein, as well as subject matter experts understanding of the connectedness and impacts of certain investments on mission outcomes (Gelsinger, 2022; IBM, 2023). The second part of the V&V effort was on the affordability comparisons. Does CRIM as constructed offer appreciable insights or reduce the burden on programs and allows them to make better decisions, faster? In an analysis conducted in spring of 2022, a comparison between the traditional “back of the napkin” math for cost avoidance for an initiative was compared to the CRIM output. In the simplified approach, the assumption goes something like, “if we double the reliability, we will halve the number of failures, meaning we have a cost avoidance of 50% * unit cost.” In the simulated result, multiple Monte Carlo simulations are run and the number of demands for a component is averaged and then multiplied by the unit price and that results in the cost avoidance. In the 2022 analysis, the results were nearly identical, which would beg the question, is “back of the napkin good enough if it’s cheaper and I don’t need specific training on a more complicated simulation?”

Furthermore, as of October 2022, none of the Air Systems Support investments could be assessed for their “quantitative effect” for “readiness either with SIMLOX models or with bespoke analysis” (Gelsinger, 2022). SIMLOX is regularly used for ready based sparing (RBS) analysis and provisioning within supportability analysis. It is also the foundation for the Navy Common Readiness Model (NCRM) and the Model-Based Product Support (MBPS) effort (Sashegyi, 2020). OPNAV has directed that “all maritime programs and systems must use the Navy Common Readiness Model (NCRM) to execute RBS” (CNO, 2022).

These quantitative results offer additional insights that were unknown previously. Why was mission capable aircraft barely noticeable in SIMLOX even with significant increases in reliability? Were those the right systems the programs should have been focused on? What else is happening in the “sustainment ecosystem” that is not captured? What are the limitations on some of our COTS products and what role does data rights or industry partnerships play in ensuring the products are build and structured for the right purpose? If preprocessing tools and data curation algorithms do not affect the underlying behaviour of the models or simulations, then the DoD runs the risk of convincing itself value is at hand when fundamental issues still persist.

A New Approach

Integrated architecture

In the prioritization and estimation development of budget submissions, teams will use a wide range of methods to determine and justify their requests. Some approved methods are as straightforward as “back of the napkin math” with the cost avoidance being fewer parts orders multiplied by the unit cost. Other methods include a slightly more sophisticated Analytical Hierarchy Process (AHP) or Multi Attribute Utility Theory (MAUT) which still depends on subjective valuation based on how a team understands the problem and dependencies (Forman & Gass, 2001). In fact, some Navy commands have adopted specific vendors and approaches aligned to this framework (Martin, 2023). The DoD requires better understanding of their decisions to make more informed decisions beyond the approaches of yesterday.

The reoccurring issue with the above approaches is that each, in their own way, are myopic in nature and fail to address the fundamental definition of “the system.” A model does



not have to be complex but it should be representative and built for a specific purpose with assumptions well-defined (Law, 2015). Each method and approach above were diagnostic in nature without the understanding of the “why” or defining system boundaries. Statements like, “accurately represent and evaluate future maintenance and supply functions for ship- and land-based naval aviation operations,” only work to undermine credibility in the M&S profession. Furthermore, without alignment on interoperability and conceptual understanding, technical solutions will limit usefulness (Tolk, 2023). Proper root case cannot be conducted because what you see may not be what you observe, which is only fostered through a transparent and inquisitive mindset.

Integrated architecture, in this sense, are all the business processes and functions that show the flow of data, the consumers and providers, and the purpose of products and services. This is the “as-is” state of understanding of a system. The resolution and fidelity can be scope with understanding of subject matter experts, but one would caution on the recency and comprehensive understanding of said experts. The architecture should not be developed solely by systems engineering, logisticians, operators, maintenance personnel, or any other single group, but in a collaborative group focused on honest outcomes. Through this process of documentation of the architecture, it is important to note that modeling is not the same cataloguing or diagramming. Semantic precision must be adhered to for the broad uses of such a model, especially for programs and teams across various stages of the life cycle, with different partners and relationships and needs and requirements.

The Future is Here

What is possible now? Most anything that is needed or desired. Consider the background and discussion above and think why could that information and those data points be integrated in the a new digitally transformed way of operating? If the traceability and dependencies and usefulness of these models, components, and guidance were linked in a digital space, one could quickly see the dependencies and choose to make a decision to invest in V&V or a more deliberate approach to adoption instead incurring technical debt. Digital transformation is not about IT, tools, data, or people only, it’s the integration of all of these elements into a way we transform to think, act, and operate differently. What is preventing a program or command from documenting their processing, architecting their business operations, identifying the investment sources, and instrumenting their workflows to understand where the biggest impacts are being observed? Consider the summary of representations of a *single* model with different *views* for a different stakeholder (Figure 6).

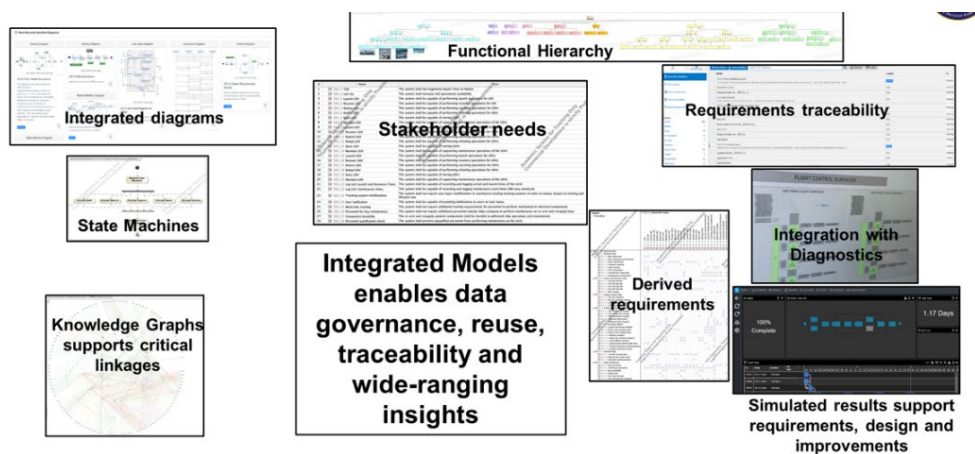


Figure 29. Digital Engineering Examples



The team could have system, mission, or business requirements, stakeholder needs, requirements traceability with triggers or thresholds for quality, and any number of simulations to assess performance. One could easily see how leveraging digital transformation for any number of investment decisions would be captured throughout the *single model* and individuals on the team could be proactive in their planning and recommendations with a common framework and understanding. External stakeholders would have confidence in what funding is being requested and the expected impacts. Furthermore, teams could run “what-if” analyses on deferment of an action, requesting a waiver to policy, or identify if an investment from another appropriation or account (CIP, CTEIP, etc.) will support their needs on a timeframe that is relevant to them. These are engineered systems, but the competency and knowledge resides in architecting and a vast amount of domain knowledge across the acquisition and operational continuum and at multiple tiered echelons. The problem space is full of wicked problems and having confidence in what must be done, what the benefits are, and where the trade space is will only be solved with a new methodology and framework that will be able to meet the demands of insights, speed, and confidence in the world ahead.

Recommendations and Future Work

The framework identified is discussed in many corners and pockets across the DoD, but understanding of the business value appears to be preventing teams from committing to the approach. It is important for each team to know their goals, objectives, and requirements are and what the pain points, user stories, and other methods of soliciting feedback are to understand their system and share that information widely. Transparency must be a foundational element of any future work in the digital and financial space. The DoD’s mission is one of national defense and individuals should not focus on what platform is in the news or on the cover our favorite periodical, but in the collective knowledge we are all doing our part for the warfighter and taxpayer. Teams should not fear the digital future as there’s already too much work to do as it is and knowing we are working on the most important thing only adds to fulfillment, retention, and innovation. It is recommended that future work in this space be on fundamental architecting and measuring various business operations and aligning investments to those areas. It is expected that requirements for investments may be duplicative without being redundant across a myriad of accounts and teams should set aggressive goals leveraging this approach.

References

- Abott, R. (2019, September 25). *Navy says it achieved 80% mission capable goal for super hornets and growlers*. Defense Daily. <https://www.defensedaily.com/navy-says-it-achieved-80-mission-capable-goal-for-super-hornets-and-growlers/navy-usmc/>
- Beckley, M., & Brands, H. (2024, April 2). How primed for war is China? *Foreign Policy*. <https://foreignpolicy.com/2024/02/04/china-war-military-taiwan-us-asia-xi-escalation-crisis/>
- Byron, M. (2021). *Enabling cost estimating in cameo systems modeler*. Dudley Knox Library.
- CNO. (2022). *Ready based sparing*.
- DAU. (2021, August 1). *Modeling & simulation overview*. AcqNotes. <https://acqnotes.com/acqnote/tasks/modeling-simulation-overview>
- DAU. (2023, August 7). *Procurement overview*. AcqNotes. <https://acqnotes.com/acqnote/acquisitions/procurement-overview>
- DoD Financial Mgmt. (1994). *Chapter 58 capital assets*.



- Draham, S. (2017, May 24). *Naval aviation readiness models list v9*.
- Fernando, J. (2024, April 1). *Opportunity cost: Definition, formula, and examples*. Investopedia. <https://www.investopedia.com/terms/o/opportunitycost.asp>
- Forman, E. H., & Gass, S. I. (2001). The analytic hierarchy process—An exposition. *Operations Research*, 49(4), 469–486. <https://doi.org/10.1287/opre.49.4.469.11231>
- Franchetti, L. (2024). *Get real get better behaviors*.
- Gebayew, C., Hardini, I. R., Panjaitan, G. H. A., Kurniawan, N. B., & Suhardi. (2018). A systematic literature review on digital transformation. *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 260–265. <https://doi.org/10.1109/ICITSI.2018.8695912>
- Gelsing, M. (2022). *Assessing the feasibility of modeling 1A4N outcomes1* [Memo]. Institute for Defense Analysis.
- Gelsing, M. (2022, April 1). *MH-60 ADTS initiative analysis*.
- Hale, R., & Lord, E. (2024). *Defense resourcing for the future*. Commission on Planning, Programming, Budgeting and Execution Program.
- Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, 58(5), 1159–1197. <https://doi.org/10.1111/joms.12639>
- Hudson, L. (2017, April 28). *VCNO directs 30-day review of T-45, F/A-18 physiological episodes*. Inside Defense. <https://insidedefense.com/inside-navy/vcno-directs-30-day-review-t-45-fa-18-physiological-episodes>
- IBM. (2023, February). *CRIM–SIMLOX sensitivity analysis*.
- Jenkins, J. (2007, September 24). *CWRIIP helps reduce costs, refocus resources*. NAVAIR. <https://www.navair.navy.mil/node/8286>
- Law, A. M. (2015). *Simulation modeling and analysis* (Fifth ed.). McGraw-Hill Education.
- Libert, B., Beck, M., & Wind, Y. (2016, July 14). 7 questions to ask before your next digital transformation. *Harvard Business Review, Business Models*. <https://hbr.org/2016/07/7-questions-to-ask-before-your-next-digital-transformation>
- Martin, J. (2023, October 12). *Press release: Decision lens software selected for use by commander, naval information forces (NAVIFOR)*. <https://www.decisionlens.com/blog/press-release-decision-lens-software-selected-for-use-by-commander-naval-information-forces-navifor>
- McGarry, B. W. (2022). *Defense primer: Planning, programming, budgeting, and execution (PPBE) process*. Congressional Review Service. <https://crsreports.congress.gov/product/pdf/IF/IF10429>
- Miller, S. (2022, February 9). *NSWC Crane leveraged NISE funds to prioritize hypersonic and digital engineering technology*. Naval Sea Systems Command. <https://www.navsea.navy.mil/Media/News/Article/2929516/nswc-crane-leveraged-nise-funds-to-prioritize-hypersonic-and-digital-engineerin/https%3A%2F%2Fwww.navsea.navy.mil%2FMedia%2FNews%2FArticle-View%2FArticle%2F2929516%2Fnswc-crane-leveraged-nise-funds-to-prioritize-hypersonic-and-digital-engineerin%2F>



Myers, M. (2022, December 29). *New data model predicts how deployments affect future readiness*. <https://www.militarytimes.com/news/your-military/2022/12/29/new-data-model-predicts-how-deployments-affect-future-readiness/>

NAVAIR. (2022, April 26). *Sustainment croup teams net two NAVAIR commander's awards*. <https://www.navair.navy.mil/news/Sustainment-Group-teams-net-two-NAVAIR-Commanders-Awards/Tue-04262022-1706>

Paust, C. (2016, March 23). *Preparing the test infrastructure for a cyber world*.

Sashegyi, M. (2020, March 31). *Model based product support heralds in new-age logistics and engineering*. DVIDS. <https://www.dvidshub.net/news/366283/model-based-product-support-heralds-new-age-logistics-and-engineering>

SPA. (2022, June 13). *Naval synchronization toolset initiative*. Systems Planning & Analysis. <https://spa.com/news-insights/naval-synchronization-toolset-initiative/>

Tolk, A. (2023). Conceptual alignment for simulation interoperability: Lessons learned from 30 years of interoperability research. *Simulation*, 00375497231216471. <https://doi.org/10.1177/00375497231216471>

Waugh, S. (2022, July). Digital strategy return on investment. *Defense AT&L Magazine, Defense Acquisition University* (July–August 2022), 6–12.



Leveraging Machine Learning and AI to Identify Alternative Parts to Increase Parts Availability and Improve Fleet Readiness

Christopher Bailey—is a Data Scientist at Govini. Prior to joining Govini, he spent 3 years working at various positions within the mortgage industry, most recently working as a Data Scientist at United Wholesale Mortgage developing and maintaining machine learning algorithms for capital and secondary markets. He holds a bachelor's degree in physics from Michigan State University and a master's of science in data science from Syracuse University.

Mihiri Rajapaksa—is an Engagement Manager at Govini. Prior to joining Govini, she spent over 5 years working as a civilian for the U.S. Navy. She started at NSWC Philadelphia as an HM&E ancillary equipment SME and then transitioned to PEO IWS 3.0. Within IWS 3, she worked in IWS 3L, the Vertical Launching Systems Program Office, for over 4 years, overseeing new construction and fleet support as well as various tech refreshes and engineering development efforts. She holds a master's of science in robotics from the University of Pennsylvania.

Abstract

As competition between the United States and near-peer adversaries intensifies, the U.S. Navy faces increasing challenges to its sea dominance. Fleet readiness, backed by superior Naval capabilities, is critical to credibly project U.S. power and deter conflict in the region.

The speed and agility of the U.S. industrial base to maintain operational availability (Ao) is foundational to readiness. However, obsolescence issues such as parts shortages plague weapon systems, negatively impacting Ao. Leveraging artificial intelligence and machine learning (AI/ML) processes to quickly identify potential alternative parts can greatly speed up the time required to identify replacement parts.

Currently, to remedy these issues, engineers must manually scour hundreds of sources and compare a multitude of technical characteristics to identify alternative parts, a time- and labor-intensive process. To address this need, this study developed an LLM-base AI model to quickly compare multiple parts, rank them based on similarity to the part under investigation, and ultimately identify feasible alternatives. The output is a prioritized ranking of parts, based on model-determined similarity of form, fit and function of the parts. The model-recommended parts are then analyzed for current stock on hand to identify the most viable parts that could also be quickly accessed.

Introduction

Traditionally, Navy engineers are notified once a part is unable to meet fleet requirements and tasked with identifying alternatives to maintain shipyard availability schedules. This manual process is time-consuming, labor-intensive, and does not always yield fruitful results, sometimes overlooking potential replacement solutions. Without the ability to compare the potential form, fit and function across a multitude of potential parts at scale, this process will continue to be a bottleneck to addressing fleet sustainment challenges. Without a way to speed up this process and ensure fleet readiness, the U.S. Navy risks diminished or loss of advantage in the maritime domain.

To ensure an accelerated and robust process for identifying parts replacements, Govini developed a repeatable and scalable methodology, which leverages a large language model (LLM) to analyze the potential form, fit and function of parts and prioritize potential alternate parts in order to provide Navy engineers a starting point for their process. The methodology examines part inventory and supply levels to heavily prioritize parts that the U.S. Navy has stock on hand for in a nearby location to further speed up part delivery to the Fleet. To accomplish this



task, the study leveraged machine learning and artificial intelligence (ML/AI) to first identify the baseline dataset of relevant parts and associated characteristics and then identify similar alternates to potentially problematic parts.

The insights from the study can aid decision-makers in the Department of the Navy (DoN), Defense Department, and broader U.S. government as they grapple with instilling processes to address the challenges extended ship and submarine lifetimes put on fleet sustainment and maintaining U.S. maritime superiority.

Key Findings

Implementing AI/ML processes is key to proactively identifying parts that pose high risk to shipyard availability schedules. Once these target parts have been identified, the LLM can be leveraged to identify potential solutions and speed up remediation actions.

- **Ohio Class Submarine parts have a lead time of up to 1,261 days.** Extremely long lead times can negatively impact shipyard availabilities due to lack of part availability once the part requirement is identified. Lead times of multiple years are untenable, and mitigations need to be identified.
- **18,007 total parts with a lead time of over 1 year.** Out of the 123,564 unique parts identified associated with the Ohio Class Submarine, ~14% of these parts have a lead time of over 1 year.
- **Of the 18,007 with lead times over 1 year, the LLM identified 10,703, which have similar parts with stock on hand.** The average similarity score of this cohort of potential alternate parts is 0.88 out of 1, which indicates a high probability of the similar parts being acceptable alternates.

Methodology

Govini's National Security Knowledge Graph (NSKG) was leveraged to identify all relevant structured and unstructured data that could be used to describe the form, fit and function of parts for a selected weapon system. The NSKG is driven by Govini's ML-powered Object Fusion data engine that continuously ingests, normalizes, and integrates new data sources with existing data catalogs. Govini analysts leveraged the vast information in the NSKG to construct the associated part landscape views across the Ohio Class Submarine through the use of ML algorithms. This comprised the baseline dataset for analysis. This study focuses on the Ohio Class Submarine as an exemplar case, but any predetermined set of parts can be utilized. The list of all the weapon system designator codes (WSDC) utilized can be seen in Appendix A.

The first phase of this study utilized an LLM-enhanced model to quantify the form, fit and function similarity between parts, leveraging all of the NSKG-derived information described in the previous paragraph. The LLM-based model was trained using input identified by subject matter experts. The relevant part identifying information that this application uses includes a wide variety of part-specific elements that describe the part in specific detail. Examples include the part's weight, material, size, description, etc. This allows for scanning for parts that share similar features that could also be quick, low-cost alternatives to long lead time parts.

In the second phase, the baseline dataset of all Ohio Class Submarine parts was evaluated to identify the top 10 long lead parts. These parts were then run through the LLM model to identify similar parts. Potential similar parts with stock on hand are prioritized for these long lead parts to provide a prioritized list of potential alternates for these long lead items.



Analysis

A training set was generated to train the model across the Ohio Class Submarine parts cohort. This model evaluated 3,106 unique types of inputs across the part cohort. The model outputs vector embeddings, which are utilized to determine which parts are similar. Vectors will be closer together if the parts are more similar. This can then be used to generate a similarity score to more easily compare various parts.

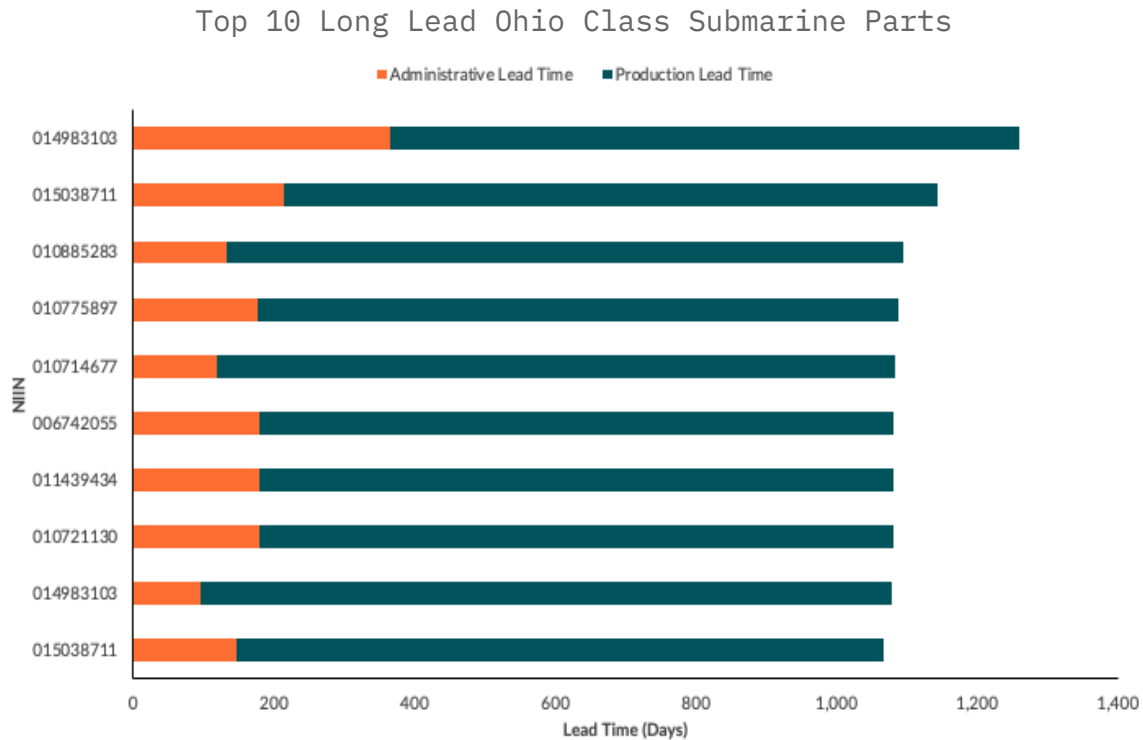


Figure 1. Top 10 Long Lead Ohio Class Submarine Parts

There are approximately 123,564 unique parts associated with the Ohio Class Submarine identified by 25 different WSDCs defined by the Defense Logistics Agency (DLA). As seen in Figure 1, some of these parts have lead times as high as 1,261 days. This means that if this part is needed to support a Submarine in a shipyard availability, it could take as long as 42 months to get the part to the Submarine and get the Submarine repaired, greatly impacting overall fleet readiness.

Running the Ohio Class Submarine parts through the LLM provides 68,569 unique parts with similar parts that have a similarity score over 90; 90 is utilized as a threshold to ensure all parts have a higher probability of being selected by engineers and subject matter experts as acceptable alternates. These lists are then further prioritized by filtering out parts with no available stock on hand. This subset represents the subset of parts that should the similar part be approved as an alternate part, have stock on hand to most quickly be utilized to leverage any schedule delays during a shipyard availability.



Top 10 Long Lead NIINs with Stock on Hand for Similar Parts

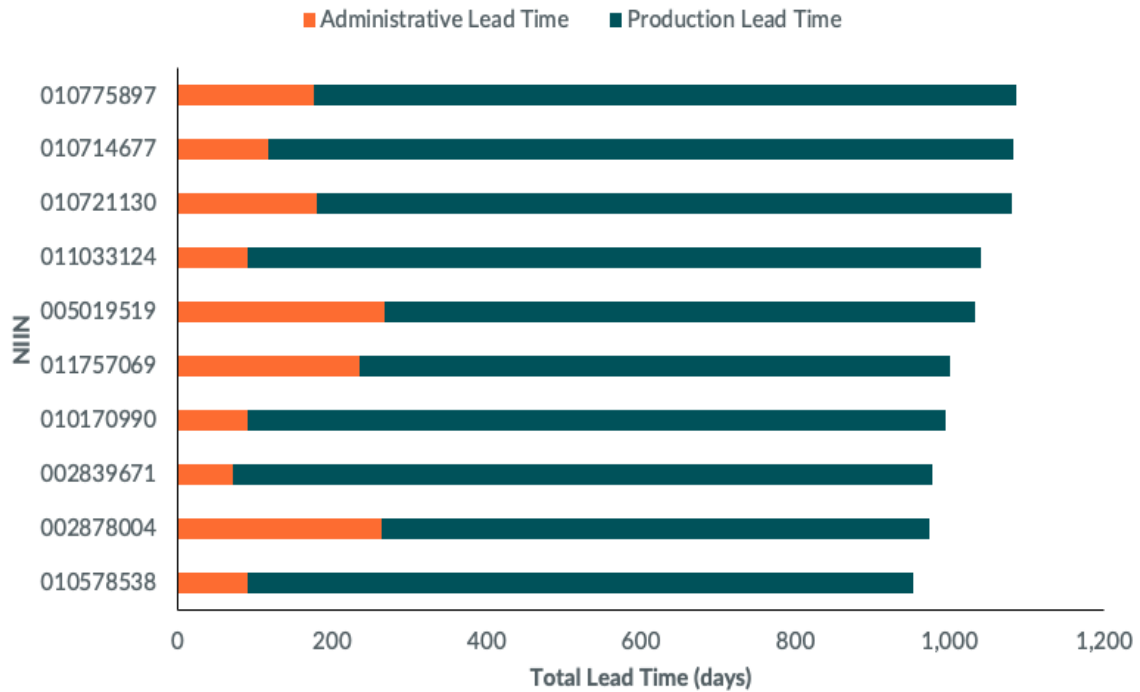


Figure 2. Top 10 Long Lead Ohio Class Submarine Parts With Highly Similar Parts With Stock on Hand

Figure 2 shows the breakdown of administrative lead time and production lead time of these top Ohio Class Submarine parts. All of these parts have multi-year lead times and could possibly cause schedule delays. The LLM outputs a number of similar parts for each of the NIINs shown in Figure 2, which can be analyzed further. This study focuses on highlighting one of these parts for in-depth analysis.

NIIN 005019519 is an Annular Ball Bearing, which is part of an auxiliary system on the Ohio Class Submarine. With a lead time of over 1,000 days, if this part is not available when it is needed, it could severely impact shipyard schedules.

Table 1. Top Similar NIINs for Annular Ball Bearing (NIIN 005019519)

Similarity Score	Similar NIIN	Similar NIIN Stock on Hand
0.987	000039270	62
0.976	006138004	13
0.983	001426059	39
0.976	010800812	15

During the detailed, manual evaluation, 18 unique technical characteristics associated with this part were compared across all parts within Govini's NSKG to identify similar parts. This detailed comparison can be seen in Appendix B. Table 1 shows the top similar NIINs for this part, which also have stock on hand. This short list can then be provided to engineers to confirm



that one of these similar parts can be used in place of the original part. The stock on hand can then be utilized to quickly solve the part availability issue.

In this study, AI/ML processes were able to seamlessly analyze the 123,564 unique parts in the Ohio Class Submarine, identifying risk factors including long part times, low stock on hand, and lack of alternates. Ultimately, the LLM identified potential alternate parts to mitigate high risks of causing operational disruption.

Implications for the Navy

With the ability to quickly identify alternative parts, the U.S. Navy can more effectively maintain shipyard availability schedules and overall fleet readiness. Leveraging AI and ML to analyze large data sets at scale will expedite a previously manual process. Tactically, utilizing these technologies will save the Navy time and man-hours, freeing up man-hour time to focus on other priorities. Strategically, faster discovery of alternate parts can mitigate overall schedule impacts. In the future, the U.S. Navy can leverage this methodology to proactively identify alternative parts for parts that are long lead critical path items. This can be leveraged during the availability planning process, instead of only after a part availability issue has been identified.

Next Steps

In order to further refine the results from this study, the team would conduct initial discussions with U.S. Navy leadership and technical subject matter experts to better understand critical parts technical characteristics. This will allow for further refinement of the LLM to ensure the model takes into account the characteristics that engineers deem more critical to speed up evaluation. Additional part data will also be ingested into the NSKG to broaden the scope of parts being evaluated as similar parts. The methodology used in this study could be applied to another baseline set of parts such as SPY-6, DDG-51, or Columbia Class Submarines as well. Expanded utilization of additional data within the NSKG would make it possible to further filter down the stock on hand based on location to determine if the stock on hand is near the shipyard where the Submarine is undergoing its availability to even further speed up access of the part to the Submarine.

Appendix A. Ohio Class Submarine Weapon System Designator Codes

List of all 25 Weapon System Designator Codes (WSDC) utilized to generate baseline dataset for this analysis. WSDCs are predefined groupings of parts and defined by DLA.

Weapon System Designator Code (WSDC)	Weapon System Name
23N	Ohio Class Submarine
86N	BQQ-6
A3N	Sonar Acoustic Missianeous
DQN	Submarine Vertical Launch System
MYN	BPS-15/16
MZN	BQQ-10
RHN	Submarine Data Processing System



SCN	SLMM Mine(s) Countermeasures
SJN	Towed Array Handling Equipment
SMN	Submarine Auxiliary System
SRN	Submarine Outfittings & Furnishings
TXN	Submarine Electrolytic Oxygen System
TYN	Submarine Interior Communication Systems
WCN	Submarine Atmosphere Control & Compressed Air/Gas System
WFN	Submarine Auxiliary & Miscellaneous Systems
WHN	Submarine Armament & Fire Control Systems
WJN	Submarine Miscellaneous Sonar & ADP Systems
WKN	Submarine Surveillance Systems
WLN	Submarine Acoustic Sonar Systems
WMN	Submarine High Pressure Air System
WPN	Submarine Atmosphere Analyzing System
WRN	Submarine Co/H2 Burner & Co2 Removal System
WSN	Submarine Ventilation
WTN	Submarine Gas Management Systems
WWN	Submarine Hydraulic Systems

Appendix B. Similar Parts Comparison for NIIN 005019519

Detailed breakdown of some of the relevant part identifying information compared to generate similarity score during this study.

NIIN	005019519	000039270	001426059	006138004	010800812
Item Name	BEARING,BALL,ANNULAR	BEARING,BALL,ANNULAR	BEARING,BALL,ANNULAR	BEARING,BALL,ANNULAR	BEARING,BALL,ANNULAR
Similarity Score		0.98749	0.982939	0.976174	0.975602
Stock On Hand		62	39.0	13.0	15.0
Bearing Seal Type	Contact	Contact	Contact	Contact	Contact
Bore Diameter	10.0 Millimeters Nominal	4.00 Millimeters Nominal	1.2500 Inches Nominal	60.0 Millimeters Nominal	30.00 Millimeters Nominal



Bore Shape	Straight	Straight	Straight	Straight	Straight
Internal Fit-Up Designation	Loose	Standard	Standard	Standard	Loose
Load Direction	Radial	Radial	Radial	Radial	Radial
Lubrication Material	Grease	Grease	Grease	Grease	Grease
Lubrication Material Document And Classification	Mil-G-81322 Mil Spec Single Material Response	Mil-G-81322 Mil Spec Single Material Response	Dod-G-24508 Mil Spec Single Material Response	Dod-G-24508 Mil Spec Single Material Response	Mil-G-81322 Mil Spec Single Material Response
Material	Steel Comp E52100 Outer Ring,Steel Comp E52100 Inner Ring,Steel Comp E52100 Ball,Steel Comp E52100 Retainer	Steel Comp E52100 Outer Ring,Rubber Synthetic Seal,Steel Comp E52100 Ball,Steel Comp E52100 Retainer,Steel Comp E52100 Inner Ring	Steel Comp E52100 Ball,Steel Comp E52100 Outer Ring,Steel Comp E52100 Inner Ring,Steel Comp E52100 Retainer,Rubb er Synthetic Seal	Steel Retainer,Rubb er Synthetic Seal,Steel Comp E52100 Ball,Steel Comp E52100 Outer Ring,Steel Comp E52100 Inner Ring	Steel Comp E52100 Inner Ring,Steel Comp E52100 Outer Ring,Rubber Synthetic Seal,Steel Comp E52100 Ball,Steel Retainer
Material Document And Classification	66 Fed Std Single Material Response Ball,66 Fed Std Single Material Response Outer Ring,66 Fed Std Single Material Response Retainer,66 Fed Std Single Material Response Inner Ring	66 Fed Std Single Material Response Outer Ring,66 Fed Std Single Material Response Inner Ring,66 Fed Std Single Material Response Ball,66 Fed Std Single Material Response Retainer	66 Fed Std Single Material Response Ball,66 Fed Std Single Material Response Inner Ring,66 Fed Std Single Material Response Outer Ring,66 Fed Std Single Material Response Retainer	66 Fed Std Single Material Response Outer Ring,66 Fed Std Single Material Response Inner Ring,66 Fed Std Single Material Response Ball	66 Fed Std Single Material Response Inner Ring,66 Fed Std Single Material Response Ball,66 Fed Std Single Material Response Outer Ring
Overall Outside Diameter	30.0 Millimeters Nominal	16.00 Millimeters Nominal	2.2500 Inches Nominal	130.0 Millimeters Nominal	62.00 Millimeters Nominal
Overall Width	9.0 Millimeters Nominal	5.0 Millimeters Nominal	0.5000 Inches Nominal	31.0 Millimeters Nominal	16.00 Millimeters Nominal



Retainer Fabrication Method	Pressed	Pressed	Pressed	Stamped	Pressed
Seal Quantity	2	2	2	1	2
Special Features	Ty 111 Cl 8 Gr 00	Nan	Nan	Nan	Nan
Standard Tolerance Designation	Abec No.1	Abec No.1	Abec No.1	Abec No.1	Abec No.1
Style Designator	Non-Loading Groove, Non-Separable	Non-Loading Groove, Non-Separable	Non-Loading Groove, Non-Separable	Non-Loading Groove, Non-Separable	Non-Loading Groove, Non-Separable
Surface Finish	Ground	Ground	Ground	Ground	Ground



Open Source Software (OSS) Transparency for DoD Acquisition

Dr. Carol Woody—is principal researcher for the CERT Division of the Software Engineering Institute (SEI). She focuses on cybersecurity engineering for building capabilities and competencies to measure, manage, and sustain cybersecurity and software assurance for highly complex software-reliant systems and systems of systems. She has been a member of the CERT technical staff for over 20 years. Dr. Woody coauthored the book *Cyber Security Engineering: A Practical Approach for Systems and Software Assurance*, published as part of the SEI Series in Software Engineering. The CERT Cybersecurity Engineering and Software Assurance Professional Certificate, a self-paced online training program, is based on research she led.

Dr. Nancy Mead—is a Fellow at the Software Engineering Institute and Adjunct Professor of Software Engineering at Carnegie Mellon University. Her research areas are software and security requirements engineering and software engineering and software assurance curricula. The Nancy Mead Award for Excellence in Software Engineering Education & Training is named for her. She has more than 200 publications and invited presentations. Her awards and honors include Life Fellow of the IEEE, Distinguished Member of the ACM; IEEE TCSE Distinguished Educator; IEEE TCSE Executive Board Member; Parnas Fellow at Lero, the Irish Software Research Center; and Speaker in the IEEE Distinguished Visitor Program. She has a BA, MS, and PhD in mathematics from NYU.

Scott A. Hissam—is a senior member of the technical staff for Carnegie Mellon University's Software Engineering Institute where he conducts research on component-based software engineering, open source software, and software assurance through his interactions with government, industry, and academic ventures. He is a founding member of the International Federation for Information Processing (IFIP) Working Group 2.13 on Open Source Software and co-organizer of a number of its annual conferences. His publications include two books (*Building Systems from Commercial Components and Perspectives on Free and Open Source Software*), papers published in international journals, and numerous technical reports. Most recently, Hissam has been focusing his time on engaging various open source software communities, open source program offices, and government program offices in best practices and their application in software-and cyber-supply chain risk management in a DoD environment.

Abstract

Caveat emptor, or "let the buyer beware," is commonly attributed to open-source software (OSS)—the onus is on the OSS consumer to ensure that it is fit for use in the consumer's context. OSS has been compared to an open market bazaar where consumers are free to browse all the source code and take a copy. But there are a few fundamental problems with such an analogy:

- The consumer must have the wherewithal and skills to comprehend the source code in a manner that allows them to use it effectively, which might exceed the skills of the myriad authors who produced that software in the first place.
- The consumer also lacks the insight into the practices exercised by the authors in the production of that source. Such practices include code quality checks, peer reviews, software testing, and secure software development practices.

The burden on the consumer is considerable. Consumers—both individuals and organizations—have access to proprietary and open-source tools to help them analyze source code as a means of understanding what is good and what is problematic if they are aware of and use those tools.

What a consumer lacks is an understanding of how OSS differs from other third-party software products. OSS contributors and maintainers may be arbiters for the OSS product, but they are not selling it. Caveat emptor, therefore, applies not only to the OSS product the consumer wants to use, but it also applies to the consumer knowing more about the OSS project, its contributors and maintainers, as well as the processes they follow. To gain that insight, transparency is needed for consumers to have access to the data and information about the OSS project that is meaningful



from a software integrity and assurance perspective and suitable for supply chain risk analysis. The Open Source Security Foundation's (OSSF's) SLSA2 and CHAOSS3 are initial steps in this space. The Software Engineering Institute (SEI), MITRE Corporation, and the Carnegie Mellon University Open Source Programs Office (CMU OSPO) are currently working on effective measures for supply chain risk management as it pertains to OSS. This collaborative effort is proposing forming a special interest group within the OSSF. This paper shares the results to date in this area of research.

Introduction

Scope of the Open Source Software (OSS) Problem Space

By some counts, there were over 53 million projects in 2022 alone at one OSS repository: GitHub (Woodward, 2022). Software from repositories and sources like these are included in practically every aspect of human endeavors touching products and services in industry, government, and academia. When faults or vulnerabilities are exploited in this software, just as in proprietary software, the impacts can be far reaching. One detailed study on the depth to which OSS is entrenched in the critical infrastructures that service daily activities is detailed in the Linux Foundation's Census II of Free and Open Source Software—Application Libraries (Nagle et al., 2022).

The reuse of software has enabled faster fielding of systems since common components can be sourced externally. However, all software comes with vulnerabilities, and attackers have expanded their capabilities to exploit them in products that have broad use. A recent report by SecurityScorecard found that 98% of their sampled 230,000 organizations use third-party software components from organizations that have been breached within the prior 2 years (Townsend, 2023).

Organizations shifting to cloud services to eliminate on-premises risk have frequently been surprised by supply chain risks inherited from service providers, which resulted from misconfigurations, unauthorized access, insecure application programming interfaces (APIs), etc. (Check Point Software Technologies Ltd., n.d.). To identify and manage this growing risk landscape, organizations must increase collaboration across the range of participants involved in the selection, installation, and monitoring of third-party software to identify and manage potential risks.

Earlier research addressed by the Software Engineering Institute (SEI) assembled practices critical to meeting this need in the Acquisition Security Framework (ASF) (Alberts et al., 2022). However, each organization has a unique technology environment, and there are no widely accepted measures for evaluating their accepted risk.

Diving into supply chain risk a little further, supply chain attacks initially appeared in third-party software that was either developed through custom contracts or purchased as commercial software. It is a bit easier to tailor a custom contract to protect against these risks, and the guidance available for supply chain risk management (SCRM) can go a long way toward developing a plan to add to a contract. Purchased software leaves the risk management to the customer to handle themselves. Establishing and implementing a SCRM plan is an expense that many organizations will not sign up for.

A recent European Union (EU) study included 1,081 organizations in 27 member states (Papaphilippou et al., 2023). In their report, the researchers noted that only 47% of the surveyed organizations in the EU had an allocated budget for information and communication technology and operational technology (ICT/OT) for supply chain cybersecurity. Researchers further observed that 52% of the organizations surveyed had a rigid patching policy that included at least 80% of their assets. On the other hand, only 13.5% of the surveyed organizations had



visibility into patching for fewer than half their assets, meaning that patching may be done by a third-party organization or not at all.

The situation is no better for users of OSS, which is the latest target of attackers in the software supply chain. For years, many considered OSS to be more secure because its code was visible and developed by “trusted” individuals. An early study provided insight into a widespread attack against both proprietary and OSS operating systems—the TearDrop and NewTear attacks—that were actually informed by OSS before vulnerability disclosure processes were widely used (Hissam et al., 2002). Contrary to popular belief, OSS has never really been secure; whether a component of OSS is secure is in the eye of the beholder—be they altruistic or nefarious.

A Sonatype survey of 621 practitioners indicated that only 28% of their organizations became aware of new open source vulnerabilities within a day of disclosure, 39% discovered them within 1 to 7 days, and 29% took more than a week to become aware of them (Sonatype Incorporated, 2023). The same survey found that 39% of the respondents’ organizations took more than a week to mitigate vulnerabilities.

Cost of an Attack

Intellias summarizes the estimated cost of supply chain attacks quite well (Fedorko, 2023). In 2023, the MOVEit vulnerability cost businesses over \$9.9 billion, with more than 1,000 businesses and over 60 million individuals affected. Furthermore, the estimated cost from just seven high-profile supply chain attacks, starting with SolarWinds, was around \$60 billion. This figure does not include the impact of government-imposed fines and legal actions related to privacy laws on both the affected organizations that rely on them. In this regard, there is little distinction between supply chain attacks resulting from compromised proprietary software and those resulting from compromised OSS.

Any successful supply chain attack can result in substantial financial loss, loss of reputation, lawsuits, and investigations. In recent articles (Birsan, 2021; O’Neill, n.d.), it was revealed that the information of 1.6 million patients was compromised by a successful MOVEit hack. Even though the vulnerability was documented in May 2023, that particular hack was not discovered until October 2023. The patch to the vulnerability had not yet been applied.

In a report about the recovery from the successful attack on Metro-Goldwyn-Mayer (MGM) casinos, interviews with management indicated a loss of \$100 million at its Las Vegas properties (Jones, 2023). The loss is covered by cyber insurance; however, the cost of that insurance has doubled or, in some cases, quadrupled in recent years. Of course, this loss estimate includes only the immediate loss of revenue and does not consider the result of class action lawsuits, some of which have already been filed. Typically, the organization plans to invest heavily in information technology (IT) after the successful attack.

Current efforts for improvement are largely focused on supply chain risk in general.

Supply Chain Problem Space

Recently, a lot of attention has been paid to the problem of strengthening the software supply chain by focusing on software bill of material (SBOM) support. Although that attention is needed since there is little information available about the software specific to a product, it is not the be-all and end-all to securing supply chains. Yes, a well-defined and completely verifiable SBOM is a key step to understanding dependencies and tracking problems within those dependencies; however, it is not the only step. Just as important is knowing and understanding the practices employed by those who created the OSS itself—the OSS project members. For example, do the project members follow community standards of care when vetting contributions to the software code (e.g., peer reviews, code quality checks)? Having such



transparency into the OSS project and the project's makeup is not unlike performing a financial health check or financial risk assessment of a commercial company for due diligence.

OSS consumers need to have the means and wherewithal to assess the products and projects they depend on. There are emerging OSS and proprietary tools to help provide insight into project health (e.g., the OSSF's Scorecard and MITRE's Hipcheck). There are also many standards, formats, and tools to help users maintain or create retroactive SBOMs that have been developed and promoted in recent years. However, to be effective, these tools must be conducive to the practices and processes employed by OSS consumers. Stakeholders in the entire OSS ecosystem and software industry in general still need novel tools, policies, and methodologies that detect and mitigate vulnerabilities found within OSS products including deviations from OSS community-recommended practices in the products built to address new uses. By defining these tools, policies, and methodologies, the software community can establish greater confidence in its existing code base as well as assess and respond to new technologies, products, and vendors that provide great benefit.

There are also emerging OSS community initiatives being established. These are designed to provide additional insight into the practices employed by OSS projects. Two of note are the OSSF's Supply Chain Levels for Software Artifacts (OSSF's SLSA) and the Eclipse Foundation's Adoptium for Reproducible Builds. Each of these is an example of OSS community-led activities to improve confidence in OSS that are used by OSS consumers in their supply chains. In many ways, at least for security and supply chains, these examples are steps forward by the OSS community to "mature" OSS projects to get closer to secure software development and distribution.

Software Risks and Challenges

Software development often lacks effective resistance to attack and controls for preventing tampering by malicious actors. Typically, software development project resources are insufficiently integrated and can be affected by cost, schedule, and compliance constraints. These characteristics leave operational environments with weaknesses and vulnerabilities that must be addressed later in the system's life cycle. A lack of cost-schedule-risk balance can lead to shortcuts resulting in software that does not conform to secure coding and secure software development practices. There is a pressing need to implement more transparent and predictable mechanisms for ensuring that software functions securely and as intended.

No software is free of risks. Defects exist even in the highest quality software. For example, best-in-class code can have up to 600 defects per million lines of code (MLOC), while average quality code has around 6,000 defects per MLOC. Some of these defects are weaknesses that can lead to vulnerabilities. Research indicates that up to 5% of software defects are security vulnerabilities (Woody et al., 2014). As a result, best-in-class code can have up to 30 vulnerabilities per MLOC. For average quality code, the number of security vulnerabilities can be as high as 300 vulnerabilities per MLOC.

Reducing the number of security vulnerabilities in code is an important part of software development. Using secure coding practices, peer reviews, and code analysis tools are important ways to identify and correct known weaknesses and vulnerabilities in code. As shown in Figure 1, there are many activities that need to be addressed across the acquisition and development life cycle to respond to defects and vulnerabilities as close to their point of origin as possible to keep costs for rework minimized.



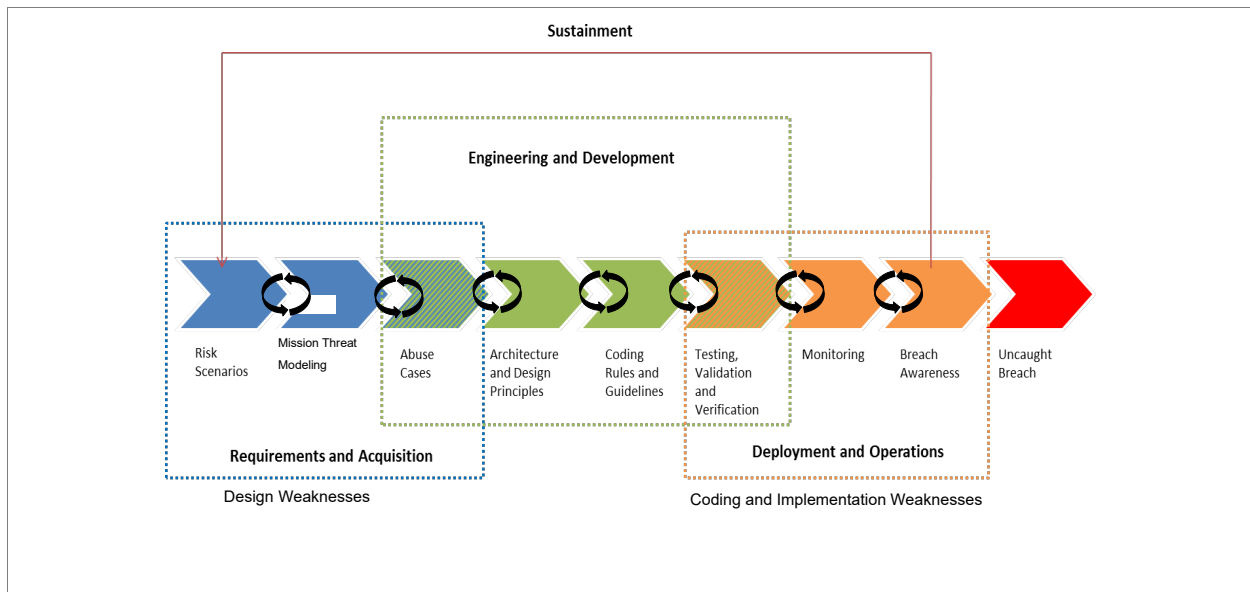


Figure 30: Activities to Address Defect and Vulnerability Identification and Mitigation

Third-party risk is a major challenge for organizations seeking to limit their exposure to cybersecurity risks from general outsourcing and supplier-provided software that is critical to system functions. There is often little transparency into the components, sources, and suppliers involved in a system’s development and ongoing operation. SBOM information can help reduce the uncertainties that result from that lack of transparency. An essential aspect of addressing supplier risk is being able to access information about supplier inputs and their relative importance and then managing mitigations to reduce risk.

Research into the various types of vulnerabilities indicates that focusing on the most prevalent with greatest risk should be considered. Security vulnerabilities that are caused by defects with memory use are significant, highly common, and greatly impactful to the risk and damage to software, data, and systems. One recent study estimates that 65% of security vulnerabilities are caused by memory unsafety (Gaynor, 2020). Studies like this make addressing memory safety issues an appealing approach. Memory safety, however, is not without limitations. Known approaches cannot eliminate memory issues; they can only reduce them. And there are still other vulnerability categories that these mitigations do not address. Therefore, memory safety is not a panacea, but it can significantly reduce memory-related vulnerabilities, which would correspond to a significant reduction of risk. Like many other software security and quality mitigations, memory-related risks need to be measured and mapped to cybersecurity concerns to establish priority.

State of the Practice of Measuring Software Assurance

Currently, it is possible to collect vast amounts of data related to cybersecurity in general. We can measure financial loss due to a successful attack and loss of confidence in a particular company as measured by the loss of customers or impacts on stock values. We can measure the elapsed time from the beginning of a successful attack until its discovery and the elapsed time from discovery until mitigation and recovery.

We can also measure specific product characteristics related to cybersecurity. Commercial companies, nonprofits, and government entities offer data-collection tools. What is not always clear is the cause-and-effect relationship between the data, vulnerabilities, and successful attacks. Also, much of the data collected reflects the results of an attack, whether



attempted or successful. Data on earlier security life-cycle activities often reflects the development processes used. Although needed, it is not diligently collected, nor is it analyzed as thoroughly as in later points of the life cycle.

As software engineers, we believe that improved software development practices and processes will result in a more robust and secure product. However, which specific practices and processes actually result in a more secure product? There can be quite a bit of elapsed time between the implementation of improved processes and practices and the subsequent deployment of the product. If the product is not successfully attacked, does it mean that it is more secure? Or does it just mean that it is a less interesting target from an attacker's perspective? Zahan (2023) concludes the following:

Security metrics are a hard problem, especially in predicting vulnerabilities or assessing the effectiveness of counter measures [Cheng 2014, Scala 2019]. We should consider that the software security field has an inherently greater amount of unexpected variation.

Consider the recently updated Cyber Security & Information Systems Information Analysis Center's (CSIAC's) The DoD Cybersecurity Policy Chart (Cybersecurity & Information Systems Information Analysis Center, n.d.). Although the authors have made a valiant effort to categorize and classify the subject policies, how can project managers be expected to sift through hundreds if not thousands of pages of documentation to find out what policy applies to them?

Certainly, government contractors have a profit motive that justifies meeting the cybersecurity policy requirements that apply to them, but do they know how to measure the cybersecurity risk of their products? And how would they know whether it has improved sufficiently? For OSS, when developers are not compensated, what would motivate them to meet these cybersecurity policy requirements? Why would they even care whether a particular organization—be it academic, industry, or government—is motivated to use their product?

Currently Available Metrics

The SEI led a research effort to identify the metrics currently available within the life cycle that could be used to provide indicators that potential cybersecurity risk exists (Woody et al., 2019). From an acquisition life-cycle perspective, there are two critical information needs:

- Is the acquisition headed in the right direction as it is engineered and built (predictive)?
- Is the implementation sustaining an acceptable level of operational assurance (reactive)?

Each step in the acquisition and development life cycle can produce useful information. However, to gather that information, the following must happen:

- The life cycle's processes and practices must be structured to gather data in a form that can be analyzed.
- Appropriate measures must be created and enforced as each process or practice is performed.

Trends in current approaches show an increased reliance on tools to perform life-cycle steps and a limited consideration of how available data from tool use will be integrated across tasks for useful analysis.

For code, the work of Capers Jones (2014) in collecting and tracking defects to evaluate quality provides an opportunity for setting expectations about code quality based on the level of expected defects. SEI research leveraged this data to project expected vulnerabilities based on defect rates (Woody et al., 2014). As development shifts further into agile increments, many of



which include third-party and open-source components, different tools and definitions are applied to collecting defects so that the meaning of this metric in predicting risk becomes obscured.

Highly vulnerable components that are implemented using effective and well-managed zero trust principles can deliver acceptable operational risk. In a similar vein, well-constructed, high-quality components with weak interfaces can be highly prone to successful attacks. Operational context is critical to the level of risk exposure. A simple evaluation of each potential vulnerability using something like a Common Vulnerability Scoring System (CVSS) score can be extremely misleading since the score without the context has limited value in determining actual risk.

The lack of visibility into the processes and methods used to develop third-party software—particularly OSS—means that measures related to the processes used and the errors found prior to deployment, if they exist, do not add to the useful information about the product. This lack of visibility into product resilience as it relates to the process used to develop it means that we do not have a full picture of the risks, nor do we know whether the processes used to develop the product have been effective. It is difficult, if not impossible, to measure what is not visible.

Consider a recent blog post by Eric Goldstein from CISA that announces a Secure by Design Alert Series and states the following:

Insecure technology products are not an issue of academic concern: they are directly harming critical infrastructure, small businesses, local communities, and American families. (Goldstein, 2023)

This, of course, is not new information. In addition to its implications for large companies producing software products, the same concerns apply to software in the supply chain and OSS. One might imagine that large companies will respond in some fashion, but how will the open-source community respond? Are open-source developers concerned with using specific languages, such as RUST, that are supportive of security goals? Are they concerned with using default settings that are inherently more secure? If so, how would consideration of these choices be communicated to customers?

Many ideas could apply to OSS, but there may be no agreement between the user and the creator of OSS. There are instances of organizations paying developers of OSS to provide software that is specific to their project. That software may or may not be covered by an agreement that would account for the cybersecurity risks. If there is no agreement between the user and the supplier, it leaves the user with the sole responsibility of defining performance measurements and determining whether they are being met. For the cybersecurity of OSS, by and large, the indicators are lagging (e.g., vulnerabilities, successful attacks) rather than leading.

When the processes used for product development are not visible, it is impossible to determine whether they have been effective in terms of thwarting future attacks. How can a process be improved when it is invisible and the measurement that should be taking place is also invisible? More to the point, how can we get leading indicators for OSS cybersecurity when the process is invisible?

Assembling What We Can Know about OSS

The OSSF Scorecard is a tool that incorporates a set of metrics that can be applied to an OSS project. The idea is that project attributes that OSSF believes contribute to a more secure open-source application are reported using a weighted approach that leads to a score.



From a metrics perspective, there are limitations to this approach:

1. The open-source community is driving and evolving which items to measure and, therefore, build into the tool. Also, it is not clear how those factors were determined, whether the set of factors is complete, or what is intended for the long-term roadmap (i.e., insufficient transparency).
2. The relative importance of each factor is also built into the tool, which makes it difficult (but not impossible) to tailor the results to specific, custom, end-user needs (Open Source Software Foundation, n.d.).
3. Many of the items measured in the tool appear to be self-reported by developers versus validated by a third party, but this is a common “attribute” of open-source projects.

Other tools, such as MITRE’s Hipcheck, have the same limitations (MITRE Corporation, n.d.). For an OSSF project, it is possible to get a score using Scorecard and scores for the individual dependency projects, but questions arise from this approach. How do those individual scores roll up into the overall score? Do you pick the lowest score across all the dependencies? Or do you apply some sort of weighted average of scores? This area needs exploration and elaboration.

Furthermore, a recent research paper (Zahan et al., 2023) described cases where open-source projects that were scored highly by Scorecard might, in fact, produce packages that have more reported vulnerabilities. From a research perspective, it is unknown whether this occurs because (1) the application received more reviews and therefore more vulnerabilities were identified, or (2) attacks on a popular application exposed it to more vulnerabilities. Needless to say, Zahan’s results are useful only for those open-source projects evaluated by Scorecard, which is applied exclusively to GitHub, and those are only a fraction of the total number of open-source applications available. All of these issues indicate that further study is needed.

Metrics by themselves are of limited value. It is the assembled picture of metrics weighed against expectations that can be valuable to decision-makers. A simple three-step process can be applied to generate an initial perspective and flag areas of concern for further consideration:

- Measure and baseline what you have, especially OSS.
- Assess how you are vulnerable and identify an improvement path.
- Integrate measurement and monitoring throughout the life cycle.

An example, as shown in Figure 2, can provide a useful starting point even if information is limited. Known information related to the project, product, available protection, and applied policies provide a useful range of sources. Mapping that information to a series of potential factors can raise red flags to show where security risk may already exist. Evaluating each of the risks against a range of acceptability ratings can identify areas of concern that will require mitigation.



	Project	Product	Protection	Policy
Identified Criteria	Long-Term Support	Forked Project		No Security Policy
	Dependencies	74 Abandoned Dependencies	No Update Tools	
	Security		4 Unfixed Critical Vulnerabilities	Workflow with Excessive Permissions
	Integrity		No Fuzz Testing	30 Unreviewed Change Sets
	Malicious Actors	Commit ID Known Malicious		
	Suitability			12 Restrictive Licenses

Red Flag

Realm of Observable Facts of OSS Projects and Products

Figure 31: Mapping Observable Information to Areas of Concern

The steps described in Figure 2 can be listed as follows:

1. Review data available.
2. Identify useful criteria.
3. Extract key data.
4. Map to acceptable criteria.
5. Evaluate red flags.
6. Identify appropriate mitigations.
7. Confirm supportability.

When obtaining information, there will be challenges about the processes used to create and maintain OSS since this is information typically not captured in available repositories. When issues arise, there are limited options at best for mandating repairs.

Future Considerations

At a March 2024 OSS Summit hosted by CISA, participants from several open-source communities gathered to discuss a path forward to improve the security and potentially reduce the risk of using OSS.

Suppliers should be proactive. Yet leadership across the supply chain continues to underinvest in software assurance, especially early in the life cycle. This lack of investment leads to design decisions that lock in weaknesses because there are no means to characterize and measure the risk they are accepting. Suppliers rush to deliver new features to motivate buyers at the expense of analyzing the code to remove potential vulnerabilities, and buyers have limited means to evaluate the risk in the products they acquire.

Even if a supplier addresses an identified vulnerability quickly and issues a patch, it is incumbent on the users of that software to apply the fix. Software supply chains are many levels deep, and too frequently the patches apply to products buried deep within the chain. Each layer must apply the patch and send an update up the chain. This can be a slow and faulty process since knowing where each specific product has been used is limited for those higher in the chain. Recent mandates to create SBOMs support improving that visibility (United States White



House Executive Office of the President, 2022), but the fix still needs to be addressed by each of the many layers that contain the vulnerability.

A better future is needed, where the tools we have available to assess software packages for supply chain risk perform comparable analyses in comparable ways, where the results of analyses are clear and can be tied to policy needs, and where the recommendations made by analyzers can be audited to provide clarity to software maintainers and security teams alike. Achieving this future means going beyond simple observation analysis and establishing a process to define standard metrics for supply chain security risk in software.

Reference Documents Used to Prepare This Paper

The following documents were used to prepare this paper:

The Measurement Challenges in Software Assurance and Supply Chain Risk Management
(Mead et al., 2023)

The SEI's response (Tucker et al., 2023) to the *Office of the National Cyber Director's Request for Information on Open-Source Software Security: Areas of Long-Term Focus and Prioritization* (United States Office of the National Cyber Director, 2023).

Legal Markings

Copyright 2024 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

References herein to any specific entity, product, process, or service by trade name, trade mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by Carnegie Mellon University or its Software Engineering Institute nor of Carnegie Mellon University - Software Engineering Institute by any such named or represented entity.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Requests for permission for non-licensed uses should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

CERT® and Carnegie Mellon® are registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM24-0391

References

Alberts, C. J., Bandor, M. S., Wallen, C. M., & Woody, C. (2022). *Acquisition Security Framework (ASF): An acquisition and supplier perspective on managing software-intensive systems' cybersecurity risk* [White paper]. Software Engineering Institute, Carnegie Mellon University. <https://insights.sei.cmu.edu/library/acquisition-security-framework-asf-an-acquisition-and-supplier-perspective-on-managing-software-intensive-systems-cybersecurity-risk/>

Birsan, A. (2021, February 9). Dependency confusion: How I hacked into Apple, Microsoft and dozens of other companies. *Medium*. <https://medium.com/@alex.birsan/dependency-confusion-4a5d60fec610>



- Check Point Software Technologies Ltd. (n.d.). *Main cloud security issues and threats in 2023*. Retrieved December 19, 2023, from <https://www.checkpoint.com/cyber-hub/cloud-security/what-is-cloud-security/top-cloud-security-issues-threats-and-concerns/>
- Cheng, Y., Deng, J., Li, J., DeLoach, S., Singhal, A., & Ou, X. (2014). Cyber defense and situational awareness. In A. Kott (Ed.), *Metrics of Security* (pp. 263–295). Springer. https://doi.org/10.1007/978-3-319-11391-3_13
- Cybersecurity and Infrastructure Security Agency. (2023). *Minimum requirements for vulnerability exploitability eXchange (VEX)* [White paper]. <https://www.cisa.gov/sites/default/files/2023-04/minimum-requirements-for-vex-508c.pdf>
- Cybersecurity & Information Systems Information Analysis Center. (n.d.). *The DoD cybersecurity policy chart*. Retrieved December 18, 2023, from <https://csiac.org/resources/the-dod-cybersecurity-policy-chart/>
- Executive Office of the President. (2022). *Enhancing the security of the software supply chain through secure software development practices* (Report No. OMB M-22-18). Office of Management and Budget. <https://www.whitehouse.gov/wp-content/uploads/2022/09/M-22-18.pdf>
- Fedorako, K. (2023, October 4). What we've learned from recent 2023 supply chain attacks: Are you prepared? *Intellias*. <https://intellias.com/supply-chain-attacks/>
- Gaynor, A. (2020, May 27). *What science can tell us about C and C++'s security*. <https://alexgaynor.net/2020/may/27/science-on-memory-unsafety-and-security/>
- Goldstein, E. (2023, November 29). CISA announces secure by design alert series: How vendor decisions can reduce harm at a global scale. *Cybersecurity and Infrastructure Security Agency*. <https://www.cisa.gov/news-events/news/cisa-announces-secure-design-alert-series-how-vendor-decisions-can-reduce-harm-global-scale>
- Hissam, S. A., Plakosh, D., & Weinstock, C. (2002). Trust and vulnerability in open source software. *IEEE Proceedings—Software*, 149(1), 47–51. <https://doi.org/10.1049/ip-sen:20020208>
- Jones, C. (2014). *Evaluating software metrics and software measurement practices, version 4* [White paper]. Namcook Analytics LLC. <https://docplayer.net/20369543-Evaluating-software-metrics-and-software-measurment-practices-version-4-0-march-14-2014-capers-jones-vp-and-cto-namcook-analytics-llc.html>
- Jones, D. (2023, November 9). *MGM Resorts anticipates no further disruptions from September cyberattack*. Cybersecurity Dive. <https://www.cybersecuritydive.com/news/mgm-resorts-no-disruptions-cyberattack/699320/?mod=djemCybersecruityPro&tpl=cy>
- Mead, N., Woody, C., & Hissam, S. (2023). *The measurement challenges in software assurance and supply chain risk management* [White paper]. Software Engineering Institute, Carnegie Mellon University. <https://insights.sei.cmu.edu/library/measurement-challenges-in-sw-assurance-and-scrm-white-paper/>
- MITRE Corporation. (n.d.). *Hipcheck*. GitHub. Retrieved December 19, 2023, from <https://github.com/mitre/hipcheck>
- Nagle, F., Data, J., Hoffman, J., Randazzon, S., & Zhou, Y. (2022). *Census II of free and open source software—Application libraries*. The Linux Foundation and The Laboratory for Innovation Science at Harvard. <https://www.linuxfoundation.org/research/census-ii-of-free-and-open-source-software-application-libraries>
- O'Neill, I. (n.d.) *Data breach notifications*. Office of the Maine Attorney General. Retrieved December 18, 2023, from <https://apps.web.maine.gov/online/aeviewer/ME/40/78e177b2-14ba-4d97-a8cf-eeaa3f1ba0b8.shtml?mod=djemCybersecruityPro&tpl=cy>
- Open Source Security Foundation. (n.d.). *Scorecard feature: Allow specifying weights for a custom aggregated score #2296*. GitHub. Retrieved December 19, 2023, from <https://github.com/ossf/scorecard/issues/2296>



- Papaphilippou, M., Moulinos, K., & Theocharidou, M. (2023). *Good practices for supply chain cybersecurity*. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/good-practices-for-supply-chain-cybersecurity>
- Ray, B., Posnett, D., Vilkov, V., & Devanbu, P. (2014). A large scale study of programming languages and code quality in Github. *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, 155–165. <https://doi.org/10.1145/2635868.2635922>
- Scala, N. M., Reilly, A. C., Goethals, P. L., & Cukier, M. (2006). Risk and the five hard problems of cybersecurity. *Risk Analysis*, 39 (10), 2119-2126. <https://pubmed.ncbi.nlm.nih.gov/30925207>
- Sonatype Incorporated. (2023). *9th annual state of the software supply chain report*. <https://www.sonatype.com/state-of-the-software-supply-chain/introduction>
- Townsend, K. (2023, February 1). 98% of firms have a supply chain relationship that has been breached: Analysis. *Security Week*. <https://www.securityweek.com/98-of-firms-have-a-supply-chain-relationship-that-has-been-breached-analysis/>
- Tucker, B., Hissam, S., Yasar, H., Yankel, J., & Schiela, R. (2023). *Response to Office of National Cyber Director request for comment on open-source software security and memory safe programming languages*. Defense Technical Information Center. <https://apps.dtic.mil/sti/citations/trecms/AD1214435>
- United States Office of the National Cyber Director. (2023). *Request for information on open-source software security: Areas of long-term focus and prioritization*. (Document No. 2023-17239). United States Federal Register. <https://www.federalregister.gov/documents/2023/08/10/2023-17239/request-for-information-on-open-source-software-security-areas-of-long-term-focus-and-prioritization>
- Woodward, M. (2022, November 17). Octoverse 2022: 10 years of tracking open source. *GitHub*. <https://github.blog/2022-11-17-octoverse-2022-10-years-of-tracking-open-source/>
- Woody, C., Ellison, R., & Nichols, W. (2014). *Predicting software assurance using quality and reliability measures* (Report No. CMU/SEI-2014-TN-026). Software Engineering Institute, Carnegie Mellon University. <https://insights.sei.cmu.edu/library/predicting-software-assurance-using-quality-and-reliability-measures-2/>
- Woody, C., Ellison, R., & Ryan, C. (2019). *Exploring the use of metrics for software assurance* (Report No. CMU/SEI-2018-TN-004). Software Engineering Institute, Carnegie Mellon University. <https://insights.sei.cmu.edu/library/exploring-the-use-of-metrics-for-software-assurance/>
- Zahan, N., Shohan, S., Harris, D., & Williams, L. (2023). Do software security practices yield fewer vulnerabilities? *IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Australia*, 292–303. <https://doi.org/10.1109/ICSE-SEIP58684.2023.00032>



Investing in the Future: Trends in the Defense Department's Science and Technology Funding

Jacob Winn—is an Associate Research Fellow, at National Defense Industrial Association's Emerging Technologies Institute. [jacob-winn@comcast.net; ETI@NDIA.org]

Cari Shearer—is a research Intern at National Defense Industrial Association's Emerging Technologies Institute [ETI@NDIA.org]

Abstract

Tracking and analyzing defense science and technology (S&T) funding has, historically, been difficult—not just for analysts, but for employees of the Department of Defense (DoD) and Congress itself. The research team collected and analyzed these data to understand how S&T funding has changed over time and to build correlations with various strategic plans and operational needs during the 21st century. The Emerging Technologies Institute (ETI) placed S&T funding levels in the context of five selected eras of strategic planning: Operation Iraqi Freedom, the Budget Control Act of 2011, the Third Offset Strategy, the 2018 National Defense Strategy (NDS), and the 2022 NDS. The research team gathered data from R-1 Budget Justifications from the Office of the Under Secretary of Defense (Comptroller) and congressional appropriations tables.

Through this lens, we identified several important trends. First, S&T budget requests from Fiscal Year (FY) 2003 to FY2015 remained remarkably consistent in inflation-adjusted terms despite significant changes to defense spending driven by the Global War on Terror and the Budget Control Act. Since then, requests have grown remarkably to \$17.5 billion in FY2024 (~+21% since FY2016). Second, ETI found that when DoD topline grows, S&T grows more slowly. However, when DoD topline shrinks, S&T shrinks more slowly. Third, the data show that Congress consistently appropriates more funding for S&T than what is requested in the President's Budget, regardless of party control of Congress or the presidency. Finally, the DoD has not requested that S&T be funded at 3% of the topline in any year in the 21st century, an often-cited goal, and congressional appropriations have not funded S&T at this level since FY2005. These findings, and other trends in the report, provide readers with the context behind past funding decisions that may be applicable to future strategic guidance.

Introduction

Looking at federal spending by examining public government budget data and explanatory materials produced by the executive branch and Congress can often provide more insight into government priorities and activities than strategy documents, lists of strategic goals, and statements made during congressional hearings. If the United States is to continue delivering cutting-edge emerging technologies across a variety of science and technology (S&T) areas in support of national defense missions, stakeholders throughout the policy, scientific, and business communities require a clear view of national priorities that is often best communicated by tracking actual spending by the government. To help achieve this goal, the National Defense Industrial Association (NDIA) Emerging Technologies Institute (ETI) analyzed national defense research and development (R&D) funding data, which typically comprises approximately 50% of all federal R&D funds (Sargent, 2022).

Defense S&T funding is a key component of the federal scientific research and development portfolio. For the purposes of this paper, S&T refers to the first three budget activities of the Research, Development, Testing, and Evaluation (RDT&E) budget title, which is typically appropriated in Title IV of annual defense appropriations acts, which the president



routinely signs into law to provide the Department of Defense (DoD) with its operating budget.¹ DoD S&T is typically executed by military services and defense agencies such as Defense Advanced Research Projects Agency (DARPA). S&T activities consist of early-stage R&D work ranging from basic research to early technology prototyping. These activities are performed by universities, small businesses, government labs, and commercial or defense industry. This accounting of defense S&T funding does not include funding under the Small Business Innovation Research/Small Business Technology Transfer (SBIR/STTR) program, defense-funded medical research within the Defense Health Agency,² chemical agents and munitions destruction, or military construction appropriations for facilities executing S&T activities (e.g., laboratories). It also does not consider additional appropriations that may have come through the CHIPS and Science Act of 2022, or other supplemental appropriations provided to the DoD. These appropriations were omitted from the analysis because, while these research and support activities could each be fairly classified as S&T, they cannot be parsed easily into the categories of basic research, applied research, or advanced technology development without examining individual SBIR/STTR awards, specific medical research efforts, or specific military construction projects.

Figure 1 outlines the eight budget activities that make up the RDT&E portfolio. The first three of these activities totaled approximately \$22.3 billion for S&T in Fiscal Year (FY) 2023 final appropriations law. These activities are the focus of this paper's analysis and are intended to develop future military capabilities by funding a broad array of research across scientific fields of study and technology sectors, ranging from speculative general scientific inquiry to technology development to meet specific military needs. The goal of this report is to better understand the impact of various policy environments announced by the DoD or Congress on real S&T spending. To do so, these numbers were put in the context of several selected "strategic environments" in the 21st century.

describe the technological maturity of the work being performed. See the discussion beginning on page 2 of CRS Report No. R44711 (Sargent, 2022) for a full explanation of the "Character of Work Structure" of the eight RDT&E budget activities.

² The Defense Health Program (DHP) received significant R&D appropriations, including the congressionally directed medical research programs. DHP was appropriated just over \$3 billion in Fiscal Year (FY) 2023 for RDT&E activities, approximately 7.5% of its total funding level. DoD had requested \$900 million. This follows a requested \$630 million and enacted \$2.6 billion in FY2022. These programs contribute to the military research enterprise but are not typically included in the tabulation of S&T funding.



Budget Activity name Commonly referred to as...	Description of funding ³	Performer of work, by dollars awarded, FY2021 ⁴	FY2023 Appropriations
Basic Research 6.1, ⁵ B.A. 1, Budget Activity 1	Funding to uncover "greater knowledge or understanding of the fundamental aspects of phenomena and observable facts"; long-term investment in scientific knowledge. Not intended to support any particular military application, though intended to lead to some scientific breakthroughs that will support future capabilities.	Intramural: ⁶ 16.76% All Extramural: ⁷ Universities: 59.77%, Industry: 14.09%, Other nonprofits: 3.96%, FFRDCs: ⁸ 2.85%, Foreign: 2.56%	\$2.9B
Applied Research 6.2, B.A. 2, Budget Activity 2	Funding for scientific research for a specific military need, such as identifying the ideal material or modality for a particular capability; medium-term focus. Supports solving a specific technological challenge; not long-term oriented.	Intramural: 31.93% All Extramural: Industry: 43.86%, Universities: 14.30%, Other nonprofits: 5.02%, FFRDCs: 4.19%, Foreign: 0.68%	\$7.8B
Advanced Technology Development 6.3, B.A. 3, Budget Activity 3	Funding for specific subsystems and components that could become prototypes for experiments and testing. Supports development that models or demonstrates the practical feasibility of a technology. Not guaranteed to lead to further development or procurement.	Intramural: 25.42% All Extramural: Industry: 46.61%, Universities: 12.84%, Other nonprofits: 11.09%, FFRDCs: 2.60%, Foreign: 1.43%	\$11.6B
Advanced Component Development and Prototypes 6.4, B.A. 4, Budget Activity 4	Funds efforts to develop full prototype components that can operate in real-world conditions. Successful technology demonstration could lead to "Milestone B" approval, the point at which a contract award permits further system development, validation, and demonstration before full procurement.		\$61.2B
System Development and Demonstration 6.5, B.A. 5, Budget Activity 5	Funds programs that have passed Milestone B approval. Technologies are being prepared to meet program requirements, but are not currently being produced at-scale. Major testing and evaluation efforts begin here to prepare for a "Milestone C" decision, which is the point at which the DoD customer determines whether or not they will support system production & deployment.	* Performers of Budget Activities 4-6 are reported together in NSF surveys. Intramural: 35.96% All Extramural: Industry: 58.47%, Universities: 2.29%, FFRDCs: 2.50%, Other nonprofits: 0.59%, Foreign: 0.18%	\$45.2B
RDT&E Management Support 6.6, B.A. 6, Budget Activity 6	Funds all efforts to "sustain and/or modernize the installations or operations" for RDT&E operations. This can include test ranges, laboratory support, military construction, and studies.		\$21.2B
Operational System Development 6.7, B.A. 7, Budget Activity 7	Funds efforts to upgrade systems that are already in production or being fielded.	Intramural: 20.14% All Extramural: Industry: 78.57%, Universities: 0.28%, Other nonprofits: 0.30%, FFRDCs: 0.68%, Foreign: 0.02%	\$41.7B
Software and Digital Technology Pilot Programs 6.8, B.A. 8, Budget Activity 8	Pilot program provides appropriations for more rapid software development and other digital technologies by allowing selected programs to use funding in this appropriation for RDT&E, procurement, and operation and maintenance as-needed. New RDT&E Program Element created beginning in FY2021 based on the recommendations in the DoD Software Acquisition and Practices (SWAP) Report by the Defense Innovation Board. ⁹	N/A ¹⁰	\$421M

Figure 1. Glossary of DoD Science and Technology Budget Activities (Sargeant, 2018)



Methodology

Each year, the executive branch submits the “President’s Budget Request” to Congress. That document and the supplemental materials that lay out justifications and explanations detail what levels of funding the administration sees as appropriate. Each fiscal year, Congress passes “authorization” legislation to authorize the administration’s departments and agencies to perform certain activities and provides “appropriations” in the form of appropriations acts to provide money for each department’s and agency’s activities at levels—higher, lower, or the same as what the executive branch requested.

The documents used for data collection were the annual Defense Appropriations Conference Reports and the DoD Comptroller’s RDT&E Programs list (Exhibit R-1) for each fiscal year from FY2001 to FY2023.³ From these documents, the extracted data were the DoD enacted topline, the requested and enacted budgets for procurement and RDT&E, including RDT&E Budget Activity 1 through Budget Activity 3, as well as Budget Activity 4 as a whole and as allocated for the services and defense-wide spending. These totals are provided in tables in the Joint Explanatory Statements for the annual defense appropriations acts. For years in which the budget activities were not totaled in those tables, the program elements within each budget activity were manually added by comparing program names to the R-1 PDF for the same year.⁴ The DoD overall topline was taken to be the sum of Titles I–IX, plus Military Construction, but excluded appropriations in other legislation such as the CHIPS and Science Act of 2022, supplemental appropriations, and other adjustments. Additionally, calculations included Overseas Contingent Operations (OCO) in the years in which it existed to place S&T funding in the context of all Pentagon expenditures (McCabe & McGarry, 2019).

To adjust for inflation, the FY2023 deflator values from the DoD’s *FY2023 Green Book* were used (Office of the Under Secretary of Defense [Comptroller], 2022). This includes inflation adjustments for the newly released FY2024 request numbers, which have been adjusted to FY2023 numbers due to the fact that the DoD has not yet released its Pentagon-specific deflators. In this report, “nominal” dollars refer to the funding stated in terms of that prior year’s dollars adjusting for inflation. When reporting a year-over-year percentage change, we calculated this using inflation-adjusted funding levels.

To understand how appropriations are affected by strategic environments, the following time periods were selected to help contextualize the analysis of spending trends:

- **The Operation Iraqi Freedom (OIF) Era (2003–2011):** Operation Iraqi Freedom constitutes the primary duration of combat operations in Iraq. Beyond 2011, trends can be much more clearly attributed to the Budget Control Act of 2011 than operations in Iraq.⁵
- **The Budget Control Act (BCA) of 2011:** The BCA was enacted amid political debate about the appropriate level of government spending. The BCA established limits for defense and nondefense discretionary spending that would trigger a budget sequestration if violated. This led to proactive congressional DoD budget cuts to remain under the caps. When an agreement on further cuts failed in 2013, a

³ The Exhibit R-1 is a document by the DoD comptroller that lists all programs in the RDT&E title’s budget request across all services and defense-wide agencies. The document is released annually and, according to the comptroller, defines each development effort with design, cost, schedule, and capability parameters. See DoD (2022).

⁴ In contrast to other years, the RDT&E appropriations tables in the FY2011 full-year continuing resolution, the DoD and Full-Year Continuing Resolution Appropriations Act, 2011, did not separate program elements by budget activity and, therefore, did not provide requested or enacted funding totals for individual budget activities.

⁵ OIF was selected without Operation Enduring Freedom (OEF) because, due to the length of OEF, it would supersede other 21st century trends.



sequestration occurred that cut defense spending and required DoD and the Office of Management and Budget (OMB) to engage in reprogramming, transfers of funding between accounts, and reductions in procurement (Kogan, 2012; Lynch, 2015). While the legislation was originally scheduled to expire in 2021, the legislation effectively ended in 2019 after the Bipartisan Budget Act increased defense spending caps (Driessen & Lynch, 2019a, 2019b; McGarry, 2015). Additionally, the BCA cannot be fully disentangled from the “earmark⁶ moratorium” established by an agreement between President Barack Obama and Former Speaker John Boehner during the FY2011 appropriations deliberations. That moratorium focused on limiting Representatives and Senators from requesting language for specific funds to be allocated to specific projects (Wong, 2011).⁷

- **The DoD’s Third Offset Strategy Era (2015–2017):** The Third Offset was an effort to draw on advanced technologies to maintain U.S. military superiority over competitors. One of the goals of this strategy was to find mechanisms for technological innovation within the DoD. Another was to identify high-priority systems and technologies for increased investment and support.
- **The 2018 National Defense Strategy (2018 NDS) Era:** The NDS provides strategic guidance to the DoD. The 2018 NDS covered the years from late 2018 to late 2022 and was characterized by a drive to rapidly develop and advance technologies amid international strategic competition.
- **The 2022 National Defense Strategy (2022 NDS) Era:** The 2022 NDS covers the period from late 2022 to late 2026. The unclassified guidance regularly refers to emerging technologies with an emphasis on near-peer deterrence as a key S&T focus. FY2022’s budget request was conceivably generated within the context of the 2018 NDS in addition to the active planning underway to release the 2022 NDS.

Finally, the data were analyzed with respect to the goal that S&T funding should be 3% of the DoD topline.⁸ This is a benchmark that is often cited by defense analysts. While imperfect, it is a helpful landmark to measure defense S&T investments that was noted to be a goal by Congress in several National Defense Authorization Acts and by the DoD in its 2001 *Quadrennial Defense Review Report* (Office of the Secretary of Defense, 2001).

⁶ For the purpose of this analysis, *earmarks* are another term for congressional additions to funding levels in excess of what was requested for a particular program or project. Those earmarks can originate from DoD informal requests or unfunded requirements lists, or they can be congressionally driven to achieve a policy goal or support a constituent.

⁷ That moratorium ostensibly expired during FY2021, though this type of earmarking re-emerged in practice almost immediately: Congress included generalized earmarks into appropriations language, leaving the DoD to determine the specific congressional intent for the funding increases through informal discussions.

⁸ This benchmark evolved from a June 1998 Defense Science Board (DSB) study, which provided several methods for computing an ideal target for S&T funding. Several defense authorization bills have recommended standards based on that report, especially the FY2003 NDAA Conference Report, which commended the DoD’s stated goal of 3%. Additionally, Section 214 of Public Law 105- 261, the FY1999 defense authorization bill, expressed the sense of Congress that S&T should grow by 2% above the rate of inflation year-over-year between FY2000 and FY2008. Notably, neither of these congressional recommendations are borne out in the numbers that came to pass during the 2000s. See Sargent (2022), pages 9, 12, 15, and 17 for more information.



Findings

S&T Funding Overview

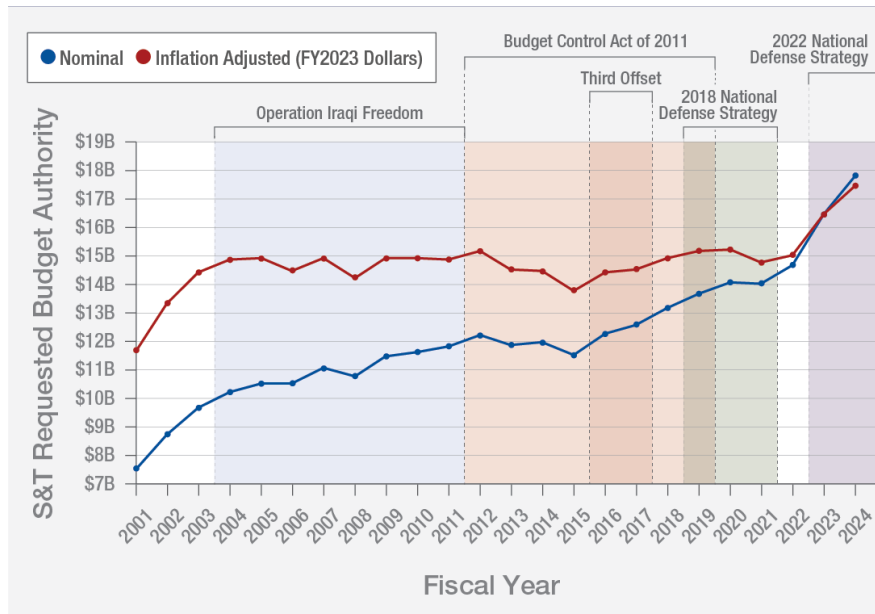


Figure 2. S&T (BA 1-3) Requested Budget Authority by Fiscal Year

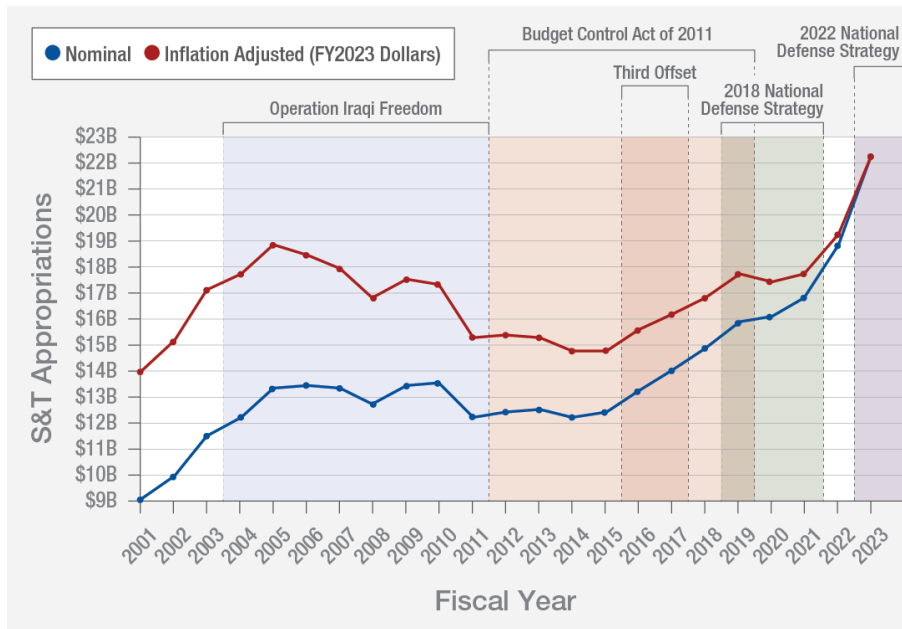


Figure 3. S&T (BA 1-3) Appropriations by Fiscal Year

In the years examined, inflation-adjusted appropriations for S&T activities had peaked at approximately \$19 billion and then declined until the most recent appropriations in FY2023, when it rose to \$22.3 billion. Since FY2001 (\$14 billion), that amounts to an increase of 60% in constant FY2023 dollars, and 146% nominally. Compared to the overall DoD topline—which has increased by over 70% in real terms and nominally by 180%—S&T has become a consistently shrinking component of the DoD budget. Coupled with the acceleration in the pace



of global technology development, this could pose risks for modernization and readiness in an environment where decision-makers must balance today's force readiness with future needs and capabilities. However, it could also be the case that S&T funding at its current level is already sufficient for procuring more of the capabilities that the warfighter will need for future missions and operations; it would make sense for procurement and later-stage modernization efforts like Budget Activity 7 to grow faster if this is the case.

Correlating the data with the selected time periods reveals other insights. While operations in Afghanistan and Iraq coincided with a rapid rise in inflation-adjusted S&T appropriations, likely due to investments driven immediately by 9/11 and the anthrax attacks, the purchasing power of these funds fell by approximately 8% while nominal funding froze at an average of \$13.3 billion from FY2005 to FY2010. The Obama Administration's Third Offset initiative appears to coincide with gradual S&T growth in inflation-adjusted terms. However, congressional additions are much more responsible than the President's Budget Request, which would more directly show the Third Offset's influence.

The end of the Budget Control Act in FY2019, influenced by similar strategic guidance on the need to invest in emerging technologies within the 2018 NDS and 2022 NDS, may have been the cause of the current period of growth in both requests and appropriations, which began to accelerate dramatically after FY2021.

S&T in the Context of the DoD Topline and RDT&E

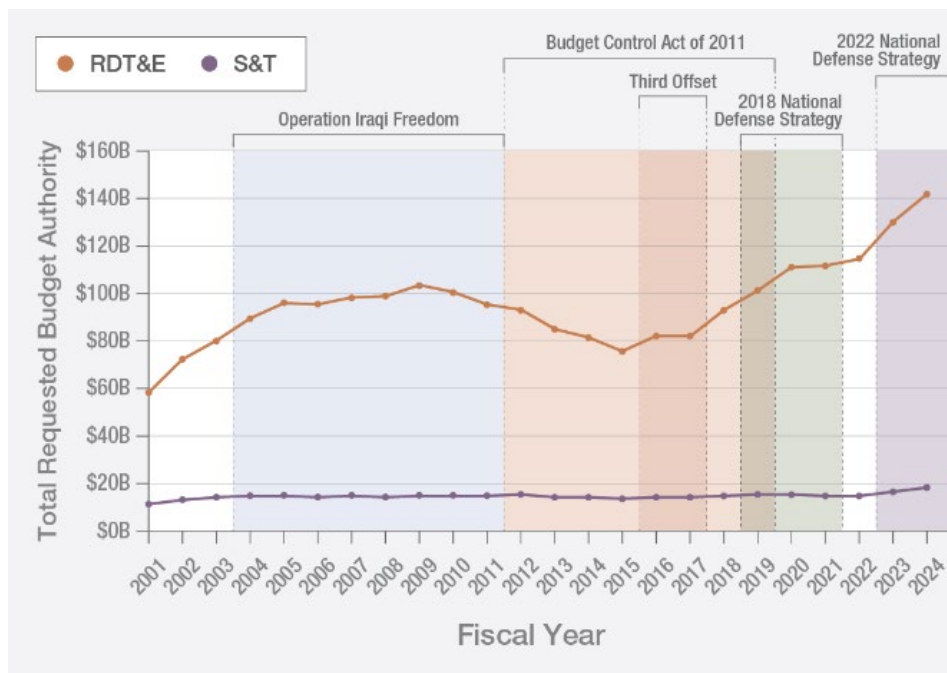


Figure 4. RDT&E and S&T (BA 1-3) Requested Budget Authority by Fiscal Year Adjusted for Inflation (FY2023 Dollars)



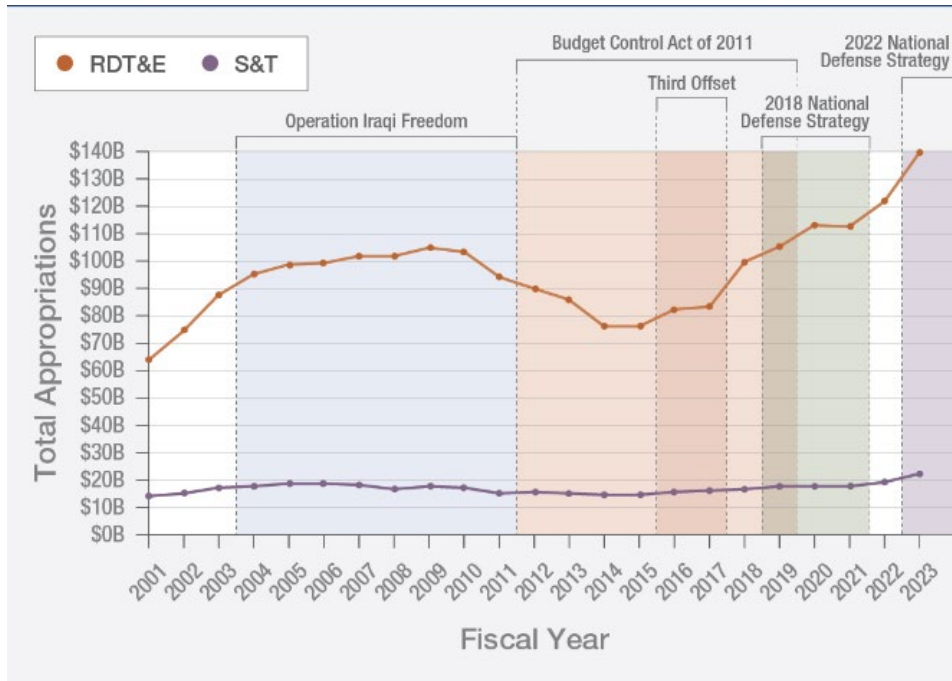


Figure 5. RDT&E and S&T (BA 1-3) Appropriations by Fiscal Year
Adjusted for Inflation (FY2023 Dollars)

When comparing the requested and enacted funding for RDT&E and S&T funding, there are two primary takeaways. First, Congress consistently appropriates more than the administration requests, regardless of the political party in power in either the executive or legislative branch. This has led to a relatively constant amount of funding in real terms for S&T, even during the Budget Control Act years when the DoD requested slightly less.

Congress also consistently appropriates more than what is requested for RDT&E more broadly, though the broader title was strongly affected by funding reductions linked to the Budget Control Act until its final years (FY2016 onward) when Congress began to increase funding again. This may be incidentally related to the larger DoD budget requests stemming from the Third Offset and 2018 NDS eras. Second, this indicates that, in a time of relative fiscal scarcity, both Congress and the DoD consistently chose to maintain S&T funding when resourcing the RDT&E portfolio. This could be due to an institutional desire to see more funding for modernization, or to Congress reacting to feedback from stakeholders in industry and at universities that advocate against cuts for these types of programs as well as for increases for specific S&T projects. A possible alternative explanation could be that decision-makers saw S&T as being particularly sensitive to budget shocks and advocated for less significant declines when compared to other programs.



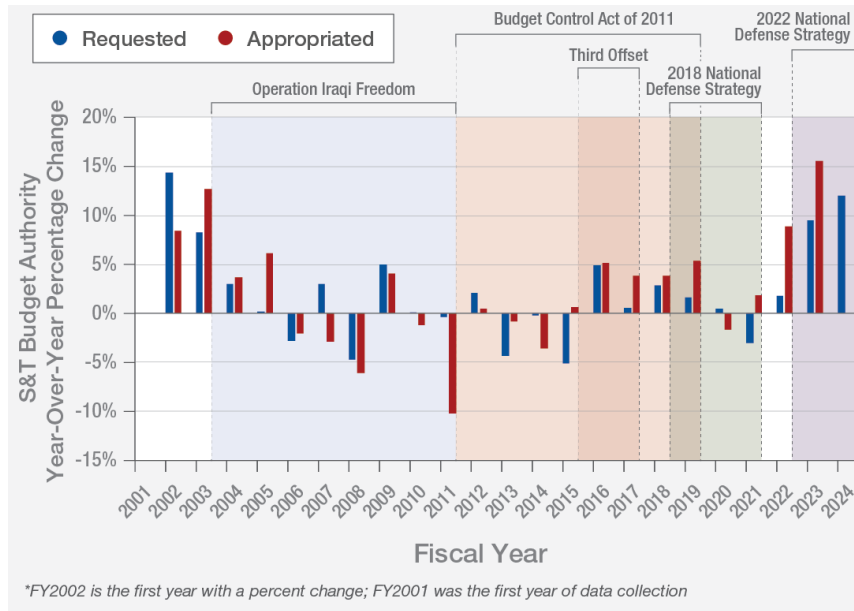


Figure 6. S&T Budget Authority, Percentage Change Year-Over-Year

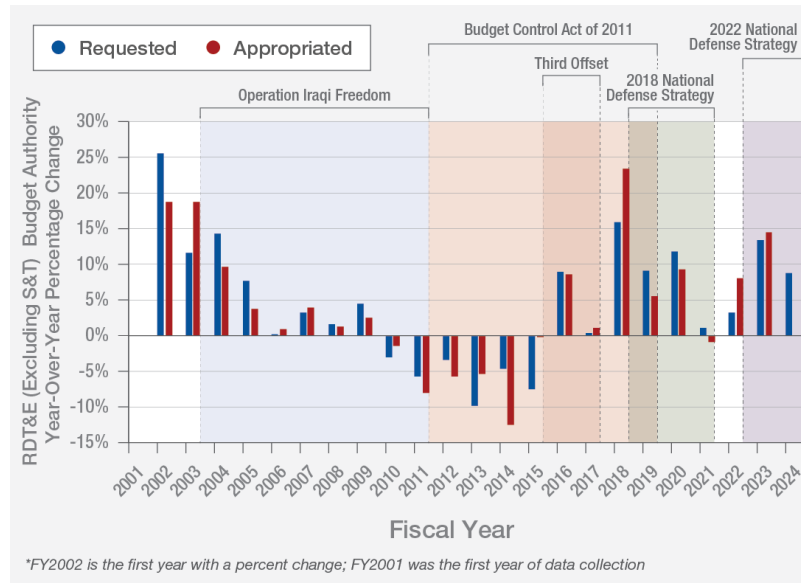


Figure 7. RDT&E (Excluding S&T) Budget Authority, Percentage Change Year-Over-Year

Looking more closely at the rate of change in S&T funding (Figure 6), budget growth and shrinkage year-over-year is both inconsistent and abrupt but relatively small in magnitude compared to swings in RDT&E funding. Compared to the broader RDT&E portfolio, DoD requests and congressional appropriations generally grow S&T funding more slowly than the rest of the RDT&E portfolio when RDT&E is growing. However, S&T funding is often protected when the RDT&E budget is flat or shrinking. This is likely because, when near-term threats seem to be shrinking, near-term prototyping and system development programs are the first to receive cuts. At the same time, S&T funding is likely maintained to hedge against long-term threats.



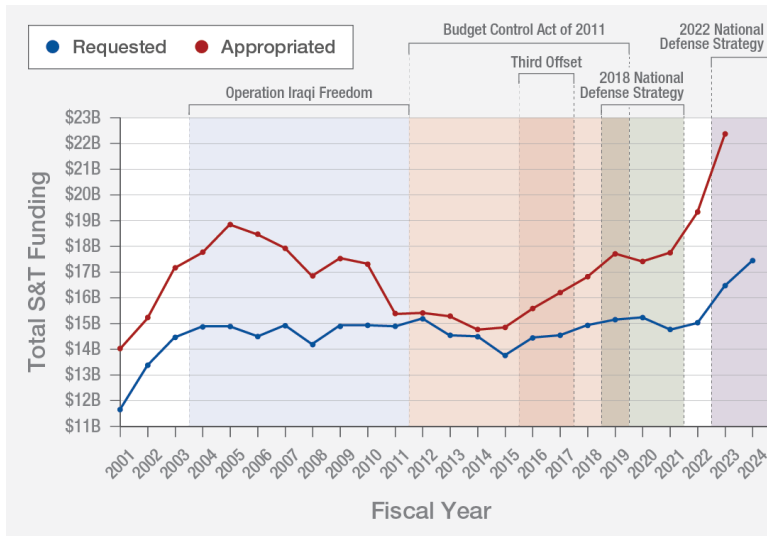


Figure 8. S&T (BA 1-3) Requested and Appropriated Budget Authority by Fiscal Year Adjusted for Inflation (FY2023 Dollars)

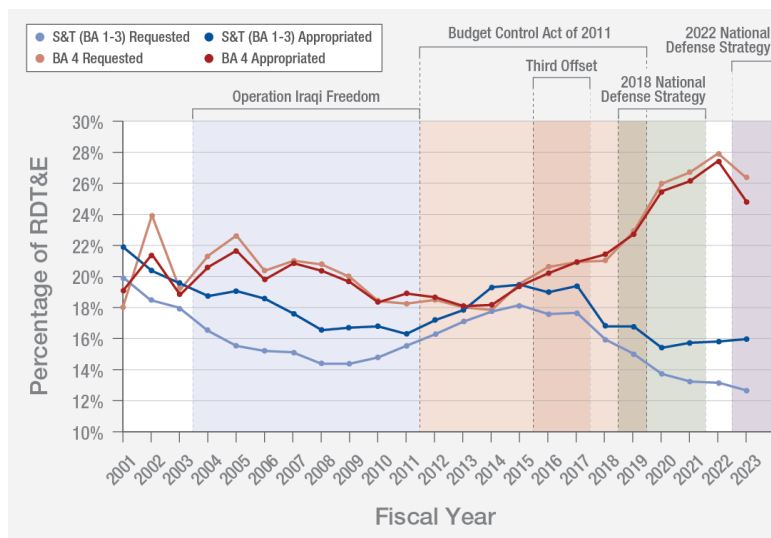


Figure 9. S&T (BA 1-3) and BA 4 Budget Authority, Requested and Appropriated, as a Percentage of Total RDT&E Budget Authority

Returning to Figure 5, these directional swings of low magnitude net out to fairly consistent S&T funding. Because most S&T funds flow out to academic and industry groups to perform the work, rapid increases and decreases cannot be easily absorbed by these institutions when research projects require multiyear investments in workforce and infrastructure, as well as stable and sustained funding to allow research projects to progress. As such, this consistency is welcome. Similarly, if S&T budgets do increase as part of a strategy to help the U.S. for longer-term research competition, policy-makers should do so consistently rather than expecting benefits from any single year increase investments.

Combining requested and enacted S&T funding onto one chart helps illustrate Congress's tendency to appropriate more S&T funding than the DoD requests. For the period examined, the amount of enacted S&T funding always exceeded the President's Budget Request in any given year no matter the strategic era. For example, while the FY2022 Budget



Request did not sufficiently protect the S&T budget topline from unanticipated levels of inflation, the relatively modest decrease in the enacted budget indicates that Congress took actions that resulted in maintaining S&T funding relatively constant in real terms. In FY2023, S&T activities received their largest-ever inflation-adjusted funding request, and Congress earmarked further funding increases for these activities that boosted the enacted funding far beyond inflation.

As mentioned earlier in this section, DoD requests for S&T as a percentage of RDT&E remained notably flat during the Third Offset era. The high increases in S&T funding seen during this era were driven primarily by congressional increases rather than any department-driven planning guidance. Conversely, in the late 2000s, the DoD consistently requested a similar amount of money in real terms, while Congress's willingness to provide such large increases over the request decreased as OIF continued.

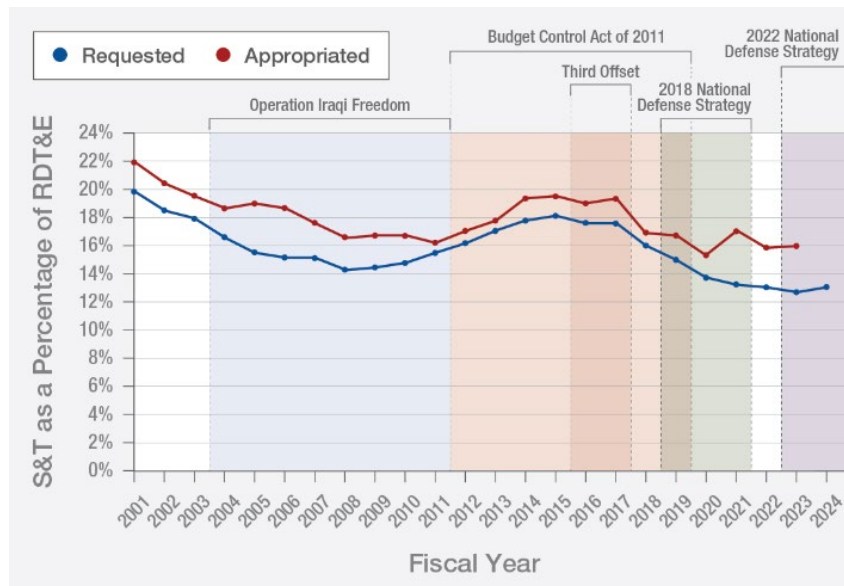


Figure 10. S&T (BA 1-3) Funding as a Percentage of RDT&E Funding

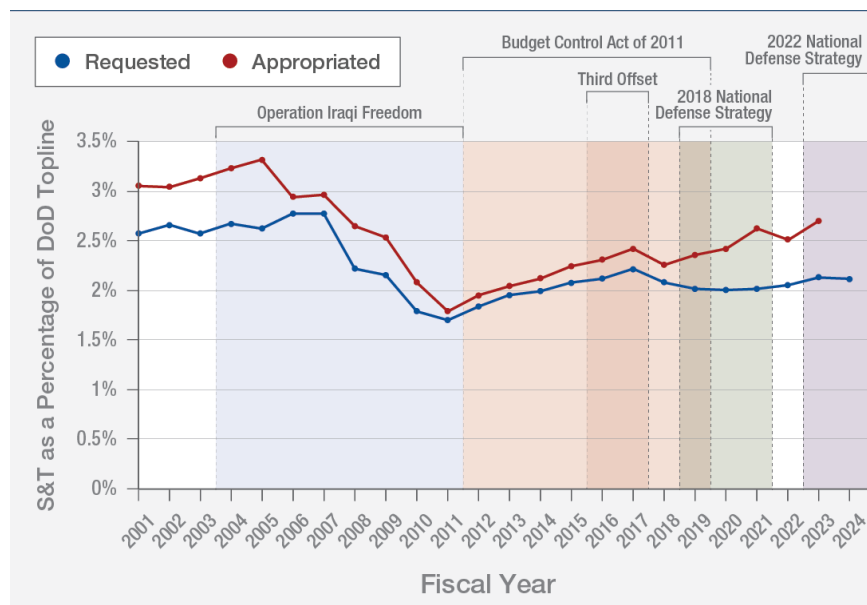


Figure 11. S&T (BA 1-3) Funding as a Percentage of DoD Topline



Even while Congress maintains S&T's funding level, S&T has shrunk by nearly a third as a share of the appropriated RDT&E portfolio over the FY2001 to FY2023 period. The share has fallen from approximately 22% of the inflation-adjusted enacted RDT&E funding in FY2001 to 16% of the funding in FY2023. This is primarily due to significant real growth in the rest of the RDT&E portfolio, which more than doubled in real terms (119%) and increased nominally by 238%. In fact, appropriated RDT&E has nearly doubled nominally since FY2017. This trend is also evident in DoD requests until FY2021, indicating that DoD leaders decided to allocate additional topline resources to near-term R&D and procurement efforts to match the perceived priority of near-term threats. However, S&T requests and appropriations after FY2021 began to grow dramatically, closer to 3% of the topline, consistent with strategic guidance on the technological nature of the great power competition with China and Russia. Because of this growth after FY2021, S&T has nominally grown by 63% over the FY2017 to FY2023 appropriations period. According to this understanding, the simplistic goal of funding S&T at 3% of the DoD topline may not be an important way of defining success. Instead, policy-makers must assess whether or not the inflation-adjusted level of funding for the S&T enterprise is achieving the scientific and technical breakthroughs that warfighters need on relevant timescales.

S&T's shrinking share of RDT&E since 2017 was examined in light of major investments in the modernization of the nuclear triad. While increasing triad modernization investments did contribute to S&T's shrinking share, S&T's share of the RDT&E title fell even when accounting for the triad's share. This is true because growth in RDT&E over the past decade has been sustained across the entire title; investments in the triad alone are not responsible for the massive increase in real funding for development activities across the title.

Another key finding is that the Third Offset era marks a divergence in S&T funding versus RDT&E Budget Activity 4. Beginning in that period, DoD requests for increased funding for bigger projects that benefit from more complete development and prototyping began to accelerate. As a result, Budget Activity 4 grew substantially as a percentage of the RDT&E title.

Separately, because the Budget Control Act appears to have increased S&T's share of RDT&E as other items in the budget shrunk, it appears that the only selected era that truly affected S&T's prioritization was the one controlled with the force of a law driving automatic cuts at the agency level, rather than any intentional actions by decision-makers. Without those types of statutory fiscal constraints, S&T budget activities only grow when the RDT&E title grows even faster, as it did in FY2023. The earmark moratorium discussed earlier also reduced congressionally directed spending on S&T during these years.

For its part, Budget Activity 4 has seen its share of the RDT&E funding grow substantially over the same period, reflecting an increased interest in prototyping activities and investments in major acquisition programs.

The DoD also seeks to influence the appropriation of additional funds through its annual unfunded requirements lists, often used by Congress to provide funding for activities that the DoD was unable to budget for. These lists rarely include S&T activities but often do include projects for Budget Activities 4 to 7. For example, of the approximately \$21.4 billion in unfunded priority requests across the DoD in FY2023, only about 0.58% was for unclassified S&T programs. Across every service and combatant command, it appears that only \$162.6 million was listed for S&T programs—nearly entirely from the Navy. Virtually all of this money was for projects within the Applied Research portion of the Innovative Naval Prototypes (INP) program element, and Congress ultimately appropriated a \$25 million increase for this purpose. That means that, out of approximately \$8 billion in congressional earmarks for S&T in the FY2023 Consolidated Appropriations Act, less than half of a percentage came from the



Unfunded Priorities List process.

The fact that unfunded requirement lists are rarely used for the S&T enterprise does not mean that the DoD does not pay attention to basic and applied research. However, the simple fact that these types of research activities provide less political value for congressional earmarking (because they do not directly support priority constituencies) may explain why the later budget activities in RDT&E have grown faster than S&T.

All told, S&T funding began to shrink from its peak percentage of the DoD topline in the mid-2000s before beginning to rebound in 2011 in Congress. S&T was funded at approximately 2% of topline, a remarkably consistent number, even in the face of significant advocacy in strategic plans such as the Third Offset and the 2018 NDS. It may be the case that, even though the language in the Secretary of Defense's Defense Planning Guidance (DPG) often instructs the services to maintain a minimum of 0% real growth in S&T with the intent to set a minimum for funding levels, in practice it becomes both the ceiling and the floor of what services will program for S&T for the next year.

It is also notable that, even though maintaining S&T funding at 3% of the DoD topline has been described as a DoD goal, the Pentagon has not requested this level of funding in any year between FY2001 and the present. This is true regardless of whether or not OCO funds are included as part of the topline. Appropriations last pushed funding to this benchmark in FY2005 and inched quite close—just over 2.7%—in FY2023. This decline below 3% occurred before the Budget Control Act and has not returned to that 3% target in the years since the BCA ended. Because, as noted above, the Budget Control Act appears to coincide with a period of S&T budget growth as a share of the topline again as other components of the DoD topline ceded resources, it appears that outright fiscal restraint may not be the reason for S&T budget activities' shrinking share.

Conclusion

This white paper is intended to be the first in a series of papers by ETI on defense S&T programs, budget requests, and appropriations. This paper seeks to describe the types of events that drive S&T funding levels. To better understand these trends, a capability to track more details of budget and appropriations data to better monitor these critical investments and the opportunities they represent to develop new defense technologies would be useful. This includes a strategy to better track congressional marks (adds and reductions), more detailed information from service and defense-wide Exhibit R-2 Budget Justification materials, other reports such as the DD Form 1414, and more detailed tracking that follows specific emerging technologies from the research phase to contracting.^{9, 10}

For the most part, the time periods analyzed do not appear to coincide with significant variation in S&T spending as a priority, even while several of these eras reflect several administrations' rhetoric and goals for defense modernization. The major exception to this rule is the Budget Control Act of 2011, which meaningfully coincides with S&T growing as a share of the RDT&E and DoD toplines due to cuts in other areas. It is notable that despite pronouncements from the DoD throughout Third Offset and 2018 NDS periods on the importance of S&T funding, it appears to have been congressional actions during periods of

⁹ The Exhibit R-2 is produced for each program within a service or defense-wide agency. It includes specific budget justification materials for each program, including a summary of resources, mission description, and justification narratives for a given program and each of its projects. See DoD (2022), Chapter 5.

¹⁰ The DD Form 1414 is a form submitted by the DoD to Congress within 60 days after enactment of a new defense appropriations act. The form establishes the DoD's baseline funding levels for each account—including adjustments, rescissions, and supplemental appropriations—which the DoD can transfer or reprogram against. Read more in McGarry (2021).



significant earmarking that push S&T dollars to dramatically higher levels relative to the topline. Both the DoD and Congress have fallen short of the stated goal to fund the S&T portfolio at a level of 3% of the DoD topline in the 21st century.

At the same time, analysis of the S&T funding portfolio alone has shown that absolute spending in S&T has been remarkably consistent in real terms—approximately \$14.5 billion per year over the course of the 21st century. However, the portfolio began to grow in inflation-adjusted terms during the Third Offset and has only increased since then in requests during the 2018 and 2022 NDS periods. It appears that something closer to \$16 to \$17 billion may be the new equilibrium for S&T funding requests. This increase across the past three administrations may be due to a strategic shift towards great power competition.

The implications of this report indicate that the high funding added by Congress for S&T is unlikely to last once the appropriated DoD topline begins to flatten. This is likely to be true even as the 2022 NDS and senior defense officials emphasize in testimony that modernization and investments in critical technology areas are crucial for maintaining a capability edge over potential adversaries.

The research and engineering community, both inside and outside of the federal government, benefit greatly from clear and consistent levels of funding. Research institutions cannot easily absorb large injections of funding that may not be sustained due to the need to hire more researchers, patiently pursue technical achievements, and build more laboratory and test infrastructure. They also cannot adapt to rapidly falling funding, which leads to project cancellations and an environment where research teams cannot take risks. For this reason, there would be some benefit in the DoD and Congress re-evaluating, and maintaining, a standard benchmark of S&T funding. That benchmark may be the “3% of topline” metric, but it could also be the case that this metric is too rigid and ignores the fact that non-S&T costs may rightfully grow faster to respond to near-term threats. Other goals, such as a commitment to protect S&T funding from inflation or to boost S&T funding at a specified rate, may also be helpful. This process will require active planning and programming efforts by the DoD as well as active legislative support by Congress, with benefits that flow across the DoD to its many modernization efforts planned for the decade ahead.

References

- DoD. (2019). *Software is never done: Refactoring the acquisition code for competitive advantage*. Defense Innovation Board.
https://media.defense.gov/2019/Apr/30/2002124828/-1/-1/0/SOFTWAREISNEVERDONE_REFACTORINGTHEACQUISITIONCODEFORCOMPETITIVEADVANTAGE_FINAL_SWAP.REPORT.PDF
- DoD. (2022, September). *Financial management regulation: Volume 2B*.
https://comptroller.defense.gov/portals/45/documents/fmr/current/02b/02b_05.pdf
- Driessen, G. A., & Lynch, M. S. (2019a). *The Budget Control Act: Frequently asked questions* (CRS Report No. R44874). Congressional Research Service.
<https://crsreports.congress.gov/product/pdf/R/R44874/19/>
- Driessen, G. A., & Lynch, M. S. (2019b). *The Bipartisan Budget Act of 2019: Changes to the BCA and debt limit* (CRS Report No. IN11148). Congressional Research Service.
<https://crsreports.congress.gov/product/pdf/IN/IN11148/7>
- Kogan, R. (2012). *How the across-the-board cuts in the Budget Control Act will work*. Center on Budget and Policy Priorities. [https:// www.cbpp.org/research/how-the-across-theboard-cuts-in-the-budget-control-act-will-work](https://www.cbpp.org/research/how-the-across-theboard-cuts-in-the-budget-control-act-will-work)
- Lynch, M. S. (2015). *Sequestration as a budget enforcement process: Frequently asked*



- questions* (CRS Report No. R42972). Congressional Research Service.
<https://crsreports.congress.gov/product/pdf/R/R42972/10>
- McCabe, E. M., & McGarry, B. W. (2019). *Overseas contingency operations funding: Background and status* (CRS Report No. R44519). Congressional Research Service.
<https://sgp.fas.org/crs/natsec/R44519.pdf>.
- McGarry, B. W. (2015). *The Defense Budget and the Budget Control Act: Frequently asked questions* (CRS Report No. R44039). Congressional Research Service.
<https://crsreports.congress.gov/product/pdf/R/R44039/21>
- McGarry, B. W. (2021, December). *DoD transfer and reprogramming authorities* (CRS Report No. IF11243). Congressional Research Service.
<https://crsreports.congress.gov/product/pdf/IF/IF11243/6>
- National Science Foundation. (2023). *2020-21 survey of federal funds for research and development*. National Center for Science and Engineering Statistics.
<https://nces.nsf.gov/surveys/federal-funds-research-development/2021>
- Office of the Secretary of Defense. (2001). *2001 quadrennial defense review report (QDR)*. DoD.
<https://history.defense.gov/Portals/70/Documents/quadrennial/QDR2001.pdf?ver=AFts7axkH2zWUHncRd8yUq%3d%3d>
- Office of the Under Secretary of Defense (Comptroller). (2022). *FY2023 green book*. DoD.
https://comptroller.defense.gov/Portals/45/Documents/defbudget/FY2023/FY23_Green_Book.pdf
- Sargent, J. F. (2018). *Defense science and technology funding* (CRS Report No. R45110). Congressional Research Service. <https://crsreports.congress.gov/product/pdf/R/R45110/6>
- Sargent, J. F. (2022). *Department of Defense research, development, test, and evaluation (RDT&E): Appropriations structure* (CRS Report No. R44711). Congressional Research Service. <https://sgp.fas.org/crs/natsec/R44711.pdf>
- Wong, S. (2011). *Senate dems give in on earmark ban*. Politico.
<https://www.politico.com/story/2011/02/senate-dems-give-in-on-earmark-ban-048623>



Capability-Based Software Cost Estimation (CaBSCE): Proposing a New Method to Estimate Software Costs

Anandi Hira—leads research to advance the state of the art in software cost estimation and supports programs with the resulting estimation and measurement processes in her current role at Carnegie Mellon University's (CMU's) Software Engineering Institute (SEI). Her most recent research objective is to develop a new, innovative, capability-based software cost estimation model. Hira received her PhD in software cost estimation under Barry Boehm at the University of Southern California (USC), where she collected software development data and calibrated the COCOMO® II software cost estimation model for Function Points. [info@sei.cmu.edu]

Abstract

The dynamic nature of software has led to exciting technology improvements as well as challenges in cost estimation. Defining specific software requirements early in the lifecycle is impossible. The Software Acquisition Pathway (SWP), defined in Department of Defense (DoD) Instruction 5000.87 (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020), requires only a high-level specification of needed capabilities—a capability needs statement (CNS) or software initial capabilities document (SW-ICD)—before entering the execution phase of software-intensive programs (Defense Acquisition University, 2020b). Most cost estimators for the U.S. government rely on size-based estimation methods, primarily source lines of code (SLOC). However, estimating SLOC requires understanding specific implementation details and can only be accurately estimated near program completion. The software cost estimation community needs a capability-based estimation method that aligns with the SWP and meets the government's need for evidence-based, flexible, and defensible estimates (Defense Acquisition University, 2020a). In response to these competing needs, the Software Engineering Institute (SEI) proposes developing a software capability-based estimation model requiring only high-level information. Using existing software development effort data, the SEI plans to analyze descriptive fields to cluster data points based on similar software functions and complexity. The resulting model would provide evidence-based rough order of magnitude (ROM) estimates for the initial stages of identifying required capabilities.

Introduction

Developers and managers from the commercial space, private sector, and government acquisition learned that defining software requirements in advance is an impossible expectation (Highsmith, 2002; Raza & Waheed, 2018). Requirements change significantly because both users and customers might not have a clear understanding of what they want and need, different stakeholders interpret requirements differently, and quickly evolving technology affects the needs of the new project (De Lucia & Qusef, 2010). These requirements issues have caused many software projects to either fail or significantly overrun their budgets and/or schedules (Shah & Patel, 2014). Several software engineers came together to draft the Agile Manifesto and its corresponding 12 principles to improve software development. Agile software development primarily addresses requirements issues by “[w]elcom[ing] changing requirements, even late in development,” showing progress, and giving customers and users the opportunity to re-evaluate requirements. This is done through short increments of working software (Beck et al., 2001). Allowing changes in requirements improves the probability of success but poses challenges for estimators. Without knowing exactly what the final product will be able to do, estimators cannot accurately estimate the program's effort, cost, and schedule required to develop what the customers and users need. In the commercial industry, many organizations



have circumnavigated this issue by implementing high-priority requirements and addressing as many requirements as the customers can afford. Since developers demonstrate working code at each increment, they continue working until the customer is sufficiently satisfied (Milani & Rossi, 2020). However, estimates play an essential role in the U.S. government's defense acquisition process, as estimates determine which programs to authorize and how to distribute annual funds across active programs.

Traditionally, the government identifies high-level capability needs, which are broken down into detailed requirements. These requirements serve as a basis for developing cost estimates that determine the feasibility of the solution and allow the comparison of alternate options. If the government both authorizes and appropriates funds for the program, the acquiring organization typically posts a request for proposal package (RFP), and organizations may place their bids towards the contract. These organizations must provide their own estimated cost and schedule, which the government evaluates to award the best-value contractors. After the contract is awarded and implementation starts, the program's estimates are updated periodically to ensure that the updated estimates reflect the learning and changes in the scope. The government encourages contractors to adopt Agile and DevSecOps (DevOps with an added focus on security, which is commonly used by defense programs) practices to "rapidly adapt to emerging user needs" and "shorten acquisition timelines" (Coonce & Alleman, 2017) and "match the speed at which new IT capabilities are being introduced in today's information age" (Defense Science Board, 2009). However, traditional waterfall acquisition lifecycle does not align with Agile and DevSecOps practices. In response, the government implemented the Software Acquisition Pathway (SWP) defined in DoD Instruction 5000.87 (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020). The SWP allows moving into the execution phase with CNS or SW-ICD, which provide high-level operational capabilities, enhancements to existing capabilities, features, interoperability needs, legacy interfaces, and other attributes relevant to define how the software solution relates to the overall threat environment (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020). Detailed requirements evolve through Agile increments. However, cost estimates are still required during the planning phase for the government's budget process. Hence, defense acquisition programs struggle with developing estimates and providing evidence to defend their budget requests.

The SEI proposes developing a capability-based software cost estimation model (CaBSCE) based on high-level information that corresponds to the capabilities provided at program initiation. CaBSCE would accommodate uncertainty in later architecture, reuse, and implementation decisions. The underlying sizing method is based on identifying clusters of software functions in common capabilities, and the cost model is calibrated with historical data from comparable efforts, thus satisfying the government's need for flexible, credible, and defensible estimates (Defense Acquisition University, 2020a) as well as the estimators' need for a cost model based on capabilities. The next section defines capability-based planning (CBP) and capabilities, which explains the level of detail made available when early estimates are needed, followed by quick summaries of predominant and relevant software cost estimation techniques and a description of how the SEI will develop CaBSCE.



Background

Capability-Based Planning (CBP)

Defining CBP

Several countries' governments (e.g., United States, Canada, Australia) have started to use CBP to "design an appropriate [military] force" and one that is "postured to adequately deal with the challenges of the future" (Taliaferro et al., 2019). Instead of focusing on who an adversary might be, where a war might occur, or how an adversary might fight (Biltgen, 2007; Hanley et al., 2006; Planeaux, 2003; Walker, 2005), the "goal is to plan for robust, flexible forces, capable of meeting a wide variety of threats, rather than an 'optimal' force for a narrow set of threats" (Titus, 2004). The "objective is to develop a flexible, adaptable, robust, and sustainable (i.e., technically manageable and financially affordable) force structure postured to address all the challenges associated with a nation's strategic defense and security environment, considering budgets and uncertainty" (Taliaferro et al., 2019). With CBP, governments and defense departments evaluate "the development and evolution of capabilities, rather than specific programs or function" (Webb, 2008) to go "from programs to portfolios of capabilities" (Bexfield & Disbrow, 2004) and "determine an efficient and effective mix of military forces" (Taliaferro et al., 2019). A capability serves as a goal that enables us to decide whether a specific technology or process supports achieving that goal (Taliaferro et al., 2019). In summary, CBP underscores the significance of cultivating adaptable military capabilities and addressing future challenges in national defense by transcending the limitations of specific defense organizations, military services, programs, or functions.

The most referenced definition of CBP comes from Paul Davis (2002):

Capabilities-based planning is planning, under uncertainty, to provide capabilities suitable for a wide range of modern-day challenges and circumstances, while working within an economic framework.

Essentially, CBP must "confront—rather than discount—uncertainty, to express risk in meaningful terms, and to weigh costs and benefits simultaneously" (Committee on Naval Analytical Capabilities and Improving Capabilities-Based Planning, 2005). Uncertainty stems from two sources: (1) the scenarios that describe the needed capabilities and (2) the "details of assumptions in those scenarios" (Committee on Naval Analytical Capabilities and Improving Capabilities-Based Planning, 2005). Additionally, CBP should consider a diverse "range of competitive options and trade-offs before making the choices necessitated by a budget" (Committee on Naval Analytical Capabilities and Improving Capabilities-Based Planning, 2005). In other words, CBP focuses on goals and outcomes and encourages innovation (Anastasios, 2014; Bexfield & Disbrow, 2004; Chim et al., 2010; Desgagné, 2009; Technical Cooperation Program, 2004) by "moving away from determining equipment solutions prematurely" (Technical Cooperation Program, 2004). This "provides a means to compare different options for achieving the same capability" (Technical Cooperation Program, 2004). Hence, it is not CBP's goal to "engineer planning processes to a fine level of detail, but rather to design an effective decision-support mechanism for regular, rigorous integration of planning process outputs" (Hanley et al., 2006). To support this high-level, generalizable framework, CBP "starts with a top-down definition through scenarios, case studies, or use cases" without eliciting detailed requirements (Alleman, 2020b).



The CBP processes require stakeholders to “think broadly about the entire scenario space of possibilities” (Davis et al., 2008). “Scenarios provide the essential link between defense policy and capability objectives” (Technical Cooperation Program, 2004). Broad, high-level scenarios help with the following:

- develop “realistic capability goals” (Technical Cooperation Program, 2004)
- provide “context and a means to share assumptions” (Hales & Chouinard, 2011) and therefore “facilitate communication” as it becomes “easier to compare options in a strategic-level framework if everyone has a fairly concrete mental image of what the evaluation cases are” (Davis et al., 2008)
- provide context for capability assessment (Titus, 2004) and identifying gaps associated with the mission area (Chairman of the Staff, 2009)
- “provide a way to test the concept against the breadth of the defense strategy” and “the spectrum of conditions to be considered” (Chairman of the Staff, 2009)
- “assess whether the capability at issue would merely be nice to have logically or would make a difference in plausible cases of concern” (Davis et al., 2008)

Therefore, to make strategic and successful decisions, governments must evaluate their ideal and existing capabilities at the highest level possible while refraining from making decisions that can overlook uncertainty, risk, and budget constraints. The generalizability of the CBP framework corresponds to the generalizable definitions of capabilities. CBP and planning at the capabilities level aim to offer a holistic and flexible perspective while ensuring success.

Defining Capabilities

Capabilities have been defined in many ways to support the high-level, flexible, and comprehensive CBP framework. The definitions can be grouped into three categories: (1) the description of objectives or high-level needs, (2) operational outcomes, and (3) the ability to produce or achieve some type of outcome.

Definitions in the first set collectively illustrate that capabilities represent objectives or high-level needs that form the fundamental basis of strategic decision making:

- “Functional approach to the articulation of broad requirements without necessarily specifying the resources that may be involved” (Chim et al., 2010)
- “Defined by an operational user and expressed in broad operational terms” (Iacobucci, 2012)
- “Foundation of defining the technical and operational requirements of the product or service produced by the project. Without the defined capabilities, those requirements have no reason for being” (Alleman, 2020b).
- “Consist of far more than just technology; in fact, technology underpins just one element—materiel—of a capability” (Hanley et al., 2006)
- “Capabilities are not the same as features and functions; they enable demands to be met without the explicit specification of the solution” (Alleman, 2020a).
- “Capabilities provide the answer to the following question: To achieve our objectives, what capabilities must we possess?” (Davis et al., 2008). In other words, “to achieve our objectives,” what must we be able to do? Capabilities include talent, expertise, materiel, and capacity, among other things.



The next category of capability definitions is related to the above, but instead of focusing on objectives and needs, it focuses on operational outcomes:

- “A description of the military operational output or outcome that a unit, force, or organization is able (and usually constituted or organized) to deliver” (Steele, 2021)
- “Define the future effects needed for agencies to meet their mission and transform into a more agile and adaptable force” (Kossakowski, 2005).
- “the combination of military equipment, personnel, logistics support, training, resources, etc. that provides Defence with the ability to achieve its operational aims” (Neaga et al., 2009)

While the first two categories of definitions focus on what can be changed or done, the last group of definitions focuses on the ability to produce or achieve some type of outcome rather than the outcome per se:

- “Ability to achieve a desired effect under specified standards and conditions through combinations of means and ways to perform a set of tasks” (Bexfield & Disbrow, 2004)
- “Further defined as the ability to contribute to the achievement of a desired effect in a given environment within a specified time and the sustainment of that effect for a designated period” (Steele, 2021)
- “Wherewithal to complete a task or produce an effect within a set of specified performance standards and environmental conditions” (Taliaferro et al., 2019)
- “The ability to achieve an objective in a military operation” (Chairman of the Staff, 2009)

At first, these definitions may seem divergent. However, they collectively describe the essence of capabilities as high-level descriptions essential for fulfilling the government’s objectives. These multifaceted definitions capture different dimensions in which achieving a solution must encompass. The individual definitions may represent a view from a time in the lifecycle, the viewer’s role, or a stakeholder’s value proposition.

Software Acquisition Pathway (SWP)

The output of CBP (i.e., identified capability needs) serves as the input to the defense acquisition lifecycle. In the waterfall acquisition lifecycle, the high-level capability needs are broken down into detailed requirements that specify how they need to be met. Traditionally, software was treated as a component of the overarching system or “an enabler of hardware systems and weapon platforms” (Defense Innovation Board, 2019). As the government realized that software is predominant and “defines our mission-critical capabilities and our ability to sense, share, integrate, coordinate, and act” (Defense Innovation Board, 2019), the government developed the SWP to provide specific guidance for software development and remove the roadblocks to Agile and DevSecOps software development practices in the defense acquisition programs. The SWP allows software development efforts to adapt quickly and meet capability needs; it also enables the “timely acquisition of custom software capabilities developed” (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020).

The SWP abandons a waterfall lifecycle model in favor of an incremental and iterative one—that is compatible with modern software development practices such as Agile and DevSecOps. The high-level capability needs replace detailed requirements before moving into the execution phase. In compliance with Agile practices, the requirements evolve through



continuous user involvement and high priority needs are addressed earlier. In the execution phase, requirements engineering, development, integration, testing, and deploying software occur incrementally; programs are required to deliver software capability releases every year or more frequently. Cost estimates are still required before moving to the execution phase and are updated annually (Office of the Under Secretary of Defense for Acquisition and Sustainment, 2020). Initial estimates are not expected to precisely capture the required cost and schedule, but they should be updated iteratively as requirements, architecture, and implementation decisions are refined (Defense Acquisition University, 2020a).

Software Cost Estimation Methodologies

Cost Estimation Criteria

Estimation plays a crucial role in the decision-making processes of the CBP framework, since “the most effective and efficient options to satisfy the requirements” and fill capability gaps “are sought” (Davis, 2002). Estimation is particularly vital when both comparing and analyzing all possible and alternative solutions within CBP. Without estimates for the potential outcomes of each choice in the list of alternatives, it becomes impossible to determine which capabilities are best suited to meet the mission’s needs and which ones fulfill a business case. As Alleman (2020) emphasizes, credible decisions in the face of uncertainty rely on estimates. Therefore, the accuracy and reliability of estimates are critical to ensuring that CBP can effectively inform decisions about which capabilities to pursue, aligning them with strategic objectives and mission requirements. The U.S. government also requires cost estimates to justify the costs of a program and distribute annual funds across active programs. For estimates to effectively support CBP and the U.S. government’s budget processes, the estimation method/model must meet the following criteria (the labels in the beginning correspond to labels used in Table 1):

1. **Early Lifecycle:** be applicable to high-level capabilities early in the acquisition lifecycle—before contractors and developers see the capability needs or requirements
2. **Defensible:** be defensible and evidence-based (based on historical data)
3. **Generalizable:** be generalizable, as the underlying cost models should be contractor agnostic (i.e., estimates should be applicable before contractors have been selected)
4. **Cost Analysis:** be able to perform cost analysis and answer questions such as, How much will it cost, and how long will it take to reach a minimal viable product (MVP)? What would a minimum viable capability release (MVCR) entail? Will required capabilities be completed by a deadline? How will the schedule and/or cost be affected by a reduced budget?

An estimate’s inability to meet the above criteria can lead to the government cutting a program’s budget, which will delay capability delivery or require descoping of capabilities. Unfortunately, as demonstrated in this section, none of the predominant, existing software cost estimation methods meet all the above criteria.

Source Lines of Code (SLOC)

Most cost estimators for government agencies still rely on size-based estimation, primarily using SLOC, due to its quantifiability, high correlation with effort (Albrecht & Gaffney, 1983; Kemerer, 1987), and large repository of historical data and cost estimation models. But estimating SLOC requires understanding specific implementation details, leading to significant changes in SLOC estimates throughout a program’s lifecycle (GAO, 2020). SLOC can only be



accurately estimated when a program is near completion (Boehm, 1981). In summary, SLOC fails to meet criterion 1—being applicable to high-level capabilities early in the lifecycle.

Story Points

Story Points are often used in Agile environments by the development team to (based on their judgment or subjective comparison to previous tasks) estimate the level of difficulty and required effort to implement the requirements. Development teams use numbers, for example, from the Fibonacci sequence, to signify the difficulty and required effort to implement requirements expressed as user stories (Cohn, 2004). However, Story Points were not developed to estimate effort or costs; they are part of an exercise used by the development team to determine which requirements they can tackle within the capacity of a development sprint. In several experiments, Jørgensen (2004) found that Story Points estimates are subject to bias with respect to the order in which projects are estimated. Jørgensen found that estimators tend to estimate larger tasks better after estimating smaller tasks. After estimating larger tasks, estimators are more likely to give optimistic (less effort) estimates for smaller tasks. Additionally, estimators may have the tendency to think that of two similarly sized projects, the second one (despite the order in which the projects are evaluated) will seem larger than the first one (Jørgensen, 2004). Since Story Points are estimated by the development team, they would not be available until the developers received the requirements of the system, which is later in the acquisition lifecycle. Hence, Story Points fail to meet criteria 1, 2, and 3.

Analogy-Based Estimation

In analogy-based estimation, program managers and subject matter experts (SMEs) typically use high-level characteristics, such as the type of program (e.g., information management system) or development process, to select an analogous program or data and apply a complexity factor to account for differences between the current and past programs (Ozkaya et al., 2008). While analogy-based estimation can be used early in the lifecycle, it still faces the same expert-based biases that Jørgensen (2004) described in his study. Therefore, analogy-based estimation fails to meet criterion 2 (i.e., being defensible and evidence-based).

Case-Based Reasoning

Academic research has studied case-based reasoning, but the studies focused on the variables available in the dataset (Idri et al., 2015; Shepperd & Schofield, 1997) rather than the functions of the program itself. Many studies found size “to be an influential factor” (Idri et al., 2015), while other studies focused on variables that would only be available or predictable at the time estimates are needed (Shepperd & Schofield, 1997). Some example variables used include programming language, number of input or output message types, and number of files changed as part of an enhancement task (Shepperd & Schofield, 1997). Given that an estimator would not know the size of the program being estimated and many of the example variables could not be used to identify similar programs across organizations, case-based estimation could not be applied early in the lifecycle (criterion 1) or be generalizable (criterion 3).

Number of Applications

The Air Force Cost Analysis Agency (AFCAA) led a study to develop cost models for DevSecOps programs that yielded cost estimating equations that can estimate a program’s fiscal year costs based on the number of applications that would be in development in that year (Bradshaw, 2022). However, these cost models do not estimate the total required effort and cost of the applications to produce the needed capabilities. Therefore, this method is unable to



perform cost analysis and answer questions like how long it will take to develop a capability (criterion 4).

Function Points

Function Points represent the size of software based on its functional processes, which consist of inputs from users, reading from or writing to memory, and outputting results. Project stakeholders can define such functional processes early in the lifecycle and would provide objective sizing that is applicable across contractors and programs. With data, cost estimation models support cost analyses like SLOC-based cost estimation models. Function Points seem to be the most promising solution because they meet all the above criteria for cost estimates. However, building historical datasets and generalizable cost models applicable to the defense acquisition solution space would require years of data collection because sufficient data is not available at present.¹ Though the International Software Benchmarking Standards Group (ISBSG) provides software development data with Function Points, the dataset primarily represents the commercial industry. Additionally, since Function Points is based on functional processes, they do not account for nonfunctional requirements (e.g., security and reliability) or algorithmic complexity (Hira, 2020)—both of which can have significant effects on software development efforts.

Estimation Methods Summary

Table 1 summarizes the existing software cost estimation methods reviewed in this section and how they fail to meet the estimation criteria required to satisfy and be compatible with the CBP framework, the government’s budgeting process, and the SWP acquisition lifecycle. While Function Points meet all the criteria, sufficient data does not exist within the defense acquisition solution space to develop generalizable cost estimation models. The current state of the art prevents defense programs from fully implementing and benefitting from modern Agile and DevSecOps software development practices.

Table 12. Summary of Predominant Software Cost Estimation Methods Failing to Meet Estimation Criteria for CBP and SWP

Criteria	SLOC	Story Points	Analogy-Based Estimation	Case-Based Reasoning	AFCAA DevSecOps	Function Points
Early Lifecycle	No	No		No		
Defensible		No	No			
Generalizable		No		No		
Cost Analysis					No	
Data Availability						No

Proposed Solution: Capability-Based Software Cost Estimation (CaBSCE)

The CBP framework, the government’s budgeting process, and the SWP acquisition lifecycle pose a difficult set of criteria for cost estimation methods to fulfill. Of the existing software cost estimation methods, analogy-based estimation, case-based reasoning, and

¹ Publication Pending: Hira, A., & Kwok, B. (2024). (in press). What is the U.S. DoD cost estimation community saying about Agile? *Journal of Cost Analysis and Parametrics*, 12. When published, the journal will be accessible here: <https://www.iceaaonline.com/publications/#journal>



Function Points are closest to fulfilling all the cost-estimation criteria. The SEI proposes taking aspects of each of these methods and developing a software cost estimation method that can meet all the estimation criteria—a capability-based software cost estimation method (CaBSCE). To develop CaBSCE, the SEI will identify clusters or groups (analogy-based estimation) of similar software functions (Function Points) calibrated with historical data from comparable efforts (case-based reasoning). For example, GPS/navigation components, despite the device they are installed on, must generally perform three functions: (1) identify the current location, (2) identify the destination, and (3) determine how to get from the current location to the destination with the use of trigonometry and artificial intelligence (AI) algorithms. Such functions would differ from compilers or text parsers, which peruse text either by comparing it to pre-specified patterns or identifying patterns in the text. This method would describe two different capability clusters, or groups, of similar software functions. These capability clusters would be a level lower than capabilities identified in CNSs and SW-ICDs (in terms of detail), but stakeholders involved at program initiation could identify the software functions required and consider alternate solutions at this level of software functions. (See Figure 1 for an example of a capability need and how software functions can be identified to satisfy it.) The associated effort ranges for each capability cluster serve as evidence-based ROM estimates. Figure 2 summarizes the proposed research methodology, and Table 2 describes how CaBSCE would satisfy the cost estimation method criteria identified in the previous section.

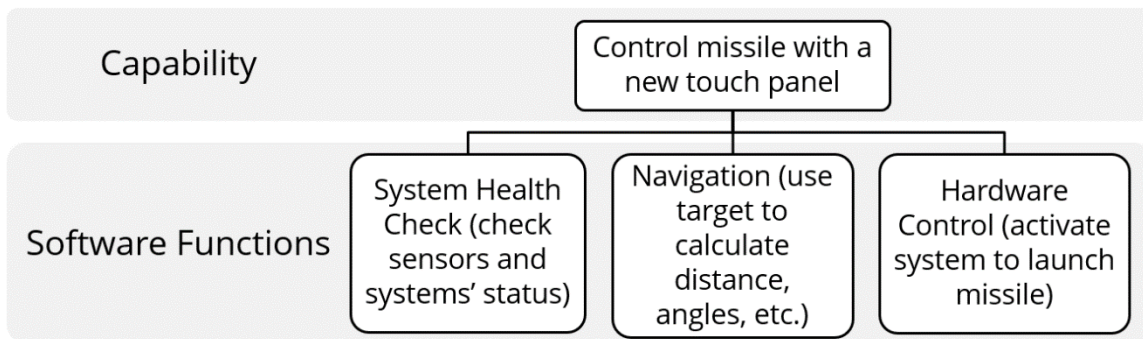


Figure 32. Example of a Capability Need and the Software Functions Required to Meet It

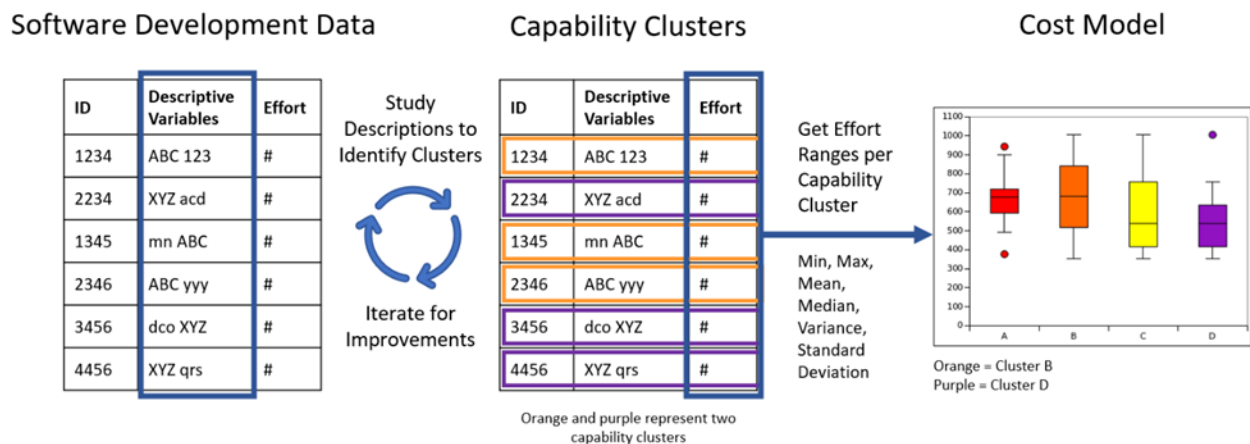


Figure 33. High-Level Research Methodology Used to Develop CaBSCE



Table 13. Description of How CaBSCE Meets the Cost Estimation Criteria

Criteria	How CaBSCE Meets the Criterion
Early Lifecycle	Though the exact solution or architectural and developmental requirements are not set yet, CaBSCE will be based on high-level software functions that can be identified within the CBP framework.
Evidence-Based	Using the clusters/groups of software functions, the effort ranges for each cluster/group will come from similar functions. The data points may represent different solutions, therefore providing an evidence-based yet capability-based estimate.
Generalizable	The datasets consist of software developed across many application types and organizations. Therefore, the resulting cost model would be generalizable across application types and organizations.
Cost Analysis	The goal is to get the required effort for completed software functions, which would allow cost estimators and program leadership to perform cost analysis.
Data Availability	The study will use existing software development datasets (ISBSG and other datasets), which provide total effort and descriptive variables for analysis.

The SEI will utilize existing software development datasets: the ISBSG and other datasets. The ISBSG dataset consists of thousands of projects that span several industries and business types and more than 32 countries. The dataset provides a few variables describing the organization type, industry, and application type. The SEI will analyze the descriptive information in the datasets to identify data points with similar software functions. (More details are provided in the next subsection.)

Classifying Software Functions

The Software Cost Estimation Methodologies section mentions several attempts to categorize data points to improve the ability to estimate in the software development lifecycle and improve estimation accuracy. Application types and domains are used to categorize software by the architecture/design typically followed (for example, client-server or database management). Many studies used application types and domains for clustering analysis (Van Hai et al., 2022) or as a categorical variable in cost estimation equations (Rosa et al., 2021). Coonce and Alleman (2017) hypothesize a standardized way of grouping the historical data on features and attributes that would support CBP and early lifecycle estimation for software. A drawback of using the application domains and types as similar and analogous data or projects is the inability to take advantage of new technology applied in another application domain or type. As an example, to estimate the first application of AI or machine learning (ML) in the satellite domain using analogy, estimators would use the most similar satellite program and add a complexity factor to account for the AI and ML application. While developing software for a satellite is evidence-based, the complexity factor or attempt to account for AI and ML would be subjective. Typically, estimators do not look outside their domain for historical data.

The SEI proposes identifying fundamental software functions that can span application domains and types. This way, expanding on the example of including AI and ML in a satellite program, estimators can pull from experience in other domains/types. This methodology can be simplified as doing a bottom-up, analogy-based estimation: Estimators and stakeholders would break down the fundamental software functions of capabilities and use analogous data for all identified functions to build up the total cost estimate. Boehm et al. (2000) developed a



comprehensive definition for software product complexity that consists of five types of operations: control, computation, device-dependent, data management, and user interface management (replicated in Table 3), which the SEI will use to identify clusters or groups of application types by software functions using the following steps:

1. Extract unique application types from the datasets.
2. Classify the unique application types, using descriptive variables as guidance, across the five software operation types and levels of complexity.
3. Identify application types with similar combinations of complexities through clustering analysis.
4. Name the clusters/groups to describe the major functions and identify common words associated with each function, using the text in the descriptive variables. Essentially, the SEI will develop a dictionary to define software functions.
5. Apply the dictionary on the datasets and extract the effort ranges for each software functions cluster, which provides comparable efforts for similar functions, serving as the evidence-based ROM estimates.

Table 14. Software Product Complexity Description from Software Cost Estimation Model COCOMO® II (Boehm et al., 2000)

	Control Operations	Computation Operations	Device-Dependent Operations	Data Management Operations	User Interface Management Operations
Very Low	Straight-line code with few non-nested structured programming operations	Evaluation of simple expressions; example: $A = B + C*(D-E)$	Simple read, write statements with simple formats	Simple arrays in main memory	Simple input forms; report generators
Low	Straightforward nesting of structured programming operators; mostly simple predicates	Evaluation of moderate-level expressions; example: $D = \text{SQRT}(B^2 - 4*A*C)$	No cognizance needed of particular processor or I/O device characteristics	Single file subsetting with no data structure changes, no edits, and no intermediate files	Use of simple GUI builders
Nominal	Mostly simple nesting; some inter-module controls; decision tables, simple callbacks, or message passing	Use of standard math and statistical routines; matrix/vector operations	I/O processing that includes device selection, status checking, and error processing	Multi-file input and single-file output; simple structural changes, simple edits	Simple use of widgets



	Control Operations	Computation Operations	Device-Dependent Operations	Data Management Operations	User Interface Management Operations
High	Highly nested structured programming operators with many compound predicates; queue and stack control	Basic numerical analysis	Operations at the physical I/O level; optimized I/O overlap	Simple triggers activated by data stream contents	Widget development and extension; voice I/O; multimedia
Very High	Reentrant and recursive coding; fixed-priority interrupt handling; task sync, complex callbacks	Difficult but structured numerical analysis	Routines for interrupt diagnosis, servicing, and masking; communication line handling	Distributed database coordination; complex triggers; search optimization	Moderately complex 2D/3D, dynamic graphics, multimedia
Extra High	Multiple resource scheduling	Difficult and unstructured numerical analysis	Device timing-dependent coding; micro-programmed operations	Highly coupled, dynamic relational and object structures	Complex multimedia, virtual reality, natural language

Figure 3 demonstrates examples of step 2 (classify the unique application types, using descriptive variables as guidance, across the five software operation types and levels of complexity) for GPS/navigation and compiler/text parser. As mentioned earlier, GPS/navigation components, despite the device they are installed on, must generally perform three functions: (1) identify the current location, (2) identify the destination, and (3) determine how to get from the current location to the destination with the use of trigonometry and AI algorithms. This understanding translates to the COCOMO® II product complexity description like this:

- Control Operations – High: AI algorithms to determine the best path to reach destination from current location will require nested programming, compound predicates, and queue and stack control.
- Computation Operations – High: calculating current and destination locations and paths between them requires numerical approximations and trigonometry.
- Device-Dependent Operations – Nominal: data to identify current location requires reading data from hardware (e.g., sensors).
- Data Management Operations – Nominal: data expressing current location, destination, and calculations for possible paths will require data structures.
- User Interface Management Operations – Very High: 2D/3D map displays for human-friendly operation.

On the other hand, compilers or text parsers peruse text either by comparing it to pre-specified patterns or identifying patterns in the text. This definition translates to the COCOMO® II product complexity descriptions as follows:



- Control Operations – High: comparing text to rules or identifying patterns in text requires nested programming, compound predicates, and queue and stack control.
- Computation Operations – Very Low: addition and subtraction may be needed to track patterns.
- Device-Dependent Operations – Very Low: simple reading and writing of inputs and outputs.
- Data Management Operations – Nominal: input and output data require data structures.
- User Interface Management Operations – Low: simple user interface is sufficient.

Step 3 then identifies application types with similar combinations of complexities through clustering analysis. For example, the software underlying scientific calculators may have similar complexity combinations as GPS/navigation. With this methodology, the SEI will define software functions with similar complexity that spans application domains.

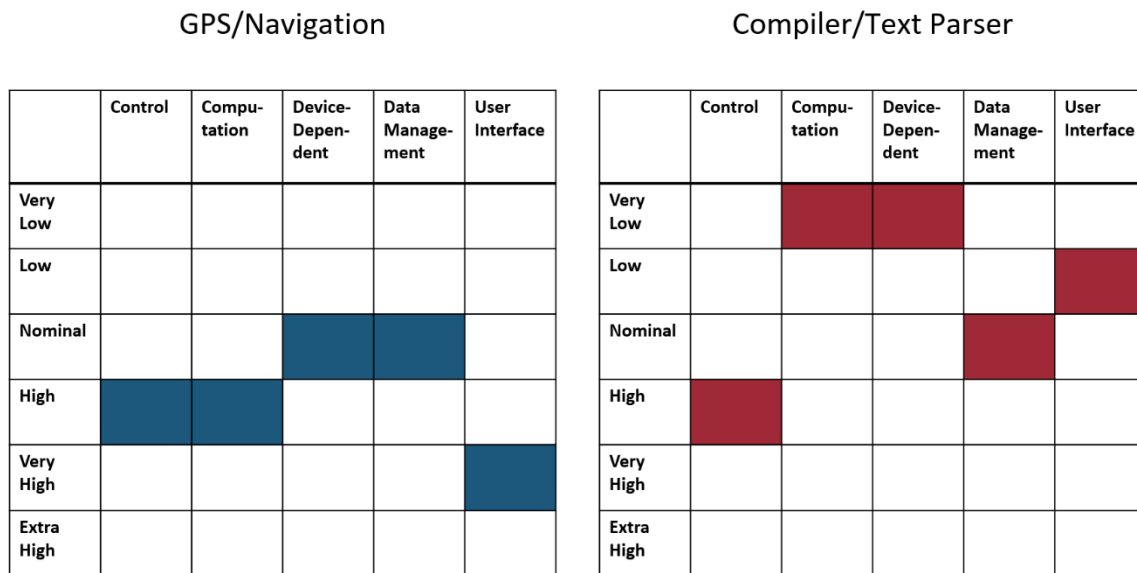


Figure 34. GPS/Navigation (Left) and Compiler/Text Parser (Right) Application Types Mapped to COCOMO® II's Software Product Complexity Description

Conclusion

Software's dynamic nature has led to significant technology improvements in an ever-evolving environment, which contributes to changing requirements throughout the software development lifecycle. Agile software development welcomes changing requirements to improve customer satisfaction and project success. While traditional government defense acquisition processes required early comprehension of the requirements for estimates and budgets, the release of the SWP brought significantly more flexibility. The SWP shortens the software acquisition lifecycle and is compatible with both Agile's and DevSecOps' incremental learning and progress. High-level capabilities replace detailed requirements before starting development. In CBP, it is important to avoid defining specific implementation solutions for capabilities so that there is space for innovation and the most cost-effective solution can be determined. Unfortunately, this change presents unique and challenging constraints on software cost

estimation needs. Existing software cost estimation methods depend on detailed requirements or do not fully utilize existing historical data.

The SEI proposes implementing CaBSCE, a software cost estimation method based on high-level descriptions of software functions and using comparable historical data for evidence-based ROM estimates. CaBSCE eliminates the weaknesses of existing software cost estimation methods while meeting the government's need for evidence-based, flexible, and defensible estimates. With CaBSCE, estimators will be able to identify analogous functions that map to the CNS or SW-ICD and use comparable efforts from historical data for defensible estimates that account for uncertainties. Programs will likely get less pushback from the government for budget requests, enabling efficient allocation of resources. The SEI will use existing software development datasets (e.g., ISBSG) and the software product complexity definition from the COCOMO® II software cost estimation model (Boehm et al., 2000) to define analogous software functions and the corresponding effort ranges. Basing the effort on software functions allows stakeholders to consider alternative solutions by adding or removing functions and pull comparable effort from various application types and domains.

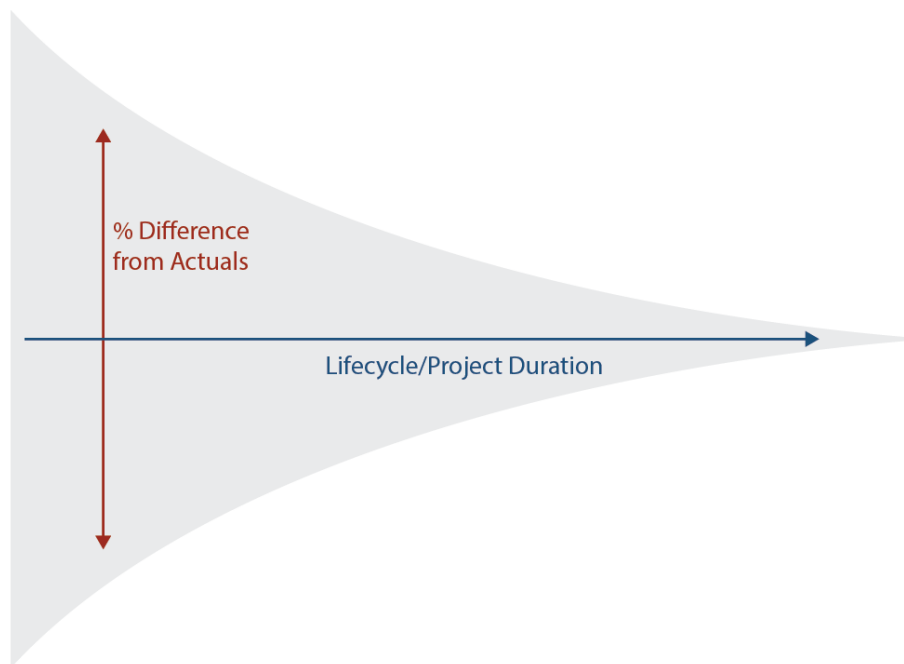


Figure 35. The Cone of Uncertainty

CaBSCE will not replace other software cost estimation methods. It will support early lifecycle estimates, when required capabilities are still being defined, and there is a substantial amount of architecture and implementation uncertainty. In 1981, Boehm developed the “*Cone of Uncertainty*” to demonstrate how uncertainty reduces over a project's lifecycle and leads to more accurate cost estimates. (See Figure 4 for a depiction of the Cone of Uncertainty, which was adapted from a graphic in Boehm's 1981 book.) CaBSCE provides an appropriate estimation method at the beginning of the lifecycle, while other estimation methods can be used as more information is learned and defined to refine the estimates. Additionally, other software cost estimation methods, (e.g., Function Points) can use the software functions grouping for more accurate estimates. Hira (2020) found that the relationship between Function Points and

effort had significantly different trends based on the underlying software functions of the projects. Hence, CaBSCE will not only fill a gap in the current software cost estimation methods but will also provide software functions grouping to improve the accuracy of existing methods.

Acknowledgments

The author thanks her colleagues at the SEI—Eileen Wrubel, Bill Nichols, Julie Cohen, Chris Miller, Sandy Shrum, Barbara White, Jeff Mellon, Nick Testa, and Shannon Gallagher—and at the U.S. Air Force—Jackson Pontsler and Joseph Florio—for supporting and reviewing this study, and Glen Alleman (Niwot Ridge Consulting) for assisting with the example of relating capability to software functions.

Document Markings

Copyright 2024 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. Requests for permission for non-licensed uses should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

DM24-0307

References

- Albrecht, A. J., & Gaffney, J. E. (1983). Software function, source lines of code, and development effort prediction: A software science validation. *IEEE Transactions on Software Engineering*, 9(6), 639–648. <https://doi.org/10.1109/TSE.1983.235271>
- Alleman, G. (2015). *Capabilities based planning*. Herding Cats. https://herdingcats.typepad.com/my_weblog/2015/07/capabilities-based-planning.html
- Alleman, G. (2019). *Capabilities based planning*. Herding Cats. https://herdingcats.typepad.com/my_weblog/2019/10/capabilities-based-planning.html
- Alleman, G. (2020a). *Compendium of resources for capabilities based planning*. Herding Cats. https://herdingcats.typepad.com/my_weblog/2020/12/compendium-of-resoruces-for-capabilities-based-planning.html
- Alleman, G. (2020b). *Eliciting capabilities*. Herding Cats. https://herdingcats.typepad.com/my_weblog/2020/03/eleciting-capabilities.html



- Anastasios, P. (2014). *Capability-based planning with TOGAF and ArchiMate* [Master's thesis, Business Information Technology School of Management and Governance].
https://essay.utwente.nl/65421/1/Papazoglou_MA_MB.pdf
- Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M. ... Thomas, D. (2001). *Principles behind the agile manifesto*. Agile Manifesto.
<https://agilemanifesto.org/principles.html>
- Bexfield, J., & Disbrow, L. (2004). Capabilities based planning: The road ahead. *Military Operations Research Society (MORS) Workshop*. Military Operations Research Society (MORS).
<https://apps.dtic.mil/dtic/tr/fulltext/u2/a443067.pdf>
- Biltgen, P. T. (2007). *A methodology for capability-based technology evaluation for system-of-systems* [Doctoral dissertation, Georgia Institute of Technology].
<https://repository.gatech.edu/entities/publication/5ddd79af-ff47-4461-aeeb-33f04bf72dfb>
- Boehm, B. W. (1981). *Software engineering economics*. Prentice-Hall.
- Boehm, B. W., Abts, C., Brown, A. W., Chulani, S., Clark, B. K., Horowitz, E., Madachy, R., Reifer, D., & Steece, B. (2000). *Software cost estimation with Cocomo II*. Prentice Hall.
- Bradshaw, K. (2022). History repeats itself: A new revolution of factories... software! *Joint IT and Software Cost Forum*. https://www.dhs.gov/sites/default/files/2023-01/History_Repeats_Itself__A_New_Revolution_of_Factories..._Software_Remediated.pdf
- Chairman of the Staff. (2009). *Capabilities-based assessment (CBA) user's guide (Version 3): Force structure, resources, and assessments directorate (JCS J-8)*. CreateSpace Independent Publishing Platform. <https://acqnotes.com/wp-content/uploads/2014/09/Capabilities-Based-Assessment-CBA-Users-Guide-version-3.pdf>
- Chim, L., Nunes-Vaz, R., & Prandolini, R. (2010). Capability-based planning for Australia's national security. *Security Challenges*, 6(3), 79–96. <https://www.jstor.org/stable/26459800>
- Cohn, M. (2004). *User stories applied: For agile software development*. Addison-Wesley Professional.
- Committee on Naval Analytical Capabilities and Improving Capabilities-Based Planning. (2005). *Naval analytical capabilities: Improving capabilities-based planning*. The National Academies Press.
<https://doi.org/10.17226/11455>
- Coonce, T., & Alleman, G. (2017). How should we estimate agile software development projects and what data do we need? *International Cost Estimating and Analysis Association (ICEAA) Professional Development & Training Workshop*. <https://www.iceaaonline.com/wp-content/uploads/2017/07/AG01-Paper-Coonce-How-Should-we-Estimate-Agile-Software.pdf>
- Davis, P. K. (2002). *Analytic architecture for capabilities-based planning, mission-system analysis, and transformation*. RAND Corporation. <https://studylib.net/doc/13665293/how-can-a-structured-representation-of-capabilities-help-...>
- Davis, P. K., Shaver, R. D., & Beck, J. (2008). *Portfolio-analysis methods for assessing capability options*. RAND Corporation. <https://www.rand.org/pubs/monographs/MG662.html>
- De Lucia, A., & Qusef, A. (2010). Requirements engineering in agile software development. *Journal of Emerging Technologies in Web Intelligence*, 2(3), 212–220.
- Defense Acquisition University. (2020a, October 2). *Cost estimation*. <https://aaf.dau.edu/aaf/software/cost-estimation/>
- Defense Acquisition University. (2020b, January 23). *Software acquisition*. <https://aaf.dau.edu/aaf/software/>
- Defense Innovation Board. (2019). *Software is never done: Refactoring the acquisition code for competitive advantage*. DoD. <https://media.defense.gov/2019/May/01/2002126690/-1/-1/0/SWAP%20EXECUTIVE%20SUMMARY.PDF>



- Defense Science Board. (2009). *Report of the Defense Science Board Task Force on Department of Defense policies and procedures for acquisition of information technology*. DoD.
- Desgagné, C. R. (2009). *Evolutionary acquisition – A complementary approach to capability based planning for the delivering of aerospace power* [Master of Defence Studies Research Project, Canadian Forces Command and Staff College].
<https://www.cfc.forces.gc.ca/259/290/295/286/desgagne.pdf>
- DoD. (2021, March). *DoD enterprise DevSecOps fundamentals*.
<https://dodcio.defense.gov/Portals/0/Documents/Library/DoDEnterpriseDevSecOpsFundamentals.pdf>
- GAO. (2020). *Cost estimating and assessment guide: Best practices for developing and managing program costs*. <https://www.gao.gov/assets/gao-20-195g.pdf>
- Hales, D., & Chouinard, P. (2011). *Implementing capability based planning within the public safety and security sector: Lessons from the defence experience* [Technical Memorandum].
<https://apps.dtic.mil/dtic/tr/fulltext/u2/a555463.pdf>
- Hanley, J. T., Jr., Fitzsimmons, M. F., Kurtz, J. H., Roark, L. M., Roske, V. P., Jr., & Cuda, D. L. (2006). *Improving integration of Department of Defense processes for capabilities development planning*. Institute for Defense Analyses. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a460395.pdf>
- Highsmith, J. A. (2002). *Agile software development ecosystems*. Addison-Wesley Professional.
- Hira, A. (2020). *Calibrating COCOMO(R) II for functional size metrics* [Doctoral dissertation, University of Southern California]. <https://digitalibrary.usc.edu/Share/hpp54fd5608060ex2w83q2het0p7dxeb>
- Iacobucci, J. V. (2012). *Rapid architecture alternative modeling (RAAM): A framework for capability-based analysis of system of systems architectures* [Doctoral dissertation, Georgia Institute of Technology]. <https://repository.gatech.edu/server/api/core/bitstreams/e948f95b-591a-48a8-b47d-3a9cc6bb23ca/content>
- Idri, A., Amazal, F. A., & Abran, A. (2015). Analogy-based software development effort estimation: A systematic mapping and review. *Information and Software Technology*, 58, 206–230.
<https://doi.org/10.1016/j.infsof.2014.07.013>
- Jørgensen, M. (2004). A review of studies on expert estimation of software development effort. *Journal of Systems and Software*, 70(1–2), 27–48.
- Kemerer, C. F. (1987). An empirical validation of software cost estimation models. *Communications of the ACM*, 30(5), 416–429.
- Kossakowski, P. (2005). *Capabilities-based planning: A methodology for deciphering commander's intent*. McLean. http://www.dodccrp.org/events/10th_ICCRTS/CD/papers/319.pdf
- Milani, M., & Rossi, M. (2020). Agile management versus budget control: Learn how to win both with Slick-Farm. *6th European Lean Educator Conference (ELEC 2019)*, 395–405. Springer.
- Neaga, I., Henshaw, M., & Yue, Y. (2009). *The influence of the concept of capability-based management on the development of the systems engineering discipline*. Loughborough University.
https://repository.lboro.ac.uk/articles/conference_contribution/The_influence_of_the_concept_of_capability-based_management_on_the_development_of_the_systems_engineering_discipline/9559580
- Office of the Under Secretary of Defense for Acquisition and Sustainment. (2020). *Operation of the software acquisition pathway* (DoD Instruction 5000.87). DoD.
https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500087p.PDF?ver=virAfQj4v_LgN1JxpB_dpA%3D%3D
- The Open Group Architecture Forum. (2018). Capability-based planning. In *The TOGAF® Standard* (Ver. 9.2). <https://pubs.opengroup.org/architecture/togaf91-doc/arch/chap32.html>



- Ozkaya, I., Kazman, R., & Klein, M. (2008). *Understanding the relationship of cost, benefit, and architecture*. Carnegie Mellon University, Software Engineering Institute.
- Planeaux, J. B. (2003). *Beyond the task force conops: The path to a capabilities-based modernization framework for the Air Force* (Research Report). Air War College, Air University, Maxwell Air Force Base. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a424623.pdf>
- Raza, S., & Waheed, U. (2018). Managing change in agile software development a comparative study. *IEEE 21st International Multi-Topic Conference (INMIC)*. IEEE.
- Rosa, W., Clark, B. K., Madachy, R., & Boehm, B. W. (2021). Empirical effort and schedule estimation models for agile processes in the US DoD. *IEEE Transactions on Software Engineering*, 48(8), 3117–3130.
- Shah, T., & Patel, S. V. (2014). A review of requirement engineering issues and challenges in various software development methods. *International Journal of Computer Applications*, 99(15), 36–45.
- Shepperd, M., & Schofield, C. (1997). Estimating software project effort using analogies. *IEEE Transactions on Software Engineering*, 23(11), 736–743.
- Steele, J. A. (2021). Capability-based planning and the Royal Canadian Air Force. In *RCAF Defence Economics* (pp. 77–131). Canadian Forces Aerospace Warfare Centre. https://cradpdf.drdc-rddc.gc.ca/PDFS/unc357/p812921_A1b.pdf
- Taliaferro, A. C., Gonzalez, L. M., Tillman, M., Ghosh, P., Clarke, P., & Hinkle, W. (2019). *Defense governance and management: Improving the defense management capabilities of foreign defense institutions: A guide to capability-based planning (CBP)*. Institute for Defense Analyses. https://www.ida.org/research-and-publications/publications/all/d/de/defense-governance-and-management_improving-the-defense-management-capabilities-of-foreign
- Technical Cooperation Program. (2004). *Guide to capability-based planning*. DoD. <https://www.hsdl.org/?view&did=461818>
- Titus, N. (2004, March). *Air Force CONOPS & capabilities based planning*. U.S. Air Force. <https://www.hsdl.org/?view&did=453278>
- Van Hai, V., Nhung, H. L., Prokopova, Z., Silhavy, R., & Silhavy, P. (2022). Toward improving the efficiency of software development effort estimation via clustering analysis. *IEEE Access*, 10, 83249–83264.
- Walker, S. K. (2005). *Capabilities-based planning – How it is intended to work and challenges to its successful implementation*. U.S. Army War College. <https://apps.dtic.mil/sti/pdfs/ADA434864.pdf>
- Webb, M. (2008). Capabilities-based engineering analysis (CBEA). In A. Minai, D. Braha, & Y. Bar-Yam (Eds.), *Unifying Themes in Complex Systems: Proceedings of the Sixth International Conference on Complex Systems* (pp. 540–547). https://doi.org/10.1007/978-3-540-85081-6_67



Acquired and Deployed but Not Adopted: Lost Mission Effectiveness Without Resilient Chat Afloat

Christopher B. Landis—is an Active Duty Commander in the U.S. Navy serving as a computer science PhD candidate at the Naval Postgraduate School. Upon graduating, he will report to the U.S. Naval Academy Cyber Science Department as a Permanent Military Professor. His education includes a BS in information technology with a second discipline in space operations from the U.S. Naval Academy and a Master of Information Technology Strategy from Carnegie Mellon University. He has served as a Surface Warfare Officer and Information Professional Officer, including three ship's company tours on a Destroyer, Littoral Combat Ship, and Aircraft Carrier. [clandis@alumni.cmu.edu]

Joshua A. Kroll—is an Assistant Professor of computer science at the Naval Postgraduate School. His research interests include trustworthy AI, technology governance, systems engineering, privacy, security, accountability, ethics, computer science and law, human factors, formal methods, data science, bias and fairness, dependability, and assurance. His education includes an AB in physics and mathematics from Harvard College and an MA and PhD in computer science from Princeton University. He has worked in both industry and academia. [jkroll@nps.edu]

Abstract

The U.S. Navy increasingly emphasizes communications resilience in distributed maritime operations. In the face of communications degradation and denial, we can improve warfighter effectiveness even using current systems when they are underused. By developing better ways to use deployed systems and applying lessons learned to new systems, we can maximize the value of system requirements and adoption of future acquisitions. Through our work on Navy communications systems' configurations, we find that some resilient systems go underused in practice, despite Navy requirements for system resilience designed into deployed systems. The Navy depends on Internet Protocol networks for conveying command and control (C2) communications. We examine the Navy's email and chat use for conveying C2 communications. We survey (n = 69) command, control, communications, and computer (C4) leadership to inform a sociotechnical analysis of how Sailors afloat use chat, considering a distributed chat architecture's resilience benefits. To ensure that acquired technologies do not go underutilized, our research results lead us to conclude that solutions must be sociotechnical: better technology alone does not solve the problem of resilient communications. Without understanding the operating environment, including operators' and their leadership's motivations, new technology solutions can go underused, limiting the anticipated gain in mission effectiveness.

Keywords: IT adoption, afloat tactical networks, chat, failure transparency, command and control communications, social computing

Introduction

Communications are critical to modern C2 in the U.S. Navy. Although afloat Internet Protocol (IP) networks provide much of the communications paths enabling C2 afloat, the Navy underuses these deployed IP systems and does not configure them to maximize their robustness in communications-degraded/denied environments (CD2Es). We offer a novel interdisciplinary approach to investigating how to overcome the organizational and technical hurdles in improving resilience in deployed capabilities. If we fail to understand why deployed



systems go underutilized, we risk falling short in the same ways when investing in any future capabilities.

The U.S. Navy multiplexes a ship's multiple networks bidirectionally over multiple satellite communications (SATCOM) paths, interconnecting ships and fleet network operations centers (NOCs) ashore, as shown in Figure 1. An underused benefit of this architecture is its ability to interconnect ships without a shore facility (Landis, 2016). Combined with ship-hosted network services, any group of ships that can exchange IP traffic can use each other's services. However, the standard practice is to route all IP traffic via a NOC ashore, making the NOC a single point of failure. Why does the Navy not use existing shoreless capabilities to increase the resilience of applications conveying C2 communications?

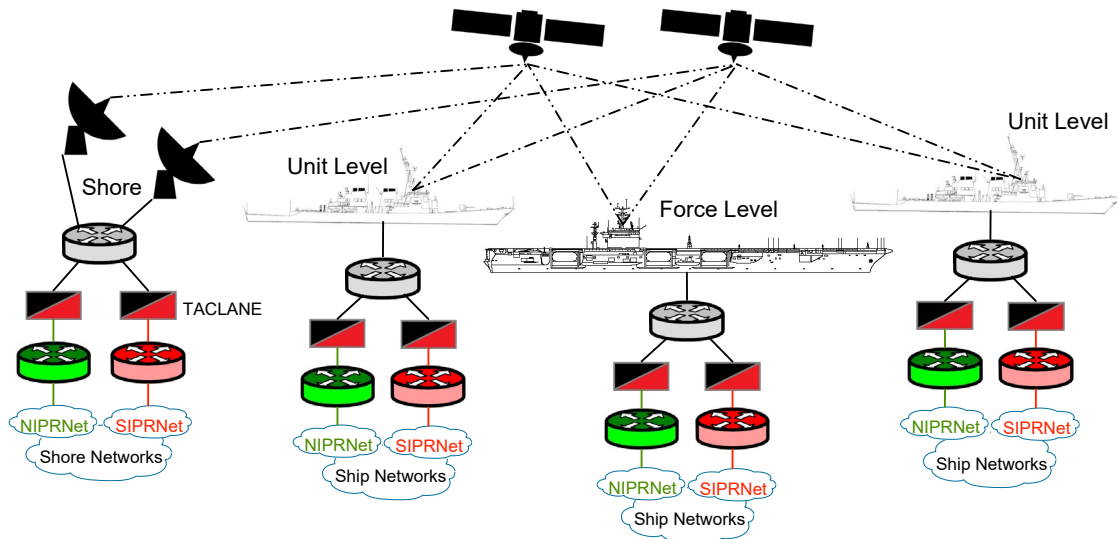


Figure 1. Conceptual Overview of Fleet Connectivity.

Note: The Navy primarily conveys unclassified communications over the Non-secure Internet Protocol Router Network (NIPRNet) and C2 communications over the Secure Internet Protocol Router Network (SIPRNet).

To answer this question, we investigate two IP network applications commonly used for C2 communications, email and chat, including their dependency on the Domain Name System (DNS). While we find technical reconfigurations may suffice for improving the resilience of email and DNS, such technical reconfigurations are insufficient for chat. We administer a survey of current and former afloat command, control, communications, and computers (C4) leadership, including embarked staffs, as part of our analysis of chat use. The survey data reveal that the most significant hurdles to using the afloat chat server include awareness of its existence, capabilities, benefits, and how best to use it. By combining our technical analysis with the results of our survey, we establish a base from which to derive interventions for improving Sailors' use of these systems. These interventions include caching all ships' DNS records on all ships, configuring fail-over or fault tolerance into ships' Exchange servers, and circulating a fleet commander-level championed how-to guide for operating a distributed chat architecture. We recommend circulating the guide as a means to overcome the structural obstacles that we find in the survey data for operating a distributed chat architecture. Enabling these applications to interconnect in a shoreless or otherwise CD2E and empowering Sailors to use these capabilities further the concept of assured C2 in distributed maritime operations, thus improving warfighting effectiveness.

Resilience relates to the concept of *failover transparency*, that the network status and configuration is transparent to the applications depending on it for connectivity (i.e., the

application's logic need not depend on the network status in order for the application's function to survive degradation or failure at the network level). For instance, for chat, users' chat clients continuing to function despite their ship losing its connectivity with shore stations. Obviously, chat communications will be unavailable between ship and shore; but the point is that users aboard the ship will continue to stay synchronized with one another, retaining the ability to exchange messages despite the loss of shore connectivity.

We review both the social and technical contexts in the Stuck in One's Ways and Technical Context sections before addressing the data involved in our analysis in the Fleet Perception of Shoreless Chat section. We explain devised interventions in the Interventions and System Developments section and conclude our remarks in the Conclusion.

Stuck in One's Ways

Understanding why a technology is not being used to its fullest capabilities requires examining both the social context and technical aspects of the systems. In improving failover transparency in a shoreless environment for C2 resilience, the challenge is more than Sailors' capabilities on one ship. Instead, the configurations' complexity drives their implementation to the level of inter-program of record (POR) cooperation. As one person at this level, a former Naval Information Warfare Systems Command (NAVWAR) commander declared, "We have got to . . . turn CANES [(Consolidated Afloat Network and Enterprise Services)] into the information warfighting platform" (Machi, 2018). But technology is only part of the solution. Even when PORs cooperatively provision all the technical aspects well, people still must operate the system proficiently for it to yield its benefit to the Navy's mission.

When introducing the Naval Tactical Data System (NTDS), the Navy trained Sailors to be the system experts rather than relying on vendor technical support (Boslaugh, 1999, p. 254). This empowered Sailors to cross the boundaries between the multiple vendors supplying NTDS components. Today, being system-of-systems experts is important for Sailors to maximize their use of deployed information systems (ISs). Improving mission assurance through well-prioritized traffic flows of afloat network applications from various PORs is best achieved when including user behavior in the analysis (Rambo, 2016). Our work builds on this toward achieving greater resilience of a high priority traffic class, C2 communications. We find that with configuration refinements to afloat DNS and email and overcoming structural obstacles, the fleet will be in a better posture to assure its C2 communications paths.

A status quo bias emerges when uncertainty in the outcome of a policy change exists (Fernandez & Rodrik, 1991). This behavioral principle also applies to cybersecurity in which policy makers and practitioners alike are uncertain of the outcome of a change in policy because of the complexity of the cyber domain or a lack of understanding in how the policy change will affect how systems react at the technical and operational levels. Like information technology (IT) practitioners, builders—Sailors can be either of these—are in the class of trained professionals that often take pride in their work. Researchers studying hindrances to the adoption of innovative and sustainable technologies in the building industry discovered that psychological factors significantly affect companies' policy decision-making (Hofman et al., 2022). A lack of information transparency functions as a barrier to adopting new methods. When builders lack the information available to them that explains why a policy change or new technique is being introduced, they are more likely to stick to their traditions. If the explanatory information is available but difficult to find, the inconvenience of searching for and processing the information exacerbates their status quo bias. In cases like this, information transparency serves as a counterbalance to builders' resistance to change, often (if well justified) persuading builders even to embrace the new policy or technique (Hofman et al., 2022). Our survey data



indicate that Sailors in the role of IT practitioners and users lack the prerequisite information about operating distributed chat architectures and their benefits. This deficit has led to their IT nonuse, corroborating this finding from the building industry as relevant to adopting IT.

Another question is whether a new system should adapt to existing processes or the processes should adapt to the new system. In neglecting to adapt its C2 communications to the full capabilities of a distributed chat architecture, the Navy has lingered in merely using basic chat functions, costing the Navy in the resilience that it could have been enjoying (Eovito, 2006). The Navy must adjust the technology's *metastructure*. Orlikowski et al. (1995) define metastructuring as:

an empirically-grounded framing of the influential actions taken by individuals when they deliberately adapt computer-mediated communication technologies and their use to particular contexts and change those contexts to accommodate use of the technology. (p. 424)

Metastructuring can serve as an explanatory model in post-facto analysis or as an organization's approach to influencing its members' IT adoption. For example, in adopting a new enterprise resource planning (ERP) tool, Apple first tried to adapt the ERP tool to its existing workflows. When Apple Chief Executive Officer (CEO) Steve Jobs heard that this effort had become too expensive and cumbersome, he took a risk, directing Apple to cut its losses and start over, this time adapting its workflows to the new ERP tool, anticipating correctly that Apple would find greater success (Wipfler, 2023). An organization's metastructuring approach to adopting new IT affects the organization's success, including how clearly it communicates its goals, requirements, and risk tolerance (*reinforcement*) and incorporates new rules and procedures in response to user feedback as minor (*adjustment*) or major changes (*episodic change*; Orlikowski et al., 1995). Our analysis indicates that a lack of such organizational metastructuring has affected Sailors' nonuse of deployed systems.

Technical Context

The current ability of DNS, email, and chat to failover transparently when shore facilities are unreachable depends on having IP connectivity and the right configuration in place before the network interruption. Any multiplexed IP path between ships can contribute to the afloat network architecture shown in Figure 1, including a Ciphertext Time-division multiple access (TDMA) Interface Processor (CT-TIP) supernet, Battle Force Tactical Network (BFTN), and commercial proliferated low-Earth orbit (pLEO) networks. Transparent failover of these applications in a CD2E furthers the concept of distributed maritime operations. We address DNS, email, and chat in turn.

DNS Afloat

For users on one ship to connect directly with other ships' applications, their terminals must resolve the other ships' services' names to IP addresses. For example, remembering to connect to chat.c3f.navy.smil.mil is easier than remembering 205.3.33.20.

¹ DNS is the service on which so much of the Internet depends. When connected to the NOC ashore, ships' DNS servers recursively query their servicing NOC's DNS server for any off-ship resource. Caching other ships' DNS records locally overcomes an unreachable shore but is

¹ This IP address is fabricated for illustration only.



not the current configuration. The fleet-wide DNS records are only in the four fleet NOCs.² To circulate and dynamically update fleet DNS records, all reachable services must be registered in the servicing NOC's DNS server and the server must allow zone transfers to ships. When ships are connected with their servicing NOC, ships' DNS servers should request zone transfers from the NOC periodically and automatically so that ships can operate without connectivity ashore. In turn, afloat DNS can support other applications without connectivity ashore.

Instead of caching all ships' DNS records on all ships, another option is to extend DNS with Service Location Protocol (SLP) adapted for ad hoc networks (Koubaa & Fleury, 2001). This would be more dynamic for a subset of interconnected ships; however, SLP incurs a delay in discovering services before applications can connect. Also, the Navy's strict configuration management generates an a priori knowledge of ship configurations that is simpler to implement manually by the POR upon system installation than by dynamic service discovery.

Email Afloat

The Navy conveys non-real-time C2 communications among commanders and staffs via email. Ships' email servers deliver all non-local email to their servicing NOC's email relay server (Landis, 2015). Once fleet-wide DNS records are available aboard each ship, improving email resilience becomes a configuration change. Enabling shoreless email requires reconfiguring ships' email servers to deliver directly to other ships when the NOC is unreachable. The Navy could improve transparent email delivery failover by one of two methods, of which both depend on having other ships' DNS mail exchange (MX) records: via multiple MX record preference levels or Exchange send connector costs.

MX Record Preference Levels

Using the basic fail-over capabilities of DNS, each ship's DNS records could consist of a lower-numbered preference (primary) MX record for the NOC and a higher-numbered preference (alternate) for its own email server, as listed in Table 1. Reportedly, 44% of about 2 million popular email-receiving domains use MX balancing or fail-over (Ruohonen, 2020). As a limitation to this approach, though, Ruohonen notes that this fail-over mechanism's lack of formal specification can yield differences in interpretation and behavior (Ruohonen, 2020). Notwithstanding, following strict configuration management—like the Navy's—can enable consistent results.

Table 1. DNS MX Record Fail-over Configuration; Lower Numbers Indicate Stronger Preference

Name	Type	Preference	Exchange
ship.navy.smil.mil	MX	10	mail.noc.navy.smil.mil
ship.navy.smil.mil	MX	20	mail.ship.navy.smil.mil

Exchange Send Connector Costs

Using the Exchange connector "cost" attribute, we can provide "fault tolerance" in email delivery (Ashalyengar21 et al., 2023). The principle behind these cost-adjusted send connectors, listed in Table 2, are these two rules:

1. If a ship's email server can establish a Simple Mail Transfer Protocol (SMTP) connection with its servicing NOC, then deliver the email there for forwarding.

² For security and better management of ships' bandwidth utilization, the Navy implements a *split horizon* DNS configuration such that all ship records resolve to the NOC when queried from outside the fleet boundary. As such, NOCs' DNS servers serve as ships' start of authority (SOA).



2. Else, try establishing an SMTP connection with the destination email server, identified by its MX record.

Whereas rule #1 is the only rule currently implemented (Landis, 2015), adding rule #2 would enable shoreless email delivery. Because ships' DNS servers would contain records only for the NOC and other ships, emails delivered by rule #2 would necessarily be destined for other ships. Other email would queue as normal until connectivity with the NOC ashore is reestablished.

Table 2. Fault Tolerant Exchange Connector Configuration

Address Space	Cost	Route Through or According To
*	10	NOC Email Relay Server
*.navy.smil.mil	10	NOC Email Relay Server
*.navy.smil.mil	20	MX Record for Recipient Domain

Note: Exchange will select the available connector with most specific address space with the lowest cost (Ashalyengar21 et al., 2023).

Chat Afloat

The Navy routinely conveys real-time C2 communications between tactical watch stations by chat. As such, the Navy has long required its chat capabilities between ships and shore stations to be resilient to CD2Es (Martin & Marcley, 2013). Eovito (2006) describes how a recommendation coming out of the U.S. Navy exercise Trident Warrior '04 to use distributed chat architectures subsequently became a Navy requirement and was tested during Trident Warrior '05. In this configuration, all afloat users connect their chat clients to their ship's chat server, whose channels are bridged with the servers on other ships or in the fleet maritime operations center (MOC) ashore, according to whoever owns the channel (i.e., fleet commander, strike group commander, etc.). When the ship loses connectivity with the shore, all shipboard users stay connected in the chat channels locally and can continue communicating with each other. Upon reconnecting with the shore, the channels automatically resynchronize the last pre-configured number of hours of history. Thus, a distributed chat architecture provides resilience for an unplanned, transient shoreless environment.

Curiously, Sailors do not use afloat chat servers in distributed chat architectures. Instead, users typically connect their chat clients only to servers ashore. This configuration is intolerant of network interruptions (i.e., all clients drop upon losing connectivity with the shore) and incurs additional bandwidth demand on SATCOM links with many afloat clients connecting ashore instead of a single server. Even if system administrators bridge afloat chat servers with off-ship servers, users still might not configure their chat clients to use their ship's chat server. Because enabling shoreless chat by a distributed architecture depends largely on afloat chat server administrators and end users' client configurations, we must understand them better, for which we offer our survey results in the Fleet Perception of Shoreless Chat section.

Our gap analysis yields different results depending on whether we frame our analysis within the Defense Acquisition Management System existing when the Navy acquired this software or today's Software Acquisition Pathway in the Defense Acquisition System's Adaptive Acquisition Framework. In the former system, the gap that we observe here occurs once the acquired chat capability reaches the operations and sustainment phase (Blanchette et al., 2010). The requirement owner (i.e., sponsor) did not reinforce the metastructure, that is, insufficiently championed the new capability and its benefits to fleet commanders such that Sailors do not use it in a distributed chat architecture. In today's terms, the Software Acquisition Pathway prompts sponsors and program managers to interact with users continuously. Specifically, it directs the sponsor to assess annually "whether the mission improvements or



efficiencies realized from the delivered software capabilities are timely and worth the investment” (DoD, 2020, p. 18). If followed, this feedback from end users to the sponsor and from the sponsor to the POR should prompt the fleet engagement necessary to expand the software’s use because it reinforces and adjusts the chat system’s metastructure (Orlikowski et al., 1995).

All ships had an Internet Relay Chat (IRC) server that can be configured to participate in a distributed chat architecture until April 2021 (*Mako 2.0: Administrator Guide*, 2021; *Mako 2.0: User Guide*, 2021) when an update of a third-party dependency broke the chat federation capability (D. H. Anunciado, personal communication, October 3, 2023). If anyone had been using the federation capability to effect a distributed chat architecture when that third-party’s software updated, they certainly would have noticed its unexpected failure. Any complaints about losing the chat federation capability did not amount enough protest to demand an immediate replacement to provide that capability, which indicates its nonuse. However, the POR is working on testing and fielding solutions, including in the next version of ships’ network services suite and with other chat software to integrate with newer communications capabilities. Any how-to configuration guide must therefore apply to newer chat software solutions.

Fleet Perception of Shoreless Chat

Of all the IP-conveyed C2 communications, tactical watch standers afloat most prevalently use chat. As described in the Chat Afloat section, using afloat chat servers to improve chat resiliently is not the norm. To discover why there seems to be little afloat chat server use, we survey current and former afloat C4 leadership, including embarked staffs (survey instrument detailed in the Appendix).³ The responses help us discover the organizational and technical factors hindering Sailors from using their ships’ chat servers. We find that Sailors are unaware of the server’s existence, capabilities, or benefits (as shown in Table 3) and conclude that greater information transparency could support the metastructuring of chat server configuration to help increase its resilient use. For example, a fleet commander-level awareness campaign would be informative and prompt greater use, improving resilience for C2 communications conveyed by chat.

Table 3. Reasons for Not Using Afloat Chat Servers, Stratified by Those that Had the Chat Federation Capability

Reason	Population n ⇒ %	Capable n
Unaware of Server’s Existence	20/64 31%	14/20
Aware of the Server’s Existence but Do Not Use It	32/43 74%	23/32
Unaware of Channel Bridging Capability	15/31 48%	10/15
Unaware of the Benefits of Its Use	11/31 35%	8/11
No Time to Receive or Provide Training on Its Use	9/31 29%	7/9
Dealing with Too Many Other Network Problems	7/31 23%	6/7
I Don’t Know	5/31 16%	2/5
Unaware of How to Connect	4/31 13%	3/4
No Operational Need (in Shipyard)	1/31 3%	1/1

³ We presented some preliminary results at the Institute of Electrical and Electronics Engineers (IEEE) 2023 Conference on Military Communications (MILCOM; Landis, 2023).



Information Professional (IP) officers fill most C4 leadership billets so we solicited survey responses from them. With installations of the current afloat network server suite architecture starting in 2013, the population for this survey is about 510.⁴ We collected 69 survey responses from those that have served on a ship at the force level (52%) or unit level (48%, $n = 63$).⁵ This sample size yields a 9.2% margin of error (± 6 of 69) at a 90% confidence level. Most respondents (70%) started their afloat tour before the POR stopped supporting the chat federation capability so they would have had the opportunity to use it so we stratify the survey data on this attribute.

Sailors Are Not Using Afloat Chat Servers

The reasons for not using the afloat chat server vary, as shown in Table 3, but the most significant hurdles include awareness of its existence, capabilities, and benefits, indicating a lack of support for reinforcing the metastructuring of resilience features into accepted technical configurations and work practices. About 30% (20, $n = 64$) are unaware of afloat chat servers. Of those in C4 leadership positions that are aware of afloat chat servers' existence, about three-fourths ($n = 43$) do not use them at all. They are unaware of the benefits of their use and untrained on their capabilities. For example, four Combat Systems Information officers (CSIOs) and a command, control, communications, computers, and intelligence (C4I) officer (P7, P34, P42, P45, and P55) report an apparent lack of benefit, one not knowing about the bridging capability. "I wouldn't want to add another chat capability separate of the existing C2 chat rooms used by FLTCDRs [(fleet commanders)] and units. I'm not sure what purposes this chat server will serve" (P7). "Even when I educated staff members about it, they wouldn't use it. The customers didn't know of bridging chat rooms to site [sic] that as a requirement nor did any of the ITs onboard know how to do that" (P34). As explained in the Chat Afloat section, a distributed chat architecture would improve the resilience of real-time tactical C2 communications. For every nonuse reason, except "I don't know," between 67% and 100% of participants had the chat federation capability, meaning they could have used it.

Responses to open-ended questions in our survey instrument reveal qualitative narratives for the sources of nonuse. Two participants (P20 and P29) ascribe Sailors' nonuse of afloat chat servers to requirements and culture.

Requirements

Chat requirements appear to have endured without a champion because the Navy treats the chat function more like a feature of existing programs rather than its own POR.

P29 claims that Sailors do not fully understand the *operational* capabilities of ships' networks, like the chat server, because no formal introduction to them exists. To make the point, P29 uses the example of the SharePoint server, which is online upon completion of the network installation but not configured or its operation introduced to the Sailors responsible for maintaining the network and its servers. The SharePoint home page not being unique for the ship does not prevent the server's use. But to get the most out of the tool in its operational context, Sailors must configure the system relative to that context. P45 concurs: "Likely wouldn't have used it, as it would have likely not been configured and Sailors are not adequately trained to configure these servers without guidance." This is akin to the lack of information transparency described in the Stuck in One's Ways section in that Sailors do not have the requisite information to configure and operate chat afloat resiliently. Further, this finding echoes a

⁴ Assuming a 10-year linear growth in installations to 170 ships, synchronized biennial officer transfers, and an average of 1 IP officer per ship, some ships having none and others several. With 6.25% annual attrition, we estimate a population size of approximately 435 of the 510 officers are available to survey in 2023.

⁵ The n value for each question varies due to branching and each question's voluntary nature, enabling participants to skip some questions and quit the survey at any time.



Government Accountability Office (GAO) finding involving IT nonuse in the U.S. Immigration and Naturalization Service (INS) in which the report blamed a lack of training (GAO, 1988).

One need not look far for the root of this. All the Navy training courses for afloat computer network administrators listed in the Catalog of Navy Training Courses (CANTRAC)⁶ share the goal of providing “the necessary knowledge and skills to perform advanced level networking system management, administration, and maintenance support,” or similar, in their descriptions. As such, Sailors responsible for managing, administering, and maintaining the network do not necessarily have the familiarity to operate all the services and software on its servers. For much of the specialized software—e.g., medical records software—expecting Information Systems Technicians (ITs) to be proficient in its operation is unreasonable. As another specialized software product, the Global Command and Control System—Maritime (GCCS-M) POR overcomes this problem by having three training courses: system administrator, operator, and watch officer. Although establishing a whole course on operating chat would be excessive, the Navy must increase Sailors’ awareness of distributed chat architectures and their benefits. Perhaps a one-page how-to guide on configuring and using the afloat chat server in a distributed chat architecture would be a more efficient and worthwhile approach.

Following this line of thinking, one may conclude that other features of in-use systems go unused because Sailors are unaware of them or their *operational* capabilities. Operational needs—as understood by the Office of the Chief of Naval Operations (OPNAV)—drive requirements. The GCCS-M POR requires Sailors to have specific training based on the nature of their interaction with the system, or the training would not be funded. The requirement for a distributed chat architecture made it into afloat tactical networks but without the associated organizational focus (metastructuring) to use its capabilities for the resilience of the Navy’s C2 communications, perhaps because no chat POR champions or oversees its use.

As a counterpoint, consider that a lack of capability requirements has not always hindered Sailors’ innovation attempts. For example, in 2011, before the Navy used dynamic multiplexing in its afloat IP connectivity, Sailors cross-connected SATCOM links based on typical throughput demands to increase the available throughput for the ship (Johnson, 2011). Sailors were watching unused transport capacity on one SATCOM link waste away while suffering through congested throughput on another SATCOM link, so they innovated to overcome the issue and improved their communications resilience. In contrast, with chat, no perceivable problem arises until the ship loses connectivity with the shore and then, the focus is on restoring the ship’s connectivity, not analyzing how to configure an application to be more resilient. Resiliency is invisible yet critical. Less motivation exists for fixing something that does not appear broken.

Culture

An aircraft carrier (CVN) communications officer (COMMO) suggests that the root of why Sailors do not use afloat chat servers combines a risk averse culture and the inexperience of C4 leadership:

The poor performance history of CANES and senior personnel’s risk averse culture have prohibited creative problem solving or the introduction of new ideas. Why haven’t more IW or IP officers advocated to use the local chat server and federate it with the Fleet’s directed chat server? Simple. The current talent pool for IP DIVOs [(division officers)] and CSIOs is plagued with LATXFRS [(lateral transfers)] that are

⁶ Access is restricted to DoD common access card (CAC) holders:
<https://app.prod.cetars.training.navy.mil/cantrac/vol2.html>



still figuring out how to be DIVOs or learning what IP is let alone the systems under their charge.⁷ (P20)

Although this sentiment is unique among participants, the survey data do not contain any reports of in-use distributed chat architectures to contradict it, even among participants that claim knowledge of the capability. Further, P20's sentiment is consistent with literature on how structural uncertainty—caused here by inexperience—leads to undue conservatism because of the possibility for unknown, undesirable outcomes (Rowe, 1994). With greater experience, like P20 demonstrates having, comes greater understanding of the communications and network architectures, which reduces structural uncertainty and the perceived risk associated with innovating to improve resilience. Because the inexperienced officers described by P20 lack time in the field, any intervention to improve chat use must consider administrators' and users' limited time to learn, configure, and teach others about using chat more resiliently.

One may argue that people sometimes do not adopt a technology because they believe that its disadvantages outweigh its benefits (Norman, 2013), but our data do not contain any indication that any Sailor believes in any disadvantage of operating a distributed chat architecture.

In the Apple ERP example (described in the Stuck in One's Ways section), the implementation team was too risk averse to upend current processes for those that promised to be better. The new ERP had too much structural uncertainty. The CEO needed to intervene through restructuring to bring about the greater good for the organization. Overcoming these junior IP officers' inexperience to prevail over their risk aversion may similarly need a CEO-level champion, like a fleet commander, who can improve the afloat chat architecture's metastructuring. A person in this authoritative position can recommend or direct using distributed chat architectures and provide justification for their use, refined requirements to PORs, and how-to guidance to fleet users.

Inherent in our recommendation for an awareness campaign is the assumption that if C4 leadership know that afloat chat servers exist and that they can achieve a distributed chat architecture that is more resilient than the prevailing configuration in which all clients connect directly to the MOC's chat server ashore, they will use this capability. This assumption is a limitation of our recommendation. The only indication in the survey data that a ship might use its chat server in a distributed chat architecture is P65's claim: "I would have ensured the entire strike group would have pointed their chat clients to the server so that it could have been utilized even if cut off from shore." Ten other participants (P3, P16, P19, P32, P35, P38, P46, P62, P63, and P67) claim they would have used it for improving communications internal to the ship. For example, P32 claims the use would be "To allow for the internal watch stations on the ship to maintain text comms even when down IP services." As an alternative use, P67 claims, "I would have liked to be able to chat to personnel aboard the ship internally. Sometimes emails are too slow especially if I know they on the computer." In part, this is why our recommendation includes a fleet commander-level champion, to encourage adoption. If a fleet commander believes operating chat in a distributed chat architecture is a good idea for C2 communications resilience, it is more likely to happen.

On Those That Use Afloat Chat Servers

Although they comprise a minority, understanding how Sailors who use afloat chat servers do so is informative to expanding their use. We list these details in Table 4 on how

⁷ "IW" is short for Information Warfare, of which IP Officer Community is a part. Officers laterally transfer between career fields and one having done so recently is relatively inexperienced in the new field. A division officer is the entry-level job for a ship's company officer.



Sailors use afloat chat servers, from the nine respondents ($n = 69$) reporting such use. We summarize the following results from the data:

- On ships that use their chat server most frequently, the survey participants do not use it.
- Each of the five participants that did not bridge their chat servers claims either to not know of the server’s channel bridging capability or that they did not devote the necessary time to figure it out.
- P6 reports the only bridged configuration but describes using the chat server for coordinating troubleshooting efforts among maintainers on the ship, a purpose that does not benefit by a distributed chat architecture.
- Except for P2, P15, and P50, participants listed in Table 4 started their tours before the installed chat software stopped supporting the chat federation capability (described in the Chat Afloat section).

Table 4. Chat Server Usage Details for Sailors Who Report Using Their Afloat Chat Server (9 of 69 Respondents)

P	Ship Class	Sailors’ Use	Participant’s Use	Purpose	User Location	Bridged	Perceived Effectiveness
P15	DDG	Routinely	Did not use	Operational	On my ship	Unknown	Moderate
P50	LSD	Routinely	Did not use	Operational	On my ship	No	Moderate
P2	DDG	Routinely	Did not use	Unknown	Unknown	Unknown	Effective
P21	CVN	Occasionally	Did not use	Operational	On and off my ship	Unknown	Moderate
P64	CG	Occasionally	Just a few times	Operational and Administrative	On my ship	No	Moderate
P6	CVN	Occasionally	Occasionally	Administrative	On my ship	Yes	Marginal
P43	CVN	Just a few times	Did not use	Administrative	On my ship	No	Ineffective
P60	CVN	Just a few times	Routinely	Administrative and Social	On my ship	No	Ineffective
P1	CVN	Just a few times	Just a few times	Proof of concept testing	On my ship	No	Moderate

Note. Cited Ship Classes: Guided missile cruiser (CG); Aircraft carrier (CVN); Destroyer (DDG); Dock landing ship (LSD)

We still find a lack of awareness of the afloat chat servers’ capabilities and benefits. These findings are consistent with the subset of participants whose ships do not use their chat servers.

Interventions and System Developments

The “adage that we design our military systems to the requirements of the last war, rather than to meet the needs of a future war” (Boslaugh, 1999, p. 355) applies somewhat differently today. For example, its POR designed the afloat network multiplexing architecture for the maximum foreseeable throughput achievable by SATCOM; however, commercially available pLEO SATCOM capabilities threaten to overload it, having grown much faster than expected. In other words, even when we try to forecast our needs for the next war, “disruptive” technologies can disrupt our best designs. As the Navy develops and deploys additional capabilities to improve C2 resilience, it should apply the lessons discovered in the survey data to maximize these systems’ resilience and use, even when facing the next disruptive technology.



Besides the one-time technical reconfigurations that we recommend in the Technical Context section, we recommend the Navy continue multiple existing lines of effort to improve the resilience of its C2 communications using existing systems. One of these efforts approaches the DNS and email shortfall that we describe but by detecting connectivity status and dynamically reconfiguring systems automatically, according to the current connectivity status. Either the one-time reconfiguration that we offer or this dynamic reconfiguring solution is viable. We posit, however, that the greater complexity inherent in dynamically reconfiguring systems at times of need (i.e., during a loss or restoration of connectivity event) carries greater fragility than maintaining the consistent configuration we offer in this paper.

Another of these efforts involves an engineering change request (ECR) to increase the number of IP paths through which ships can communicate on SIPRNet directly without having to route through a NOC ashore (Stoffel, 2018). This is a necessary step to improving the resilience of C2 communications and recommend expanding it to other network enclaves. Without this piece of the solution, distributed chat architectures afloat will still improve resilience but be limited by the afloat network architecture.

U.S. Navy engineers are developing Communications as a Service (CaaS) to enable an *anything over anything* capability. It enables encoding IP traffic for conveyance over paths not traditionally recognized as IP paths. For example, one CaaS server can convey IP traffic to another CaaS server in J-series messages over a tactical data link (TDL).

CaaS developers are integrating new chat software that uses CaaS for its connectivity into afloat networks. The capability promises to be more resilient than previous chat capabilities because of its ability to use many communications paths, not just native IP paths. Will this become yet another unused capability akin to the distributed chat architectures that the fleet first used in Trident Warrior 2005 (Eovito, 2006)? What metastructuring actions will the Navy take to help ensure its success? Learning from the lessons observed in Sailors' (non)use of afloat chat servers, the Navy should ensure that Sailors—maintainers and operators—are aware of the new chat capability, its benefits, and how best to use it to apply those benefits to bolstering the resilience of C2 communications.

Conclusion

The U.S. Navy's dependence on IP networks for conveying C2 communications requires its leadership's intervention to use the full capabilities of applications to maximize their resilience. We analyze DNS, email, and chat and provide technical recommendations for each of their configurations—applicable to afloat tactical networks—for improving their transparent failover. We conclude that all ships should have all ships' DNS records cached locally so that the system can support other applications when unable to resolve queries from the NOC's server ashore. Improving the resilience of email (i.e., getting it to work) without the NOC's email relay server requires implementing an MX record fail-over or a fault-tolerant send connector configuration on ships' Exchange servers.

Effecting a distributed chat architecture requires both a technical configuration change and additional metastructuring of resilient chat as a capability, including a championed awareness campaign. Our survey ($n = 69$) of C4 leadership informs a sociotechnical analysis of why most Sailors do not use chat servers afloat and how some Sailors use them, bearing in mind the resilience benefits of a distributed chat architecture. Survey data reveal that the most significant hurdles to using afloat chat servers include awareness of their existence, capabilities, and benefits, and how best to use their capabilities, indicating a lack of metastructure reinforcement. Promisingly, the relatively new Software Acquisition Pathway in the Defense Acquisition System's Adaptive Acquisition Framework directs an annual assessment mechanism that forms the very structure to reveal gaps like this to sponsors (DoD, 2020). Raising



awareness will involve providing greater information transparency into why ships have afloat chat servers and how to operate them in a distributed chat architecture by having a fleet commander-level champion promote their use and circulate guidance. This circular combination of receiving feedback and implementing it throughout the fleet comprise an effective metastructuring approach (Orlikowski et al., 1995). Whether the Adaptive Acquisition Framework's annual assessment mechanism will yield tighter links between development and use is the subject of future research.

Finally, we describe the relevant interventions and systems currently in development and limited deployment. The energy and excitement around a new system with all its whiz-bang shininess can effect a metastructure episodic change that shakes people out of their routines such that they are willing to try something different or invent new practices to get at the new capabilities in the new system, as was ultimately the case for the Naval Tactical Data System (NTDS; Boslaugh, 1999). None of the new systems in development to which we refer amass enough gain in capability to stimulate such an episodic change. We must therefore understand the whole-system context for adopting each new system into Sailors' complex sociotechnical ecosystem through metastructuring reinforcement and adjustment. As the U.S. Navy emphasizes C2 communications resilience in distributed maritime operations and communications-degraded/denied environments (CD2Es), exploiting current systems—for better ways to use them and to apply lessons learned to new systems—can help keep new systems from the same nonuse fate and improve warfighter effectiveness.

References

- Ashalyengar21, American-Dipper, chrisda, DCtheGeek, mattpennathe3rd, Ajayan1008, v-albemi, Duncanma, get-itips, msdmaguire, balinger, & SharS. (2023, February). Send connectors in Exchange server. <https://learn.microsoft.com/en-us/exchange/mail-flow/connectors/sendconnectors?view=exchserver-2016>
- Blanchette, S., Jr., Albert, C., & Garcia-Miller, S. (2010, December). *Beyond technology readiness levels for software: U.S. Army workshop report* (Technical Report No. CMU/SEI-2010-TR-044). Software Engineering Institute, Carnegie Mellon University. <https://apps.dtic.mil/sti/citations/tr/ADA535517>
- Boslaugh, D. L. (1999). *When computers went to sea: The digitization of the United States Navy*. IEEE Computer Society.
- Department of Defense. (2020, October). *Operation of the software acquisition pathway*. <https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500080p.PDF>
- Eovito, B. A. (2006, June). *An assessment of joint chat requirements from current usage patterns* [Master's thesis, Naval Postgraduate School]. <https://hdl.handle.net/10945/2753>
- Fernandez, R., & Rodrik, D. (1991). Resistance to reform: Status quo bias in the presence of individual-specific uncertainty. *The American Economic Review*, 81(5), 1146–1155. <https://www.jstor.org/stable/2006910>
- Government Accountability Office. (1988, April). *Immigration service: INS' technology selection process is weak, informal, and inconsistently applied* (Report No. GAO/PMED-88-16). <https://www.gao.gov/products/pemd-88-16>
- Hofman, B., de Vries, G., & van de Kaa, G. (2022). Keeping things as they are: How status quo biases and traditions along with a lack of information transparency in the building industry slow down the adoption of innovative sustainable technologies. *Sustainability*, 14(13). <https://doi.org/10.3390/su14138188>



- Johnson, E. (2011). USS Enterprise SHF cross-connect configuration increases allocated bandwidth by 100 percent. *CHIPS*.
<https://www.doncio.navy.mil/CHIPS/ArticleDetails.aspx?ID=2298>
- Koubaa, H., & Fleury, E. (2001). A fully distributed mediator based service location protocol in ad hoc networks. *IEEE Global Telecommunications Conference (GLOBECOM '01)*, 5, 2949–2953. <https://doi.org/10.1109/GLOCOM.2001.965968>
- Landis, C. B. (2015). Email queues can be a false indicator. *CHIPS*.
<https://www.doncio.navy.mil/CHIPS/ArticleDetails.aspx?ID=5930>
- Landis, C. B. (2016). A new era of afloat IP services. *CHIPS*.
<https://www.doncio.navy.mil/CHIPS/ArticleDetails.aspx?ID=8349>
- Landis, C. B. (2023, October). Afloat resilience of IP tactical communications: Chat systems and sailors [DTIC citation AD1215296, restricted distribution].
<https://milcom2023.milcom.org/program/restricted-technical-program>
- Machi, V. (2018). Navy beefing up at-sea enterprise network. *National Defense*, 102(772), 30–31.
<https://www.jstor.org/stable/27022127>
- Mako 2.0: Administrator guide* [Restricted distribution]. (2021, April). PEO C4I.
- Mako 2.0: User guide* [Restricted distribution]. (2021, April). PEO C4I.
- Martin, O., & Marcley, A. (2013, July). *Mako chat support* (Notice of Intent No. N66001_SNOTE_0007EFDE). NIWC Pacific.
<https://sam.gov/opp/08d1ade640cd6fba0a02462113dbc880/view>
- Norman, D. A. (2013). *The design of everyday things* (Revised and expanded). Basic Books.
- Orlikowski, W. J., Yates, J., Okamura, K., & Fujimoto, M. (1995). Shaping electronic communication: The metastructuring of technology in the context of use. *Organization Science*, 6(4), 423–444. <https://doi.org/10.1287/orsc.6.4.423>
- Rambo, M. B. (2016, September). *System of systems engineering and integration process for network transport assessment* [Master's thesis, Naval Postgraduate School].
<https://hdl.handle.net/10945/50470>
- Rowe, W. D. (1994). Understanding uncertainty. *Risk Analysis*, 14(5), 743–750.
<https://doi.org/10.1111/j.1539-6924.1994.tb00284.x>
- Ruohonen, J. (2020). Measuring basic load-balancing and fail-over setups for email delivery via DNS MX records. *2020 IFIP Networking Conference (Networking)*, 815–820.
<https://ieeexplore.ieee.org/abstract/document/9142814>
- Stoffel, A. N. (2018, April). *Automated Digital Networking System (ADNS) system design description (SDD) PT routing ECR-ADNS-00928* [Restricted distribution]. PEO C4I, PMW 160.
- Wipfler, G. (2023, June). *2023 spring quarter graduation—June 16, 2023 (full)* [29:20–43:23]. <https://www.youtube.com/watch?v=Xuw8zgsPRTQ>



Appendix Afloat Chat Server Survey

We received Institutional Review Board (IRB) approval to conduct this survey and Navy Survey Office authorization with control number NSPM23.14, expiring June 14, 2025.

We solicited survey participants via the Navy IP Officer's milSuite page with cross-posts into IP Officer Teams channels. The solicitation advertised that the survey supports research on improving the resilience of C2 communications in afloat tactical networks for the purpose of helping the Navy overcome sociotechnical barriers to using capabilities inherent within fielded systems. With 5–15 minutes to complete, Sailors could take this survey once for each CANES ship in which they served in an IT leadership position. The questions are as follows:

1. Have you ever or are you now serving aboard a CANES ship? *An answer of "no" or "I don't know" to this question ended the survey.*
2. What is the date range in which you served aboard this ship?
3. What is that ship's class?
4. What was/is your job title aboard that ship?

The remainder of the questions involve branching so we illustrate this flow in Figure A1.



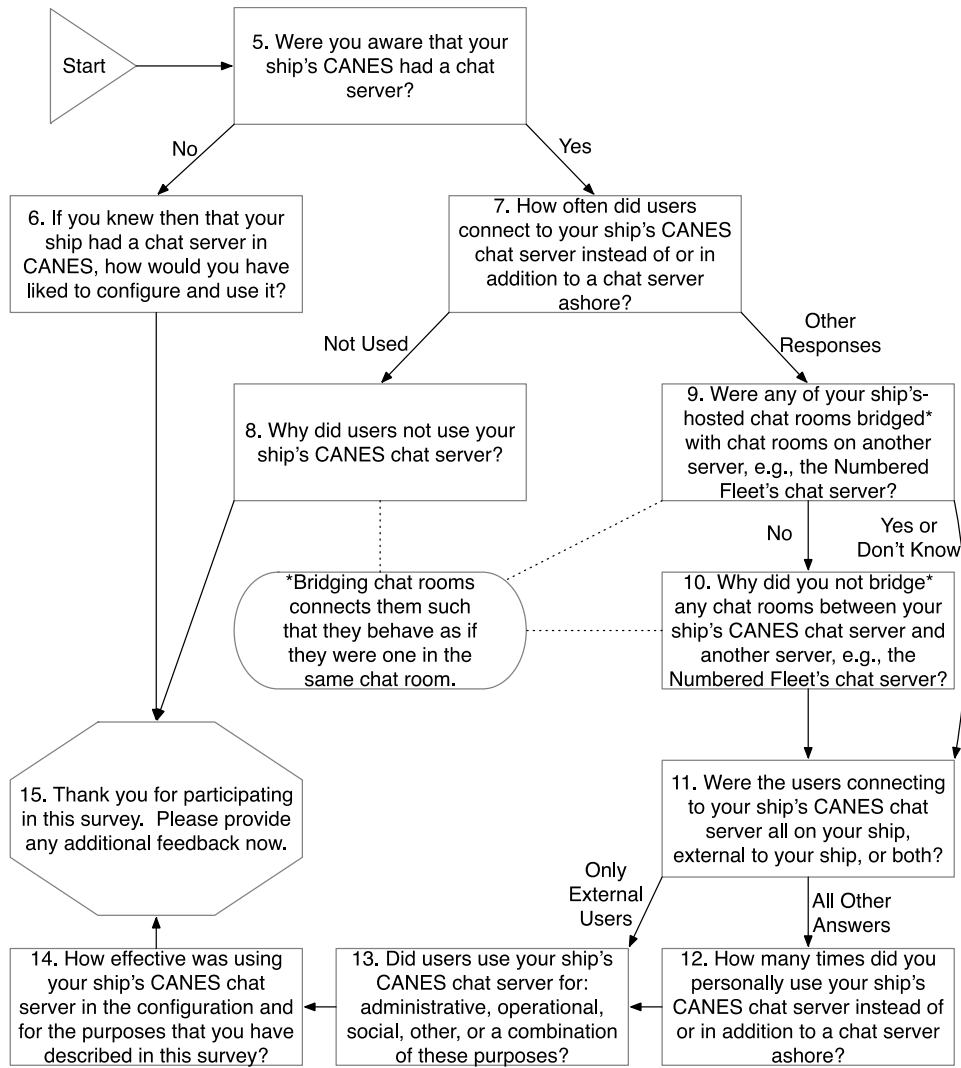


Figure A1
Chat Server Survey Questions 5–15

A Quantitative Analysis of the Relationship Between DoD Contractors' Financial Health and Their Merger and Acquisitions Spending Using Panel Data Regression

Corey D. Mack—is a Program Manager for the Defense Common Ground System Global Network Systems Branch at Hanscom Air Force Base. He received his Bachelor of Science in electrical engineering from the Rochester Institute of Technology, his Master of Science in acquisitions and program management from the Air Force Institute of Technology. [corey.mack.1@us.af.mil]

Clay Koschnick—is an Assistant Professor of Systems Engineering in the Department of Systems Engineering and Management at AFIT. He received his BS in operations research from the United States Air Force Academy, his MS in operations research from the Georgia Institute of Technology, and his PhD in industrial and systems engineering from the University of Florida. His research interests including financial and risk analysis, engineering economics, and econometrics. [clay.koschnick@us.af.mil]

Jonathan Ritschel—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He received his BBA in accountancy from the University of Notre Dame, his MS in cost analysis from AFIT, and his PhD in economics from George Mason University. Ritschel's research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis. [jonathan.ritschel.1@us.af.mil]

Michael J. Brown—is an Assistant Professor of Cost Analysis in the Department of Systems Engineering and Management at AFIT. He holds a BS in economics from the United States Air Force Academy, a MS in cost analysis from AFIT, and an MA and PhD in economics from the University of Tennessee. Brown's research interests include recruitment and retention, econometric analysis, and governmental acquisitions. [michael.brown.105@us.af.mil]

Brandon M. Lucas—is an Assistant Professor of Systems Integration & Cost Analysis in the Department of Systems Engineering and Management at AFIT. He holds a BA in history from the University of Texas at Austin, an MA in international relations and ME in teacher education from the University of Oklahoma, an MS in cost analysis from AFIT, and a PhD in economics from George Mason University. Lucas' research interests include profit analysis, cost and economic analyses, and incentive structures. [brandon.lucas@afit.edu]

Abstract

This paper examines the relationship between a prime contractor's financial health and its mergers and acquisitions (M&A) spending in the defense industry. It aims to identify financial characteristics of prime contractors that are associated with M&A spending which can provide the Department of Defense (DoD) indications of future M&A activity. These indications can help decision makers better understand the competitiveness of the defense market and develop policies that benefit the overall health of the defense industrial base. The study uses panel data regression models on 40 companies between 1985 and 2021. The company's financial health is assessed using common financial ratios while controlling for key economic factors. The results show a significant relationship between efficiency and M&A spending, indicating that companies with lower efficiency tend to spend more on M&As. However, there was no significant relationship between M&A spending and a company's profitability or solvency. These results were consistent with previous research. However, the effect of liquidity was the opposite of the expected result, possibly due to the defense industry's different view on liquidity compared to previous research.

Keywords: Mergers, Acquisitions, Financial Ratios, Financial Health, Defense Industrial Base

Introduction

The U.S. Department of Defense (DoD) expected budget reductions at the end of the Cold War. This posed a significant problem for many defense contractors who relied heavily on DoD contracts to stay in business. For instance, Lockheed Martin regularly had up to 90% of its annual revenue from defense contracts (Mahoney, 2021). The industry's feast-or-famine market



was compounded by the fact that only a few large contractors could be financed simultaneously (Higgs, 1990). To address the challenges faced by the defense industrial base (DIB) during the budget drawdown, the DoD encouraged the consolidation of its contractors' assets through M&As.

The DIB is a vast set of more than 100,000 companies providing goods and services to support the DoD's mission (Defense Industrial Base Sector, n.d.; Peters, 2021). These companies include prime contractors, sub-contractors, suppliers, small businesses, and foreign and domestic contractors. Prime contractors in particular play a significant role in the defense industry. Prime contractors act as the primary system integrators for the DoD's complex weapon systems (Susman & O'Keefe, 1998) and can also be identified by the number or value of the contracts they receive from the DoD (Bernal, 2022). A recent report shows the top five contractors (all primes) securing 25%–50% of the annual defense contracts (Hartung, 2021). These primes often subcontract or collaborate with non-prime contractors to complete a project (U.S. Small Business Administration, n.d.). Thus, prime contractors have a central role in the structure of the DIB.

In 1993, DoD policymakers organized a meeting between defense industry leaders and government officials to encourage consolidation and reduce the assets the DoD needed to support and maintain (Deutch, 2001). The primary intent was reducing tangible assets like properties, plants, and equipment where savings would be split between the government and its partners. Consequently, the consolidation recommendation triggered a new wave of M&As. Tellingly, the top six contractors increased their share of total defense contract obligations by more than 20% between 1990 and 2014 (Ellman & Bell, 2014). This concentration is even more pronounced at the sector level with the top six contractors awarded nearly 70% of aircraft manufacturing contracts—up from roughly 30% in 1990 (Ellman & Bell, 2014)

The government promotion of M&As within the DIB in 1993 and its subsequent growth paralleled an expansion of M&A activity across the general economy (Institute of Mergers, Acquisitions, and Alliances, 2024). Businesses globally have used consolidation efforts to change ownership, increase product and service variety, add to their current asset mix, foster alliances, maximize shareholder value, and improve firm performance (Mboroto, 2013). Companies also resort to M&As in industries where circumstances prevent typical organic growth (Tikhomirov et al., 2019). While non-defense businesses conduct M&As for reasons similar to defense companies, the defense industry also has its unique considerations. For example, the DIB struggles with organic growth due to a monopsony with the DoD as the primary customer acting as a gatekeeper for access to other customers (Driessnack & King, 2004).

The multitude of reasons for consolidation is also reflected in how companies combine. While the terms merger and acquisition are often used interchangeably, they have subtle differences (Kovacic & Smallwood, 1994). A merger occurs when two businesses in similar lines of business combine their organizations with mutual consent, often under a new name to reflect the new partnership (Corporate Finance Institute Team, 2022). On the other hand, an acquisition happens when one company, usually larger, buys out another company, and the acquiring company completely takes over the target company's operations (Corporate Finance Institute Team, 2022). Acquisitions often involve hostile takeovers where the buyer purchases 51% or more of the target company's shares potentially leading to the dissolution of the target company (Inoti, 2014). In amicable cases, the target company may retain its name but operates under the new parent company's hierarchy—as in Lockheed Martin's acquisition of Sikorsky in 2015.



Furthermore, the Federal Trade Commission (2013) has identified different types of consolidation: two of which are horizontal and vertical M&As. A horizontal M&A occurs when two companies consolidate within the same line of business—i.e., two aircraft manufacturers; horizontal M&As usually involve a merger of primes rather than an acquisition (Deutsch, 2001). In contrast, vertical consolidation happens when a manufacturer combines with a supplier in its chain.

To justify M&A activity, acquirers often report synergies as the primary reason for consolidation to regulating agencies (Amano, 2022; Blonigen & Pierce, 2016; Tikhomirov et al., 2019). These synergies can lead to benefits such as rapid access to technology and products, an extended customer base, and enhanced market positions (Mboroto, 2013). Synergy can be further classified into operational and financial synergy (Dewi & Mustanda, 2021). Consolidations can cause a reduction in capability, capacity, and depth of competition that will severely affect national security (Freling & Hastings, 2022).

Due to the increased M&A activity over the last few decades and the government's interest in maintaining a healthy and competitive defense industry, understanding the financial circumstances of the defense companies involved in M&As is important. Financial ratio analysis has been applied to various industries, like randomly sampled markets (Amano, 2022; Gozali & Panggabean, 2019), banking (Rashid, 2021), fuel industries (Mboroto, 2013), and even households (DeVaney, 1994). These industries have utilized financial ratios to analyze company failures, acquisitions, and the results of post-consolidation synergy. However, little research has been conducted on the relationship between financial ratios and M&A spending within the U.S. defense sector. This research was undertaken to close that gap.

Monitoring top contractors' acquisition activities is crucial to ensure the fair allocation of defense contracts. Analyzing past M&A activity and company performance can inform future policy decisions that can promote industry health and competitiveness. Specifically, this research studies the relationship between financial ratios and the M&A spending of prime contractors over the period 1985–2021 using panel data models. The M&A activity includes both horizontal and vertical consolidation as prime-on-prime and prime-on-non-prime M&As are included.

Literature Review

Financial Ratios

Financial ratios are mathematical calculations that analyze the relationship between two or more financial variables using fractions or percentages (Suthar, 2018). Ratios provide insights into a company's financial health and allow for meaningful comparisons between companies of different sizes (Barnes, 1987). Financial ratios are commonly grouped into one of the following categories: profitability, efficiency, liquidity, and solvency (Budiantoro et al., 2022). Financial reporting practices are standardized for organizations through federal tax codes, the Federal Reserve System, and the Securities Exchange Commission (SEC) allowing for the calculation of financial ratios (Horrigan, 1968).

Profitability is the first measure of a company's financial health and reflects the effective management and productive use of resources (Burja, 2011). Profitability ratios can be further categorized as margin and return ratios. Margin ratios include gross or net profit, cash flow, and operating profit margin; while return ratios include return on assets (ROA), return on invested capital (ROIC), and return on equity (ROE).

The second category of financial ratios involves solvency (or leverage) ratios which measure a company's financial stability and ability to meet its long-term debts and financial obligations. Companies with assets that are greater than the sum of their liabilities are



considered solvent (U.S. Code, 2011); insolvency is an important indicator of company failure and often becomes an essential determinant in bankruptcy declaration (Ghosh & Chaudhuri, 2017). Common ratios used to study solvency include interest coverage, debt to assets, and debt to equity.

The next financial health category is efficiency which measures how effectively managers of a company utilize their assets to generate sales (Adedeji, 2014). Efficiency can be analyzed qualitatively or quantitatively. While qualitative measures focus more subjectively on company goals, quantitative measures can provide more objective data across companies using financial ratios (Zietlow et al., 2018). Some common examples of quantitative efficiency measurements include accounts receivable turnover, fixed asset turnover, and sales to inventory.

Finally, liquidity ratios measure a company's ability to meet its short-term financial obligations and is a vital indicator of its financial standing (Kritsonis, 2005; Tikhomirov et al., 2019). Investors and creditors assess the risk of lending money to or investing in the company (Beaver, 1966). Some common ratios used to measure liquidity include the current, quick, net working capital, and days sales outstanding.

Limitations of Financial Ratios

While valuable for analyzing a company's performance, financial ratios have several limitations. First, the numerical value of a ratio can vary depending on the specific values used from the financial statements. This discrepancy arises from multiple profit figures disclosed in income statements such as operating profit, net profit before interest and taxation, and net profit after taxation (Frecknall-Hughes et al., 2007). Thus, companies facing financial troubles may manipulate ratios to meet creditor and investor expectations, rendering them unreliable until a crisis occurs (Lev, 1969; Wilcox, 1971).

Furthermore, the choice of which ratios to analyze can be challenging due to the numerous ratios available; it is impractical to examine all of them in a single study (Murphy et al., 1996). Yet, it is important to note that not all ratios measure a company's performance equally well. To obtain a comprehensive understanding, it may be necessary to consider a combination of ratios alongside other economic factors (Gâdoiu, 2014). Moreover, the usefulness of financial ratios can vary depending on their application or the specific sample being analyzed, and researchers sometimes overuse ratios, leading to over-fitting and overstated predictability (Palepu, 1986)

Another limitation is that the importance of different financial ratios may vary across industries. Different industries prioritize certain categories of ratios based on their specific needs. For example, creditors may emphasize debt payment ratios more while managerial accounting practices focus more on profitability measures (Horrigan, 1968). Examining how financial institutions evaluated debt ratios for creditworthiness in industries reliant on debt financing, such as defense contractors in the 1990s, can provide valuable insights (Beaver, 1966; Deutch, 2001).

Despite these limitations, financial ratios remain valuable for assessing a company's health. Ensuring consistency in the calculation process can minimize variations caused by different financial statement values. Selecting the most relevant ratios for a specific industry or context is vital and combining them with other economic considerations can enhance their value. Financial ratios have been widely used for decades and have demonstrated their descriptive abilities in assessing companies across various industries. By avoiding known biases and building on previous studies, accurate modeling procedures can be developed to give interested parties a more accurate picture of a company's financial health over time.



Financial Ratios and Acquisition Activity

M&As occur for various reasons and the financial health of the companies involved provide insight into the possible reasons. Previous studies have indicated that some companies choose to be acquired strategically to avoid bankruptcy or other distress (Officer, 2007; Pastena & Ruland, 1986). Other researchers found that acquisitions caused by distress were less common (Camerlynck et al., 2005; Higson & Elliott, 1993).

Assuming struggling companies were the targets, successful companies would be the presumed acquirers. A company's profitability could be one measure of its health or success. The company may save its excess funds, pay debts, distribute dividends, or acquire businesses with these profits. However, according to research by Yang et al. (2019), successful companies with strong operating performances were less likely to pursue external acquisitions. Although, companies with higher growth opportunities may rely more on external investments to fund their expansion (Yang et al., 2019).

The relationship between liquidity and solvency concerning acquirers has been a subject of debate among researchers. Some argue that companies with excess cash prefer to utilize their funds instead of seeking external financing, while others contend that acquisitions financed with additional debt are more favorable, particularly in the defense industry (Bruner, 1988; Deutch, 2001; Myers, 1984). Since Myers' and Yang's studies, there has been a significant time gap, making it challenging to determine any potential shifts in the utilization of liquid assets for M&As. However, recent research suggests that higher liquidity tends to increase the likelihood of M&As (Erel et al., 2021; Shleifer & Vishny, 1992). Furthermore, a company's acquisition strategy may involve transitioning from internal assets to external debts and eventually equity financing, with variations based on the company's experience (Fourati & Affes, 2013). The free cash flow theory concept is also discussed, suggesting that companies with substantial liquidity may engage in self-interested, low-benefit acquisitions (Jensen, 1996; Yang et al., 2019).

Various ratios have been studied to determine the likelihood of a company becoming an acquisition target. Cudd and Duggal (2000) discovered that when there are imbalances between sales and resources within target firms, their acquirers can invest the excess resources more profitably in their projects or finance the acquired firms' projects at a lower cost of capital. Belkaoui (1978) found that non-liquid asset ratios best-predicted takeovers in Canadian industries. Additionally, comparing a company's ratio to the industry average has proven helpful in predicting failures and acquisitions (Barnes, 1990; Camerlynck et al., 2005).

Financial Ratios to Predict Failure

Revisiting the idea that targets may be in economic distress leads to two research efforts (Ghosh & Chaudhuri, 2017; Horrigan, 1968) which extensively outlined past studies (i.e., Altman & Hotchkiss, 2006; Beaver, 1966; Merwin, 1942; Smith & Winakor, 1935). These studies used financial ratios to predict company failure years before it occurred. Failure was defined as bankruptcy in these studies, not acquisitions. Bapat and Nagale (2014) compiled a list of 35 ratios used in these and other studies to evaluate three methodologies for bankruptcy prediction. They found that 24 of the ratios had statistical relevance in each of their three models predicting the failure of a company. At least one ratio fell into their performance categories of leverage, operating cash flow, liquidity, profitability, activity, and market ratios.

Many models and their included financial ratios have proven somewhat effective in predicting company failures. Successful models have seen anywhere from 60%–90% accuracy, depending on the analyzed periods and the financial ratios used (Bapat & Nagale, 2014; Ghosh & Chaudhuri, 2017). Although failure and target prediction models can provide valuable insight into a company's future status, not all researchers agree on their predictability or methodology. Palepu (1986) suggested that while these studies claimed high accuracy, they suffered



methodological flaws. The flaws in the results were caused by using non-random sampling and arbitrary cutoffs, which created biases and made it difficult to interpret the results. Palepu (1986) and Jensen and Ruback (1983) agree that it is difficult and perhaps impossible for the market to predict the fate of a company. These flaws were also addressed by Powell (2001), who studied several models reporting high accuracy in acquisition target detection. He made an interesting claim that companies targeting profitable investments should not use statistical models, but the same models could effectively predict future takeovers (Powell, 2001).

Economic Factors

The economic conditions influencing a company's financial performance occur at different levels of the economy. Macroeconomics refers to the behavior of the overall, national economy. Microeconomics refers to the economic behavior of individual businesses. A third, intermediate level (also commonly included in microeconomics and less commonly referred to as mesoeconomics) refers to regional or organizational economics (Lambooy, 1990); a unique market structure (i.e., the defense industry) and its behavior would fall into this category.

M&As can be influenced by macroeconomic factors such as national productivity, inflation, and interest rates. Research has found that national productivity, measured by gross domestic product (GDP), can significantly impact M&As. Companies are more likely to engage in M&As during strong economic conditions and less likely to do so during weak economic conditions (Ji, 2016). Cordeiro (2014) observed that the number of M&As sharply declined in 2007 due to the subprime crisis and the subsequent recession in the United States, yet it rose globally in 2014 after the economy recovered. Additionally, Carbonara and Caiazza (2009) and Wang (2009) found that national productivity growth was the most significant economic factor influencing M&As in Italian and Chinese markets.

Previous studies have shown that financial ratio analysis is most effective when comparing companies within similar industries (Barnes, 1990; Beaver, 1966; Cudd & Duggal, 2000; Edmister, 1972). In this research, the sample only studies defense contractors, so this will not be an issue. By examining the mesoeconomic level, we can better understand how industry-specific disruptions can affect certain areas of the economy and individual businesses. This understanding can lead to shifts in M&A activity. For example, the government's budget drawdowns and pro-consolidation policy recommendations from 1993 to 1998 threatened contractor revenue, prompting some companies to seek M&As to secure DoD contracts. These conditions are unique to the defense sector and cannot be captured at a macroeconomic level.

Lastly, size was a prevalent microeconomic influence concerning company failure and M&A activity. There are various proxies for the size of a company, like market capitalization, sales, resources, or employees—each capturing different aspects of the firm. Market capitalization is market-oriented and forward-looking, while total sales are more related to the product market and not forward-looking, and total assets measure the firm's total resources (Dang et al., 2018). Nevertheless, the correlation among these proxies is high (Shalit & Sankar, 1977). The growth of a company demonstrated unusual findings in that while larger firms tend not to fail as frequently (Beaver, 1966), growth-to-asset imbalances can increase the likelihood of failure and becoming an acquisition target (Camerlynck et al., 2005; Cudd & Duggal, 2000; Palepu, 1986; Yang et al., 2019). Companies often became acquisition targets when their assets were insufficient to maintain their observed growth and acquirers typically had higher asset growth rates than their targets (Camerlynck et al., 2005).

In consideration of the existing literature linking a company's financial health (as measured by financial ratios) to failure and M&A activity, this research seeks to further explore the financial characteristics associated with M&A spending for DoD prime contractors. While there has been some previous research on this topic, little research has been done within the



defense sector. As the government has a vested interest in the health and competitiveness of defense contractors, this paper seeks to provide insight into M&A spending by analyzing its association with profitability, liquidity, solvency, and efficiency ratios.

Hypothesis Development

The observed relationship between profitability and liquidity ratios and M&A activity is varied. Camerlynck et al. (2005) found acquirers often had higher profitability than their industry peers, yet Yang et al. (2019) found that companies with higher profitability tended to rely on internal investments over M&As. Similar results were seen for liquidity. Some research found that cash-rich or liquid firms were more likely to attempt acquisitions (Bruner, 1988; Erel et al., 2021; Jensen, 1996; Myers, 1984) while Camerlynck et al. (2005) found acquirers often reported below industry-average liquidity and were highly leveraged.

For solvency and efficiency ratios, the relationship with M&A activity shows more consensus. Most researchers suggest that larger, less efficient firms seek to acquire smaller, more efficient firms to improve the acquiring company's efficiency (Inoti, 2014; Jensen & Ruback, 1983). Thus, efficiency and M&A activity appear to have an inverse relationship. Although an optimal size proxy has not been discovered (Lev & Sunder, 1979), this study utilized total sales to measure a contractor's size.

Economic factors have been shown to be associated with M&A activity and thus should be controlled. National productivity is a proxy for an economy's overall health. Increases in national productivity, or GDP, have been associated with increased M&A activity in American, Chinese, and Italian markets (Carbonara & Caiazza, 2009; Wang, 2009). Industry-specific factors have also influenced M&A activity (Cordeiro, 2014; Palepu, 1986). The U.S. defense budget can be used as a proxy for the strength of the defense industry; the theory is that as budgets decrease, as they did in the 1990s, M&A activity will increase (Office of the Under Secretary of Defense [OUSD] for Acquisition and Sustainment, 2022). Finally, a firm's size is used to represent an individual company's potential for M&A activity. Research has shown that as a company grows, its acquisition spending increases (Dang et al., 2018; Shalit & Sankar, 1977). These macro and microeconomic variables are used as controls in studying four hypotheses on the relationship between M&A spending and a company's financial health:

Hypothesis 1 (H1): Defense contractors with higher profitability are associated with increased spending on M&As.

Hypothesis 2 (H2): Defense contractors with greater liquidity are associated with increased spending on M&As.

Hypothesis 3 (H3): Defense contractors that are highly leveraged are associated with increased spending on M&As.

Hypothesis 4 (H4): Companies with lower efficiency are associated with increased spending on M&As.

Table 15. Summary of Expected Results

Increase in	M&A Spending
Profitability	Increases (+)
Solvency	Increases (+)
Efficiency	Decreases (-)
Liquidity	Increases (+)
National Productivity	Increases (+)
Defense Budget	Decreases (-)
Contractor Size	Increases (+)



Methodology

This analysis required the identification of prime contractors as well as the collection of their annual 10-K reported financial data and M&A activity. The financial data was used to calculate the financial health ratios. At the same time, M&A activity was used as the response variable for the models. This study used various databases, and this section will define the processes used to build the final dataset. The collected data was used to develop the independent variables (IVs) to test the hypotheses. A panel data model was utilized since the experiment studies M&A activity on a cross-section of contractors over several years.

Sample

Several reports were critical in identifying whom the DoD classified as prime contractors within its weapons categories (OUSD for Acquisition and Sustainment, 2022; U.S. General Accounting Office, 1998a, 1998b). Between these reports, 42 prime contractors were identified for 1990, 1998, and 2020. The 42 initial contractors fall into the “DoD Identified Contractors” category in Table 2. Although the DoD provided the initial foundation for prime contractor identification, several contractors have entered or exited their market through M&A activity. The entry and exit of contractors required additional contractors to be added as primes or removed from the analysis. Additionally, some data could not be obtained due to financial reporting requirements and database limitations.

Several M&As transpired between the release of the GAO (1998a, 1998b) and the OUSD for Acquisition and Sustainment (2022) reports causing some primes to be absent in either report. It is assumed that if a company acquired a prime contractor, the acquirer becomes a prime themselves, as indicated by “M&A of DoD Prime.” Models with and without the acquirer were assessed to ensure the acquirer’s data did not influence the results. M&As also spurred two spin-off companies that still serviced their previous sectors within the industry. All defense industry spin-offs identified in this research were acquired or no longer considered primes by the OUSD (OUSD Acquisition and Sustainment, 2022). However, they still needed to be tracked in this analysis due to their strong influence within their sectors.

Nine DoD-identified prime contractors were subsidiaries of a larger parent company. Since financial data for a parent company is rarely reported at the subsidiary level, limited financial data could be found. Therefore, eight parent companies were added to the analysis as a proxy for the prime contractor they represent. Similarly, private companies’ financial data is rarely made public. Although databases like Techsalerator, Mattermark, Crunchbase, and PitchBook provide private company financials, seven private prime contractors were omitted from this study due to database access limitations. Finally, most of the databases used in this study have limited financial data for companies that merged before 1995. Since Northrop merged with Grumman in 1994, Grumman’s financial data could not be obtained.

The dataset became unbalanced due to contractors being acquired, merging into new companies, or companies being generated via spin-offs. M&A activity can cause contractors to come and go, making it hard to balance the dataset. The *p1m* package from the statistical modeling software R was used for this analysis and was capable of using unbalanced data. Balancing the dataset would reduce the sample size to only seven contractors.

Table 2 summarizes the data inclusion and exclusion process for this research. The final sample includes 40 companies from 1985 to 2021, totaling 683 observations¹.

¹ These values reflect the entire dataset used for this assessment. The most stringent model tested 22 companies, over 27 years, totaling 352 observations. The model removed one or more of the following: parent companies, years, or non-GAO identified prime contractors. The final model was robust to these changes. Therefore, the entire dataset was utilized.



Table 16. Contractor Sample

Contractor Inclusion / Exclusion Criteria	# of Contractors	Observations
DoD Identified Contractors	42 Initial	419*
M&A of DoD Prime	5 Added	120
Spin-Off Companies	2 Added	29
Parent of Subsidiary	8 Added	115
Subsidiary	9 Removed	0
Private Companies	7 Removed	0
Missing Financials	1 Removed	0
Total	40 Contractors	683

*Refers to the observations of the 25/42 initial DoD-identified contractors. The remaining 17 were removed per the table.

Data

To conduct this research, two areas of analysis were required. The first involved gathering financial data from each company, which must be disclosed in an annual 10-K report for publicly traded companies. In priority order, the sources for 10-K data were Mergent, Yahoo Finance, Compustat, SEC's EDGAR, and S&P Capital IQ². Missing values were cross-checked with another source for consistency. The financial data gathered from these sources were used to calculate a company's financial health via ratios. These ratios were implemented as IVs to test the hypotheses. The second area of research required collecting each company's M&A activity from Yahoo Finance, SEC's EDGAR Archive, Mergr, and Mergent Online. M&A spending was analyzed as the dependent variable (DV) for this study.

Modeling Considerations

Panel data models were used to test the relationship between financial ratios and annual M&A spending. While pooled ordinary least squares (POLS) models can be used for a cross-section of entities over time multiple periods, it pools the data into one large cross-section ignoring the fact that some observations came from the same entity. Conversely, fixed and random effects models can often improve upon the pooled cross-section parameter estimates by controlling for the unobserved heterogeneity.

Heterogeneity refers to the differences amongst the contractors in the sample which can be either observed or unobserved. Observed characteristics can be directly modeled by including independent variables; conversely, unobserved heterogeneity comes from unobserved group characteristics (Armstrong, 2021). Unobserved heterogeneity can include characteristics such as company culture, management styles, and employee attitude. Additionally, unobserved heterogeneity can be time-invariant (i.e., the unobserved characteristics of a company that do not change over time) or time-variant (i.e., the unobserved characteristics that do change over time). Both fixed and random effect models attempt to account for the time-invariant unobserved heterogeneity by exploiting the fact that the panel data includes multiple observations from the same entity and using a composite error term of time-invariant and idiosyncratic components. The underlying assumptions regarding the unobserved heterogeneity is different between the two models. Fixed effects (FE) modeling assumes that time-invariant, unobserved heterogeneity is correlated with one or more of the IVs. FE demeans each variable across time for each entity effectively removing the unobserved time-invariant heterogeneity and leaving only the idiosyncratic error. Thus, the net effect of the IVs on the DV can be explored. Conversely,

² Yahoo Finance, Compustat, and EDGAR were only used when Mergent was missing information in a handful of cases. S&P Capital was used to confirm a company's private status. The Mergent database has several different storage mechanisms, such as Archive, Horizon, and Online. Mergent Online was used exclusively for this study.



random effects (RE) assumes that the time-invariant unobserved heterogeneity is uncorrelated with the IVs. While RE also demeans the data, it does so in such a way that does not completely remove the time-invariant unobserved heterogeneity.

The distinction between fixed and random effects is whether the time-invariant unobserved entity effect embodies elements correlated with the regressors in the model—not whether these effects are stochastic (Greene, 2008). Choosing between fixed or random effects often depends on the assumptions of the research and the nature of the data. Fixed effects models are almost exclusively used in econometric research since the time-invariant component of the error term is often correlated with one or more IVs (Hilmer & Hilmer, 2014).

Without an extensive deep dive into each contractor’s management structures, processes, and other company-specific characteristics that are generally constant over the period of time studied, it is hard to rule out that the time-invariant unobserved heterogeneity has zero correlation with any predictor variables—meaning fixed effects models would be preferred. Additionally, fixed effect models are often preferred for small (Borenstein, 2009) and non-random samples (Dougherty, 2011)—both of which were used in this study. The sample size was 40 contractors manually selected due to their status as defense contractors. Due to this, both fixed and random effects models were considered, and a Hausman test was used to determine the best-fit model.

Additionally, time fixed effects were assessed for inclusion in the model. These are the effects that remain constant across the entities for a given time period but may have a different effect in different time periods (e.g., the attitude of a government regulatory body may change from year to year but is the same for each entity in a given year). Modeling time fixed effects is similar to controlling for unobserved heterogeneity in fixed and random effects models; but instead it focuses on demeaning the data across the time dimension instead of the contractor dimension. Controlling for time-fixed effects may correct potential biasedness in the estimates of the model parameters. A Breusch-Pagan Lagrange Multiplier (LM) test was used to determine the need for time-fixed effects.

It is important to test model assumptions to support the quality of the model. First, the presence of unit roots is considered. A dataset is considered stationary if its statistical properties remain constant over time. However, in a time-series dataset, some variables may have upward or downward trends over time, indicating a unit root’s presence. To evaluate unit effects, the Dickey-Fuller test is utilized.

Next, serial correlation and cross-sectional dependence within the errors were tested. These type of correlations may be a problem for macro-panels with long time series. While the distinction between micro and macro panels is subjective, Baltagi (2021) defines macro panels as having more than 10 to 20 years and a moderate number of entities, 7–200, which this assessment satisfies. Serial correlation refers to the dependence between observations over time within the same unit was assessed using a Breusch-Godfrey/Wooldridge. Cross-sectional dependence is the simultaneous dependence between observations across different units (Armstrong, 2021). When the errors of multiple observations within a certain period are correlated, it can lead to biases and inefficient estimates of the model parameters. Breusch-Pagan LM and Pesaran Cross-Sectional Dependence tests were used to assess the model for cross-sectional dependence.

Finally, heteroscedasticity occurs when the error term’s variance in the model is not constant across all IV levels resulting in inefficient estimates of the model parameters; heteroscedasticity was assessed using a Breusch-Pagan test. If heteroscedasticity exists, robust covariance matrix estimation can correct it. Different methods of correcting for assumption violations can depend on whether the model uses random or fixed effects. For



example, White’s heteroscedasticity-consistent covariance estimators are commonly used for random effects, and the Arellano method of clustered standard errors for fixed effects (Arellano, 2009; Greene, 2008).

Variables and Model

To avoid the issues surrounding statistical overfitting, the models in this study used one ratio per financial health category and economic level. The variables chosen for this study were selected based on the frequency and statistical relevance of the ratios in previous research. The variables and model structure are reported in Table 3.

Table 17. Variable Definitions and Model Structure

IVs:	Definition
Profitability (P)	ROA (%) measures a company’s profitability by dividing net income by average total assets. Average total assets are calculated by averaging the current and previous year’s total assets
Solvency (S)	The debt-to-equity ratio (%) measures a company’s solvency by comparing its long-term debt to shareholder equity
Efficiency (E)	Total asset turnover (%) measures a company’s efficiency by comparing its revenue to its average total assets
Liquidity (L)	The current ratio (%) measures a company’s liquidity by comparing its current assets to its current liabilities
Control Variables:	
National Productivity (GDP)	Represents macroeconomic factors as measured by U.S. GDP in billions of dollars (Office of Management and Budget, n.d.-a)
Defense Budget (DB)	Represents defense industry factors as measured by the U.S. defense budget in billions of dollars (Office of Management and Budget, n.d.-b)
Size (Sz)	Represents the contractor’s size measured by year-end sales in billions of dollars
DV:	
M&A Spending (MA)	Represents the millions of dollars a contractor spends on annual M&As
Model:	
$MA_{it} = \beta_0 + \beta_1 P_{it} + \beta_2 S_{it} + \beta_3 E_{it} + \beta_4 L_{it} + \beta_5 GDP_{it} + \beta_6 DB_{it} + \beta_7 Sz_{it} + \epsilon_{it}$	

Empirical Results and Discussion

Initially, it was believed that a company’s unobserved heterogeneity and IVs were correlated—e.g., its internal processes and company culture could be related to the company’s observed financial ratios. Although fixed effects models are usually preferred in econometric research, the Hausman test frequently indicated that random effects models better fit the data. Random effects models may be better than fixed effects models in industries where market randomness affects the correlation between a company’s unobserved characteristics and financial health. The fixed effects models still regularly outperformed Pooled (OLS) models until the most stringent criteria for prime classification was placed on the model. As the sample size decreased, this may have caused a lack of statistical power within the fixed and random effects models.

After running a series of increasingly more exclusive criteria on the contractor selection, it was noted that the general trends remained constant throughout each model. There were a few cases in which the signs of an estimate would change. However, the term was statistically insignificant in these cases and contained large standard error margins. Thus, the remainder of



this research utilized the entire dataset, which includes the acquirers of primes, spin-off companies, and parent companies.

Table 4 reports the relationship between the contractor’s financial health and M&A spending. The Hausman and F-tests reported that the best-fit model was the random effects, followed by fixed effects, and then the pooled OLS model. Although time-fixed effects were found via a Breusch-Pagan LM test, they were intentionally left off the results reported in Table 4 to assess the control variables’ effects and provide easily interpretable results. While time-fixed effects can be used in fixed and random effects models, how they are reported differs. After demeaning for time-fixed effects that are constant for each contractor in a fixed effects model, the result of the variable is zero. In a random-effects model, the quasi-demeaned data remains and is not fully reduced to zero, making interpretation of the variables difficult. This difficulty in interpretation only applies to the national productivity and defense budget variables since they are constant for each contractor. The remaining variables are company specific.

Table 18. Empirical Results

IVs	Pred.	Models		
		Random- Effects	Fixed- Effects	Pooled OLS
		Coef. (Std. Error)	Coef. (Std. Error)	Coef. (Std. Error)
Intercept		694.527** (239.487)	--	799.454*** (222.721)
Profitability	+	4.847 (3.784)	1.792 (2.945)	7.700* (3.907)
Solvency	+	0.030 (0.101)	0.015 (0.126)	0.026 (0.100)
Efficiency	-	-3.381** (1.057)	-2.826* (1.548)	-3.633*** (0.941)
Liquidity	+	-0.853** (0.265)	-0.728 (0.671)	-0.945*** (0.247)
National Productivity (GDP)	+	0.039 (0.028)	0.037 (0.026)	0.036 (0.027)
Defense Budget	-	-1.093 (0.839)	-1.184 (0.934)	-1.103 (0.027)
Size (Sales)	+	11.402*** (3.350)	22.254 (14.752)	10.682*** (2.747)
	R ²	0.073	0.057	0.098
	Adjusted R ²	0.063	-0.013	0.088

Significant at: *** 0.001, ** 0.01, • 0.1

These results point to some interesting findings. First, M&A spending increases with increased profitability (H1), solvency (H3), GDP, and firm size match their respective hypotheses and economic theory. As predicted in the hypothesis development section, efficiency (H4) and defense budget increases negatively impact M&A spending. It can be seen that not all variables were significant. Also, several had large standard errors, suggesting that some coefficients are not statistically different from zero. This is seen in the case of solvency, which has a p-value of 0.77 or higher in all models. If the significance is ignored, all estimates’ signs match the hypotheses across modeling techniques, except for liquidity (H2). It appears as if efficiency has the strongest relationship with M&A spending. Although, profitability and liquidity can also be contributors depending on the modeling processes used.



Interestingly, liquidity (H2) had one of the greatest effects on M&A spending, yet the sign was the opposite of the expected result. While financial ratios have been used to analyze various industries, not much ratio analysis has been conducted in the defense industry. The defense industry may not give as much significance to liquidity due to its relationship with the DoD. For example, the defense industry may choose to place less weight on liquidity due to its relationship with its primary customer—the DoD. Evidence shows that the DoD offers subsidies to troubled contractors through research and development contracts, loan guarantees, tax breaks, and possible strategic selection for new contracts (Higgs, 1990). Knowing a contractor may have a fail-safe in the event of a near collapse, liquid assets may not be viewed as they would have been in markets without these subsidies.

Conclusion

This research investigated the relationship between a company's financial health and M&A spending. The study utilized common industry ratios to assess a company's profitability, solvency, liquidity, and efficiency as measures of financial health. Panel data regression models were employed, revealing a significant association between a company's efficiency and M&A spending, supporting Hypothesis 3. Although the regression models did not indicate a significant relationship between M&A spending and a company's profitability or solvency, the signs of the parameter estimates aligned with prior research and Hypotheses 1 and 3. However, the unexpected opposite sign on liquidity contradicted Hypothesis 4. The defense industry may perceive liquidity differently than the industries examined in previous studies.

While the primary focus of this study was not predicting future consolidation activities, the model's findings can provide indicators or warnings to DoD policymakers. These indicators enable more precise targeting of policies to promote or discourage consolidation. Additionally, insights can be gained regarding the impact of specific actions, such as budget reductions or contract awards, on M&As within the industry. The control variables used in this research, namely national productivity, defense budgets, and firm size, exhibited signs consistent with economic theory. Increased national productivity and firm size, along with decreased defense budgets, could contribute to increased M&A activity.

Future research can further enhance the model's specification by conducting a more comprehensive analysis of a company and the factors influencing its unobserved heterogeneity. This would help determine whether a fixed or random effects model should be utilized. Although this study did not rule out the assumptions of either model, additional investigations can contribute to this determination.

Expanding the observation frequency from annual to quarterly financial reports could enhance the model's power. Broadening the sample beyond the GAO-identified contractors investigated in this study can augment the model's statistical power. During the initial data collection, uncertainties arose concerning the number of contractors that needed to be excluded due to their private status, ownership by larger companies, or unobtainable financial information.

Lastly, while some researchers have examined the financial performance of companies before and after M&A in other markets, limited research has been conducted within the defense sector. Although previous studies have provided fragments of the puzzle, a comprehensive assessment of defense contractors' performance before and after M&As remains elusive. Further exploration of this area is warranted to gain a holistic understanding of the dynamics involved.

This research demonstrates that companies with higher profitability and solvency, along with lower efficiency and liquidity, are more likely to engage in M&A spending. The findings hold implications for DoD policymakers, enabling more targeted policies, and shedding light on the



influence of various actions, such as national productivity, defense budgets, and firm size, on industry consolidation. However, future research should delve deeper into the model specification, increase observation frequency, expand the sample size, and explore the financial performance of defense contractors in the context of M&A.

References

- Adedeji, E. A. (2014). A tool for measuring organization performance using ratio analysis. *Advances in Social Sciences Research Journal*, 1(8), 89–97. <https://doi.org/10.14738/assrj.18.572>
- Altman, E. I., & Hotchkiss, E. (2006). *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt* (3rd ed). Wiley.
- Amano, Y. (2022). Do acquiring firms achieve their mergers and acquisitions objectives? Evidence from Japan. *Accounting & Finance*, 62(2), 2905–2945. <https://doi.org/10.1111/acfi.12909>
- Arellano, M. (2009). Practitioners' corner: Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434. <https://doi.org/10.1111/j.1468-0084.1987.mp49004006.x>
- Armstrong, D. (2021). *R: Learning by example*. <https://quantoid.net/files/rbe/book-2021/>
- Baltagi, B. H. (2021). *Econometric analysis of panel data*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-53953-5>
- Bapat, V., & Nagale, A. (2014). Comparison of bankruptcy prediction models: Evidence from India. *Accounting and Finance Research*, 3(4), 91–98. <https://doi.org/10.5430/afr.v3n4p91>
- Barnes, P. (1987). The analysis and use of financial ratios: A review article. *Journal of Business Finance & Accounting*, 14(4), 449–461. <https://doi.org/10.1111/j.1468-5957.1987.tb00106.x>
- Barnes, P. (1990). The prediction of takeover targets in the U.K. by means of multiple discriminant analysis. *Journal of Business Finance & Accounting*, 17(1), 73–84. <https://doi.org/10.1111/j.1468-5957.1990.tb00550.x>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4. <https://doi.org/10.2307/2490171>
- Belkaoui, A. (1978). Financial ratios as predictors of Canadian takeovers. *Journal of Business Finance & Accounting*, 5(1), 93–107. <https://doi.org/10.1111/j.1468-5957.1978.tb00177.x>
- Bernal, K. (2022, May 18). Prime and sub-government contractors. ExecutiveGov. <https://executivegov.com/articles/prime-and-sub-government-contractors>
- Blonigen, B. A., & Pierce, J. A. (2016). Evidence for the effects of mergers on market power and efficiency. *Finance and Economics Discussion Series*, 2016(082). <https://doi.org/10.17016/FEDS.2016.082>
- Borenstein, M. (Ed.). (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Bruner, R. F. (1988). The use of excess cash and debt capacity as a motive for merger. *Journal of Financial and Quantitative Analysis*, 23(2), 199–217.
- Budiantoro, H., Artanti, D. V., & Lapae, K. (2022). Analysis of liquidity ratio, profitability, solvency, and activity before and after mergers or acquisitions in mergers or acquisitions companies listed on the Indonesia stock exchange (2015–2019 period). *Jurnal Ekonomi*, 11(02).
- Burja, C. (2011). Factors influencing the companies' profitability. *Annales Universitatis Apulensis Series Oeconomica*, 2(13), 215–224. <https://doi.org/10.29302/oeconomica.2011.13.2.3>
- Camerlynck, J., Ooghe, H., & Langhe, T. D. (2005). Pre-acquisition profile of privately held companies involved in take-overs: An empirical study. *Small Business Economics*, 24, 169–186.
- Carbonara, G., & Caiazza, R. (2009). Mergers and acquisitions: Causes and effects. *Journal of the American Academy of Business*, 14(2), 188–194.



- Corporate Finance Institute (CFI) Team. (2022, March 4). Merger vs acquisition. <https://corporatefinanceinstitute.com/resources/knowledge/deals/merger-vs-acquisition/>
- Cordeiro, M. (2014). The seventh M&A wave. Camaya Partners. <https://camayapartners.com/wp-content/uploads/2016/06/The-seventh-MA-wave.pdf>
- Cudd, M., & Duggal, R. (2000). Industry distributional characteristics of financial ratios: An acquisition theory application. *The Financial Review*, 35(1), 105–120. <https://doi.org/10.1111/j.1540-6288.2000.tb01409.x>
- Dang, C., Li, F., and Yang, C. (2018). Measuring firm size in empirical corporate finance. *Journal of Banking and Finance*, 86, 159–176.
- Defense Industrial Base Sector (n.d.). Cybersecurity and Infrastructure Security Agency (CISA). Retrieved July 28, 2022, from <https://www.cisa.gov/defense-industrial-base-sector>
- Deutch, J. (2001). Consolidation of the US defense industrial Bbase. *Acquisition Review Quarterly*, 8(3).
- DeVaney, S. A. (1994). The usefulness of financial ratios as predictors of household insolvency: Two perspectives. *JFCP Research Journal*, 5.
- Dewi, N. P. L. K., & Mustanda, I. K. (2021). Comparative study of financial performance before and after acquisition. *American Journal of Humanities and Social Sciences Research (AJHSSR)*, 5(3), 151–161.
- Dougherty, C. (2011). *Introduction to econometrics* (4th ed). Oxford University Press.
- Driessnack, J. D., & King, D. R. (2004). An initial look at technology and institutions on defense industry consolidation. *Defense Acquisition Review Journal*, January–April 2004. <https://apps.dtic.mil/sti/citations/ADA423534>
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis*, 7(2). <https://doi.org/10.2307/2329929>
- Ellman, J., & Bell, J. (2014). Analysis of defense products contract trends, 1990–2014 (No. HQ0034-12-A-0022). Center for Strategic and International Studies. [https://www.csis.org/analysis/analysis-defense-products-contract-trends-1990-2014](https://csis-website-https://www.csis.org/analysis/analysis-defense-products-contract-trends-1990-2014)
- Erel, I., Jang, Y., Minton, B. A., & Weisbach, M. S. (2021). Corporate liquidity, acquisitions, and macroeconomic conditions. *Journal of Financial and Quantitative Analysis*, 56(2), 443–474. <https://doi.org/10.1017/S0022109019000978>
- Federal Trade Commission. (2013, June 11). Competitive effects. <https://www.ftc.gov/advice-guidance/competition-guidance/guide-antitrust-laws/mergers/competitive-effects>
- Fourati, H., & Affes, H. (2013). The capital structure of business start-up: Is there a pecking order theory or a reversed pecking order? —Evidence from the panel study of entrepreneurial dynamics. *Technology and Investment*, 04(04), 244–254. <https://doi.org/10.4236/ti.2013.44029>
- Frecknall-Hughes, J., Simpson, M., & Padmore, J. (2007). Inherent limitations in using financial ratio analysis to assess small and medium sized company performance [Working paper]. https://www.researchgate.net/publication/228484206_Inherent_limitations_in_using_financial_ratio_analysis_to_assess_small_and_medium_sized_company_performance.
- Freling, S. A., Hastings, A. (2022, February 23). DoD signals increased scrutiny of gov con M&A and renewed interest in background IP rights. Covington. <https://www.insidegovernmentcontracts.com/2022/02/dod-signals-increased-scrutiny-of-gov-con-ma-and-renewed-interest-in-background-ip-rights/>
- Gâdoiu, M. (2014). Advantages and limitations of the financial ratios used in the financial diagnosis of the enterprise. *Scientific Bulletin—Economic Sciences*, 13, 87–95.
- Ghosh, S. K., & Chaudhuri, A. (2017). Bankruptcy prediction through soft computing based deep learning technique. Springer Singapore. https://doi.org/10.1007/978-981-10-6683-2_3



- Gozali, R. & Panggabean, R. R. (2019). Mergers and acquisitions, their impact on liquidity, profitability, activity and solvency: Evidence from the Indonesian stock exchange. *Innovation: Scientific Journal of Management Science*, 6(1), 13–28.
- Greene, W. H. (2008). *Econometric analysis* (6. ed). Pearson, Prentice Hall.
- Hartung, W. D. (2021, September 13). Profits of war: Corporate beneficiaries of the post-9/11 pentagon spending surge. Watson Institute of International and Public Affairs. <https://watson.brown.edu/costsofwar/papers/2021/ProfitsOfWar>
- Higgs, R. (Ed.). (1990). *Arms, politics, and the economy: Historical and contemporary perspectives*. Holmes & Meier.
- Higson, C. J., & Elliott, J. (1993). Are acquired firms distinctive? The effect of choice of control. London Business School, 170–193.
- Hilmer, C. E., & Hilmer, M. J. (2014). *Practical econometrics: Data collection, analysis, and application*. McGraw-Hill Education.
- Horrigan, J. O. (1968). A short history of financial ratio analysis. *American Accounting Association*, 43(2), 284–294.
- Inoti, G. G. (2014). Impact of acquisitions on the financial performance of the acquiring companies in Kenya: A case study of listed acquiring firms at the Nairobi securities exchange. *Journal of Finance and Accounting*, 2(5). <https://doi.org/10.11648/j.jfa.20140205.12>
- Institute for Mergers, Acquisitions, and Alliances (IMAA). (March 15, 2024). M&A statistics. <https://imaa-institute.org/mergers-and-acquisitions-statistics/>
- Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *The American Economic Review*, 76(2), 323–329.
- Jensen, M. C., & Ruback, R. S. (1983). The market for corporate control: The scientific evidence. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.244158>
- Ji, X. (2016). How the GDP will affect M&A deals in U.S. Southern Illinois University Carbondale. https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1998&context=gs_rp
- Kovacic, W. E., & Smallwood, D. E. (1994). Competition policy, rivalries, and defense industry consolidation. *Journal of Economic Perspectives*, 8(4), 91–110. <https://doi.org/10.1257/jep.8.4.91>
- Kritsonis, A. (2005). Assessing a firm's future financial health. *International Journal of Management, Business, and Administration*, 8(1).
- Lambooy, J. G. (1990). Meso-economics and organizational ecology. In K. Peschel (Ed.), *Infrastructure and the space-economy* (pp. 254–273). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-75571-2_15
- Lev, B. (1969). Industry averages as targets for financial ratios. *Journal of Accounting Research*, 7(2), 290–299. <https://doi.org/10.2307/2489971>
- Lev, B., & Sunder, S. (1979). Methodological issues in the use of financial ratios. *Journal of Accounting and Economics*, 1(3), 187–210. [https://doi.org/10.1016/0165-4101\(79\)90007-7](https://doi.org/10.1016/0165-4101(79)90007-7)
- Mahoney, C. W. (2021). Acquire or expire: Publicly traded defense contractors, financial markets, and consolidation in the U.S. defense industry. *Defense and Peace Economics*, 32(3), 325–342. <https://doi.org/10.1080/10242694.2019.1667216>
- Mboroto, S. N. (2013). The effect of mergers and acquisitions on the financial performance of petroleum firms in Kenya [Master's thesis, University of Nairobi]. <https://www.semanticscholar.org/paper/The-effect-of-mergers-and-acquisitions-on-the-of-in-Mboroto/bc16b59136247323c2db8a47b863335afbcf0604>



- Merwin, C. L. (1942). Financing small corporations in five manufacturing industries, 1926–36. NBER. <https://www.nber.org/books-and-chapters/financing-small-corporations-five-manufacturing-industries-1926-36>
- Murphy, G. B., Trailer, J. W., & Hill, R. C. (1996). Measuring performance in entrepreneurship research. *Journal of Business Research*, 36(1), 15–23. [https://doi.org/10.1016/0148-2963\(95\)00159-X](https://doi.org/10.1016/0148-2963(95)00159-X)
- Myers, S. C. (1984). The capital structure puzzle. *The Journal of Finance*, 39(3). <https://doi.org/10.2307/2327916>
- Office of Management and Budget. (n.d.-a). Table 1.2—Summary of receipts, outlays, and surpluses or deficits as percentages of gdp: 1930–2029 [Data set]. The White House. Retrieved December 22, 2022, from <https://www.whitehouse.gov/omb/budget/historical-tables/>
- Office of Management and Budget. (n.d.-b). Table 5.1—Budget authority by function and subfunction: 1976–2029 [Data set]. The White House. Retrieved December 22, 2022, from <https://www.whitehouse.gov/omb/budget/historical-tables/>
- Office of the Under Secretary of Defense (OUSD) for Acquisition and Sustainment. (2022). State of competition within the defense industrial base. https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/reference_list_other_print_sources.html
- Officer, M. S. (2007). The price of corporate liquidity: Acquisition discounts for unlisted targets. *Journal of Financial Economics*, 83(3), 571–598. <https://doi.org/10.1016/j.jfineco.2006.01.004>
- Palepu, K. G., (1986). Predicting takeover targets: A methodological and empirical analysis. *Journal of Accounting and Economics*, 8(1), 3–35.
- Pastena, V., & Ruland, W. (1986). The merger/bankruptcy alternative. *American Accounting Association*, 61(2), 288–301.
- Peters, H. M. (2021). Defense primer: U.S. defense industrial base (No. IF10548, Version 9). Congressional Research Service. <https://crsreports.congress.gov>
- Powell, R. G. (2001). Takeover prediction and portfolio performance: A note. *Journal of Business Finance & Accounting*, 28(7 & 8), 993–1011. <https://doi.org/10.1111/1468-5957.00402>
- Rashid, C. A. (2021). The efficiency of financial ratios analysis to evaluate company's profitability. *Journal of Global Economics and Business*, 2(4), 119–132.
- Shalit, S. S., & Sankar, U. (1977). The measurement of firm size. *The Review of Economics and Statistics*, 59(3), 290–298. <https://doi.org/10.2307/1925047>
- Shleifer, A., & Vishny, R. W. (1992). Liquidation values and debt capacity: A market equilibrium approach. *The Journal of Finance*, 47(4), 1343–1366. <https://doi.org/10.1111/j.1540-6261.1992.tb04661.x>
- Smith, R. F., & Winakor, A. H. (1935). Changes in the financial structure of unsuccessful industrial corporations (Vol. 51). University of Illinois, Urbana.
- Susman, G. I., & O'Keefe, S. (Eds.). (1998). *The defense industry in the post-cold war era: Corporate strategies and public policy perspectives*. Pergamon.
- Suthar, K. U. (2018). Financial ratio analysis: A theoretical study. *International Journal of Research in all Subjects in Multi Languages*, 6(3), 61–64.
- Tikhomirov, A. F., Kalchenko, O. A., & Bogacheva, T. V. (2019, October 24). Assessment of the synergistic effect from solvency changes in M&A transactions based on a dynamic model. *Proceedings of the 2019 International SPBPU Scientific Conference on Innovations in Digital Economy*. <https://doi.org/10.1145/3372177.3373282>
- U.S. Code, Title 11, Pub. L. 95–598, title I, §101, Nov. 6, 1978, 92 Stat. 2549 (2011). <https://www.govinfo.gov/content/pkg/USCODE-2011-title11/html/USCODE-2011-title11.htm>



- U.S. General Accounting Office. (1998a). Defense industry consolidation: Competitive effects of mergers and acquisitions (Testimony GAO/T-NSIAD-98-112). <https://www.gao.gov/products/t-nsiad-98-112>
- U.S. General Accounting Office. (1998b). Defense industry: Consolidation and options for preserving competition (GAO/NSIAD-98-141). <https://www.gao.gov/assets/nsiad-98-141.pdf>
- U.S. Small Business Administration. (n.d.). Prime and subcontracting. Retrieved January 11, 2023, from <https://www.sba.gov/federal-contracting/contracting-guide/prime-subcontracting#section-header-0>
- Wang, J. (2009). The macro determinants of M & A timing in China. *International Journal of Business and Management*, 3(9), 141–146. <https://doi.org/10.5539/ijbm.v3n9p141>
- Wilcox, J. W. (1971). A simple theory of financial ratios as predictors of failure. *Journal of Accounting Research*, 9(2). <https://doi.org/10.2307/2489944>
- Yang, J., Guariglia, A., & Guo, J. (Michael). (2019). To what extent does corporate liquidity affect M&A decisions, method of payment and performance? Evidence from China. *Journal of Corporate Finance*, 54, 128–152. <https://doi.org/10.1016/j.jcorpfin.2017.09.012>
- Zietlow, J., Hankin, J. A., Seidner, A., & O'Brien, T. (2018). Developing financial reports and ratios: Making sense of the numbers. In *Financial Management for Nonprofit Organizations* (3rd ed., pp. 257–332). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119419242.ch7>



Reimagining Defence: Architecting Estonia's Force Management for the Uneasy Future

Ivo Peets—a Senior Officer in the Estonian Defence Forces, specializes in defence planning, capability development, and military technology. With experience from international combat and training missions, he is currently advancing his expertise through doctoral research on comprehensive defence planning methodologies for small states at the Naval Postgraduate School and Tallinn University. He is also responsible for coordinating the long-term force planning of the Estonian Defence Forces.

Abstract

This paper examines the Estonian Defence Forces' evolution from 1992 to 2023, during which the defence budget expanded over 100-fold, from just under \$12 million to beyond \$1.4 billion. This financial growth signifies an increased commitment to national defence but presents acquisition and force management challenges. Through analyzing the development of Estonia's defence enterprise and the country's experiences, this article outlines how Estonia is refining its defence resource management by developing a cohesive acquisition system. It highlights Estonia's progress towards sectoral maturity, confronting challenges unseen in the initial development phase. By incorporating the author's experience and proposals, the article aims to enrich the discourse on defence management. The findings underscore the importance of adaptable systems in maintaining defence capabilities amid rapid expansion, providing valuable lessons for similar transformations worldwide.

Keywords: defence planning, force development, force planning, Estonia, force management, acquisition, comprehensive national defence, PPBE

Introduction

Over the last decade, Estonia, like many other European countries, has faced increasing tension due to Russian aggression against its neighbors, necessitating a significant overhaul of its national defence planning and strategies (Flanagan et al., 2019, pp. 1–8). The security environment has emphasised the need to develop and modernise military capabilities to ensure Estonia's ability to respond effectively, resulting in additional resources for the defence sector.

To understand the challenges faced in the field today, it is necessary to understand how Estonia's present system has evolved. In the period leading up to Estonia's NATO accession in 2004, Estonian defence planning primarily drew on the experiences of established Western countries, as there was no tradition and history of Estonian defence planning due to the 50-year Soviet occupation after World War II.

Significant emphasis was placed on adopting the United States' practices as a model, a concept reinforced by the foreign advisors and advisory teams supporting Estonia (Luik, 2019; Kask et al., 2003; Murumets, 2007). By 2004, over 10 years had passed since the end of the Cold War, and NATO and its member states' defence policies and planning had reached the phase of implementing new conceptual starting points and principles (Monaghan, 2022). At the same time, NATO countries were implementing significant changes in defence planning (NATO Research and Technology Organization, 2003), which resulted in "hesitations" and left the defence planners searching for new forms and content. Against this backdrop, one can understand why Estonian defence planners faced difficulties adopting and comprehending the methods to their finest detail – part of the solution either relied solely on theory or did not meet Estonia's needs. Considering Russia's existential threat, which has consistently been a central focus of Estonia's defence policy and military strategy, the defence planning approaches common in the early 21st century appeared either inappropriate or excessively cumbersome,



resulting in their quick classification as superfluous. These short-term compromises have resulted in enduring challenges, with issues whose adverse effects were not visible in the past.

Specifically, initial assessments of how forces and capabilities could be managed, including the role of acquisition, have diverged significantly from the early 2000s compared to the realities faced by the end of 2023. One major factor is that the defence budget's growth (Figure 1) from under \$12 million in 1991 to over \$1.4 billion by 2024, marking over a hundredfold increase in three decades, has enabled enhanced equipment procurement and force expansion. This expansion of resources has not only enabled the procurement of more equipment and the sustenance of a more significant force but has also introduced new requirements and showed a gap between assumptions and actual needs.

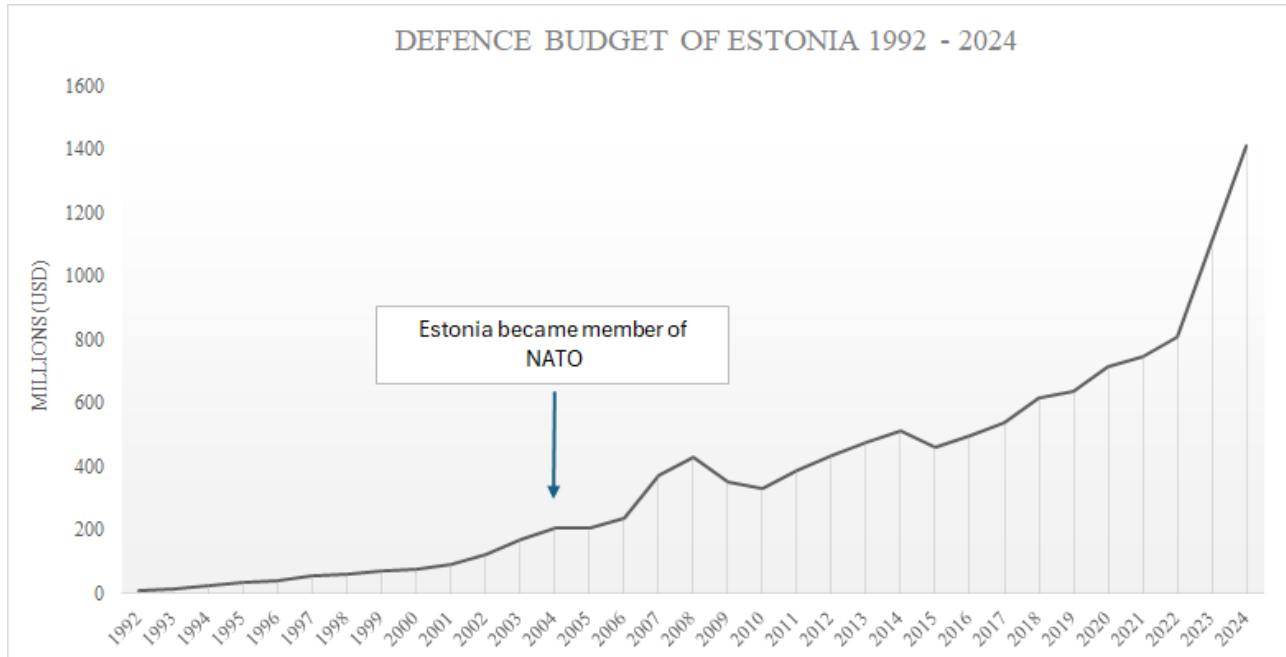


Figure 1. Defence Budget of Estonia, 1992–2024

The Estonian Centre for Defence Investment (ECDI) initiation in 2017 highlights the acquisition system's gradual progression towards maturity. The ECDI, responsible for consolidating defence procurements, has streamlined processes and fostered better civil–military cooperation. However, this centralization comes after two decades post-reestablishment, when Estonia's Defence Force relied heavily on foreign aid, with minimal focus on comprehensive planning for equipment life cycles or replacing aging assets.

The ECDI, under the Ministry of Defence's jurisdiction, now manages procurements and infrastructure, addressing the growing volume and complexity of both procurements and real estate. Established to enhance transparency and efficiency in procurement activities, the ECDI has unified management processes to prevent duplication and conserve resources. This restructuring has markedly improved the oversight of procurement activities within the Ministry of Defence, signifying a progression towards a more mature acquisition system. This advancement towards maturity emphasizes the need for better equipment and capability life-cycle management and necessitates a more refined force management system. Such a system must align closely with force requirements, integrating all stakeholders—including the Ministry of Defence, the ECDI, and the Estonian Defence Forces—thereby ensuring that procurement



activities are fully synchronized with the nation's and defence forces' strategic and operational needs.

This paper examines and explains the existing challenges within the Estonian defence planning, acquisition system and sheds light on forthcoming reforms. A review of official documentation and working notes produced during the formulation of national defence development plans over the past decade informs the content of this document. It incorporates practical insights from the author's leadership in the force development project and participation in defence planning.¹ Additionally, it encompasses notes from personal interviews conducted with officials from the Ministry of Defence and the Defence Forces Headquarters and incorporates selective outcomes from the author's doctoral research.

This blend of sources provides a comprehensive foundation for the analysis and conclusions presented herein, ensuring they are well-grounded in documented evidence and firsthand experience. While qualitative and quantitative academic methodologies underpin this analysis, the intent is to produce something other than a traditional academic paper. Instead, it is a waypoint to describe the current state of Estonia's defence planning and acquisition and offer context for possible future developments grounded in the logic prevailing in the Estonian context. The aspiration is that this paper facilitates the sharing of Estonia's experiences, rendering this evolving system comprehensible to allies and partners and engaging a broader audience by being transparent and relatable. The author believes there is much to gain from openness, the ability to benchmark and learn from others, and potentially providing lessons in return.

Strategic Planning Framework

National defence aims to preserve Estonia's independence and sovereignty, as well as the undivided integrity of its land, territorial waters, airspace, and constitutional order. To achieve this goal, the entire society and the state's resources and reserves are mobilized based on the principles and procedures established by law (Parliament of Estonia, 2015).

While Estonia adheres to the principles common in the defence planning systems of other NATO member states, specific details are unique to its environment stemming from its prevailing laws. Defence planning is a component of the nation's strategic planning, which only sometimes allows for the direct transference of principles used in other countries. Estonia's strategic planning framework is derived from the State Budget Act, which outlines the general principles and types of strategic development documents in sections 19 and 20. This is further complemented by the Government of the Republic's regulation No. 117 of 2019, detailing the procedure for the preparation, implementation, reporting, evaluation, and amendment of sectoral development plans and programs (Government of Estonia, 2019).

Within the State Budget Act context, strategic development documents include the foundational principles of policy, strategic objectives, sectoral development plans, and programs. Metaphorically speaking, the foundational principles of policy and strategic objectives set the direction, the sectoral (or ministry) development plan marks the destination, and the program outlines the route to achieving the goal. The foundational principles of policy are adopted by the Riigikogu (Parliament), the sectoral development plan by the Government of the Republic, and the program (ministry action plan) by the responsible minister or ministers (Government Office of Estonia, 2021).

¹ The Force Development project was initiated by the Chief of Defence (CHOD) in 2020 to review the methodologies employed in defence and force planning. Its mandate includes coordinating the development of force structure inputs essential for the national defence development plan. This initiative ensures that planning processes are efficient and aligned with established objectives.



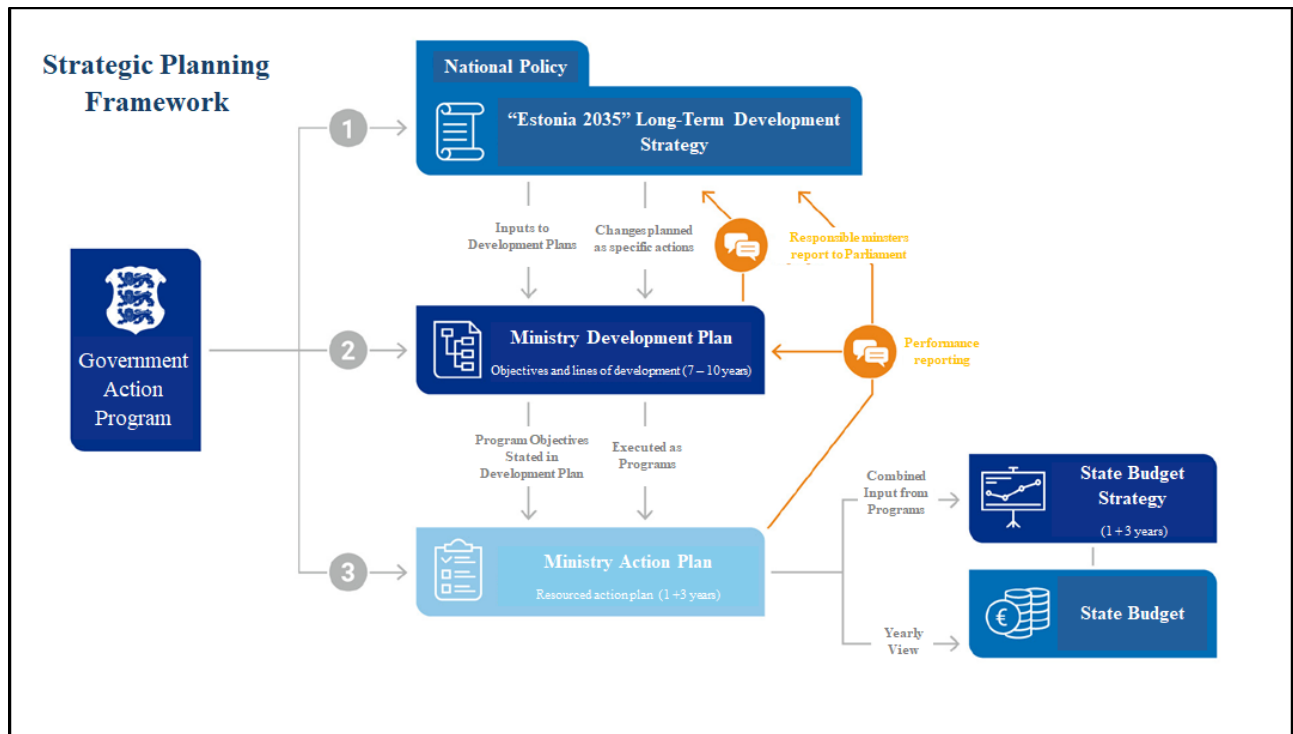


Figure 2. Estonian Strategic Planning Framework (Government Office of Estonia, 2021)

Drafting Estonia’s strategic development plans for defence and the national defence action plan encompasses the entire society, utilizing the state’s resources and reserves. This broad-based approach facilitates the cross-utilization of limited resources in addressing various threat scenarios and necessitates additional coordination.

Comprehensive National Defence

Defence planning in Estonia functions within the broader scope of national strategic planning, incorporating a comprehensive national defence approach. This approach signifies the integration of defence planning into the nation’s overarching strategic framework, engaging diverse stakeholders across the nation (Figure 3). The initiation of national defence is predicated on identifying existential threats, with the output being the mobilization of national resources to counteract these threats. The foundational document is the national security policy, which outlines broad objectives and establishes the nation’s posture against existential threats. This document is refined into specific political guidance, translating into actionable strategies.

The national defence strategy is an essential component that addresses existential threats to Estonia, with numerous sub-strategies linked to the primary strategy, each corresponding to particular threat vectors. For instance, the military strategy addresses military-specific threats (NATO, 2021), while other strategies like crisis management cover non-military threats. Grounded in the national security policy, these sub-strategies are designed to be developed and implemented in coordination, allowing Estonia, as a small nation, to be more resilient against existential threats by sharing resources – trading more efficient use of resources for higher coordination requirements between stakeholders.

In the military context, defence planning involves formulating a military strategy and allocating resources to develop a suitable force structure. This planning is executed in the context of defined capabilities, which will be detailed later. The ultimate goal in addressing military existential threats is to devise a force structure that aligns with the national security

policy and the derived political guidance, thus ensuring that strategic objectives are fulfilled through a capable and ready military force.

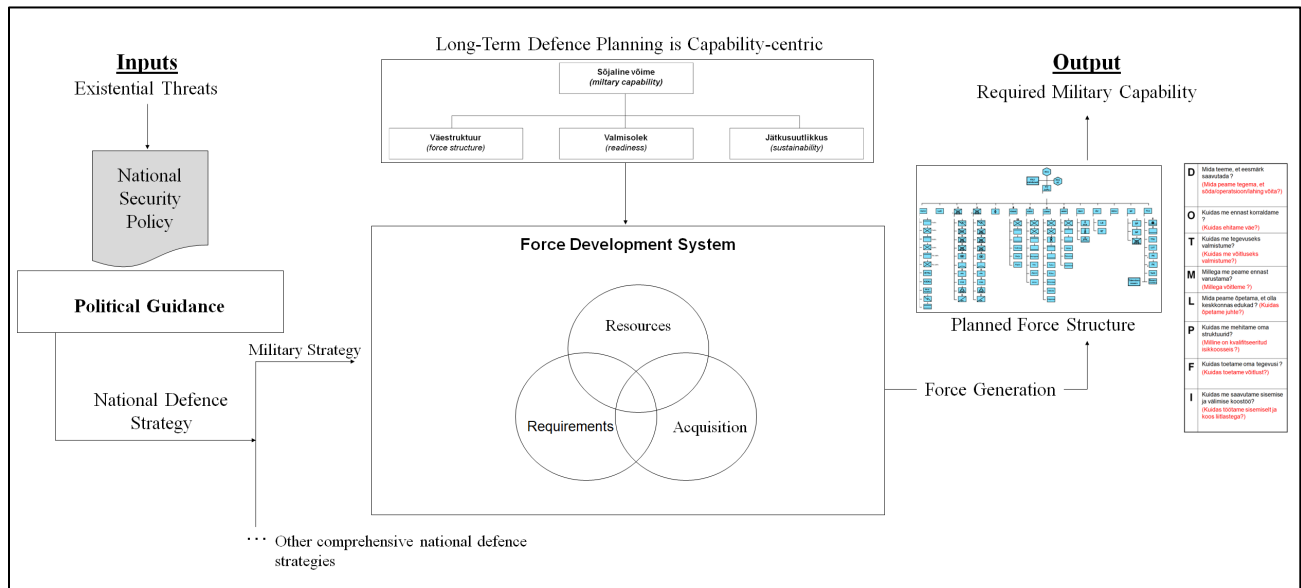


Figure 3. National Defence Model

Drawing insights from the development of three National Defence Development Plans (NDDPs),² the author identifies three essential areas: planning, strategy, and management. These are fundamental to national defence and require detailed explanations. The planning, encompassing defence planning and its associated processes, alongside strategy formulation, have both achieved a mature state and thus are not the main focus of this paper. However, they will be briefly outlined for context.

The emphasis of this paper is on management, reflecting its significant role in national defence. Effective management is critical for integrating national defence objectives with practical aspects of procurement, highlighted by the creation of the ECDI. However, based on recent experience, centralized procurement is only part of the solution that is required to establish a mature force management system. **Top of Form**

Defence Planning

Defence planning is a political and military activity nations undertake to ensure the capabilities necessary for their defence. In the Estonian context, this involves building the country's defence capabilities based on the current security strategy, focusing on the systematic development of the nation's military capabilities. This process forms a long-term strategic outlook for military defence, linking the country's military needs and resources. It sets priorities for the development of national defence and identifies the resources required for capability development (personnel, equipment, infrastructure, etc.). Defence planning, as a logic to align resources and objectives, is guided by existing constraints and future possibilities with clear military objectives, integrating the main directions of a security policy with measurable goals, activities, capabilities, and units.

² The NDDP is the central capability planning document of the national defence. The plan's objective is to identify, based on existing threat scenarios, necessary non-military and military capability developments for the next decade.

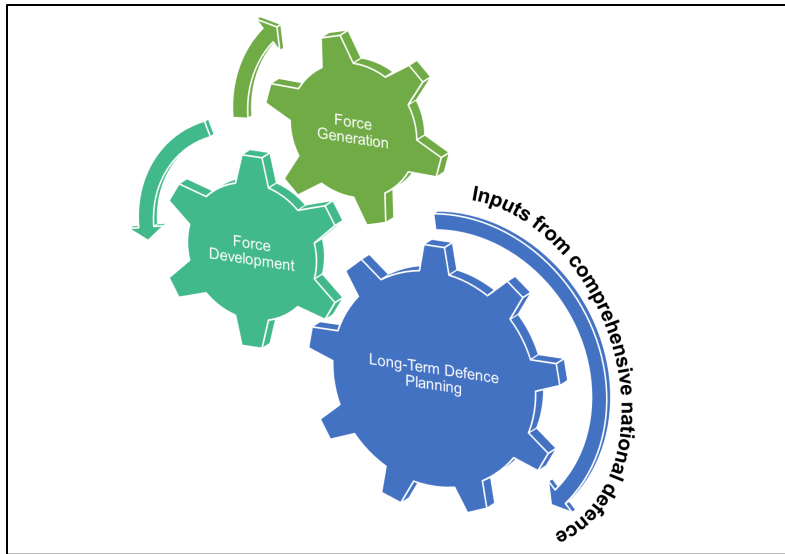


Figure 4. Key Processes for Developing Estonia's Military Defence Capability

It has been established that Estonia's military and national defence development is based on three interrelated processes (Figure 3): long-term defence planning, force development, and force generation. These processes ensure the effective coordination of defence activities across different periods, enabling the country to adapt to changing security conditions and secure its future security.

During the long-term defence planning process, the country's military threats, the capabilities needed to counter these threats, and the required resources are specified. This resulted in the decision to develop the military force structure over a 10-year period. Force development then focuses on developing the decided force structure and its military capabilities, including the procurement of decided resources and the precise allocation of resources. The force generation process regulates forces' actual organization, equipping, and training to ensure their readiness and effectiveness.

Significant experience has been accumulated in implementing these three processes separately, but more experience is needed in integrating these processes and identifying clear interrelations.

The Long-Term Defence Planning Process

Long-term defence planning aims to clearly define the military capabilities to be developed over the next 10 years, along with the associated force structure and composition. This process produces the military section of the NDDP, which the Ministry of Defence leads in compiling.

Long-term defence planning is conducted every 4 years to specify military capability needs and update the target level of the force structure being developed. The NDDP or its components may also be updated based on specific events. However, clarifying the military capability requirements that underpin the force structure is an ongoing activity.

In classical terms, Estonia's military strategy is articulated through two primary documents: the NDDP and the Defence Forces Capability Description.

The Defence Forces Capability Description aims to describe the military component (ends-ways) precisely. It includes the force structure to be developed in the following 10-year timeframe, while the NDDP provides a resources (means) perspective and bridges military

objectives with non-military ones. This combination offers a comprehensive approach to implementing military strategy, linking strategic goals and methods with the resources necessary for their achievement. This integrated approach facilitates the coordination of military requirements and objectives (ends) with the overall directions of national defence development, ensuring that military planning and resource allocation occur within a unified system, reflecting the broader logic of force development, where the Defence Forces define military requirements. However, the ministry, policymakers, and the public decide to allocate resources.

Planning, Programming, and Budgeting

The Planning, Programming, and Budgeting System (PPBS) was introduced in the U.S. Department of Defense (DoD) in the 1960s (Enthoven & Smith, 1971). This approach was also applied to developing Estonia's defence planning methodology in the early 2000s, as Kask et al. (2003) outlined. However, as noted by officials working at the Ministry of Defence, it has never been strictly adhered to in detail. Thus, the PPBS's U.S.-centric nature has encountered challenges within the Estonian context before, notably due to its close association with the U.S. Congress's annual budget allocation procedures. This method contrasts with Estonia's approach,³ where the defence budget is determined as a percentage of GDP and approved by the Riigikogu (Parliament). This fundamental difference underscores the need for a more customised adaptation of the PPBS principles to suit Estonia's specific defence planning and budgeting requirements.

The discrepancy between the U.S. and Estonian budgeting processes has necessitated focusing on activities not directly applicable in the Estonian context, where decisions are made while formulating long-term development plans. Considering the vast differences in scale and complexity—with the U.S. defence budget significantly more extensive than Estonia's—direct transplantation of the PPBS model is impractical. Therefore, a foundational review of the PPBS principles is essential to discern which elements could beneficially inform Estonia's defence planning and resource management principles.

The Commission on PPBE Reform⁴ underscores the importance of a strategy-driven resource allocation process characterised by detailed analysis and collaborative execution (Commission on PPBE Reform, 2024). This feedback is particularly relevant for Estonia, which, despite already having a maturing defence planning system, lacks a fully established force management framework. Together with the Commission's report, integrating methodologies inspired by the evolution of PPBS, including the Joint Capabilities Integration and Development System (JCIDS Manual, 2021) and acquisition processes (DoD, 2020), into Estonia's defence planning architecture could mark a critical step towards establishing a robust force management system.

Force Development

Given the nature of defence planning and the necessity for effective resource allocation, insights from the Commission on PPBE Reform (2024) in the United States present a valuable perspective for enhancing Estonia's existing defence planning processes. By selectively embracing these principles, Estonia can substantially refine its defence planning and budgeting processes, achieving greater strategic coherence and operational agility. This selective

³ In addition, there are multiple issues that are unique to the U.S. context, such as the historic independence that services have maintained. This has led to stovepiped development and the necessity to emphasize jointness in planning and execution between services. This issue is not relevant in the Estonian context, as all services are de facto under the single authority of the Chief of Defence.

⁴ Congress established the Commission on PPBE Reform in the National Defense Authorization Act for Fiscal Year 2022 to conduct a comprehensive assessment of all four phases of the PPBE process, with a specific focus on budgetary processes affecting defence modernization.



integration strategy—leveraging the U.S. experience as a blueprint for developing a tailored Estonian force management system.

When adopting principles used by other nations, it is crucial that they are well-documented, readily accessible, and have been thoroughly examined in academic literature, with significant consideration given to their impact on the development of the field in a specific country. The force development system employed by the DoD meets these criteria (Defense Systems Management College, 2022), fundamentally consisting of three main components: resources, requirements, and acquisition (Figure 5).

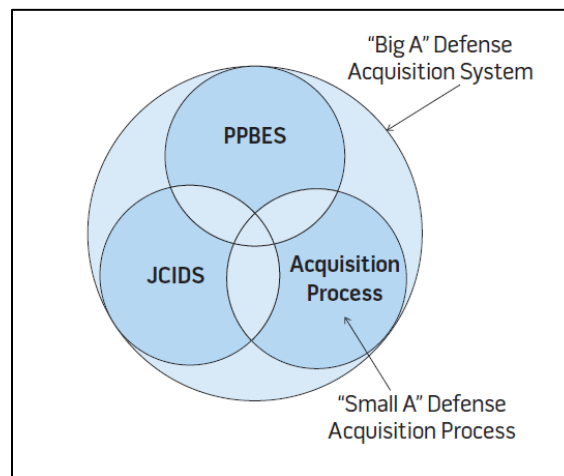


Figure 5. Components of Generic Force Development System (Kadish et al., 2005)

Force development does not occur in a vacuum; it is based on guidelines, including the objectives for force structure, developed during the long-term national defence planning process. It is important to note that, compared to the PPBE system implemented in the United States, which encompasses both long-term defence planning and annual budget distribution, these processes are temporally staggered in Estonia – the detailed budget division crucial for force development occurs annually, but the long-term defence planning and the long-term development plan (NDDP) that underlie the force development plan are prepared every 4 years. It must be noted that requirements for force structure are updated and evaluated continuously, although this system faces similar management challenges we will discuss later in detail. Additionally, Estonia's comprehensive approach to security and broad-based national defence requires the inclusion of stakeholders responsible for developing non-military capabilities, widening the process scope.

Force Planning

Force planning, by definition, means designing a force that meets military needs, and this process encompasses the entire defence force and its smaller components. In Estonia, the term force planning has been loosely used and not even officially defined as a term before 2023, highlighting the previous neglect, but growing importance, of the process and its nuances in the future.

Force planning is a process of transforming abstract capabilities into a concrete, developable force structure; it is about designing capability carriers that meet the criteria set by capability requirements. In the context of defence planning, the force planning process is a process to design a force structure (Davis, 2002, pp. 9–11) that meets the needs of national defence, encompassing capabilities decided for development during the defence planning process.



Force planning (as interpreted in the context of Estonian defence planning) combines two views: the force view and the capability view of the force structure (Figure 6). The force view focuses on the stovepipe view of a collection of capabilities organized by specific functions and roles tied to a higher purpose and objective. Examples are the services and units in the services, like battalions, brigades, etc., which address how these units are organized. It can be defined as a vertical view of the force structure. In contrast, the capability view concentrates on these structures' specific military capabilities, considering various operational requirements and functions. It can be defined as a horizontal view of the force structure. This distinction between the two perspectives necessitates an integrated approach in force planning, where both aspects—structural organisation and operational capabilities—are interwoven. Such an approach ensures that the force structure is designed comprehensively, considering current and future operational needs.

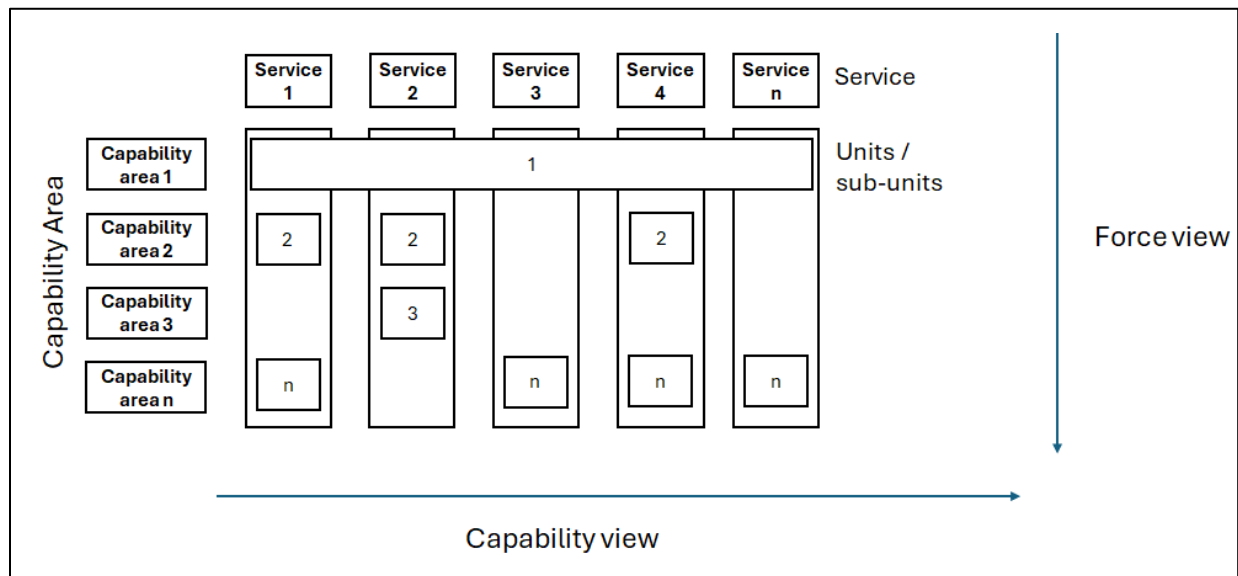


Figure 6. Force Structure Perspectives – Capability View and Force View

Force planning's (force view) detailed time horizon is deemed to be no more than 10–15 years; anything beyond that is not feasible, and the number of alternatives makes it too resource-demanding to manage.

The first 10 years focus on aligning the force's target level with resources. At the same time, the subsequent 5-year period informs about opportunities and constraints that help planners make more informed decisions. Capability planning (capability view) concentrates on one or multiple services. It can extend up to 50 years into the future,⁵ being more flexible and considering changing factors such as technological trends and demographic changes. Integrating the capability perspective helps reduce fragmentation among types of forces, providing an overview of how to adapt and maintain the force structure cost-effectively and keep it up to date to respond to potential threats efficiently.

⁵ An example of this concept is viewing capability as multiple alternatives for existing or required capability. For mobility capability, this perspective would consider the entire life cycle of existing trucks and project how technology might influence potential alternatives over a 50-year span, including the integration of new technologies. Therefore, the decision to conduct a mid-life upgrade in 10–20 years would depend on factors such as the availability of alternative technologies, like transitioning from diesel to all-electric trucks. These decisions on capability will affect the entire force structure, as capability is viewed as a whole and not dependent on where it is located within the force structure.

Force Management

The importance of and need for force management have been highlighted for Estonia, especially after the surge in defence expenditures post-2022 and the significantly increased share of resources in the domain. Moreover, there has been an emergence of demands for the readiness of the force structure and the administrative load resulting from the mobilization of resources.

It had become apparent that overseeing over 40,000 military personnel and assets valued in the tens of billions of euros necessitates an approach markedly distinct from the management of force structure initiated in the 1990s, when the yearly defence budget was merely a hundredth of its current size. Thus, the defence forces and the defence sector are progressing towards a new level of maturity, which presupposes a different treatment of certain aspects. This directly influences the roles of various parties, including the role of the Chief of Defence (CHoD) and the processes, a realization that has already begun. The following section outlines principles of how force management will be implemented to comprehend the changed situation. Descriptions are intended to be less detailed and describe the main principles, as details are still to be decided.

Force management focuses on organising activities related to creating, maintaining, and employing combat-capable units. Thus, it must be comprehensive, combining three main components: requirements, resources, and acquisition (Figure 7).

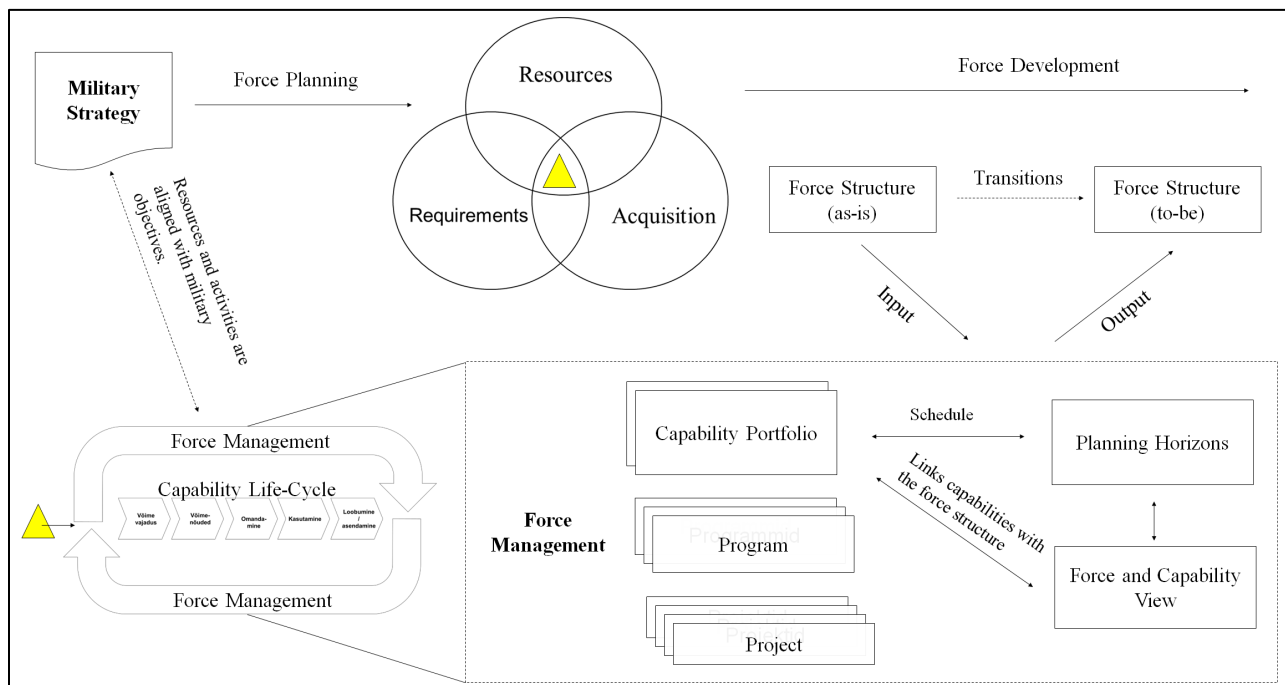


Figure 7. Proposed Force Management System

The CHoD is responsible for developing the country's military capabilities. Its role encompasses the management and direction of strategic planning, ensuring that the development of the Defence Forces aligns with the national defence's overall needs and objectives.

Centralised management by capability portfolios is critical to ensure that changing requirements are operationally addressed across the entire force structure and that decisions are consistent with operational requirements, available resources, and activities in developing

and integrating capabilities into the force. Force management is organised by capability portfolios, with centralised management.

Force management encompasses activities related to achieving the objectives of the existing, planned, and future force targets. This involves orchestration to unify activities across defence and force planning, as well as force development and implementation, encompassing the entire force development system. At its core lies cooperation between internal and external parties. For internal participants, it represents guidance for systematically organising their activities, while for external participants, it offers structured input and coordinated engagement opportunities (Figure 7).

Force management is closely associated with military requirements, resource allocation, and acquisition processes, integrating project management principles, processes, tools, and techniques. It aims to establish a solid foundation for effectively implementing the organisation's strategic goals and decisions. Moreover, force management includes planning activities and their practical arrangement, linking resources and actions to the strategy, and creating conditions for achieving planned military objectives.

Force management within Estonia's defence sector is, at the time of writing, in the process of being structured into distinct capability portfolios. These comprehensive groupings collect related programs and projects for specific military capability areas. Each portfolio is designed to address a range of capabilities, ensuring that all aspects of the force structure—from operational readiness to logistics—are methodically developed and enhanced. Capability areas, such as logistics, communications, etc., organize the portfolios. Within each area, a series of programs and projects are initiated to meet the current and future needs of the force. This methodical organisation prevents overlapping efforts and ensures that resources are allocated efficiently and effectively. The scope and content of these portfolios are dynamic; they are regularly reviewed and adjusted in line with the evolving requirements of the force structure. The specific details and configurations of the portfolios are revisited and defined in each new version of the NDDP, which acts as a guiding document for Estonia's defence planning and capability development. This process underscores the adaptive nature of force management, ensuring that Estonia's defence capabilities remain aligned with strategic objectives and operational demands. If similar or identical capabilities are present in multiple portfolios, a decision must be made to consolidate them into a chosen portfolio. For instance, all transportation means utilized within the force structure would belong to one portfolio (e.g., mobility portfolio). In contrast, the communications portfolio would not include transportation means but would encompass communication solutions compatible with the transportation means.

Capability programs are enduring, managing the entire development cycle of a specific capability, including all related systems and resources. The scope and content of a program are defined according to current and future needs. The methodology for managing programs and projects within force management emphasizes an integrated approach, enabling the achievement of synergies that are not possible through the independent management of individual projects.

Projects are established to implement a specific part of a program and are characterised by a defined goal, tasks, start, and end. Several projects can be underway for one program at a time, and projects may also cross programs, fulfilling the needs of a specific portfolio. Projects can also be ongoing and routine, usually associated with maintaining a capability. Programs and projects are methodically structured according to a capability perspective, ensuring a strategic correlation with the existing, anticipated, and future force structures alongside their requisite resources.



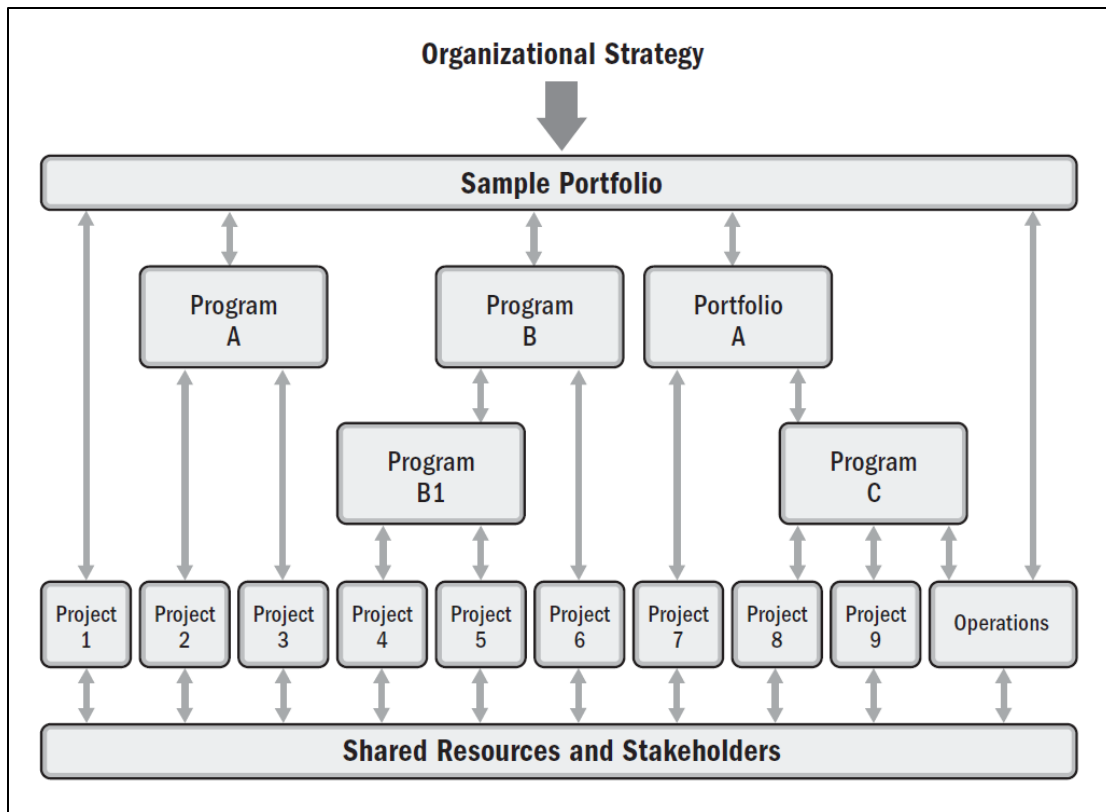


Figure 8. Example of Portfolio, Programs, Projects, and Operations (Project Management Institute, 2021)

The structure of the capability portfolio sets the foundation for aligning related programs, projects, and operations with the strategy’s goals and the allocated resources involving relevant stakeholders.

Relationships between capability programs and projects:

1. The primary goal of the capability portfolio structure is to ensure that all associated programs and projects are closely related and aligned with the organization’s strategic objectives and allocated resources. This structure guarantees that all activities support each other and are aimed at achieving common goals.
2. Each capability portfolio program consists of interconnected, specific projects designed to complement each other and support the main program’s objectives. This approach creates synergy and efficiency, which would be challenging if projects were managed separately.
3. In addition to long-term programs, the portfolio includes temporary projects to achieve specific, short-term goals. These projects may focus on developing new technologies, specific innovations, or improving operational processes, contributing to the overall portfolio strategy.

Strategic Planning and Management:

1. The planning and management of capability portfolios within the defence forces and the Ministry of Defence’s jurisdiction are conducted in alignment with the organization’s broader strategy. This ensures that each capability program and project’s objectives,

resources, and action plans are aligned with the organization's strategic goals and priorities.

2. The capability portfolio structure allows flexibility and adaptability to respond to changing demands and challenges while ensuring the optimal use of resources and achievement of goals.

Resource Management:

1. The allocation and management of resources within a capability portfolio are critical to ensuring that each program and project has the financial, human, and technological resources to achieve its objectives.
2. Effective resource management and allocation within the capability portfolio contribute to overall organizational efficiency and capability. This ensures that all projects and programs are directed towards creating maximum strategic value and supporting the organization's overall mission and vision.

The fundamental principle guiding portfolio management is aligned with the methodologies established by global standards (Project Management Institute, 2021, pp. 11–18). This approach aims to harmonize unique defence requirements with the world-renowned methodologies of the Project Management Institute, leveraging best practices to optimize and elevate the underlying logic of these initiatives.

Conclusion

The last 10 years have represented a critical juncture for Estonia, characterised by an unmistakable intensification of regional tensions that prompted an exhaustive revision of its defence enterprise. This period has been distinguished by deliberate efforts to bolster military capabilities and witnessed a remarkable expansion of Estonia's defence budget. This financial trajectory signifies Estonia's dedication to fortifying its national defence. However, this financial growth also introduced new challenges, particularly in terms of managing and planning the force.

Thus, the emphasis on management within the national defence framework is at the heart of Estonia's future defence evolution. Establishing a centralized procurement system through the ECDI has profoundly influenced management's role in bridging the gap between defence objectives and the practicalities of procurement and capability development.

Estonia is strategically aligning its defence planning and budgeting framework with agile and coherent principles by selectively adopting recommendations from the PPBE Reform Commission. Integrating advanced methodologies such as the Joint Capabilities Integration and Development System (JCIDS) and refining acquisition processes are key steps toward establishing a robust force management system. As Estonia enters a new era in defence planning, comprehensive integration of requirements, resources, and acquisition, along with dynamic capability portfolio management, are crucial for building an adaptable and operationally ready defence force for an uneasy and uncertain future.

Disclaimer:

A.I. tools were used to enhance the quality of translation of text from Estonian to English. Tools used include Grammarly, ChatGPT, and Google Translate.

The ChatGPT prompt used was: translate estonian text into english, use defence planning and military terms, do not add any new ideas, only direct translation is allowed.



References

- Commission on PPBE Reform. (2024). *Final report*. Retrieved from <http://ppbereform.senate.gov/finalreport>
- Davis, P. K. (2002). *Analytic architecture for capabilities-based planning, mission-system analysis, and transformation*. RAND Corporation. https://www.rand.org/pubs/monograph_reports/MR1513.html.
- Defense Systems Management College. (2022). *A guide to DoD program management business processes* (Version 1.1). Defense Acquisition University.
- DoD. (2020). *The defense acquisition system* (DoD Directive 5000.01). Office of the Under Secretary of Defense for Acquisition and Sustainment. <https://www.esd.whs.mil/DD/>
- Enthoven, A. C., & Smith, K. W. (1971). *How much is enough? Shaping the defense program, 1961–1969*. Harper & Row, Publishers.
- Flanagan, S. J., Osburg, J., Binnendijk, A., Kepe, M., & Radin, A. (2019). *Deterring Russian aggression in the Baltic states through resilience and resistance*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2779.html
- Government of Estonia. (2019). *Procedure for the preparation, implementation, reporting, evaluation, and amendment of sectoral development plans and programs*. RT I, 12.23.2019, 5. <https://www.riigiteataja.ee/akt/123122019005>
- Government Office of Estonia. (2021). *Strategic planning manual*.
- JCIDS Manual. (2021). *Manual for the operation of the joint capabilities integration and development system*.
- Kadish, R. T., Abbott, G., Cappuccio, F., Hawley, R., Kern, P., & Kozlowski, D. (2005). *Defense acquisition performance assessment report*. Office of the Deputy Secretary of Defense.
- Kask, A., Murumets, J., & Young, T.-D. (2003). *Approaching the need for defence reform: Background and outlines of suggested Estonian defence planning system*. Estonian National Defence College.
- Luik, J. (2019, April 22). *Estonia's long road to NATO*. International Centre for Defence and Security. Retrieved from <https://icds.ee/en/estonias-long-road-to-nato/>
- Monaghan, S. (2022, June). *Resetting NATO's defense and deterrence: The sword and the shield redux*. Center for Strategic & International Studies. Retrieved from <https://www.csis.org/analysis/resetting-natos-defense-and-deterrence-sword-and-shield-redux>
- Murumets, J. (2007). *Renewed national defence planning management: Capability based planning, programming, budgeting, and execution system for small states*. Estonian Military Academy.
- NATO. (2021). *NATO warfighting capstone concept*. <https://www.act.nato.int>
- NATO Research and Technology Organization. (2003). *Handbook on long term defence planning*. Retrieved from <https://apps.dtic.mil/sti/citations/ADA414193>
- Parliament of Estonia. (2015). *National Defence Act*. RT I, 03.14.2023, 31.
- Project Management Institute. (2021). *A guide to the project management body of knowledge (PMBOK® guide)* (7th ed.).



Defense Civilian Training Corps (DCTC) Integrated Curriculum and Development Pilot-0 Implementation

George P. Korfiatis, PhD—is currently a professor of environmental engineering at Stevens Institute of Technology. From 2006 to 2016, Dr. Korfiatis served as Provost and University Vice President at Stevens Institute of Technology, acting as chief academic officer of the university. Previously Dr. Korfiatis served as Dean of the Charles V. Schaefer, Jr. School of Engineering at Stevens, and Founding Director of the Stevens Center for Environmental Engineering, where he managed research programs valued at more than \$70 million, many of them funded by the DoD. Dr. Korfiatis has co-authored nine U.S. patents and was a co-founder of two Stevens enterprise companies, PlasmaSol Corp. and HydroGlobe LLC, both later acquired by Fortune 500 companies. He is a Fellow of The National Academy of Inventors. (gkorfiat@stevens.edu).

Michael Orosz, PhD—directs the Decision Systems Group at the University of Southern California's Information Sciences Institute and is a Research Associate Professor in USC's Sonny Astani Department of Civil and Environmental Engineering. Dr. Orosz has over 30 years' experience in government and commercial software development, systems engineering and acquisition, applied research and development, and project management and has developed several successful products in both the government and commercial sectors. Dr. Orosz is currently investigating best practices and lessons learned when introducing agile and DevSecOps processes into traditional waterfall-focused DoD acquisition programs. (mdorosz@isi.edu).

Dinesh Verma, PhD—served as the Founding Dean of the School of Systems and Enterprises at Stevens Institute of Technology from 2007 through 2016. He currently serves as the Executive Director for the Systems Engineering Research Center (SERC), a U.S. Department of Defense–sponsored University Affiliated Research Center (UARC) focused on systems engineering research, and for the Acquisition Innovation Research Center (AIRC). During his 20 years at Stevens, he has successfully proposed research and academic programs exceeding \$200 million in value. Prior to this role, Dr. Verma served as Technical Director at Lockheed Martin Undersea Systems, in Manassas, VA. In addition to his publications, Dr. Verma has received three patents in the areas of life-cycle costing and fuzzy logic techniques for evaluating design concepts. (dverma@stevens.edu).

Contributing Authors—Philip Anton, AIRC Chief Scientist; Karen Thornton, AIRC Fellow; John Willison, AIRC Fellow; Mark E. Krzysko SES, Deputy Director, Acquisition Policy & Innovation (API), OUSD (A&S); Garry Shafovaloff, Defense Acquisition University

Abstract

The FY 2023 DoD Consolidated Appropriations Act, enacted on December 29, 2022, provides funding for DCTC pilot program implementation with the purpose to prepare college students for “Department of Defense careers relating to acquisition, digital technologies, critical technologies, science, engineering, finance, and other civilian occupations determined by the Secretary of Defense,” under the auspices of OUSD(A&S). The DoD commissioned the Acquisition Innovation Research Center (AIRC) to (a) convene a DCTC Execution Planning Panel to provide recommendations for program implementation, and (b) to develop and execute a pilot DCTC program. Pilot-0 has been implemented as a 2-year program at four land grant universities namely, North Carolina A&T State University, Purdue, University of Arizona, and Virginia Tech in the fall semester 2023, attracting a cohort of 89 rising junior scholars enrolled in 45 different majors. The DCTC Integrated Curriculum and Development (ICAD) philosophy is founded on the principle of seamlessly integrated multidimensional learning experiences and includes the following five elements: (1) a core curriculum; (2) project-based cohort summer internships; (3) action (immersive) learning experiences; (4) innovation capstone projects; and (5) a culture of care. This paper presents our experiences in designing and implementing the five ICAD elements within the DCTC Pilot-0 program execution. This includes the unique integration of the five ICAD elements, the challenges and initial lessons learned, program assessment metrics and process, and the results of the first round of assessment of the scholar learning outcomes. This integrated



education and development approach is consistent with aligning, engaging, and preparing students for a robust career in public service earlier in their academic journey.

Introduction

There is an overwhelming recognition in recent years by the highest levels of U.S. national leadership of the need for significant targeted investments in the people of the defense enterprise to achieve a highly skilled, innovative, modern, and resilient civilian workforce. The rationale, justification, and urgency of this can be found in many reports including: *2022 Defense Business Board*, *POTUS 2022 National Security Strategy*, *POTUS 2021 Management Agenda Vision*, *2022 National Defense Strategy (SECDEF)*, and *2022–2026 DASD (CPP) Human Capital Operating Plan*. This highly critical need has compelled the U.S. Congress and OUSD(A&S) to fund and launch the DCTC pilot program.

DCTC is an integrated education and development program with a mission “*to develop a highly skilled talent pipeline to fill critical skill gaps in DoD occupations related to acquisition, digital technologies, critical technologies, science, engineering, and finance. Investing in high-performing students through a prestigious scholarship and accelerated career development program to keep pace in the great power competition.*” In September 2022, AIRC commissioned the DCTC Execution Planning Panel to develop recommendations for program implementation. The panel report provided 30 recommendations, which serve as a guide to the philosophy, design, and implementation of the DCTC program (Korfiatis et al., 2023). Although the original plan was to start Pilot-1 implementation in the fall of 2024, upon request of the sponsor, AIRC proceeded with Pilot-0 implementation starting in the fall semester of 2023. Pilot-0 is being implemented as a 2-year program at four land grant universities namely, North Carolina A&T State University, Purdue, University of Arizona, and Virginia Tech, attracting a cohort of 89 rising junior scholars enrolled in 45 different majors. AIRC manages the design, development, and implementation of the DCTC pilot program through four main working groups (WGs) namely, the Curriculum WG, the Student Management WG, the Strategic Partnering WG, and the Strategic Communications WG.

The overarching goal of the pilot is to develop, test, optimize, and transition the DCTC program that will become the catalyst in developing the next generation of a highly skilled DoD civilian workforce and leaders in future areas of critical acquisition-related needs. This goal is achieved by:

- a. Partnering with top tier academic institutions and DoD organizations to recruit talented (STEM and non-STEM) undergraduate students and create a unique learning and skill development ecosystem.
- b. Providing students with scholarships and stipends for a DoD service commitment.
- c. Guiding scholars through a rigorous, cohort-based educational and experiential development program.
- d. Helping scholars transition into satisfying and rewarding DoD careers.

The scholar preparation is achieved through the unique DCTC Integrated Curriculum and Development (ICAD), which is the focus of this paper.

The DCTC Integrated Curriculum and Development (ICAD)

The DCTC-ICAD philosophy is founded on the principle of multidimensional learning experiences that, by design, are seamlessly integrated. The ICAD includes five major elements, depicted in Figure 1, and they include a core curriculum, project-based cohort summer



internships, action (immersive) learning experiences, a culture of care, and innovation capstone projects. In designing the ICAD, the AIRC Curriculum WG placed an equal emphasis and value on the importance of classroom-based learning and experiential learning acquired by working on real-life, DoD-relevant projects. Consistent with this philosophy, the AIRC Curriculum WG has selected four guiding attributes that are interwoven throughout all ICAD elements and are intended to sharpen the DCTC scholar skills and qualities in leadership, effective teamwork, critical thinking, and creativity and innovation.

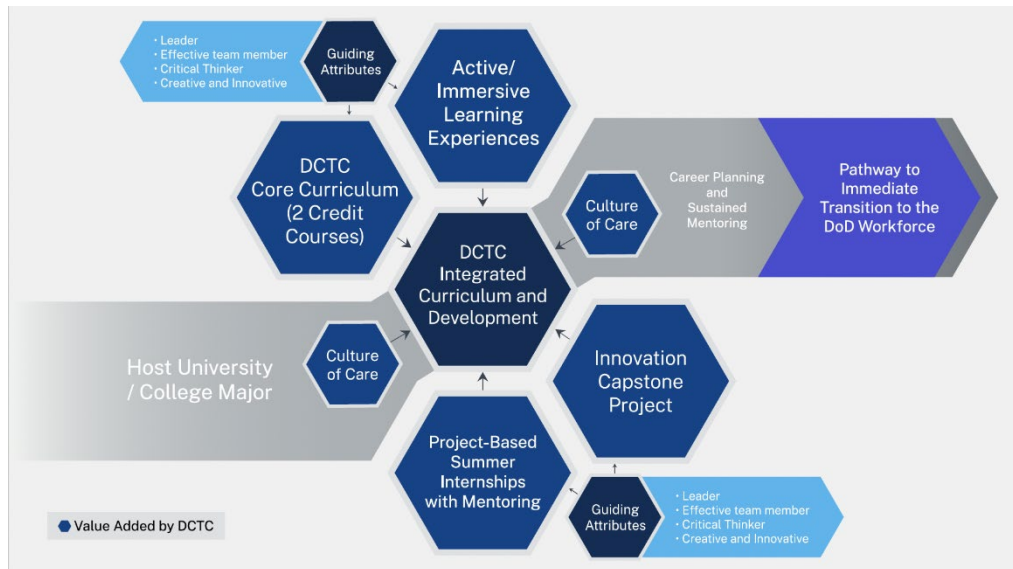


Figure 1. Elements of the DCTC Integrated Curriculum and Development

1. The Core Curriculum: The Pilot-0 core curriculum is composed of four two-credit courses given in the junior and senior years. They are:

- ***DCTC-301 (Fall 2023): Fundamentals of Civilian Service in the DoD***

This course is intended to introduce the students to fundamental principles of the DoD mission, structure, operations, and culture and give them a sense of what it takes to become a successful DoD employee. Topics include, but are not limited to, DoD organization, authorities, and responsibilities; mission and structure of military services; the DoD innovation and acquisition ecosystem and what it takes to develop and deliver an operational capability; critical technologies and their importance to the DoD strategic mission; and the importance of multidisciplinary, diverse teams in providing innovative solutions to future DoD challenges.

- ***DCTC-302 (Spring 2024): Exploration of the DoD Acquisition Environment***

This course builds on DCTC-301 and is intended to introduce the students to the fundamental roles and analytical processes applied throughout the acquisition process. The topics are presented from the perspectives of various stakeholders in the acquisition process including end users, program managers, vendors, contracting professionals, and members of the integrated product teams. The course also introduces students to the practice of self-awareness and provides exercises to strengthen resilience skills, which will improve their capacity to handle unexpected challenges and adapt to change. In this course, faculty use an active learning method of teaching, including role-playing exercises, games, and case studies.



- **DCTC-401 (Fall 2024): Overview of DoD Missions and Community Functions**

This course builds on DCTC-301 and DCTC-302 and places emphasis on emerging threats and security concerns that shape DoD strategic drives and mission, the role the DoD community plays in supporting the U.S. national security mission, and the structure, functions, and culture of the broader defense community. It also examines relationships with both internal and external stakeholders—including other DoD departments and agencies, the U.S. Congress and other government organizations, U.S. global allies, the defense industry, and academia. In addition, the course explores the vital role that technology plays in the acquisition process and the challenges and opportunities that the acquisition community faces now and in the near future. Also, in this semester student groups will start working on their innovation capstone project, sharpening their innovation, team-building, and leadership skills.

- **DCTC-402 (Spring 2025): Driving Institutions to Success**

This is the final course in the DCTC program and builds on the prior three courses with emphasis on personnel management, leadership, project management, how to overcome adversity, and how to navigate and be effective within the DoD organizational bureaucracy. In addition, students will undertake a deep dive into the 14 DoD critical technologies and will develop an understanding of the important role these technologies play in supporting the U.S. national security mission. In addition, students will have the opportunity to focus on critical leadership skills development in preparation for a career in the DoD acquisition community. Finally, student groups will complete and present their innovation capstone project to the DoD sponsors.

Table 1. DCTC-301 Course Syllabus

1	Week	Modules	JUNIOR YEAR
2	Fall		DCTC-301
3	0	Orientation	Full Disclosure: What am I getting into? Opportunities, expectations, principles, ethics, rewards, and sense of purpose from government employment.
4	1	Government: Authorities and Policies, Structures, Ethics, and Rewards	Gov't overview: Legislative & Executive Branches. Pres. Budget Requests. Authorization (NDAA, HASC, SASC w/ Service subcommittees) and Appropriations (Acts, HAC-D, SAC-D). GAO Audits & bid protest. CRS. CBO.
5	2		Executive Branch: POTUS, OMB. President's Budget Request. Agencies. Authorities, responsibilities, limitations (checks & balances)
6	3		DoD: Organization, authority, responsibility, Titles 10 and 50, 4th Estate (OSD, Defense Agencies), Military Departments and Services.
7	4		Legislation and Policy Sources: Legislative: Acts, NDAA's, committee reports, JES. US Code. Code of Federal Regulations. FAR & DFARS. DoD policy: WHS archive (DoDDs, DoDIs, memos), CJCS, FMR, Service Policy.
8	5	Military Structures	Military structures, responsibilities, and culture
9	6	Military Structures	Military operations: CCMDs, chain of command
10	7		Allied considerations
11	8	Workforce and Team Basics	Personnel basics: Hiring, billets, career fields, MIL/CIV/CTR, Acquisition workforce demographics and training
12	9		Team building, IPTs, operation, cooperation, and compromise
13	10	Innovation and Acquisition	Acquisition basics: Requirements, financial resources, technology, intelligence, and acquisition
14	11		Government-industry-academia collaborations and partnerships
15	12	Tech. Literacy and Implications for Business	Data literacy
16	13		Data, data sources, and decision analytics
17	14		AI Literacy
18	15		Other Tech. Literacy: Current priority modernization areas (hypersonics, quantum, 5G, etc.)



The following steps are taken to ensure maximum uniformity on the content and delivery of each course among the four pilot universities:

1. The AIRC Curriculum WG develops a detailed lesson outline for each weekly lecture, which contains the learning objectives, sample quizzes, and class discussion topics and provides them to the university instructors teaching the course ahead of time. In addition, the WG provides the faculty with media collateral (videos/podcasts) relevant to the material covered in the lecture.
2. The AIRC Curriculum WG meets with the university course instructors on a weekly basis to discuss various issues including lessons learned and to make decisions for going forward.

A solicitation was sent to various DoD organizations to participate in sponsoring the DCTC scholar internships for the summer of 2024. The solicitation resulted in an overwhelming response with a total of 37 internship project proposals submitted by DoD organizations. The requirements that must be satisfied by the sponsoring organizations for the DCTC internship program are shown in Table 3.

Table 2: DCTC-301 Learning Objectives that Students Must Demonstrate

Week	Modules	Scholar Learning Objectives (SLO)	GUIDING ATTRIBUTES
1	MODULE 1 (Lectures 1-4) Orientation Government: Authorities and Policies, Structures, Ethics, and Rewards	<input type="checkbox"/> SLO-0: Basic understanding of the opportunities, expectations, principles, ethics, rewards, and sense of purpose with government employment. <input type="checkbox"/> SLO-1: Basic understanding of DoD structure, organization, authorities, responsibilities and relationships with the executive and legislative branches <input type="checkbox"/> SLO-2: Basic knowledge of DoD related policies and policy development process	<input type="checkbox"/> Leadership <input checked="" type="checkbox"/> Critical Thinking <input type="checkbox"/> Effective Team Member <input checked="" type="checkbox"/> Creativity and Innovation
2			
3			
4			
5	MODULE 2 (Lectures 5-7) Military Structures	<input type="checkbox"/> SLO-3: Basic understanding of military services, missions, structures, responsibilities, operations and culture	
6			
7			
8	MODULE 3 (Lectures 8-9) Workforce and Team Basics	<input type="checkbox"/> SLO-4: Basic understanding of personnel management, training, and teaming	
9			
10			
11	MODULE 4 (Lectures 10-11) Innovation and Acquisition	<input checked="" type="checkbox"/> SLO-5: Understanding of the DoD innovation and acquisition ecosystem, structures, and opportunities to innovate	
12			
13			
14	MODULE 5 (Lectures 12-15) Tech. Literacy and Implications for Business	<input checked="" type="checkbox"/> SLO-6: Technological literacy in DoD critical areas including priority tech modernization areas, and the implications for business functions	
15			

Semester Critical Thinking Team Project
SLO-7: The students demonstrate that can function in a team setting and they have a good grasp of team dynamics and functions including roles and responsibilities, conflict management and team effectiveness

3. **Action/Immersive Learning Experiences (ALEs):** ALEs entail “hackathon”-type, scholar team competitions and projects on rapid, DoD-related problem solving and are aligned with the DCTC curriculum. ALEs are designed to help scholars hone their skills in areas such as
 - a. How to be creative and innovate within the DoD innovation ecosystem
 - b. How to listen and communicate effectively and how to pitch their ideas
 - c. How to function effectively as a member of a cross-functional team



- d. How to appreciate leadership and the attributes, responsibilities, and values of leaders
- e. How to think critically and understand trade-offs
- f. How to solve problems within the bounds of ethical conduct

ALEs are executed over 2-3 consecutive full days. Pilot universities have the option to use the CDAO-offered ALE learning activities or create their own to fulfill DCTC minimum curricular requirement for one ALE per year.

Table 3: DCTC Requirements for Developing and Executing Summer Internships

Requirement	Description
Address Critical DoD Skills	Critical DoD skills are further developed in a project setting; applies elements from DCTC curriculum
Multi-disciplinary	Requires efforts of a multi-disciplinary team effort and collaboration
Cohort-based	Cohorts of DCTC scholars from a university or across DCTC pilot universities work as a team
Challenging problem	Problem space should represent a sample of DoD's compelling problems and mission sets
Representative work	Work represents the type of work assigned to a junior DoD civilian in that organization
Mentor network	Availability and role of a mentor network for both project and general scholar development
Support staff/processes	Support staff and processes facilitate a top caliber experience
Pathway to assignment	Proposing organization(s) have the ability to offer post-graduation employment
Culture of care	Environment and experience embody the DCTC culture of care

AIRC has created the DCTC Acquisition Game to support the objectives of ALEs. The game provides scholars an immersive experience into the basic processes of the defense acquisition system and is composed of three phases: a technical solution phase, a contracting approach phase and a program management phase. These phases allow scholars to gain an understanding of the various decision points, the key information needed, the critical thinking required and the challenges that can be



experienced through the life cycle of a typical acquisition program. The DCTC scholars start using the acquisition game in DCTC-302.

4. Culture of Care: The Culture of Care element of ICAD helps scholars to sharpen their personal qualities, traits, values, and attributes and develop a reliable compass to navigate an increasingly uncertain, complex, and volatile world. The Culture of Care relies on a mentor network composed of experienced leaders from the DoD, other federal agencies, academia, and industry, who act as role models for DCTC scholars, and mentor them throughout their studies and beyond. The first members of the DCTC culture of care mentor community are the DoD leaders who sponsor the summer internship projects and will host and mentor the scholars in their organizations. Also, the university instructors who teach the courses are experienced in various aspects of DoD functions and serve as student mentors. Mentoring and coaching scholars on shaping their personal leadership skills and qualities in areas that include the following:

- Passion for what you do and ability to inspire others
- Integrity and professional ethics
- Balanced temperament
- Engendering respect and loyalty
- Thought leadership
- Fair, just, unbiased, transparent, clear values
- Compassion and empathy for others
- Mental toughness
- Resiliency and adaptability
- Balance between family and job obligations

Awareness of these values and leadership attributes is included within the core curriculum courses. For example, resiliency and adaptability are embedded in DCTC-302.

5. Innovation Capstone Projects: Innovation Capstones are multidisciplinary undergraduate team projects where students engage to formulate solutions to meet the needs of customers in the Department of Defense. These span the two senior year semesters. Students pursue projects using technical, managerial, and entrepreneurial tools and methods. Each Innovation Capstone project is formulated by two student teams. The “*Solution Team*” focuses on the application of engineering and science to support conceptualization, modeling and design, technical analysis, fabrication, integration, and testing of prototyped solutions in response to the identified need. The “*Business Team*” addresses business dimensions on each project such as ensure the requirements are well understood; develop and track cost and schedule estimations; conduct a competitive analysis, if necessary; formulate and manage risks; conduct a supply chain and manufacturability assessment to ensure an ability to scale the solution, if necessary; and formulate an intellectual property assessment, if necessary. Each capstone project team will be funded by the DCTC program to cover costs of supplies, materials, and incidentals as required and not exceeding \$10,000 per project.

With support from the DCTC/AIRC team, the pilot universities have multiple pathways to formulate topics for Innovation Capstone projects namely:

- Engage with any DoD organization to formulate a project/need. Some project ideas have been developed by the AIRC team, but the universities are not constrained by these ideas.



- Leverage the project ideas that have been identified by NSIN (National Security Intelligence Network). This will be facilitated by a DCTC partnership with NSIN that allows their existing model to identify potential capstone projects and DoD organizations.
- Extending the DCTC summer internship project. The DoD organizations believe there are some summer project-based internships that will need a second phase of research and prototype development. The senior year capstone project will be the perfect transition from the summer into senior year capstones.

The participating universities agree to assign faculty advisors as required to guide and manage both the technical and business teams for each topic. The faculty advisors and the scholar teams have the following major responsibilities:

- Establish and maintain periodic contact with the government customer and innovation capstone managers by phone calls, emails, virtual meetings, and personal visits as required.
- Write a project plan in accordance with template provided by AIRC.
- Conduct research, development, and business operations necessary to meet project objectives.
- Present periodic design reviews to government points of contact as agreed in the project plan.
- Present a final project review containing the following:
 - A final status review of all project elements to the government customer and innovation capstone managers.
 - Final deliverables including a) final project review documents, including a final project plan and “quad chart”, b) copies of a final report (as required by the university), c) all relevant technical/business data and analyses, d) demonstrations as required, and e) short executive summary of the project in video format.

In addition, each team is given the opportunity to enter their projects in a year-end competition.

DoD Critical Skills Areas

AIRC has recently completed a research study (Ramirez-Marquez, 2024) to identify and develop a list of what various DoD organizations and project reports consider as areas of critical skill and capability needs in the broader acquisition space. The study identified 80 published reports relevant to critical skills and talent development in DoD acquisition and used Natural Processing Language (NPL) techniques to perform analyses and extract information on skills and capabilities described in the documents. Based on these analyses, the investigator developed summaries of the report conclusions and plausible clusters of skill sets were identified through meta-analysis across the 80 reports. The study has identified 14 skill and capability sets that are considered critical in developing the future DoD acquisition workforce (Table 4). This list serves as a guide in developing and executing the DCTC-ICAD.



Table 4. DoD Acquisition Workforce Critical Skills and Capabilities

Critical Skills and Capabilities	
Leadership Skills, Adaptability, and Strategic Thinking – communications and stakeholder engagement; collaborative training, knowledge sharing, and cross-capability training; adaptive communication, learning, and skill-capability development.	Agile Procurement – develop skills in agile procurement processes to acquire and deploy new capabilities; enhance capabilities to source and integrate cutting-edge solutions.
Cognitive Skills in Cyber Space – analysis and evaluation of cyber exploits and risks; design and risk; using encryption algorithms and security tools; cyber operations.	Innovative Ecosystem – foster an environment that encourages innovation, experimentation, and rapid prototyping; foster professionals who can bridge gaps between technical and strategic aspects.
Digital Tools and Transformation – role of automation and AI in business operations and warfare; workforce enhancement and efficiency.	Policy and Intervention – develop effective policies and interventions to improve digital inclusivity and address digital skill difficulties.
Data Analytics and Insights – impact on organizational performance; decision making.	Enhanced Technical Proficiency – artificial intelligence; autonomy; cyberspace; software; digital literacy in general.
Cybersecurity and Network Design	Communication Skills – oral; written; interpersonal.
Resilience and Adaptability – enhance resilience to withstand crises; adapt to changing circumstances; develop skills to cope with stress.	Interdisciplinary Knowledge and Skills – encouraging individuals to acquire knowledge and skills from multiple disciplines, fostering a holistic approach to problem-solving and innovation.
Public-Private Partnerships; International Engagement	Ethical and Social Responsibility – understanding the ethical implications of technology and developing a sense of responsibility toward society.

The DCTC-AIRC team is in the process of validating and refining this list of critical skills and capabilities with partner DoD organizations that are hosting the scholar summer internships.

Consistent with the philosophy of placing equal value and emphasis on classroom facilitated learning and experiential learning, and the mission to prepare scholars for future leadership positions in the DoD workforce, the DCTC-ICAD is designed to address the full spectrum of knowledge, skills, and abilities (KSAs) in critical areas. KSAs have been used by the DoD for hiring purposes. They involve the understanding acquired by study or experience, the ability to perform a job because it has been practiced, and the physical or mental power to execute tasks. Attributes bring to play higher order personal characteristics and qualities that are more difficult to acquire by study. Attributes can be thought of as qualities that are made up by a collection of KSAs in combination with a collection of critical interpersonal characteristics leading to the concept of knowledge, skills, abilities, and attributes (KSAs). Figure 2 shows how the five elements of the DCTC-ICAD are designed to provide a holistic preparation of scholars for leadership positions in the DoD by combining KSAs with shaping their interpersonal leadership skills.



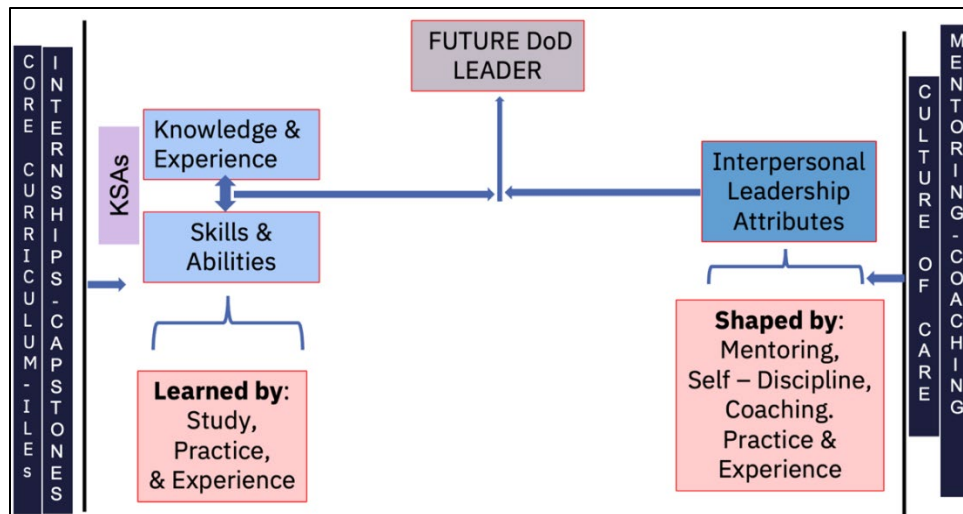


Figure 2. KSAs and Interpersonal Leadership Attributes

Cohort-0 Scholar Academic Profile

Table 5 shows the diversity of the DCTC Cohort-0 scholars' fields of undergraduate study mapped against the OUSD(A&S) critical functions namely, auditing, business-financial management and cost estimating, contracting, engineering and technical management, life-cycle logistics, program management, and test and evaluation. Approximately 60% of the scholars are pursuing STEM fields and 40% non-STEM degrees.

The student selection for Cohort-0 was performed by each university based on the following selection requirements:

- U.S. citizenship required
- Cumulative Grade Point Average (GPA) higher than 3.0 on a scale 0–4
- Eligibility to obtain a security clearance
- Personal interview assessment by a selection committee

Table 5. Undergraduate Scholars' Majors Mapped to A&S Functions

STEM Disciplines (<i>A&S Functions</i>) (60% of Cohort-0)	Non-STEM Disciplines (<i>A&S Functions</i>) (40% of Cohort-0)
Engineering & Technology (<i>PM, SE, Test and Evaluation, Sustainment, Energy</i>) <ul style="list-style-type: none"> ▪ Aeronautical and Astronautical Engineering ▪ Aerospace Engineering ▪ Chemical Engineering ▪ Industrial and System Engineering ▪ Mechanical Engineering 	Business and Management (<i>Policy, Contracts, Budgets, PM, Acquisition, Sustainment, Industrial Base, Supply Chains</i>) <ul style="list-style-type: none"> ▪ Accounting ▪ Business Information Technology ▪ Business Management ▪ Economics and Finance ▪ Management ▪ Marketing ▪ Supply Chain Management ▪ Technology Leadership & Innovation
Human Factors, Life Sciences, Medicine, and Health (<i>Human Factors, T&E, Health Systems and Data, Chem-Bio Defense, Environment, Animal Systems</i>)	Public Policy (<i>Policy, NSS/NDS, Leg. Affairs, Public Affairs</i>) <ul style="list-style-type: none"> ▪ Political Science ▪ Public Health & Political Science



<ul style="list-style-type: none"> ▪ Biochemistry & Biology ▪ Biological Engineering ▪ Biomedical Engineering ▪ Biosystems Engineering ▪ Human Nutrition ▪ Psychology ▪ Veterinary Science 	
Computer and Information Systems (<i>Software, Data Science, T&E, SE, Sustainment, Logistics</i>) <ul style="list-style-type: none"> ▪ Computer & Information Technology ▪ Computer Engineering ▪ Computer Graphics Technology ▪ Computer Modeling & Data Analytics ▪ Computer Science ▪ Electrical and Computer Engineering ▪ Software Engineering 	National Security (<i>Policy, Requirements, Threats, Intel.</i>) <ul style="list-style-type: none"> ▪ Chinese ▪ National Security & Foreign Affairs
Design and Construction (<i>Installations & Environment, Housing, Real Property, Construction</i>) <ul style="list-style-type: none"> ▪ Building Construction ▪ Construction Management Technology ▪ Industrial Design ▪ Landscape Architecture ▪ Geo-Science ▪ Urban Planning 	Law (<i>Policy, Contracts, A&S, Leg. Affairs, OGC</i>) <ul style="list-style-type: none"> ▪ Criminal Justice ▪ Criminology ▪ Digital Criminology

Table 6 shows the average scholar GPA for each of the university cohorts. The overall GPA average across the four universities is 3.64, which demonstrates that the DCTC program is capable of attracting very high performing, bright students.

Table 6: Average GPA for Each University Scholar Cohort

University	Number of Students	Average GPA
North Carolina A&T	20	3.68
Purdue	20	3.67
University of Arizona	19	3.65
Virginia Tech	30	3.56

DCTC Program Assessment

The AIRC-DCTC team has adopted an evidence-based process of assessment, evaluation, and continuous improvement for the ICAD modeled after the Accreditation Board of Engineering and Technology (ABET) assessment guidelines (www.ABET.org). The process consists of three loops of assessment:

- Loop I: Closes at the Course Level
- Loop II: Closes at the Program Level
- Loop III: Closes at the DCTC Mission Level

The process is shown in Figure 3 and consists of the following: (a) The course loop (I) of assessment is the sole responsibility of the course instructor, who evaluates student mastery of the course learning objectives and proposes course improvements; (b) the program loop (II) of assessment relies on input from various stakeholders such as scholar employers, the program external advisory board, and program alumni to evaluate how well the five elements of ICAD



program meet the program educational and development objectives; and (c) the DCTC mission loop (III) is informed by the results of the program evaluation and maps the program evaluation results to the DCTC mission to assess if the stated mission is achieved. Continuous improvement initiatives are implemented as necessary.

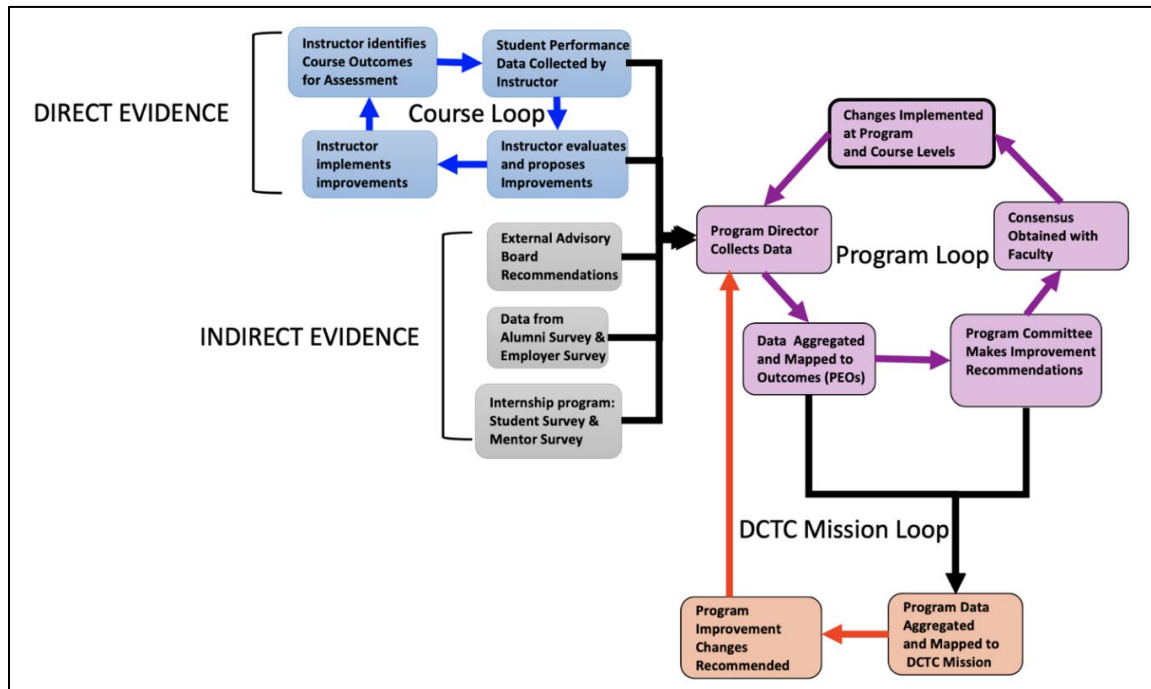


Figure 3. DCTC-ICAD Program Assessment Flow Chart

The following definitions are used in the assessment process:

Student Learning Objectives (SLOs): These are the learning objectives at the course level developed and approved by the DCTC curriculum committee. An example for DCTC-301 is shown in Table 2.

1. **Student Outcomes (SOs):** These are the skills and abilities the students are expected to have mastered at the time of graduation from the DCTC program. DCTC scholars' performance is monitored and evaluated to ensure that student outcomes are met. The SOs for the DCTC-ICAD program are ability to identify critical DoD challenges and formulate innovative solutions by applying fundamental principles.
2. Ability to communicate effectively with a range of audiences within the national defense and security ecosystem.
3. Ability to recognize ethical and professional responsibilities and make informed judgments while considering impacts of their work in global, political, environmental, economic, and societal contexts.
4. Ability to navigate with ease through the national defense and security broader enterprise by understanding the culture, structure, operations, and functions of the federal and private defense and security communities and their missions, roles, and responsibilities.
5. Ability to function effectively on cross-functional teams whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives.



6. Ability to acquire and apply new knowledge as needed using appropriate learning strategies including pursuing graduate degrees.

Program Educational Objectives (PEOs): These are broad statements that describe what graduates are expected to attain within a few years of graduation. PEOs must be consistent with the DCTC mission, which is aligned with DoD needs. PEOs are periodically reviewed and modified to meet the various DoD constituency needs in critical areas. The PEOs for the DCTC-ICAD program are:

PEO-1: Benefitting from the ICAD program, DCTC scholars will advance much faster in their DoD careers and will reach higher levels of achievement than non-DCTC graduates in similar fields that are employed at the same time with only a bachelor's degree.

PEO-2: DCTC scholar graduates will fill gaps in critical areas and will have a positive impact on the DoD workplace through multidisciplinary/cross-functional collaboration, teamwork, innovation, and leadership.

PEO-3: DCTC program graduates will effectively navigate important contextual factors in their careers, including the historical, regulatory, political, policy, economic, ethical, public relations, and technological aspects of DoD future challenges.

Assessment of DCTC-301

The AIRC team has performed an evaluation of DCTC-301, the first course of Pilot-0 phase of the program, completed in the fall 2023 semester. The following outcomes were of particular interest to the curriculum committee:

1. Student level of satisfaction with respect to the course content, instructional methods, and instructional materials used.
2. Student level of satisfaction with the instructor and the way the material was presented in the classroom.
3. Instructors' assessment of how well students demonstrated their level of mastery of the course learning objectives.
4. Students' opinions on how well they think they mastered each course learning objective.
5. Degree of uniformity of 1 through 4 above among the four universities participating in the Pilot-0.

The assessment process involves four steps: (1) the administration of an online student survey with questions designed to assess items 1, 2 and 4 above; (2) a direct assessment of student outcomes by each instructor on the level of mastery of the SLOs by the entire class; (3) a report by each instructor on lessons learned and proposed improvements; and (4) analysis of the data and course evaluation performed by the curriculum committee. This process results in evidence-based implementation of continuing improvement initiatives at the course level.

The student survey contained eight questions about course content and quality, six questions about instructor performance in class, and a self-assessment of student mastery of each course learning objective. The student response rate ranged from 63% to 100% across the four universities with an average rate of 84.3%. The survey was anonymous and invited the students to make written comments.

Student Course Evaluation. For the eight course assessment areas, the students rated each area on a Likert scale of 5 (strongly agree) to 1 (strongly disagree), producing the results shown in Table 7.



Table 7. Student Course Evaluation Ratings

Statement	Average Rating	Range
The material was instructive	4.2	4.0-4.6
The class discussions helped me understand the material	4.2	3.4-4.8
The assignments/quizzes were graded promptly	4.3	3.6-4.8
Quiz questions related highly to the course material	3.8	3.0-4.6
The videos were motivating and helpful in better understanding the lecture topics	3.7	3.0-4.5
The class group project was helpful in understanding DoD issues	4.1	3.5-4.5
The guest speakers were effective in making me understand the material better	4.3	4.0-4.7
Overall, the course was a great experience	4.1	3.7-4.8

The students also submitted 32 comments addressing issues with the course. Figure 4 shows the variability of the rating scores for the last statement across the four universities (color-coded).

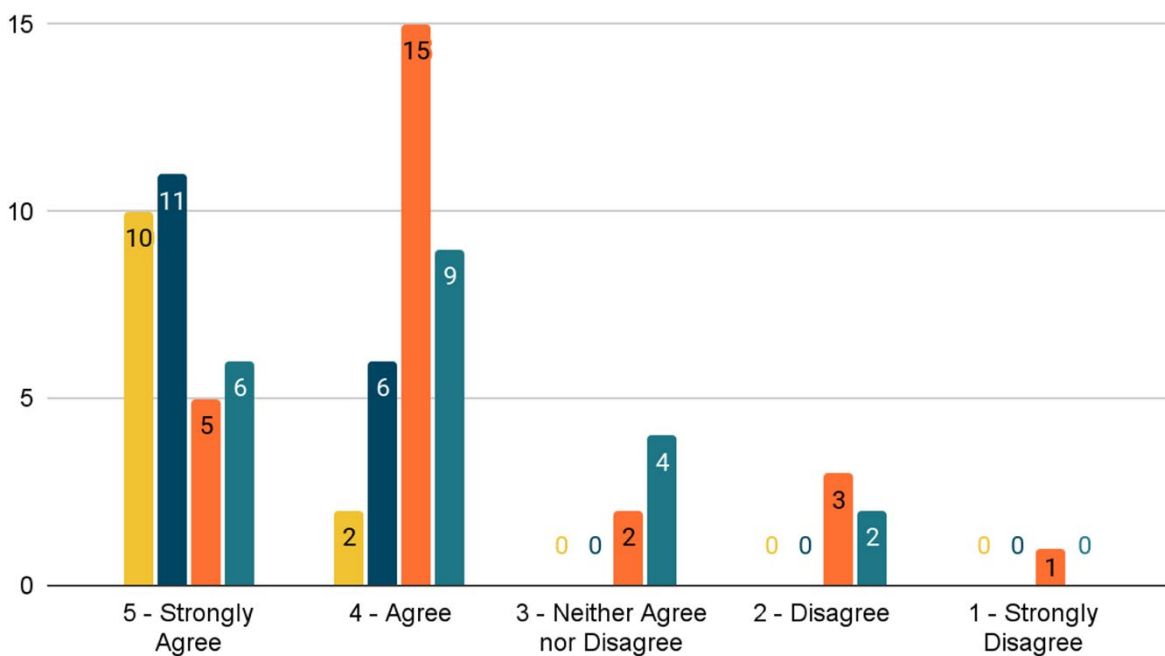


Figure 4. Student Ratings on the Statement "Overall, the course was a great experience."

Student Instructor Evaluation. For the six instructor assessment areas, the students rated each area on a Likert scale of 5 (strongly agree) to 1 (strongly disagree), producing the results shown in Table 8.



Table 8. Student Instructor Evaluation Ratings

Statement	Avg. Rating	Range
The course objectives and grading policy were clearly explained	4.4	3.6-4.9
The instructor was prepared for class	4.6	4.0-5.0
The instructor was successful at communicating the material	4.6	4.2-5.0
The instructor was successful at engaging the students in class discussion	4.5	4.3-4.8
The instructor was available to the students	4.7	4.4-5.0
Overall, the instructor was very effective	4.6	4.1-4.9

The students also submitted 14 comments addressing issues with the course instructors.

Figure 5 shows the variability of the rating scores for the last statement across the four universities.

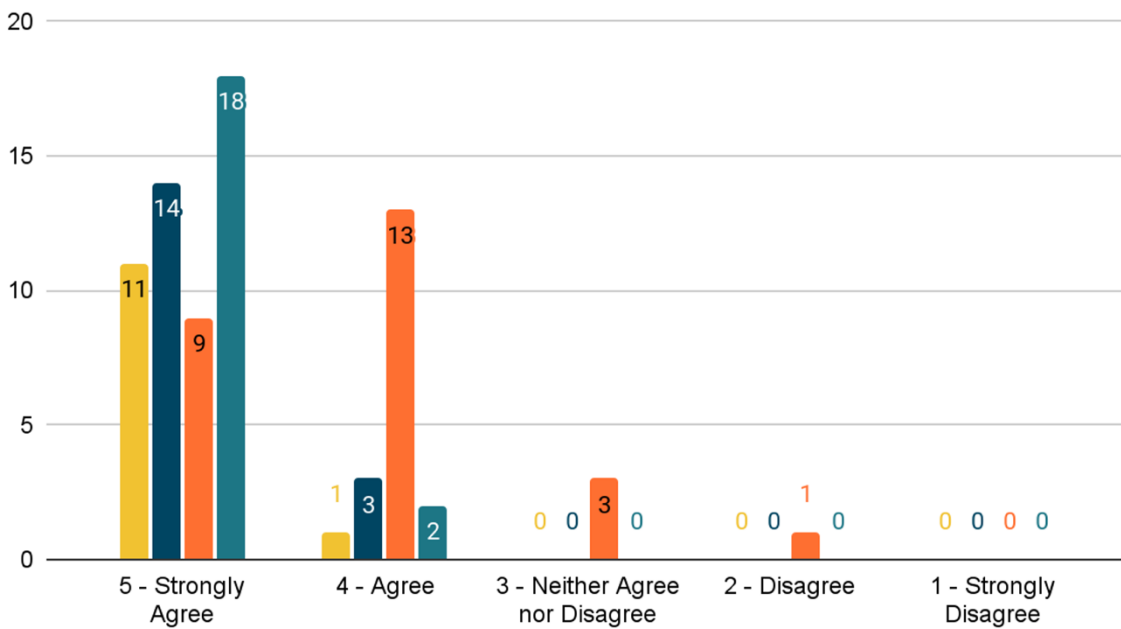


Figure 5. Student Ratings on the Statement "Overall, the instructor was very effective."

Evaluation of Student Learning Outcomes (SLOs). The SLOs for the class were evaluated by the course instructors at each university based on student performance as tested by various assessment instruments including quizzes, essays, project presentations, and class participation.



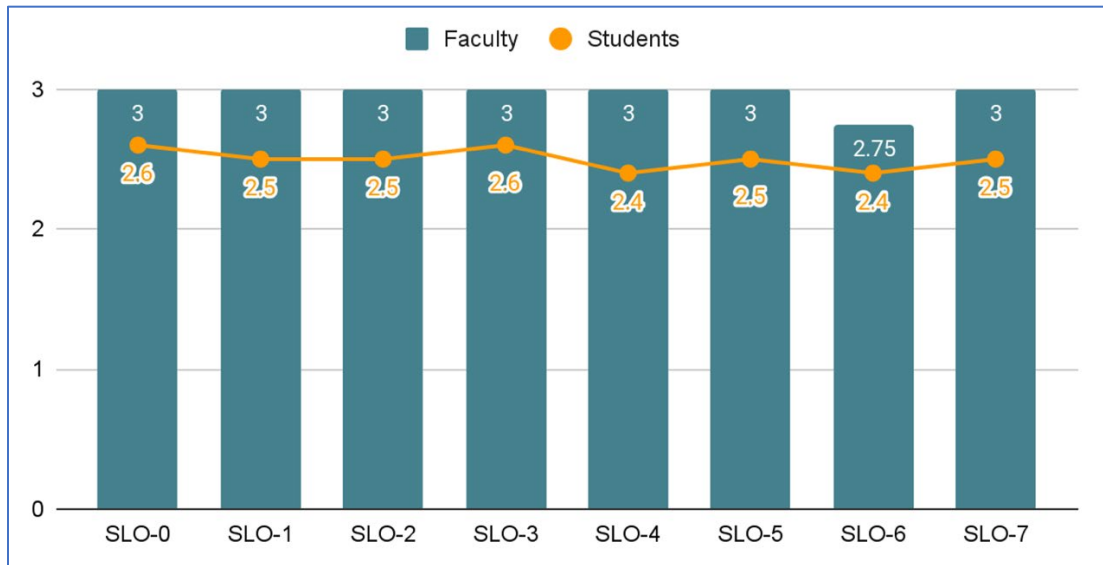


Figure 6. Assessment of SLOs by Instructors and Students

In addition, the survey asked students to rate their perception of the mastery they achieved in each of the eight learning objectives. A scale of 3 (high), 2 (medium) and 1 (low) was used representing the 75th, 50th, and 25th percentiles respectively. Figure 6 shows the responses of the instructors' SLOs assessment plotted along with the students' SLO mastery self-evaluation. The average faculty rating of all SLOs across the four universities was 2.97, where the student self-evaluation was very consistent with an average rating of 2.5.

The results of the DCTC-301 assessment presented herein, along with several recommendations made by both students and course instructors, have resulted in several changes to be made with the purpose of improving both this course as well as the remaining three courses of the program. Some of these changes have already been implemented in the DCTC-302 course offered in the spring 2024 semester.

Conclusion

The DCTC is a highly selective program attracting bright and highly motivated students and preparing them to seamlessly enter the DoD acquisition workforce and have impactful careers leading into future leadership positions. The DCTC pilot program provides a rare opportunity to develop, test, evaluate, and continuously improve a novel, innovative integrated curriculum and development program that places equal weight on classroom learning and on learning through cohort-based problem solving and project execution on real, current, and future DoD challenges. The ICAD program is designed to develop a talent pool with the knowledge, skills, abilities, and personal attributes required to maintain the United States in a global leadership position for the next generations. The vision is that DCTC scholar cohorts will form a distinctive, cohesive, and strongly bonded community during their studies and throughout their professional careers and will carry their DCTC affiliation as a badge of honor and distinction for life.

Acknowledgment


The authors acknowledge financial support from the U.S. Department of Defense through SERC/AIRC on research task WRT-1080 contract no. HQ0034-19-D-0003.




References


- DASD(CPP). (2022). *Fiscal year 2022 – fiscal year 2026 DoD civilian human capital operating plan*. https://www.dcpas.osd.mil/sites/default/files/DoD%20HCOP%20FY22-26_3%20June%202022%20%28003%29.pdf
- Defense Business Board. (2020). *Strengthening defense department civilian talent management*. <https://dbb.defense.gov/Portals/35/Documents/Reports/2022/DBB%20FY22-03%20Talent%20Management%20Study%20Report%2018%20Aug%202022%20-%20CLEARED.pdf>
- DoD. (2022). *National defense strategy*. <https://media.defense.gov/2022/Oct/27/2003103845/1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF>
- Korfatis, G. P., et al. (2023). *Defense Civilian Training Corps (DCTC) implementation recommendations*. DCTC Execution Planning Panel Report, Acquisition Innovation Research Center (AIRC).
- POTUS. (2022). *National security strategy*. <https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf>
- POTUS. (2021). *Management agenda vision*. https://assets.performance.gov/PMA/Biden-Harris_Management_Agenda_Vision_11-18.pdf
- Ramirez-Marquez, H. E. (2024). *Leveraging NPL for skill identification and talent management in the Department of Defense* [Report]. Acquisition Innovation Research Center (AIRC).

APPENDIX A: Core Curriculum Modules and Learning Objectives for DTC-302, 401, and 402

 DCTC-302: Exploration of the DoD Acquisition Environment															
Week	Modules	Scholar Learning Objectives	GUIDING ATTRIBUTES												
1	MODULE 1 (Lectures 1-11) • Acquisition Principles, Functions and Processes	SLO-1: Understanding project management, risk management, and basic acquisition processes and functions	Semester Critical Thinking Team Project SLO-3: The students demonstrate they can function in a team setting and they have a good grasp of team dynamics and functions including roles and responsibilities, conflict management and resolution and team effectiveness	Leadership											
2		SLO-2: Understanding basic systems engineering principles and value of the systems engineering process to DoD acquisition			Critical Thinking										
3						SLO-3: Understanding the value chain of technology research and development, test and evaluation, production, and fielding	Effective Team Member								
4								SLO-4: Understanding the way the DoD industrial base operates, logistics, supply chain management and risks							
5									SLO-5: Understanding science and technology implications for project management and contracting, intellectual property and operations and support of the military						
6										SLO-6: Understanding security threats and the regulatory and ethical responsibility of personal and DoD security and security clearances	Creativity and Innovation				
7												SLO-7: Orientation on key ethics requirements, restrictions, and philosophy of government employees			
8													MODULE 2 (Lectures 12-13) • Cross-Functional Implications		
9														MODULE 3 (Lectures 14-15) • Security and Ethics	
10															
11															
12															
13															
14															
15															



 DCTC-401: Overview of DoD Missions and Community Functions				
Week	Modules	Scholar Learning Objectives	GUIDING ATTRIBUTES	
1	MODULE 1 (Lectures 1-5) • DoD Mission	SLO-1: Understanding the fundamental national security mission and objectives: Why are we here?	Leadership	
2		SLO-2: Understanding DoD culture, strategic drivers, concerns and challenges		
3		MODULE 2 (Lectures 6-15) • Communities Roles and Equities	SLO-3: Understanding the decomposition into communities and their roles, responsibilities, authorities, and equities	Critical Thinking
4			SLO-4: Understanding the DoD relationships with various communities including allies, defense industry, academia and intelligence	
5			SLO-5: Understanding how communities differ and how those differences can cooperate (or not!) to support national security	
6	SLO-6: Understanding the impact of cooperation with various communities on the acquisition process		Effective Team Member	
7				
8				
9				
10				
11				
12				
13	Service, rewards, endurance, and perseverance; Farewell Keynote; Innovation Capstone Presentation and Competition	Creativity and Innovation		
14				
15				

 DCTC-402: Driving Institutions to Success				
Week	Modules	Scholar Learning Objectives	GUIDING ATTRIBUTES	
1	MODULE 1 (Lectures 1-3) • Workforce	SLO-1: Understanding DoD personnel management policies and processes, and career planning	Leadership	
2		SLO-2: Understanding DoD culture, strategic drivers, concerns and challenges		
3		MODULE 2 (Lectures 4-5) • Leadership and Management	SLO-3: Acquire knowledge and experience on how to work effectively in a cross-disciplinary team environment, manage conflicts and be a good leader as well as a good follower	Critical Thinking
4	SLO-4: Understand the basic principles of managing people			
5	MODULE 3 (Lectures 6-8) • Institutions and Bureaucracies		SLO-5: Acquire knowledge on how to navigate and be effective amidst organizational bureaucracy	
6				
7	MODULE 4 (Lectures 9-14) • Critical DoD Technologies	SLO-6: Acquire knowledge about the importance of selected emerging technologies and why they are critical to DoD's current and future national defense strategy.	Effective Team Member	
8				
9				
10				
11				
12				
13				
14	Service, rewards, endurance, and perseverance; Farewell Keynote; Innovation Capstone Presentation and Competition	Creativity and Innovation		
15				



Design Revision of Walters Probes—High Sensitivity Differential Temperature, CT2, Probes for Atmospheric Optical Turbulence Characterization

Douglas H. Nelson—is motivated to provide our troops with everything they require to defend our country. He has demonstrated operations, technical, teaching and leadership expertise in over 35 years' experience with the U.S. Army (as an active-duty Armor Officer), United States Military Academy (alumnus and former faculty member), Los Alamos National Laboratory (LANL), Raytheon, The Boeing Company, the U.S. Naval Postgraduate School (alumnus and former faculty member), the U.S. Army Space and Missile Defense Command and Teknicare, Inc. This has included a more than 30-year professional and personal relationship with Dr. Donald L. Walters, the namesake of this paper's subject. [douglas.nelson@teknicare.com]

Kyle S. Edwards—graduated from UCLA in '08 with a BS in atmospheric sciences and has worked since graduation for NAVAIR at NAWCWD Point Mugu Geophysics. Starting there initially as an intern and weather observer calling fog on the airfield, he worked his way up to being a project forecaster, meteorologist, and now Geophysics Branch Head where he supports a wide breadth of Sea Range operations, dabbles in directed energy projects, and gets blamed for bad weather to this day. [kyle.s.edwards.civ@us.navy.mil]

Justin Lee—graduated from Cal State Long Beach in 2019 with a master's in computer science. He joined the Point Mugu Sea Range as a Computer Scientist soon after, but was quickly drawn to geophysics by their work in the field of directed energy. Now, he supports range operations launching weather balloons and participating in atmospheric measurement campaigns while still picking up the occasional software project. Justin enjoys long walks on the beach with a BLS transmitter and catching the sunrise with the MZA DELTA. [Justin.d.lee47.civ@us.navy.mil]

Roberto M. Garcia—lives to exploit technology and understand the atmosphere. Garcia graduated from the U.S. Naval Academy in 1981 with a BS in oceanography, serving 20 years in Surface Warfare, starting in Anti-Submarine Warfare and ending in combat systems testing for Anti-Air Warfare. Garcia began live-fire T&E of DoD weapons and systems in 1987 and has been involved in the testing of 13 different weapons or combat systems on eight ship classes. Garcia recently retired as Geophysics Branch Head providing atmospheric and oceanographic support to test programs, and as Lead for Directed Energy T&E at the Point Mugu Sea Range. [surfwar81@yahoo.com]

Jim Johnson—graduated in 1998 with a bachelor's degree in engineering with a mechanical emphasis from LeTourneau University. He has held a variety of design engineering positions in areas of mechanical, optical, aerospace, defense, and robotics. He's been with Teknicare Inc. for seven years wearing a number of hats including opto-mechanical engineer, PDM administrator, mechanical project lead, structural analyst, test data analyst, and controls engineer. He holds a SWCP certification in the use of SolidWorks, and is skilled with software such as Nastran/FEMAP, SolidWorks Simulation, KiCAD, MatLab, Python, Visual Basic. [jim.johnson@teknicare.com]

Steven West—is a Teknicare, Inc. Subject Matter Expert in ultra-low noise coherent analog circuits and sensors and laser propagation testing. He has direct and extensive experience in design, development and test, propagation, effects testing, and fielding of laser devices and technology since the mid-1970s at the Air Force Phillips and Air Force Research Laboratories and The Boeing Company. Systems include extremely low-level photon counting and detection, coherent detection, and characterization of atmospheric effects on laser beam propagation. He is Chief Engineer of the Airborne Laser Effects and Atmospheric Characterization Experiment (ABLEACE). West performed circuit and performance analysis of the Walters Probe. [steve.west@teknicare.com]

Bryan L. Kelchner—leads the Teknicare Engineering Team with years of experience guiding electro-optical and laser design for multiple projects nationally. Kelchner is a graduate of the U.S. Air Force Academy and Air Force Institute of Technology, with BS degrees in physics and astronautical engineering and an MS in solid state physics. Kelchner's proven valuable in roles such as Beam Control & Fire Control Lead, Flight Test Director, Chief Engineer in early development, low and high-power integration,



and testing for a variety of programs including: Airborne Laser Program (ABL), Advanced Tactical Laser (ATL), CCS-Bright and Innovative Research for Optical Support Services (IROSS).
[Bryan.Kelchner@teknicare.com]

Abstract

As the Department of Defense (DoD) and industry continue to advance the development of high energy laser (HEL)–based directed energy weapon (DEW) systems, the need for modernized test and evaluation (T&E) and experimentation capabilities for DEW has become critical. This will allow airborne, ship-based, and land-based HEL systems to be tested against air and surface targets in a well-understood atmospheric environment. T&E capabilities in maritime test arenas and land environments will be enhanced by instrumenting open-air test ranges with advanced sensor systems, including atmospheric optical turbulence measurement systems, which play a critical role in HEL device beam and fire control. Teknicare, Inc. has conducted extensive reverse engineering of an existing differential temperature sensor previously used in data collection at various locations including Starfire Optical Range, Los Alamos National Laboratory, Nevada Test Site, Nellis Air Force Base, San Nicolas Island & Point Mugu NAWC on the PMSR as well as China Lake NAWC. This effort has included detailed requirements analysis as well as research of related designs, schematics, theses and reports conducted at Naval Postgraduate School by Dr. Donald L. Walters and his students. The devices, now known as Walters Probes, make use of precision fine wire thermocouples separated at a known distance to provide a measurement of temperature difference. The associated electronics provide necessary amplification and sampling rate to ensure a measurable ensemble average is obtained to determine CT2 at the Walters Probe location. These systems are capable of low noise operation and hence measurement of CT2 values that render atmospheric optical turbulence at extreme values indicative of terrestrial neutral events, therefore they can easily measure values expected in the near maritime environment. Revisions to the design to enhance reliability, sustainability, operability and maintainability have been undertaken to ensure the Walters Probes meet T&E mission OPTEMPO requirements.

Motivation

The Department of Defense (DoD) and industry continue to advance the development of high energy laser (HEL)–based directed energy weapon (DEW) systems. Consequently, there is a need for modernized test and evaluation (T&E) and experimentation capabilities for DEW that has become critical. These capabilities will allow airborne, ship-based, and land-based HEL systems to be tested against air and surface targets in a well-understood atmospheric environment. T&E capabilities in maritime test arenas and land environments will be enhanced by instrumenting open-air test ranges with advanced sensor systems, including atmospheric optical turbulence measurement systems, which play a critical role in HEL device beam and fire control.

Background

Teknicare, Inc. has conducted extensive reverse engineering of an existing differential temperature sensor previously used in data collection at various locations including Starfire Optical Range, Los Alamos National Laboratory, Nevada Test Site, Nellis Air Force Base, San Nicolas Island, and Point Mugu NAWC on the PMSR as well as China Lake NAWC. This effort has included detailed requirements analysis as well as research of related designs, schematics, theses and reports conducted at Naval Postgraduate School by Dr. Donald L. Walters and his students. The devices, now known as Walters Probes, make use of precision fine wire thermocouples separated at a known distance to provide a measurement of temperature difference. The associated electronics provide necessary amplification and sampling rate to ensure a measurable ensemble average is obtained to determine C_T^{-2} at the Walters Probe location. These systems are capable of low noise operation and hence measurement of C_T^{-2}



values that render atmospheric optical turbulence at extreme values indicative of terrestrial neutral events, therefore they can easily measure values expected in the near maritime environment. Revisions to the design to enhance reliability, sustainability, operability and maintainability have been undertaken to ensure the Walters Probes meet T&E mission OPTEMPO requirements.

Theory

The index-of-refraction structure parameter is a mean-square statistical average of the difference in the index of refraction between two points in space which are separated by the distance r_{12} (Walters, 1981). It is defined by

$$C_n^2 = \langle (n_1 - n_2)^2 \rangle / r_{12}^{2/3},$$

where angled brackets stand for an ensemble average. The differences in index of refraction stem from density fluctuations induced by the velocity fluctuations in the atmosphere. These differences are caused by the mixing in a turbulent velocity field of passive contaminants such as heat and moisture. C_n^2 is quite difficult to measure directly. It is usually more convenient to measure the temperature structure parameter C_T^2 , which is related to C_n^2 (neglecting humidity effects, which may have some contribution in maritime environments) by

$$C_n^2 = (79 \times 10^{-6} P/T^2)^2 C_T^2$$

where P is the atmospheric pressure in millibars and T is the atmospheric temperature in Kelvins (Tatarski, 1961). C_T^2 is defined in a similar manner as C_n^2

$$C_T^2 = \langle (T_1 - T_2)^2 \rangle / r_{12}^{2/3}.$$

This temperature structure parameter is commonly measured using a pair of fast response temperature probes as was done by Dr. Don Walters with his original devices shown in Figure 1.



Figure 36. Original Walters Probe with a Fixed Separation of 50 cm Between Thermocouples

Note: The rear plate is removed providing access to the device electronics. The horizontal arms each extend to a thermocouple used in tandem to measure temperature differences and calculate C_T^2 through an ensemble average. This configuration includes the option of power from an AC source. For transport, a protective assembly slides down each arm, is locked with a nylon screw and capped as shown on the right side.

Reverse Engineering

The original devices were documented only in an indirect manner (Holdaway, 2000; Hoover, 1991; Richardson, 1997; Roper, 1992) Related devices were described in theses and



technical papers written by Dr. Don Walters' students. However, the devices acquired by Teknicare, Inc. had no direct documentation. Therefore, reverse engineering was necessary to determine function of these devices. It is important to note that these devices were used to gather data for research purposes in specific climatic conditions (mostly desert) and specific time durations (short term calibration of other turbulence measurement systems). As such, Teknicare, Inc.'s Senior Electrical Engineer had the daunting task of reverse engineering the devices to determine a jumping off point for modernization.

Requirements Development

In order to modernize these devices, it was necessary to understand their projected functional and non-functional requirements. A variety of operational environments were anticipated to include near maritime (on the shore) as well as possibly maritime (on a buoy). In addition, deployment duration was to be extended beyond the short-term calibration times used in the past. Requirements were developed through a stakeholder analysis with NAVAIR Pt. Mugu Geophysics Branch as well as the Teknicare, Inc. Senior Combat Systems Engineer. Both parties are experts in atmospheric characterization particularly when applied to laser propagation.

Design Revisions

A dual approach was taken to apply interim revisions in anticipation of modernized design.

First, the existing devices were improved against the maritime environment by sealing possible ingress locations where possible. In addition, the operational procedure was modified to include a gentle rinse of the thermocouples exposed to salt spray with distilled water. This was done each time the devices were rotated out of the measurement cycle on ~a daily basis.

Second, a single device was radically altered to provide a variable thermocouple separation capability. This alteration provided the developer a capability to determine if there was a more optimum thermocouple separation than that of the original design. It also provides the user with the option of exploring the nature of atmospheric optical turbulence in, say, the maritime environment to determine whether or not it conforms to the accepted Kolmogorov theory.

Interim Variable Separation Test Prototype

A simple approach was taken to provide a means of effectively varying the separation between the thermocouples so that they would remain fixed during a given test series. This approach consisted of fitting tubing with fittings that allowed freedom of movement and routing of the appropriate circuitry from the electronics box to the thermocouple location at each end. The arrangement is shown with thermocouple arms extended in Figure 2. The arrangement with the thermocouple arms folded for transport is shown in Figure 3.





Figure 37. Interim Variable Separation Test Prototype with Thermocouple Arms Extended to maximum Separation

Note: The thermocouples are protected by the caps on the end of each arm. Batteries (9V Lithium) are internal to this version. The BNC data connector is visible at the bottom of the electronics box. The electronics box is exposed during a work in progress with the cover plate removed.

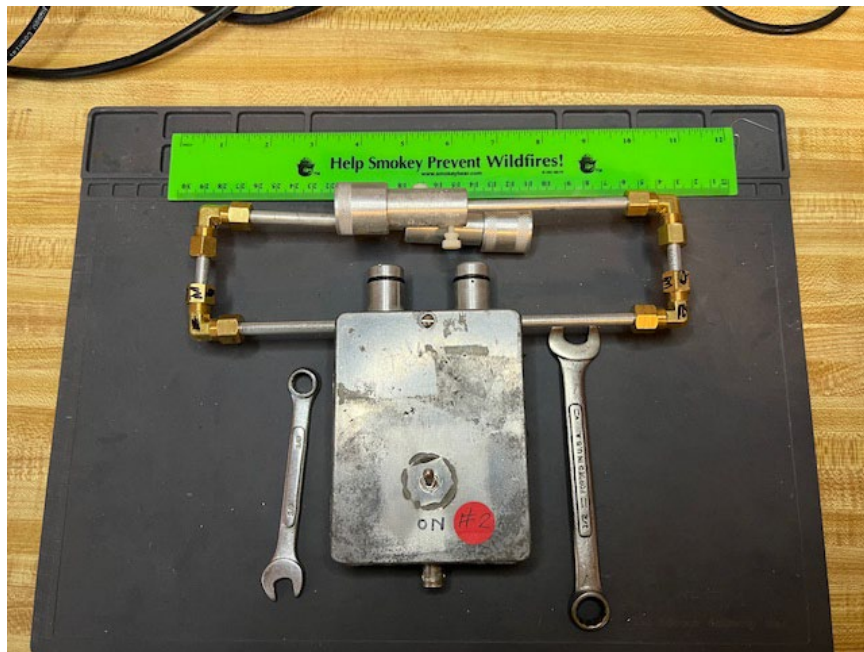


Figure 38. Interim Variable Separation Test Prototype with Thermocouple Arms Folded

Note: The thermocouples are protected by the caps on the end of each arm. The BNC data connector is visible at the bottom of the electronics box. The electronics box is now face up with the ON/OFF switch visible.

Summary/Conclusions

The modernization of the Walters Probe design is a continuing process. This work serves as a progress report along with an associated T&E report on the Interim Variable Separation Test Prototype (Nelson et. al., 2024). This test was conducted successfully and is detailed in that work.

Acknowledgments

The authors wish to express our thanks to NAVAIR Pt Mugu Geophysics Branch and the president and CEO of Teknicare, Inc. for their support in this effort.

References

- Holdaway, A.M. (2000). *Investigations of a single point temperature probe for measurement of atmospheric turbulence* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. <https://hdl.handle.net/10945/28319>
- Hoover, C.R. (1991). *Turbulence profiles and outer length scale determination in the atmosphere using balloons* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. <https://hdl.handle.net/10945/9361>
- Nelson, D., Edwards, K., Lee, J., Garcia, R., Johnson, J., West, S., & Kelchner, B. (2024). Test & evaluation of revised concept walters probes—High sensitivity differential temperature, CT2, probes for atmospheric optical turbulence characterization. *21st Annual Acquisition Research Symposium, Naval Postgraduate School*.
- Richardson, D. J. (1997). *Solar heating effects on balloon-borne microthermal probes for the airborne laser program* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. <https://hdl.handle.net/10945/8977>
- Roper, D. S. (1992). *Investigation of systematic effects in atmospheric microthermal probe data* [Master's thesis, Naval Postgraduate School]. NPS Archive: Calhoun. <https://hdl.handle.net/10945/23938>
- Tatarski, V.I. (1961). *Wave propagation in a turbulent medium*. McGraw-Hill.
- Walters, D. L., & Kunkel, K. E. (1981). Atmospheric modulation transfer function for desert and mountain locations: the atmospheric effects on r_0 . *J. Opt.Soc. Am.* 71, 397–405.



Test and Evaluation of Revised Concept Walters Probes— High Sensitivity Differential Temperature, CT2, Probes for Atmospheric Optical Turbulence Characterization

Douglas H. Nelson—is motivated to provide our troops with everything they require to defend our country. He has demonstrated operations, technical, teaching and leadership expertise in over 35 years' experience with the U.S. Army (as an active-duty Armor Officer), United States Military Academy (alumnus and former faculty member), Los Alamos National Laboratory (LANL), Raytheon, The Boeing Company, the U.S. Naval Postgraduate School (alumnus and former faculty member), the U.S. Army Space and Missile Defense Command and Teknicare, Inc. This has included a more than 30-year professional and personal relationship with Dr. Donald L. Walters, the namesake of this paper's subject.
[douglas.nelson@teknicare.com]

Kyle S. Edwards—graduated from UCLA in '08 with a BS in atmospheric sciences and has worked since graduation for NAVAIR at NAWCWD Point Mugu Geophysics. Starting there initially as an intern and weather observer calling fog on the airfield, he worked his way up to being a project forecaster, meteorologist, and now Geophysics Branch Head where he supports a wide breadth of Sea Range operations, dabbles in directed energy projects, and gets blamed for bad weather to this day.
[kyle.s.edwards.civ@us.navy.mil]

Justin Lee—graduated from Cal State Long Beach in 2019 with a master's in computer science. He joined the Point Mugu Sea Range as a Computer Scientist soon after, but was quickly drawn to geophysics by their work in the field of directed energy. Now, he supports range operations launching weather balloons and participating in atmospheric measurement campaigns while still picking up the occasional software project. Justin enjoys long walks on the beach with a BLS transmitter and catching the sunrise with the MZA DELTA. [Justin.d.lee47.civ@us.navy.mil]

Roberto M. Garcia—lives to exploit technology and understand the atmosphere. Garcia graduated from the U.S. Naval Academy in 1981 with a BS in oceanography, serving 20 years in Surface Warfare, starting in Anti-Submarine Warfare and ending in combat systems testing for Anti-Air Warfare. Garcia began live-fire T&E of DoD weapons and systems in 1987 and has been involved in the testing of 13 different weapons or combat systems on eight ship classes. Garcia recently retired as Geophysics Branch Head providing atmospheric and oceanographic support to test programs, and as Lead for Directed Energy T&E at the Point Mugu Sea Range. [surfwar81@yahoo.com]

Jim Johnson—graduated in 1998 with a bachelor's degree in engineering with a mechanical emphasis from LeTourneau University. He has held a variety of design engineering positions in areas of mechanical, optical, aerospace, defense, and robotics. He's been with Teknicare Inc. for seven years wearing a number of hats including opto-mechanical engineer, PDM administrator, mechanical project lead, structural analyst, test data analyst, and controls engineer. He holds a SWCP certification in the use of SolidWorks, and is skilled with software such as Nastran/FEMAP, SolidWorks Simulation, KiCAD, MatLab, Python, Visual Basic. [jim.johnson@teknicare.com]

Steven West—is a Teknicare, Inc. Subject Matter Expert in ultra-low noise coherent analog circuits and sensors and laser propagation testing. He has direct and extensive experience in design, development and test, propagation, effects testing, and fielding of laser devices and technology since the mid-1970s at the Air Force Phillips and Air Force Research Laboratories and The Boeing Company. Systems include extremely low-level photon counting and detection, coherent detection, and characterization of atmospheric effects on laser beam propagation. He is Chief Engineer of the Airborne Laser Effects and Atmospheric Characterization Experiment (ABLEACE). West performed circuit and performance analysis of the Walters Probe. [steve.west@teknicare.com]

Bryan L. Kelchner—leads the Teknicare Engineering Team with years of experience guiding electro-optical and laser design for multiple projects nationally. Kelchner is a graduate of the U.S. Air Force Academy and Air Force Institute of Technology, with BS degrees in physics and astronautical engineering and an MS in solid state physics. Kelchner's proven valuable in roles such as Beam Control & Fire Control Lead, Flight Test Director, Chief Engineer in early development, low and high-power integration,



and testing for a variety of programs including: Airborne Laser Program (ABL), Advanced Tactical Laser (ATL), CCS-Bright and Innovative Research for Optical Support Services (IROSS).
[Bryan.Kelchner@teknicare.com]

Abstract

As the Department of Defense (DoD) and industry continue to advance the development of high energy laser (HEL)–based directed energy weapon (DEW) systems, the need for modernized test and evaluation (T&E) and experimentation capabilities for DEW has become critical. This will allow airborne, ship-based, and land-based HEL systems to be tested against air and surface targets in a well-understood atmospheric environment. T&E capabilities in maritime test arenas and land environments will be enhanced by instrumenting open-air test ranges with advanced sensor systems, including atmospheric optical turbulence measurement systems, which play a critical role in HEL device beam and fire control. Teknicare, Inc., in conjunction with NAWC WD Pt Mugu, has conducted extensive improvement and testing of reverse engineered differential temperature sensors at various locations including San Nicolas Island and Point Mugu NAWC on the PMSR as well as China Lake NAWC. This effort has included research of related designs, schematics, theses, and reports conducted at Naval Postgraduate School by Dr. Donald L. Walters and his students. The devices, now known as Walters Probes, make use of precision fine wire thermocouples separated at a known distance, including an added variable separation capability, to provide a measurement of temperature difference. The associated electronics provide necessary amplification and sampling rate to ensure a measurable ensemble average is obtained to determine CT_2 at the Walters Probe location. These systems are capable of low noise operation and hence measurement of CT_2 values that render atmospheric optical turbulence at extreme values indicative of terrestrial neutral vents, therefore they can easily measure values expected in the near maritime environment. Testing and evaluation have been conducted to determine reliability, sustainability, operability, and maintainability revisions to be undertaken to ensure the Walters Probes meet DEW T&E mission OPTEMPO requirements.

Motivation

The Department of Defense (DoD) and industry continue to advance the development of high energy laser (HEL)–based directed energy weapon (DEW) systems. Consequently, there is a need for modernized test and evaluation (T&E) and experimentation capabilities for DEW that has become critical. These capabilities will allow airborne, ship-based, and land-based HEL systems to be tested against air and surface targets in a well-understood atmospheric environment. T&E capabilities in maritime test arenas and land environments will be enhanced by instrumenting open-air test ranges with advanced sensor systems, including atmospheric optical turbulence measurement systems, which play a critical role in HEL device beam and fire control.

Background

Teknicare, Inc. has conducted extensive reverse engineering of an existing differential temperature sensor previously used in data collection at various locations including Starfire Optical Range, Los Alamos National Laboratory, Nevada Test Site, Nellis Air Force Base, San Nicolas Island, and Point Mugu NAWC on the PMSR, as well as China Lake NAWC. This effort has included detailed requirements analysis as well as research of related designs, schematics, theses, and reports conducted at Naval Postgraduate School by Dr. Donald L. Walters and his students. The devices, now known as Walters Probes, make use of precision fine wire thermocouples separated at a known distance to provide a measurement of temperature difference. The associated electronics provide necessary amplification and sampling rate to ensure a measurable ensemble average is obtained to determine C_T^{-2} at the Walters Probe location. These systems are capable of low noise operation and hence measurement of C_T^{-2} values that render atmospheric optical turbulence at extreme values indicative of terrestrial



neutral events, therefore they can easily measure values expected in the near maritime environment. Revisions to the design to enhance reliability, sustainability, operability and maintainability have been undertaken to ensure the Walters Probes meet T&E mission OPTEMPO requirements. Progress on this design effort is detailed in an associated report (Nelson et al., 2024).

Theory

The index-of-refraction structure parameter is a mean-square statistical average of the difference in the index of refraction between two points in space which are separated by the distance r_{12} (Walters, 1981). It is defined by

$$C_n^2 = \langle (n_1 - n_2)^2 \rangle / r_{12}^{2/3},$$

where angled brackets stand for an ensemble average. The differences in index of refraction stem from density fluctuations induced by the velocity fluctuations in the atmosphere. These differences are caused by the mixing in a turbulent velocity field of passive contaminants such as heat and moisture. C_n^2 is quite difficult to measure directly. It is usually more convenient to measure the temperature structure parameter C_T^2 , which is related to C_n^2 (neglecting humidity effects, which may have some contribution in maritime environments) by

1

$$C_n^2 = (79 \times 10^{-6} P/T^2)^2 C_T^2$$

where P is the atmospheric pressure in millibars and T is the atmospheric temperature in Kelvins (Tatarski, 1961). C_T^2 is defined in a similar manner as C_n^2

$$C_T^2 = \langle (T_1 - T_2)^2 \rangle / r_{12}^{2/3}.$$

This temperature structure parameter is commonly measured using a pair of fast response temperature probes as was done by Walters with his original devices.

Note that as long as the distance r_{12} falls within the inertial subrange, between the outer scale, L_0 , and the inner scale, l_0 , the definition above of C_T^2 should remain valid if Kolmogorov theory applies. The inner scale is nominally a few mm. The outer scale is theoretically $\sim 1/2$ the probe height (~ 2 m), in this case ~ 1 meter. This concept will be put to the test with the Interim Variable Separation Test Prototype.

T&E of Initial Selected Revisions

Over several support test series in coordination with NAVAIR Pt Mugu Geophysics Branch, the original probes were deployed with improved seals with a qualitative improvement in performance. However, there were insufficient hours to fully quantify the effect of these improvements to the probes.

The operational revision of gently rinsing the thermocouples with distilled water upon cycling the probes off duty also appeared to provide a lower failure rate and increased quality of data. However, insufficient hours to fully quantify the effect were logged after these improved procedures were undertaken.

¹ Per Robi Garcia, the fixed value of 79 which appeared in the original development Kolmogorov formula has been compared to empirical results and may be adjusted to as high as 80 in maritime environments.



T&E of Interim Variable Separation Test Prototype

A test of the Interim Variable Separation Test Prototype was undertaken in June 2023 with an original probe used as a control device. The original probe had thermocouples at a separation of 50 cm. The test location and predominant wind direction are shown in Figure 1.



Figure 39. Test Location in June 2023 for the Interim Variable Separation Test Prototype

Note: Conducted on San Nicolas Island (left) on a test pad in the vicinity of Thousand Springs (center) with winds predominately from the NW so some terrestrial influence prior to reaching the test setup location for the Interim Variable Separation Test Prototype and the original control probe (right).

Results

The Interim Variable Separation Test Prototype was set to several separations as shown in Figure 2.

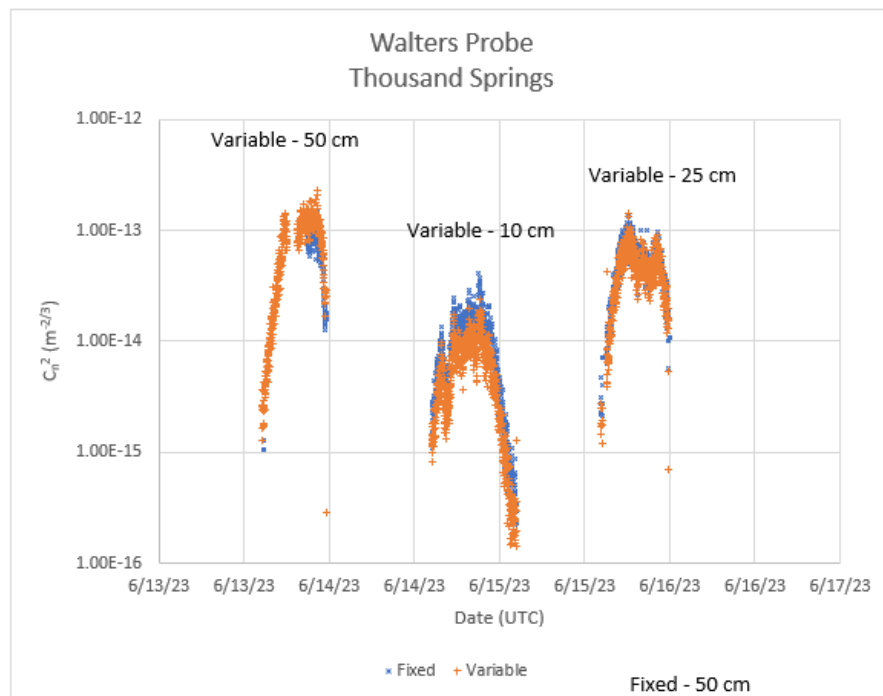


Figure 40. Results for Comparison of Interim Variable Separation Test Prototype with Original Probe.

Note: Original probe thermocouple separation was 50 cm. Interim Variable Separation Test Prototype was varied to match the original on the first day, then set at 10 cm separation on the second day, and then set at 25 cm separation on the third day. In all cases, the thermocouple separations of all probes fall within the expected inertial subrange. There was a high standard applied to this data. If there were any dropouts due to moisture in electronics, etc. over an ensemble averaging time period, then that data was not counted. However, for the data logged there is a strong visual correlation, as expected, regardless of separation.

The original probe thermocouple fixed separation was 50 cm. The Interim Variable Separation Test Prototype was varied to match the original on the first day, then set at 10 cm separation on the second day, and then set at 25 cm separation on the third day. In all cases, the thermocouple separations of all probes fall within the expected inertial subrange. There was a high standard applied to this data. If there were any dropouts due to moisture in electronics, etc. over an ensemble averaging time period, then that data was not counted. However, for the data logged there is a strong visual correlation, as expected, regardless of separation. This provides evidence that the Interim Variable Separation Test Prototype operates on sound principles and can be used to conduct optimal separation or theoretical turbulence studies.

In Figure 2 it is seen that atmospheric optical turbulence levels varied from day to day despite the agreement between the two style probes. Figure 3 shows the ambient temperature at the site during the test series that, in part, drove the terrestrial influenced turbulence levels.

Note that the level of temperature follows the general trend of atmospheric optical turbulence level for each day. The predominant wind direction during the test series resulted in a significant terrestrial influence on turbulence level. The atmospheric flow took a path that allowed surface heating and at least some convective flow of the air prior to measurement by the co-located sensors. For a rise in ambient temperature, the surface heating was likely more prevalent resulting in higher levels of atmospheric optical turbulence measured on days with higher temperature.

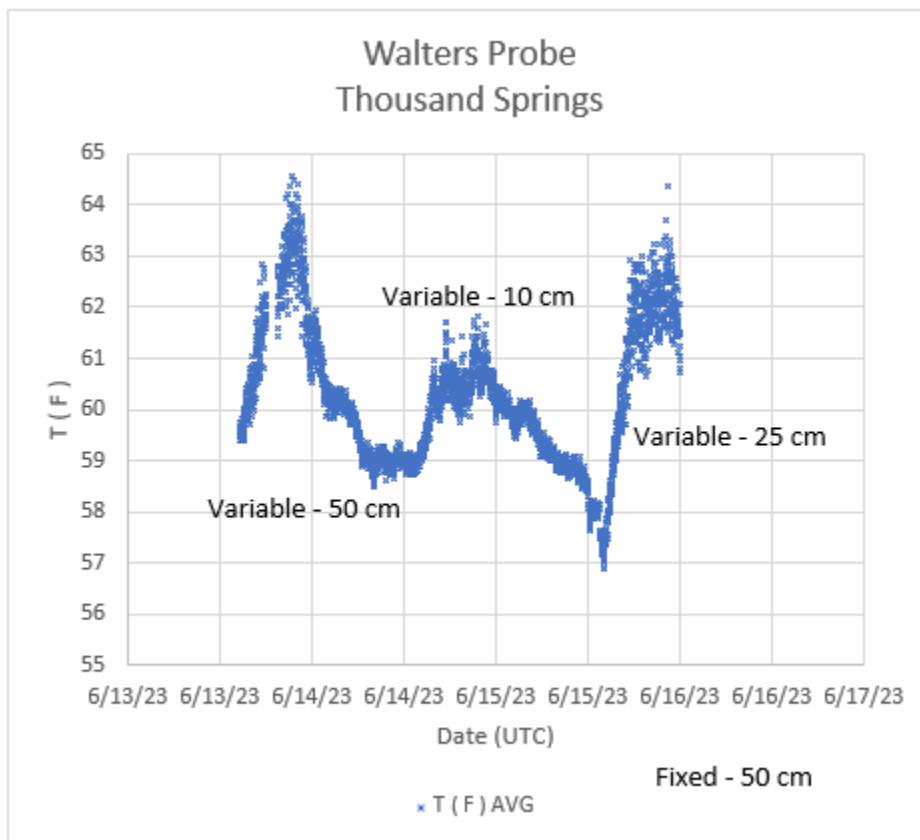


Figure 41. Ambient temperature at the Test Pad Where the Two Probes Were Taking Measurements

Note: That the level of temperature follows the general trend of atmospheric optical turbulence level for each day. The predominant wind direction during the test series resulted in a significant terrestrial influence on turbulence level and hence the importance of the ambient temperature given likely surface heating.



Summary and Conclusions

The ability to verify requirements for improvements to the original Walters probes in a challenging (maritime) test environment has been invaluable. The addition of seals and the rinsing of thermocouples during operations provided longer life and greater sensitivity to the deployed systems. In addition, the variable separation feature provides the ability to explore optimal probe separation distances for test operations as well as investigate theoretical premises in extended environments. These tests were a mere step along the way but provide confidence that a low noise atmospheric optical turbulence measurement solution is feasible for longer durations and extended environments.

Acknowledgments

The authors wish to express our thanks to NAVAIR Pt Mugu Geophysics Branch and the president and CEO of Teknicare, Inc. for their support in this effort.

References

- Nelson, D., Edwards, K., Lee, J., Garcia, R., Johnson, J., West, S., & Kelchner, B. (2024). Design revision of Walters probes—High sensitivity differential temperature, CT2, probes for atmospheric optical turbulence characterization. *21st Annual Acquisition Research Symposium, Naval Postgraduate School*.
- Tatarski, V. I. (1961). *Wave propagation in a turbulent medium*. McGraw-Hill.
- Walters, D. L., & Kunkel, K. E. (1981). Atmospheric modulation transfer function for desert and mountain locations: The atmospheric effects on r_0 . *J. Opt. Soc. Am.* 71, 397–405.



Leveraging Digital Transformation Innovating the Acquisition Workforce – Product Support Edition

Jason Thomas—serves as the Digital Transformation Lead at the Naval Air Warfare Center Aircraft Division (NAWCAD) in Patuxent River, MD, where he is responsible for establishing, integrating, and implementing the Command's Digital Transformation strategy. Primary lines of effort include, but are not limited to, MBSE, MBE, Sustainment, Data Analytics, Digital Infrastructure, Workforce Development, and Test and Evaluation across NAWCAD and its Echelon IV Commands. He is a Joint and Information Warfare qualified Aerospace Engineering Duty Officer in the Navy Reserve with experience in Space Operations, Surface, Air, and Information Warfare, and is a Systems Engineering PhD student at NPS. [Jason.j.thomas34.civ@us.navy.mil]

Abstract

The DoD acquisition workforce has undergone some changes and updates in recent years. The Back-to-Basics initiative focused on revamping and streamlining the core functional areas and emphasized foundational knowledge and practitioner levels of experience. The addition of digital engineering courses is starting to modernize the curricula for already-underway initiatives to better prepare and establish core knowledge to advance digital transformation. Even with both of these thrust areas, one major component of the lifecycle and knowledge seems to be left behind, logistics and sustainment. Back-to-Basics added emphasis for the 12 integrated Product Support elements, but not much has been updated or shown to be advancing the field or expectations for the new generation of logistics professionals. This paper discusses for improvement, modernizing and trailblazing for logistics and sustainment to reside shoulder-to-shoulder with its engineering and testing counterparts. Additional discussion will highlight where logistics plays a more integrated role in the digital framework as well as where engineering and test can learn and leverage these newfound explorative formations.

Keywords: logistics, product support, sustainment, modeling, analytics, digital transformation, lifecycle, total ownership costs, workforce, acquisition

Introduction

In 1990, the Defense Acquisition Workforce Improvement Act (DAWIA) was established as a recommendation resulting from President Reagan's Executive Order (EO) 12526, which created a "blue ribbon commission on defense management" to "study the issues surrounding defense management and organization" (DAU, 2024a; Reagan, 1985). The Defense Authorization Act of fiscal year (FY) 1991 established the Defense Acquisition University (DAU) as one of its most important provisions (Layton, 2007). The Department of Defense (DoD) acquisition workforce, both civilian and military personnel, receive the majority of their training and development through DAU. Prior to 2022, there were 16 career fields, with an individual receiving certification through a three-tiered certification program. For more than 30 years, training was a one-size-fits-all approach and delivered early in an individual's career without a deliberate focus on continuous learning and development. The DoD reevaluated training and developed a new framework to foster a culture of lifelong learning for current and future acquisition professionals. This new initiative, Back-to-Basics (B2B), took effect February 1, 2022 (DAU, 2022). With the modernization, career fields went away, Functional Areas were established, and courses were added, updated, or removed to fit the new modern acquisition era.

The new era of acquisition is moving towards an emphasis on interoperable systems, data analytics throughout operations and business, and digital operating models, specifically digital transformation (DX). Outside of the DoD, many companies are undergoing DX to improve or rethink how their organization is structured and actively realigning their organization to make



the best use of this new way of operating. The concept has been described as anything from digitizing current processes to implementing digital tools, and even how a business or organization rethinks its entire structure and approach by leveraging a digital ecosystem and operating model. While there are differing opinions on an exact definition, DX is considered the adoption and integration of digital technologies to improve efficiency and streamline business operations across all sectors of a business and fundamentally altering how value is brought to customers (Gebayew et al., 2018; Hanelt et al., 2021; Libert et al., 2016). Examples of DX technologies are Artificial Intelligence (AI), cloud computing, autonomy, and advanced model and simulation (M&S; Waugh, 2022). Other benefits may be considered modernization of current methodologies, like what Model-Based Systems Engineering (MBSE) does for systems engineering or rapid prototyping for accelerating design and manufacturing. The benefits of these transformations are increased automation, increased traceability, organization agility, strengthened testing of system designs and breadth of testing, optimization, early verification and validation (V&V), and a higher level of support integration (Khandelwal, 2020; McDermott et al., 2020; Waugh, 2022).

In the engineering technical domains, the DX research is plentiful but disparate. In other technical domains, the gap of research is wider and sparser. Within the product support and logistics community, DX is often interpreted as model-based product support (MBPS) or linking databases, where the emphasis is placed on using tools to perform a subset or discrete set of analyses (Sashegyi, 2020). Alternatively, DX is also discussed in terms of model-based methods and where sustainment should be considered but is not discussed to the necessary level with a more holistic view of integration and feedback mechanisms for the product or system across the enterprise (Draham, 2017; Gaska, 2019).

Within the new engineering and technical management Functional Area and curricula for DAU B2B, the digital landscape is plentiful. New courses include a focus on mission engineering, systems thinking, digital literacy, data analytics, multiple digital engineering courses, value management and engineering, data rights, and more. In DAU currently, there are only two courses on modern topics under digital transformation, a 4-hour course on Digital Product Support fundamentals, and a 4-hour course on data analytics fundamentals for product support (DAU, 2024c). If product support, logistics and sustainment are to be truly integrated into the future of a digitally transformed, modern acquisition profession, the training must showcase the integral nature of these important concepts, not only in the logistics Functional Area, but integrating product support more throughout acquisition training.

What is Product Support

According to the Office of the Secretary of Defense for Sustainment (OSASD[S]), product support is, “the package of support functions required to field and maintain the readiness and operational capability of covered systems, subsystems, and components, including all functions related to covered system readiness” (OASD[S], 2024). This definition is consistent in DAU and other sources as well. Decomposing this statement, we see a few main components, support functions, readiness and operational capability, and “covered systems.” To better understand the scope and complexity of product support, these components need to be discussed and understood further.

Readiness and Operational Capability

According to the DAU glossary, Readiness is defined as a “State of preparedness of forces or weapon system or systems to meet a mission or to engage in military operations. Based on adequate and trained personnel, material condition, supplies/reserves of support system and ammunition, numbers of units available, etc.” (DAU, 2024d). “Preparedness of forces” is a topic of discussion in multiple circles, including at the Service Chief level, where “readiness



is equated with availability and “capability” took a lesser role (Brown & Berger, 2021). Is the count of aircraft, tanks or assets a true measure of “being ready,” moreover, ready for what and when? Are we measuring a “fight today,” “fight tonight,” or “fight tomorrow” perspective? What are the time horizons or epochs that we’re focused on? Additionally, do we need a quantity of assets or the right quantity, at the right place, at the right time to perform a set of missions to achieve a particular objective that aligns to national security? Operational capability is an even-handed counterpart to the availability discussion. One of the challenges with “availability” and “capability” is that traditional mindsets place “availability” with the logisticians and “capability” with engineers because we artificially self-impose a view of equating “availability” with maintenance and parts and “capability” with mission threads and system performance and design.

Covered Systems

Covered systems include the specific weapon system of interest (aircraft, missile, tank, submarine), as well as the support equipment, computer hardware, test equipment, etc. to fully and completely support the safe operations of the system. Referring to Secretary of Defense Mattis, the *Redefine Readiness or Lose* article states, “the former secretary rightly concluded the U.S. military was failing to deter adversary hybrid activities, losing the gray zone competition, and losing its warfighting advantage” (Brown & Berger, 2021), the idea being that the services have been so focused on availability of primary weapon systems that the U.S. advantage has eroded in deterring our adversaries with continued “grey zone” conflict and the island build-up in the South China Sea. Strategic guidance is not explicitly part of product support, but understand those strategic documents, like the National Defense Strategy, allows product support professionals to know what and how to prepare and address the full complement of “covered systems.” In a briefing at the Association of the U.S. Army’s Global Force symposium, U.S. Army Futures Command head Gen. James Rainey stated that he believes that the effectiveness of towed artillery is near its end (Roque, 2024). This realization, one of these “covered systems” has an era or period of usefulness that needs to be evaluated and addressed proactively, is important. Systems engineers and product support professionals plan for 10, 20, 30 years of service life for a weapon system, but in the business of defense, the enemy gets a vote, too, and our systems may become less effective, even if the performance of the system is unchanged. This may be the networks, weapon system or other “covered systems” that are the responsibility of sustainment and logistics professionals, and a modern way to develop and train the workforce to this reality requires a digital approach.

Support Functions

The support functions of product support are typically understood to be defined as the 12 Integrated Product Support (IPS) elements (DAU, 2024b). These include:

- 1. Product Support Management** is the development and implementation of product support strategies to ensure supportability is considered throughout the system life cycle through the optimization of the key performance outcomes of reliability, availability, maintainability, and reduction of total ownership costs.
- 2. Design Interface** is the integration of the quantitative design characteristics of systems engineering (reliability, maintainability, etc.) with the functional logistics elements (i.e., IPS Elements). This includes Reliability and Maintainability (R&M), testability, legal requirements, Human Systems Integration, and more.
- 3. Facilities and Infrastructure** are the permanent and semi-permanent real property assets required to support a system, including studies to define types of facilities or facility improvements, location, space needs, environmental and security requirements, and equipment.



4. **IT Systems Continuous Support (formerly computer resources)** encompasses the facilities, hardware, software, firmware, documentation, manpower, and personnel needed to operate and support mission critical information technology (IT) systems hardware/software systems.
5. **Maintenance Planning and Management** establishes maintenance concepts and requirements for the life of the system. It includes, but is not limited to, levels of repair, repair times, testability requirements, support equipment needs, manpower skills, facilities, interservice, organic and contractor mix of repair responsibility, site activation, etc.
6. **Manpower and Personnel** involves the identification and acquisition of personnel (military & civilian) with the skills and grades required to operate, maintain, and support systems over their lifetime.
7. **Packaging, Handling, Storage, and Transportation (PHS&T)** is the combination of resources, processes, procedures, design, considerations, and methods to ensure that all system, equipment, and support items are preserved, packaged, handled, and transported properly, including environmental considerations, equipment preservation for the short and long storage, and transportability.
8. **Supply Support** consists of all management actions, procedures, and techniques necessary to determine requirements to acquire, catalog, receive, store, transfer, issue and dispose of spares, repair parts, and supplies.
9. **Support Equipment** is made up of all equipment (mobile or fixed) required to support the operation and maintenance of a system. This includes ground handling and maintenance equipment, tools, metrology and calibration equipment, and manual and Automatic Test Equipment (ATE).
10. **Sustaining Engineering** spans those technical tasks (engineering and logistics investigations and analyses) to ensure continued operation and maintenance of a system with managed (i.e., known) risk. Sustaining Engineering involves the identification, review, assessment, and resolution of deficiencies throughout a system's life cycle. Sustaining Engineering both returns a system to its baselined configuration and capability and identifies opportunities for performance and capability enhancement.
11. **Technical Data Management** consists of recorded information of scientific or technical nature, regardless of form or character (such as equipment Technical Manuals [TMs] and engineering drawings), engineering data, specifications, standards and Data Item Descriptions (DIDs).
12. **Training and Training Support** consists of the policy, processes, procedures, techniques, training devices, and equipment used to train civilian and military personnel to acquire, operate and support a system.

Product Support and Digital Transformation

The next generation of product support professionals must be more thoughtfully considered in the digital landscape as it relates to training and development. In supportability analysis, there is an expression of Design for Support, Design the Support, and Support the Design (Dallosta & Simcik, 2012). This holds true in training and workforce development, particularly as the DoD implements its digital transformation. For example, designing for support in engineering has included the use of MBSE, design the support has meant the specific tools and software to deliver the products and services, and support the design includes the software licenses, the training, and instructions to deliver effective digital designs in that MBSE solution. The Assistant Secretary of the Navy for Sustainment (ASD[S]) released a Strategic Plan which highlighted the need to deliver sustainable logistics to support DoD mission requirements,



deliver cost-effective materiel readiness to meet the DoD's warfighting requirements, enable effective, affordable, and sustainable warfighting capability, and optimize warfighter logistics (Lowman, 2023). In order to deliver on this strategic plan, product support must be included more foundationally in digital transformation initiatives, particularly the future training of the workforce.

IPS-Digital Integration

Deliberate actions must be taken to ensure product support is engineered and integrated into the digital thread of weapon systems development and business systems and become the new normal of acquisition and sustainment. Some options to more clearly call out the IPS elements with digital integration are listed below.

- 1. Product Support Management** – all support functions, costs, manpower, contracts and requirements more clearly integrated into digital management models, such as a model-based program manager model akin to a MBSE instantiation of the weapon system. This allows the holistic view of needs early for partnership planning, funding requests, maintenance concepts and more.
- 2. Design Interface** – identification and instantiation of the “-ilities” (i.e., maintainability, reliability, testability, interoperability, etc.) in the system weapon and mission engineering models for effectiveness and demands on the design.
- 3. Facilities and Infrastructure** – analytics and data modeling of throughputs, demands and realistic delivery of products through our facilities and support structures.
- 4. IT Systems Continuous Support** – redundant services to ensure product support and sustainment activities can continue if a disruption occurs, including when and where data needs to flow to ensure the warfighter is supported always.
- 5. Maintenance Planning and Management** – detailed understanding of work packages and task analyses integrated into digital tools for evaluation and tracking to plan. Are the plans optimized and packaged appropriately based on lessons learned and active data feedback? How effective are the repairs, how skilled are the personnel, and can we integrate training with effectiveness of the maintenance plans? See Fisher's (2019) *Moneyball for Maintainers* for additional information.
- 6. Manpower and Personnel** – What new rates or skill sets are needed and the most effective at performance of the duties? What do we understand about manning documents, ascension and rotations, and how does that affect our manpower documents and assumptions?
- 7. Packaging, Handling, Storage, and Transportation (PHS&T)** – similar to service life design considerations and how are strategies conceptualized? What works, what does not, and are there special considerations for afloat, ashore or coalition practices with respect to these functions should be evaluated? There are a number of overlapping considerations with the “-ilities” in systems engineering, and reinforcing that connection only helps the design and support of the system.
- 8. Supply Support** – Specific analysis on forecasting, optimal planning, just-in-time support, vendor producibility, condition-based maintenance, and other critical topics should be expanded upon when training on supply support. Those key concepts help reinforce or redirect our understanding to then develop a more realistic sustainment support strategy.
- 9. Support Equipment** – Discussions on condition-based maintenance, software updates and network connectivity, shared assets and more are assisted with digital transformation tenets and practices. What is capable in austere environments or at high operational tempo, and how does that affect training or maintainability?



- 10. Sustaining Engineering** – one of the most important elements of integrated logistics, and it should be discussed more in the engineering and technical management functional area. Specific training on in-service repairs and requests and how local policy for deviations may affect throughput or ability to disposition support requests and the impact on manuals, publications and how does data availability help or hinder this element.
- 11. Technical Data Management** - Modern workforce training should include product lifecycle management (PLM) capabilities and limitations to ensure what is required to utilize this important digital capability. The training should highlight the tools available but also the significant preparation work, data mapping, use case definition and other factors before integrating this key enabler. The engineering and technical management functional area should also have training in this area as it applies to airworthiness or safety of operations considerations.
- 12. Training and Training Support** – Digital integration and workforce development is primed for this IPS element given the intersection and overlap with many of the other elements. This area should include how people learn and new methods of conveying training, as well as what do personnel “need to know,” compared to “ought to know” and when each are relevant. The media in which people learn will increase or educe effectiveness, especially for the complexity of the work. Product support training in this element should be considered at each phased and design review for applicability, maturity, and effectiveness for the best solution and approach and not simple re-use of previous methods.

Whether the digitally transformed training for these elements are with DAU or another institution, specific courses should be developed or updated based on the above recommendations to truly modernize workforce training and development in this functional area. The training should be on each element as well as training where the elements overlap and interplay with one another. It is also important for other functional area professions to cross-train in the logistics functional area so interoperability and partnering occurs at a more professional and academic level, before the practical applications of program implementation and support create a “fog of acquisition and sustainment.”

Other sources of professional development

In addition to DAU, there are other sources of training and professional certification for the product support community, including the International Society of Logistics (SOLE). Founded in 1966, SOLE (2024) is “a non-profit international professional society composed of individuals organized to enhance the art and science of logistics technology, education and management.” SOLE offers three certification levels, with the latest version of qualifications released in 2012. For the highest level of certification, the Certified Professional Logistician Program, the exam topics cover concepts of systems management, principles and functions of management, system design and development, formal design review, acquisition product support, production support, distribution and customer support, customer support, and equipment phase-out and disposition. One can see the overlap with the 12 IPS elements such as Design Interface, Product Support Management, Supply Support, PHS&T, and Support Equipment. What appears to not be included are system performance, operational use and analysis and lifecycle management of the product. The certification focus emphasizes certain areas and is gapped in others. A holistic view of what constitutes product support in training must be addressed for the profession to atrophy is key areas.

Other certifications in logistics include the Certified Supply Chain Professional (CSCP), Certified in Planning and Inventory Management (CPIM), Certified Professional in Supply Management (CPSM), the Council of Supply Chain Management Profession (CSCMP) and



many others (Indeed, 2023). The glaring observation with the list is that the primary focus is on the Supply Support IPS element, with other elements either minimized, with the potential to align Program Management Professional to Product Support Management, or left out completely. This indicates a lack of awareness or connection to other specific training and focus that align to Product Support, including reliability, quality, human systems, maintainability, or testability.

The DoD acquisition workforce at-large now has access to modern digital learning platforms such as LinkedIn Learning and the U.S. Air Force's Digital University to augment their learning paths and curricula (*LinkedIn Learning*, 2024; USAF, 2024). LinkedIn Learning has a number of courses that would assist product support professionals, including IT certifications, data analysis, business and program management, and supply chain analysis. Digital University "provides anytime access to Silicon Valley accredited technology training & fosters a community of learners for tomorrow's warfighter." Courses that would assist the product support community include supply chain management, IT, and data analysis. What is lacking from these platforms and offerings are the specialized training the product support workforce could leverage, including system design, PLM, Instructional System Design (ISD), facility management, operational use cases for data analysis and modeling, data rights, and more. It is important to note that the focus of these training offerings focus on IT systems and analysis rather than model-based methods for product support, logistics or a deeper, more integrated approach for how product support and sustainment activities of our weapon systems should be addressed.

Policy Considerations

The DoD released a new Digital Engineering Instruction in December 2023 to establish policy, assign responsibilities, and provide procedures for implementing and using digital engineering in the development and sustainment of defense systems (USD[R&E], 2023). However, this guidance gives only passing reference to logistics and product support with inclusion of "product lifecycle management" and "sustainment of a system" in a couple of areas. In the *Requirements for the Acquisition of Digital Capabilities Guidebook*, the only coverage to product support or sustainment is in regards to purchasing the data rights to support the weapon system and operation and support of requisite IT (DoD CIO, 2022). Policy documents are intended to be high-level and not tactical in nature; however, more clarity could prove useful for multiple communities on integrating their perspectives into a more holistic solution. Policy documents, including those regarding digital transformation and digital engineering, should be deliberate in referencing and integrating product support to strengthen the ties across the lifecycle. Equally, sustainment and product support policies should include reference to where the design and operational performance development and trade-offs have an impact on overall system availability and effectiveness. This bi-directional linkage only serves to reinforce the acquisition workforce to think, act, and operate differently in the acquisition and sustainment of our weapon systems and capabilities. The digital landscape makes this alignment a reality and apparent in the day-to-day operations at all levels of planning and execution. It is the only way to truly reach digital acquisition and sustainment.

Recommendations and Future Work

The product support community, made up of operations experts, logisticians, business financial professions and more, has not updated their workforce development and training curricula to keep pace with the digital revolution. Some of the course ware is updated but remains limited in breadth and depth to properly cover the scope of product support and cross-cutting dependencies. Specific areas for consideration are provided and recommended to change the status quo of previous program experience to a truly new approach. For example, we rely on people who use the systems to develop the training, which may be sufficient as a



secondary aspect that complements the need for a primary, deliberate focus by trained professionals in systems or maintenance training. The acquisition community spends a lot of time developing the weapon system specification but considerably less time specifying the sustainment of that weapon system, including system degradation, system training and continuous feedback and continuous certification assessment utilizing advance analytics and digital methods. This research makes the following recommendations:

- 1.) Develop new courses in DAU that discuss and address the IPS elements for the new modern approach to the acquisition community
- 2.) Develop courses in the logistics functional area that address specific applications of the IPS elements and digital transformation tenets
- 3.) Align product support more intentionally with digital engineering policy, guidance and instructions
- 4.) Develop a new product support certificate program or master's track with universities specifically targeted for the acquisition workforce similar to the Master of Science in Engineering Management and Systems Engineering curricula for the Engineering and Technical Management functional area.

Future work in this space includes the need for prioritization of what aspects of product support are the biggest contributors or degraders of overall availability and capability of our weapon systems. This includes both system level and integrated system performance within an integrated operational capability. Additional insights as to what areas of product support are either obsolete or contributing less to mission outcomes than previous system designs and support constructs. These areas should be addressed with advanced analytics, conceptual and data modeling, artificial intelligence and other capabilities within digital transformation to keep pace with the necessary decision superiority to support and defend our interests around the world. Workforce development must keep pace and be infused with these necessary skills to ensure the transformation is supported and sustained. Product support is primed to leverage digital transformation to innovate its workforce and broader reach to deliver decisive combat power in warfighting, strengthen our navy warfighters, and reestablish a foundation of being ready for the demands of the future.

References

- Brown, C., & Berger, D. H. (2021, March 15). *Redefine readiness or lose*. War on the Rocks. <https://warontherocks.com/2021/03/redefine-readiness-or-lose/>
- Dallosta, P. M., & Simcik, T. A. (2012). *Designing for supportability: Driving reliability, availability, and maintainability in...*
- DAU. (2022). *Getting back-to-basics*. <https://www.dau.edu/back-to-basics>
- DAU. (2024a). *Defense Acquisition Workforce Improvement Act*. <https://www.dau.edu/glossary/defense-acquisition-workforce-improvement-act>
- DAU. (2024b). *Integrated product support (IPS) elements—Overview*. <https://www.dau.edu/acquipedia-article/integrated-product-support-ips-elements-overview>
- DAU. (2024c, April 5). *Digital and data search in DAU for LOG*. https://www.dau.edu/search?search=data&f%5B%5D=content_type%3Acourses&f%5B1%5D=courses_credentials_workshops%3ALife%20Cycle%20Logistics
- DAU. (2024d, April 5). *Readiness definition*. <https://www.dau.edu/glossary?combine=&title=R>



- DoD CIO. (2022). *Requirements for the acquisition of digital capabilities guidebook*.
- Draham, S. (2017, May 24). *Naval aviation readiness models list v9*.
- Fisher, K. (2019, February 18). *Moneyball for maintainers*. AMCL.
<https://www.marinecorpslogistics.org/post/moneyball-for-maintainers>
- Gaska, M. (2019). *Proceedings of the 10th model-based enterprise summit (MBE 2019)* (Presentation NIST AMS 100-24). National Institute of Standards and Technology.
<https://doi.org/10.6028/NIST.AMS.100-24>
- Gebayew, C., Hardini, I. R., Panjaitan, G. H. A., Kurniawan, N. B., & Suhardi. (2018). A systematic literature review on digital transformation. *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 260–265.
<https://doi.org/10.1109/ICITSI.2018.8695912>
- Hanelt, A., Bohnsack, R., Marz, D., & Antunes Marante, C. (2021). A systematic review of the literature on digital transformation: Insights and implications for strategy and organizational change. *Journal of Management Studies*, 58(5), 1159–1197.
<https://doi.org/10.1111/joms.12639>
- Indeed. (2023, July 31). *11 supply chain certifications to advance your career*. Indeed Career Guide. <https://www.indeed.com/career-advice/career-development/supply-chain-certifications>
- Khandelwal, H. (2020, November 22). *Digital thread implementation presents both a challenge and an opportunity* [Social media]. LinkedIn. <https://www.linkedin.com/pulse/digital-thread-implementation-presents-both-challenge-khandelwal/>
- Layton, E. (2007). *The Defense Acquisition University: Training professionals for the acquisition workforce 1992-2003*. DAU.
- Libert, B., Beck, M., & Wind, Y. (2016, July 14). 7 questions to ask before your next digital transformation. *Harvard Business Review*. <https://hbr.org/2016/07/7-questions-to-ask-before-your-next-digital-transformation>
- LinkedIn learning: Online courses for creative, technology, business skills*. (2024).
<https://learning.linkedin.com/cx/get-started>
- Lowman, C. (2023). *Assistant secretary of defense for sustainment strategic plan*.
- McDermott, T., Salado, A., Van Aken, E., & Henderson, K. (2020). *A framework to categorize the benefits and value of digital engineering*. SERC.
- OASD(S). (2024). *Welcome to ODASD (product support)*.
https://www.acq.osd.mil/log/PS/about_PS.html
- Reagan, R. (1985, July 15). *Executive order 12526—President's blue ribbon commission on defense management*. <https://www.reaganlibrary.gov/archives/speech/executive-order-12526-presidents-blue-ribbon-commission-defense-management>
- Roque, A. (2024, March 27). Towed artillery has reached 'end of the effectiveness,' Army four-star declares. *Breaking Defense*.
<https://breakingdefense.sites.breakingmedia.com/2024/03/towed-artillery-has-reached-end-of-the-effectiveness-army-four-star-declares/>
- Sashegyi, M. (2020, March 31). *Model based product support heralds in new-age logistics and engineering*. DVIDS. <https://www.dvidshub.net/news/366283/model-based-product-support-heralds-new-age-logistics-and-engineering>



SOLE. (2024). *SOLE - About SOLE*. <http://www.sole.org/about.asp>

USAF. (2024). *Digital university*. <https://digitalu.af.mil/>

USD(R&E). (2023). *Digital engineering* (DoD Instruction 5000.97).

https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500097p.PDF?ver=bePlqKXaLUTK_lu5iTNREw%3d%3d

Waugh, S. (2022, July–August). Digital strategy return on investment. *Defense AT&L Magazine*, 6–12.



The Acquisition Game: Learning Through Play

Matthew MacGregor—is a Fellow at the Acquisition Innovation Research Center. He previously worked for the MITRE Corporation as an Acquisition SME and spent 21 years as a military and civil service program manager (PM) across the space, command and control, weather, and aircraft portfolios. His last PM assignment was as the F-35 Deputy Program Manager and his last five years in government were at the Pentagon, where he served in multiple acquisition leadership roles. His passion is reforming our current industrial age acquisition system for the 21st century. [mmacgreg@stevens.edu]

Tory Cuff—is a Fellow at the Acquisition Innovation Research Center. She has expertise involving the Commission on Planning, Programming, Budgeting, and Execution; workforce initiatives; and acquisition management policies and practices. She was the Senior Advisor for Agile Acquisition for the Office of the Undersecretary of Defense for Acquisition and Sustainment (OUSD[A&S]), where she led the Single Appropriation Pilot for Software and Digital Technology. She played a leading role in building the Air Force's Kessel Run software factory, where she pioneered ways to deliver software to classified networks. [vcuff@stevens.edu]

Abstract

The DCTC Acquisition Game immerses players into the basic processes of the defense acquisition system. It is composed of three phases: a technical solution phase, a contracting approach phase and a program management phase. Traversing these phases allows players, who may have limited to no experience in acquisition, to gain an understanding of the various decision points, the key information needed, the critical thinking required, and the challenges that can be experienced throughout the life cycle of a typical acquisition program. The various decision points include the selection of a technical solution (which emulates a design review), the approval of a contracting approach (which emulates a source selection), and a leadership review (which emulates the acquisition strategy approval process). As the game unfolds, detailed requirements and contract rules allow players a small glimpse into the complexities of the acquisition process. Players must demonstrate critical thinking during the technical solution phase process to ensure the operational user's minimum requirements are being satisfied. The game is configured so that no one solution can meet the criteria for a satisfactory solution. Instead, the players, who play as part of teams, must add technical enhancements (which emulate system modifications) to meet the necessary capability level. Players experience challenges through the program management phase of the game, selecting cards that represent both positive and negative developments that can occur through the life cycle of a typical acquisition program. These challenges serve not only to illuminate the complexity of the acquisition process but also to teach players about the specific challenges they will likely encounter throughout their acquisition career. The overall intent of exposing players to these various elements is to promote learning through fun. As part of the Defense Civilian Training Corps curriculum, scholars are expected to gain increasing familiarity and knowledge through repeated play, setting the foundation for higher-level learning objectives.

Introduction

The civilian defense acquisition career field currently faces a situation where, in less than five years, more than 31% of the current workforce will be retirement-eligible (Figure 1; DoD, 2021). This shift will precipitate the need to onboard an increasing number of new civilian employees to backfill the inevitable exodus. The Department of Defense (DoD) has already received new authorities, such as Direct Hiring Authority to recruit experienced members of the civilian population and help mitigate the challenge, **but these have proven inadequate**(DoD, 2024).

In 2020, Congress created the Defense Civilian Training Corps (DCTC) and subsequently directed the Under Secretary of Defense for Acquisition and Sustainment



(USD[A&S]) to lead the effort. The USD(A&S) has leveraged the Acquisition Innovation Research Center (AIRC) to design DCTC as a civilian talent pipeline that minimizes typical hiring frictions by facilitating summer internships and security clearances. In the inaugural DCTC cohort, 87 undergraduate scholars

¹ at four pilot universities (North Carolina A&T, Purdue University, University of Arizona, and Virginia Tech) are engaged in a multidisciplinary, active-learning, and acquisition-oriented curriculum along with summer internship projects at DoD organizations to prepare them for an acquisition career.

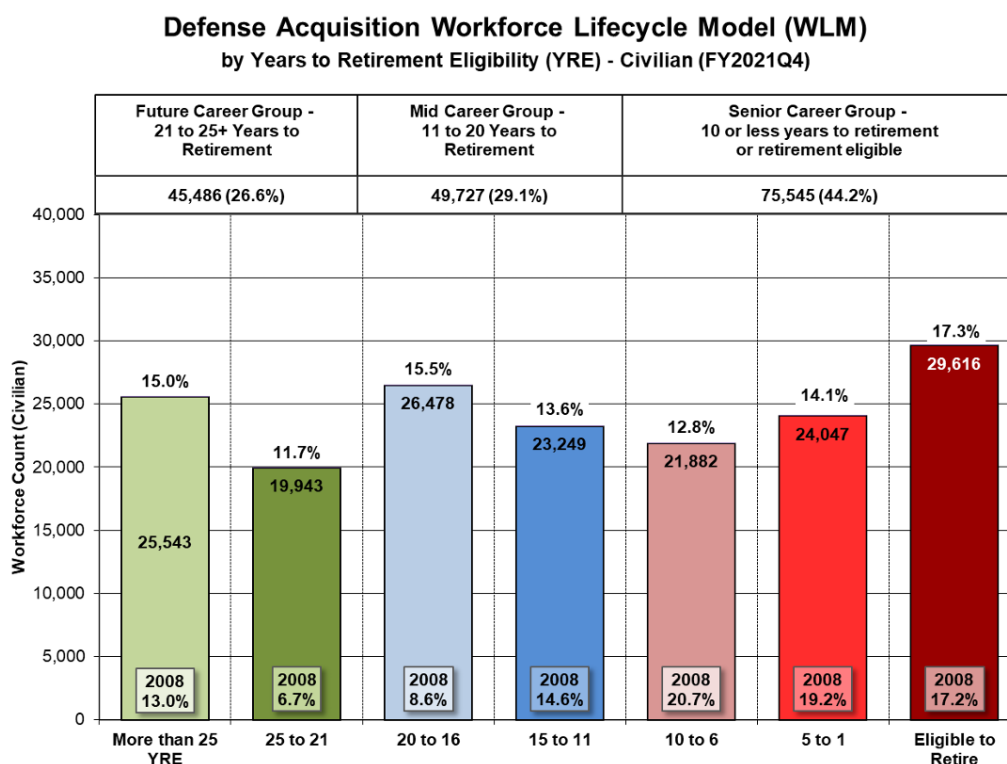


Figure 1. Acquisition Workforce—Years to Retirement Eligibility

The development and integrated curriculum approach is a hallmark of the DCTC program, which is designed to provide scholars with the skills and experience needed to jump-start their professional careers and be productive members of the DoD acquisition community immediately upon graduation (OUSD[A&S], n.d.). However, as has been demonstrated through multiple studies, developing a competent acquisition workforce is a challenge (Levine, 2019).

In a 2016 study, 64% (of 250 government respondents) stated that it takes 10 years or more to become fully proficient in acquisition with 90% of that group relying on colleagues and references for their training—leaving only 10% of the respondents citing formal acquisition training as their primary avenue for learning and developing acquisition skills (Murphy & Bouffard, 2017). Admittedly, this study is now seven years old, and acquisition training organizations like the Defense Acquisition University have overhauled their approach to acquisition training with a “back to basics” model (DAU, n.d.). However, there is still good reason to believe that the generation graduating college now (Gen Z) and future generations will

¹ Students who are competitively selected to join DCTC are referred to as “scholars.” Scholars are held to the highest standards academically, professionally, and in the community.



demand different training models than exist today. One expert noted that Gen Z has a preference for “interactive and immersive learning experiences such as video games, virtual reality, and other digital tools” and often prefers a “more collaborative and peer-to-peer learning environment” (Ávila, 2023). Jeff Koses, Senior Procurement Executive at the General Services Administration (GSA), said that the government needs to be “rethinking [acquisition] training to match the way these digital natives are accustomed to receiving and consuming information” (McCabe & Laurent, 2015).



The DCTC curriculum team has embraced interactive approaches in the classroom by blending traditional lecture/seminar-style learning with multiple in-class exercises that can be completed as a group. The reception thus far from the scholars and professors is that the interactive lessons evoke the most interest and maintain their attention for longer. In the most recent semester (a course designated “DCTC 302”), the team also introduced to The Acquisition Game, which emulates the acquisition process at a basic level.

Games and Acquisition

One notable author on acquisition-focused gaming, Dr. Dan Finkenstadt, defined gamification as “the means of acquiring new skills or knowledge infused and enhanced with game-like elements.” His research and experiences in exploring gamification for defense training and education revealed three primary game modalities: (1) Serious/Simulation Gaming, (2) Exposure Gaming, and (3) Engagement Gaming. He observed serious games are more realistic and focused on “performing real-world tasks in a simulated operating environment with the intent to sharpen skills”; exposure games test and improve the “skills and abilities” of specific work roles less the environment realism; and engagement games are more focused on “introducing curriculum subject matter . . . in an alternate universe/setting to evoke a sense of increased interest and engagement.” Summarily, Finkenstadt asserts that games primarily seek to make learning fun and that “gamified learning as augmentation may be the best approach for most situations and curricula.”

This finding is consistent with another group of professors who teach political science at the University of Albany and whose research concluded that students found lectures difficult to understand and less effective compared to interactive approaches or active learning. According to one study cited, the “traditional lecture format should be used together with active learning to achieve specific . . . learning outcomes.” Based on these findings, the professors devised “mini-games” to help students improve their understanding in a new and engaging way where active participation was required (Asal et al., 2018).



Gamification research reinforces the DCTC curriculum philosophy as it relates to teaching complex and sometimes arcane acquisition topics and developing specific skills needed to be successful. The Acquisition Game is a combination of Finkenstadt’s exposure and engagement gaming as it seeks to practice certain skills and abilities while also introducing curriculum subject matter in a more accessible way using what is essentially a mini game.

Game Development Approach

The AIRC fellows who served as DCTC game developers, admittedly game design novices, initially planned to design a video game that incorporates a host of complex acquisition factors with detailed personae representing typical program office roles. However, the team quickly realized, that the initial design would quickly outpace the curricular learning objectives and convey more complexity up-front than was desired.

We pivoted by taking a page from the agile software techniques that we have advocated and taught across DoD, and adopted an iterative approach to building a Minimum Viable Product. Under this approach, a game that conveys an appropriate level of acquisition knowledge while also providing an enterprise perspective could be quickly developed and provided to the DCTC scholars. Our overarching goal was to build a game that helps scholars understand, in a simple way, the experience of navigating the entire acquisition life cycle, thereby providing a foundation to make other curricular lessons easier to absorb and contextualize.

We strove to convey four learning objectives through the game. Although there are many acquisition functions that could be covered within those objectives, we deemed four primary elements—user needs, systems engineering, contracting, and program management—to be essential to understanding the acquisition process from a 10,000-foot view.

Table 1. Learning Objectives Supported by the Acquisition Game

Recognize the operational user’s role in the DoD acquisition process.	Explain the general elements that go into an acquisition strategy.
Identify and discuss the major components and processes of contracting and criticality of market research.	Recognize the broader group of stakeholders who have equity in acquisition outcomes.

In developing the board game, we decided to compress the four primary elements into three phases: determining the technical solution, selecting a contracting approach, and managing the program scholars receive the requirement in the game instructions, but then immediately face the complex user engagement considerations of how to meet those requirements. As the game board turns to the contracting approach phase, scholars are introduced more directly to different elements of the contracting process. Details were written on the playing cards and incorporated into game play to support learning objectives and foster absorption of some of the more arcane aspects of government contracting.

The number of different scenarios in the typical execution of an acquisition program is vast, so we had to design the management phase of the board to reflect these complexities. We crafted a series of playing cards, each representing an event that could reasonably happen in execution. These included both positive and negative outcomes such as contractor management turnover or conversely having a successful testing event. Scholars select different cards at random introducing an element of chance into the game to simulate the reality that certain factors are outside the control of the acquisition professional.



The game also includes several decision points or major activities to simulate the reality of a typical acquisition program: a Leadership Review to emulate an Acquisition Strategy Review (or Panel); a Protest Status to emulate the potential for a contract to be protested; a Requirement Change to emulate the reality of an Engineering Change Proposal due to a shifting user need or a new threat; and finally a Fielding Decision to represent user acceptance of the system and satisfaction of criteria (i.e., Initial Operating Capability) to operationally field the capability as the culminating action of the game.






Figure 2. The Acquisition Game Phases

Game Play

The Acquisition Game is played in the context of a recent wildfire that has devastated the Arctic and Siberia, leaving many stranded without adequate resources (water, food, and shelter). The players represent a team supporting DoD operators who are coordinating a response and providing military situational awareness capabilities.

Scholars start the game with a set allocation of chips that represent individual units of cost, schedule, and user satisfaction (Figure 3). The goal of the game is to acquire the maximum number of user satisfaction chips without exhausting the cost or schedule chips. At the conclusion of the game, cost and schedule chips can be exchanged for user satisfaction chips at a 2:1 ratio. The player with the most chips wins. For more details on the game board, see Appendix A.

Starting Balance. Each player or team begins with:

-  **10** Schedule Chips
-  **10** Funding Chips
-  **8** User Satisfaction chips

Exchanging Chips. If, during the course of the game, you are low on any particular chip, **the exchange rate is 2:1. You must give up 2 of the same type.** ***If you only have 1 chip, it cannot be used in an exchange.***

Figure 3. Allocation of Chips

Scholars begin gameplay by reviewing the provided set of user requirements (Figure 4). The complexity introduced here is that some requirements are mandatory and some are discretionary. It is up to each team to decipher, from the language, which are most important. The bolded text provides hints.



The requirement process includes tradeoffs that must be considered to deliver the most viable solution available. Relevant requirement information is below:

- A typical operation is executed in 2 weeks with 24-hour coverage.
- The typical coverage range **required** for a humanitarian mission from the main operations center to affected areas is 50 miles.
- Roughly 60% of humanitarian operations occur in degraded weather conditions.
- The budgeted cost of a typical humanitarian operation (excluding logistics) including all operational costs is \$5 million.
- A 10 sq/m resolution image or better is **required** to provide the information needed to generate viable courses of action. Image storage requires at least 1TB through a mission.
- Statutorily **required** to maintain 85% availability or better of its sensing capability.

Figure 4. System Requirements

After reviewing the requirements, each team considers different technical solutions that could satisfy the user’s needs (Figure 5). This process is not dissimilar to the Analysis of Alternatives that an acquisition program might support. Another complexity introduced here is that no solution will be able to fully satisfy the requirements. Instead, there will be a need for additional system capability, which we term Technical Solution Enhancements (Figure 6).

These additional features represent some typical tradeoffs that would occur during a Preliminary or Critical Design Review, where government, contracting teams, and users discuss the various options and the associated cost and schedule implications. To emulate reality, there are cost and schedule penalties (handing in chips to the banker) for adding enhancements (Figure 6). Another element of the Technical Solution Phase is to choose a development approach that encompasses decisions about how engaged the government plans to be throughout the engineering and design process (Figure 7).

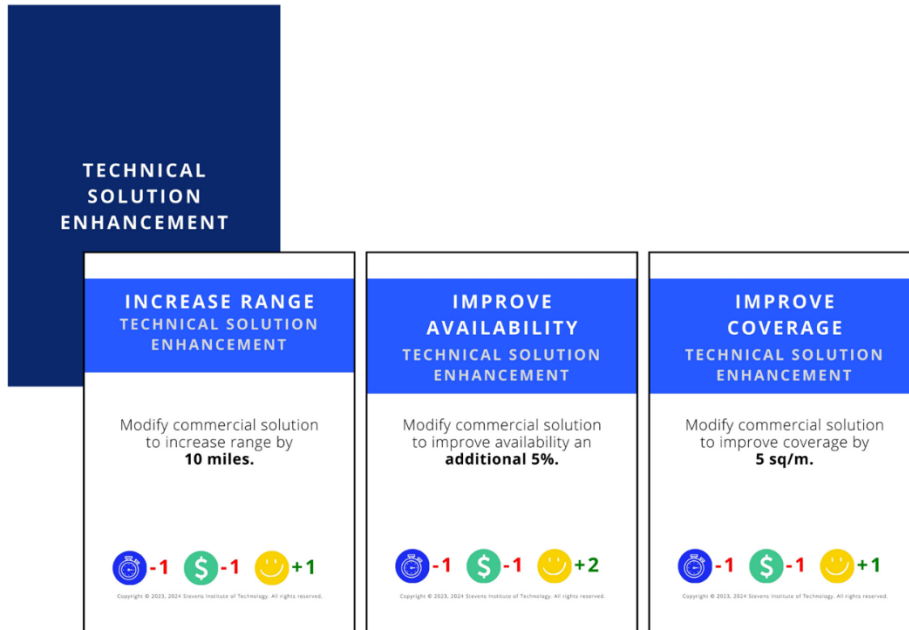
Solution	Qty (100 sq/m)	Availability	Resolution	Range	Weather Tolerance
UAV Swarm	100	85%	10 sq/m	40 miles	High
UAV Fleet	10	80%	5 sq/m	100 miles	Moderate
Satellite	N/A	75%	20 sq/m	N/A <i>meets requirement</i>	Low

*NOTE: **No solution meets all the requirements, you will need to combine a technical solution with a solution enhancement(s) cards.**

Figure 5. Technical Solutions



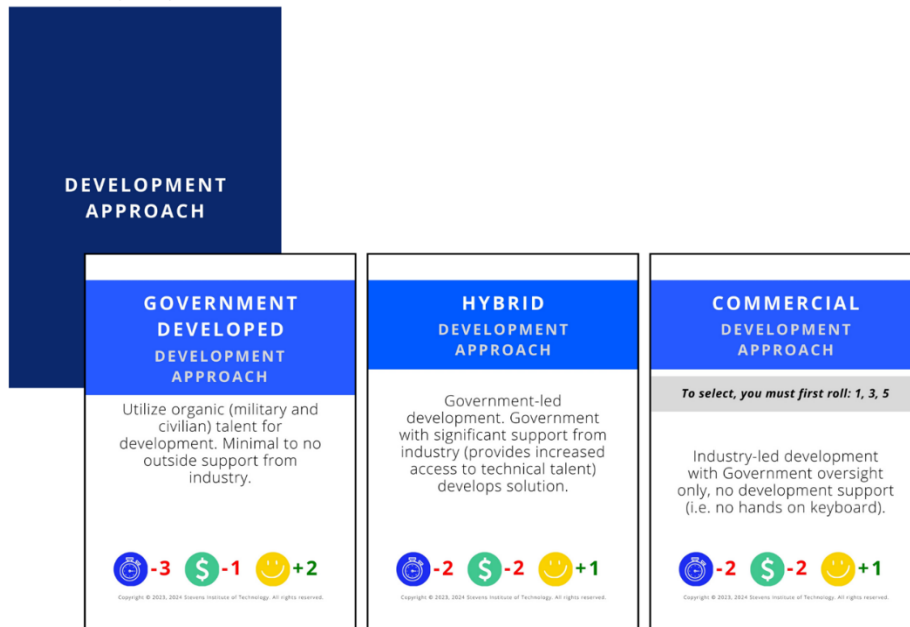
Technical Solution Enhancements
(Front)



Technical Solution Enhancements
(Back)

Figure 6. Technical Solution Enhancement Cards

Development Approaches
(Front)



Development Approaches
(Back)

Figure 7. Development Approach Cards



In the contracting phase of the game, there are three primary sets of choices to be made. These include selecting a market research approach, a contracting methodology, and a contract type (Figures 8–10). These various decisions represent the process a contracting officer will engage with the broader program team. As with any acquisition program, there are certain benefits and negatives associated with each decision, which in real life leads to optimized and sub-optimized contracting strategies. The cards in this phase also represent the potential interdependencies between an earlier decision and a future one. For example, making the choice to use Federal Acquisition Regulation (FAR) Part 12 as the contract approach is predicated on the decision to select a commercially oriented development approach. While by no means comprehensive, this phase of the game provides scholars with an important overview of the steps and decisions involved.

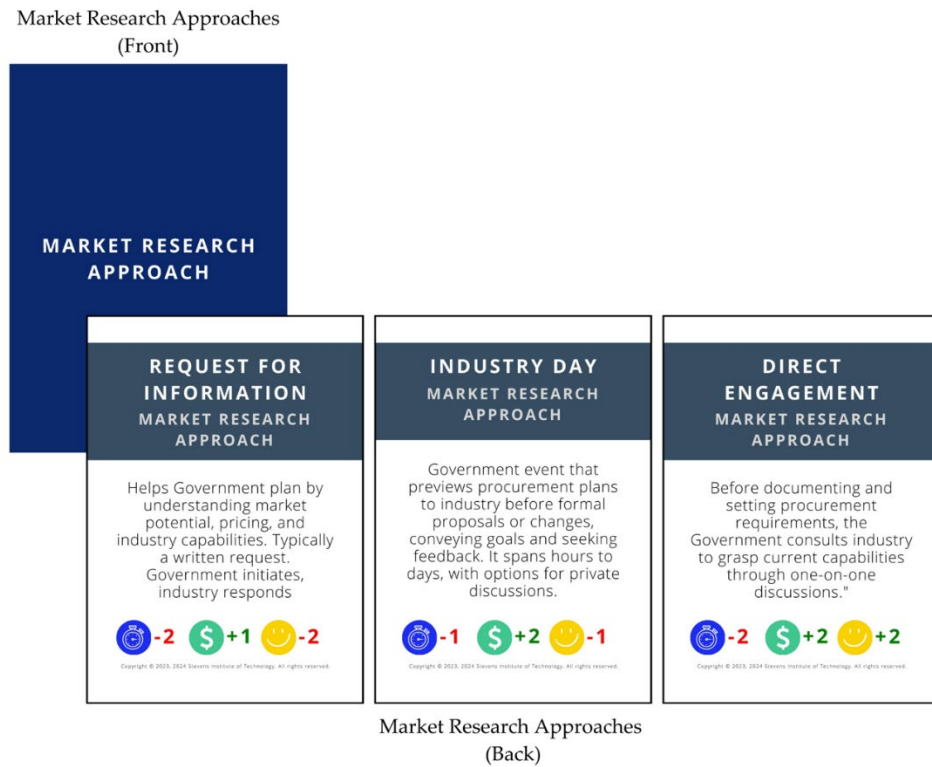


Figure 8. Market Research Approach Cards



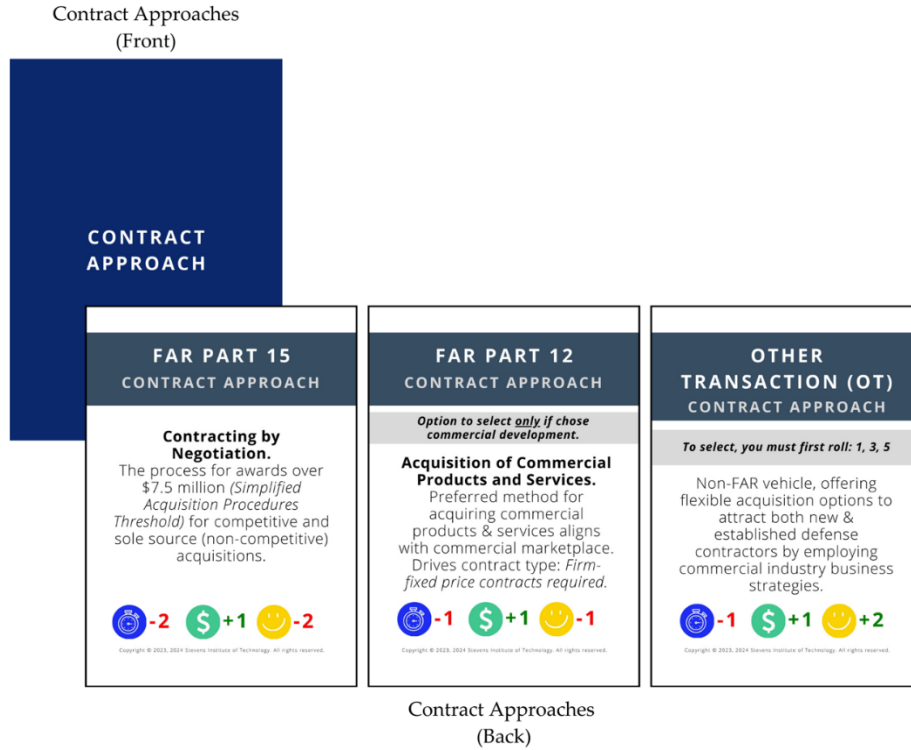


Figure 9. Contract Approach Cards

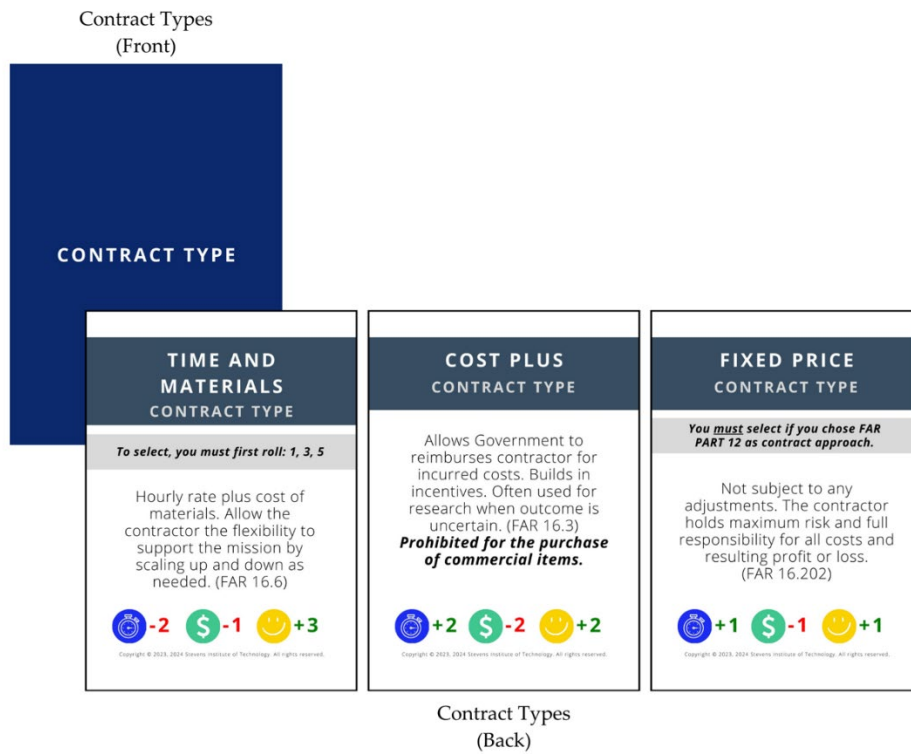


Figure 10. Contract Types Cards



In the management phase of the game, there are no decisions or choices to make. While this is not a direct emulation of the execution of an acquisition program, it does introduce scholars to scenarios they are likely to encounter. The number of events that teams face in the game is dependent on their roll of the dice and what spaces they land on, which have mandatory card selections ranging from zero to two.

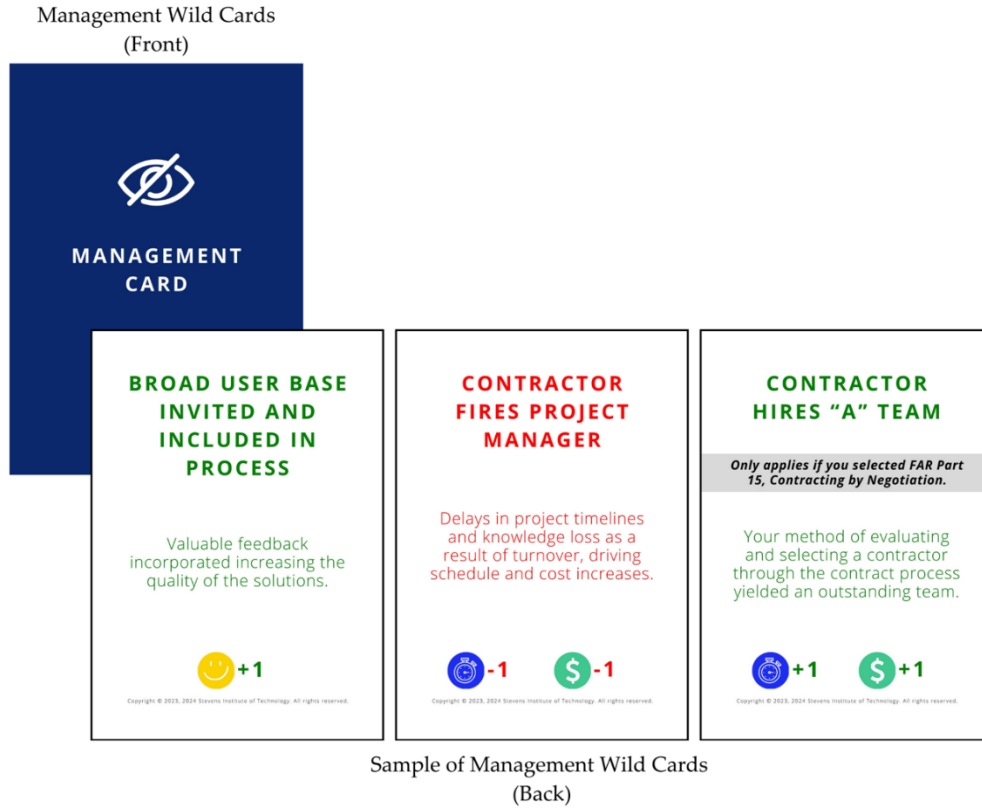


Figure 11. Management Card Example

In the game's final move, each team encounters a Fielding Status event that is decided by the dice roll. The roll of certain numbers will indicate success, other numbers will result in a penalty. This game play directly represents the process that acquisition professionals face when attempting to operationally field a capability. When the user imposes subjective criteria, initial success is not always guaranteed at the end of the process (Figure 12).



Figure 12. Fielding Status

Feedback

The Acquisition Game has been played in multiple venues, including with a broad range of senior acquisition professionals from across the Services. Their feedback was positive but came with serious suggestions for improving the fidelity of the game. The sponsoring officials from the Office of the Undersecretary for Acquisition and Sustainment (OUSD/A&S), have also played the game and observed that game play can meaningfully serve the DCTC curriculum.

Recently, the DCTC team attended Nexus, an event co-sponsored by the National Contract Management Association (NCMA) and the Defense Acquisition University, to play the Acquisition Game with conference participants from across government and industry and invite their feedback. Twelve participants responded to the survey after playing the game more than once. The feedback (five responses from industry and seven from government) was overwhelmingly positive, with the large majority finding it to be helpful across different measures. When asked what they liked most about the game, the general response was that it was realistic, fun, and interactive. This demonstrates some level of success in achieving our initial goals.

Table 2. Survey Results from the NEXUS Conference

Survey Question	Result
The game accurately portrays the DoD acquisition process.	11 Agreed or Strongly Agreed
Indicate the level of benefit you believe the game offers for new hires.	6 chose Very Beneficial with 3 choosing Beneficial
Playing the game helped me understand some terms and concepts that I need to explore further.	10 Agreed or Strongly Agreed
Overall, the game was an instructive experience.	10 Agreed or Strongly Agreed



The most important feedback has come from the DCTC scholars. After playing the game multiple times as part of classroom learning, the feedback has been highly positive. Some of the statements (see below) we received from scholars reaffirmed the importance of interactive learning in the DCTC experience.

“With the Acquisition Game, learning the acquisition process is not only educational but also enjoyable. Working through the procurement process can be daunting but with the acquisition game’s realistic scenarios, we were able to use critical thinking, strategic decision-making, and teamwork to create strategies and learn hands-on.”—**Faith Jones, DCTC Scholar at Virginia Tech.**

“The Acquisition Game is an engaging board game that transforms the complex world of government acquisitions into an accessible and enjoyable learning experience. It cleverly combines strategy and education, allowing players to immerse themselves in a realistic scenario and acquisitions in a playful yet informative way.”—**Sangmuk Kang, DCTC Scholar at Virginia Tech.**

“The Acquisition Game is a riveting board game which highlights the intricacies of the DoD acquisition process and delivers it to players in a fun, competitive, and never before seen format as they traverse various acquisition pathways to accomplish the mission at hand. Every choice matters, and by the end of the process you’ll have not only had fun, but also attained a deeper understanding of the U.S. Department of Defense.”—**Marco Antonio Cortes Esparza, DCTC Scholar at University of Arizona.**

“The Acquisition Game merged teamwork, realistic challenges, and government complexities seamlessly.”—**Katlind Michele Nearing, DCTC Scholar at University of Arizona.**

“The Acquisition Game was very insightful and fun once we got the hang of things. It was a struggle to figure out what was the objective, but when we did it all came together and made sense. It helped me understand how satisfaction, time, and money play a part in the real world.”—**Tamara Daye, DCTC Scholar at North Carolina A&T**



“Participating in The Acquisition Game gave me a better understanding of the addition and procurement process regarding acquisition. The game allowed me to collaborate with my fellow scholars and learn how to balance competing priorities and navigate difficulties effectively. Engaging with peers showed me the importance of teamwork



when addressing challenges in a DoD perspective.”—Justin Reid, DCTC Scholar at North Carolina A&T

Next Steps / Potential Future States

The DCTC curriculum team has accumulated a list of improvements from the various feedback sources and we will continue iterating on The Acquisition Game to refine aspects and build in additional layers of complexity for more advanced players. Additionally, we are working with the Army Gaming Studio to build out a video game version, which will provide options for injecting additional acquisition considerations including personae, role-playing different functions, and making risk management a greater focus point.

The DCTC curriculum team has also begun designing related games such as an Industry Game that emulates the decisions a company makes when deciding to work with the government and/or bid on a request for proposal. There are also plans to build a PPBE (Planning, Programming, Budgeting and Execution) Game that emulates the defense resourcing process for DCTC scholars to play in the final semester of their curriculum.

The Acquisition Game experience has given those involved in the DCTC curriculum a better sense of the potential for gaming in defense acquisition training and we expect it to become a core part of the DCTC curriculum to prepare scholars to enter the acquisition workforce and begin contributing to the many challenges that require solutions.

References

- Asal, V., Jahanbani, N., Lee, D., & Ren, J. (2018, October). *Mini-games for teaching political science methodology*. Cambridge University Press.
https://www.cambridge.org/core/services/aop-cambridge-core/content/view/2CB21D4F2FAC326F9FE0B663D23F5393/S1049096518000902a.pdf/minigames_for_teaching_political_science_methodology.pdf
- Ávila, P. M. (2023, January 27). *From baby boomers to generation Z: Understanding the learning styles of different age groups*. LinkedIn. <https://www.linkedin.com/pulse/from-baby-boomers-generation-z-understanding-learning-m%25C3%25A9ndez-%25C3%25A1vila/>
- DAU. (n.d.). About. <https://www.dau.edu/back-to-basics>
- Department of Defense. (2021, September 30). *Defense acquisition workforce key information*. [https://www.hci.mil/docs/Workforce_Metrics/FY21Q4/FY21\(Q4\)OVERALLDefenseAcquisitionWorkforce\(DAW\)InformationSummary.pdf](https://www.hci.mil/docs/Workforce_Metrics/FY21Q4/FY21(Q4)OVERALLDefenseAcquisitionWorkforce(DAW)InformationSummary.pdf)
- Department of Defense. (2024). *DoD acquisition workforce hiring authorities as of 1 February 2024*. <https://www.hci.mil/docs/DoD%20Acquisition%20Workforce%20Hiring%20Authorities%20as%20of%20February%202024.pdf>
- Levine, P. (2019, May 6). Building a 21st century defense acquisition workforce. *War on the Rocks*. <https://warontherocks.com/2019/05/building-a-21st-century-defense-acquisition-workforce/>
- McCabe, K., & Laurent, A. (2015, April 3). *The real problem with acquisition training*. Government Executive. <http://www.govexec.com/contracting/2015/04/real-problem-acquisition-training/109138/>



Murphy, C. & Bouffard, A. (2017, April 26). *Understanding defense acquisition workforce challenges*. Mitre. <https://www.mitre.org/sites/default/files/publications/17-1585-understanding-defense-acquisition-workforce-challenges.pdf>

Office of the Under Secretary of Defense for Acquisition and Sustainment. (n.d.). *Defense civilian training corps*. <https://dctc.mil>



Improving Comptroller Benchmarks on Program Spending

Philip S. Antón, PhD—is the Chief Scientist of the Acquisition Innovation Research Center (AIRC) in the Systems Engineering Research Center (SERC) led by the Stevens Institute of Technology. He assesses DoD needs, envisions and develops innovative research, and facilitates application to DoD practices, policies, and workforce development. For 23 years, Dr. Antón was a Senior Information Scientist at the RAND Corporation. From 2011–2016, Dr. Antón directed Acquisition Policy Analysis for USD(AT&L). His PhD and MS are in Information and Computer Science from UC Irvine, specializing in Computational Neuroscience and Artificial Intelligence. His BS in engineering is from UCLA. [panton@stevens.edu]

Odd J. Stalebrink, PhD—is an Associate Professor of Public Administration at the School of Public Affairs at the Pennsylvania State University. He specializes in public budgeting and financial management. He is a member in the Association for Budgeting and Financial Management (ABFM); the National Tax Association (NTA); the Association for Public Policy Analysis and Management (APPAM); and the European Accounting Association (EAA). He holds a PhD in public policy from the School of Public Policy, George Mason University. His master's and bachelor's degrees are in Business Administration from the Jönköping International Business School (JIBS), Sweden. [ojs10@psu.edu]

Douglas J. Buettner, PhD—is the AIRC Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah's Systems Engineering Program. His doctoral research at the University of Southern California's Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer's Dilemma. Dr. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu]

Abstract

A common tool for overseeing program execution is to compare spending against linear benchmarks to identify programs that may be falling behind or unable to fully use their funding. These benchmarks identify candidates for further investigation and potential budgets reduced and reallocated. Pressures to meet benchmarks can drive bad behaviors, such as premature spending before good prices and intellectual property rights can be negotiated. This paper analyses business theory, program manager observations, and historical trends of DoD obligation and expenditure rates to assess ways to improve these benchmarks. Regressions of historical obligation data find that recent spending has an underlying linear trend, but temporal variables, theory, and execution realities indicate that S-shaped curves are better benchmarks. Also, benchmarks should be adjusted when Congress provides Continuing Resolutions (CRs) in lieu of full appropriations at the start of the fiscal year. Also, as expected by theory, historical expenditure patterns for Research, Development, Test, and Evaluation (RDT&E), Procurement, and Operation and Maintenance (O&M) funds follow S-shaped curves rather than the linear profiles in DoD benchmarks. Recommendations are provided, including adjustments for variable effects on obligations, S-curve profiles for improved benchmarks, and leveraging improved DoD data environments to switch to plan-based benchmarking.

Introduction

What gets measured gets managed – even when it's pointless to measure and manage it, and even if it harms the purpose of the organization to do so.

Peter Drucker

A common management tool for overseeing program execution is to track spending overtime against benchmarks to monitor progress and identify any programs that may be falling behind. In the federal government, program and contract spending involve two basic steps:



- *Obligations*, which commit funds from the U.S. Treasury for payment of goods and services, such as on a contract (see GAO, 2005, p. 70).
- *Expenditures* (also called *outlays* or *disbursements*), which are the actual financial payments (money) from the U.S. Treasury to liquidate an obligation, for example upon receipt of goods or services under a contract (see GAO, 2005, pp. 73–74).

The Benefits and Dangers of Benchmarks and Metrics

Program managers should attend to items that are important enough to measure. The problem is that we get exactly what we measure when we enforce and incentivize the metric. People will spend—one way or another, and often regardless of unintended side effects—if we measure execution against a metric and especially if we apply enforcement consequences and incentives. See, for example, anecdotal and survey evidence in Marsalis (2022) and Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform (2023).

Below is a short overview of the principles, theory, and realities of setting, using, and enforcing benchmarks and metrics. This is just a short overview, but it is important to begin by reminding ourselves what these metrics result in—good and bad.

Benefits of Benchmarks

Benchmarks can help identify performance issues. Some programs do not obligate or expend all their authorized and appropriated funding. Thus, monitoring the level of funding can be a quick way to identify programs that eventually may not fully execute their spending. With the recognition that such benchmarks are but one source of needed information, they can help focus attention on more likely candidates.

Monitored benchmarks can ensure that attention is paid to managing financial resources. When not emphasized to the extreme, monitored benchmarks like these can motivate program managers to properly plan and track spending along with associated risks.

Concerns on Using Benchmark

Untailored benchmarks may not align with program realities and plans. Programs have spending needs based on plans as well as events and decisions that need to be made during the spending period. Some programs may plan to obligate funds as soon as they are authorized and appropriated while others may have good reasons to obligate late in the spending period. Untailored (a priori) fixed benchmarks may be out of sync with such plans. Also, programs may not know in advance when they need to obligate. For example, programs, especially in Research, Development, Test, and Evaluation (RDT&E) (which often involves systems that have never been developed before), may need to address issues that arise during the spending period, address unforeseen costs or schedule challenges, or to change priorities based on new threats or technology issues and opportunities. Again, fixed benchmarks would only reflect such spending profiles from random chance.

Benchmarks can drive undesirable behaviors or effects. It is well known in business and psychology that enforced or even monitored benchmarks will drive behaviors to achieve the benchmarks despite negative consequences (Behn, 2008; Marsalis, 2022; Norden, 1970). This is a real concern in the Department of Defense (DoD) to avoid wasting resources in programs—either by not spending what could be used elsewhere, or by wasting it through less-than-prudent (but entirely understandable from an incentives perspective) spending.

Input (consumption) benchmarks lack prioritization and thus require additional information before acting on below-target programs. Finally, spending benchmarks like



these are simply input metrics (as opposed to output or outcome metrics)⁶⁴ that lack measures of the *value* of the spending. A program (and thus the DoD and taxpayers) may get better value from early spending, or it may get higher value by giving a program more time to obligate (or a contractor more time to execute). Instead, programs appear mostly driven to avoid unspent funds (which does happen in non-trivial amounts) and to identify potential resources for new urgent priorities that arise during the spending period. Thus, they can only identify potential candidates for further (deeper) assessment to understand a program's status. Without this added consideration, these metrics can devolve into blind bureaucratic taking of resources with undesirable outcomes.

Benchmarks restrict agility at the program level. Taken together, these concerns can reduce administrative flexibilities at the program level, pressuring, and restricting spending decisions within the year(s) of execution. Increased agility in meeting DoD program outcomes requires a willingness to delegate decisions while providing clearly defined goals and objectives. This should be accompanied by appropriate accountability to program managers to fully utilize their appropriated funds or advise leadership early on if they will not be able to fully execute these funds and thus make them available for reprogramming for other purposes. Such a willingness to delegate, decentralize, and utilize administrative flexibilities at the program level could form a base-level of reform for the larger PPBE system, wherein program planning and execution agility is increased. See Stalebrink (forthcoming) for discussion of these concepts across all levels of PPBE.

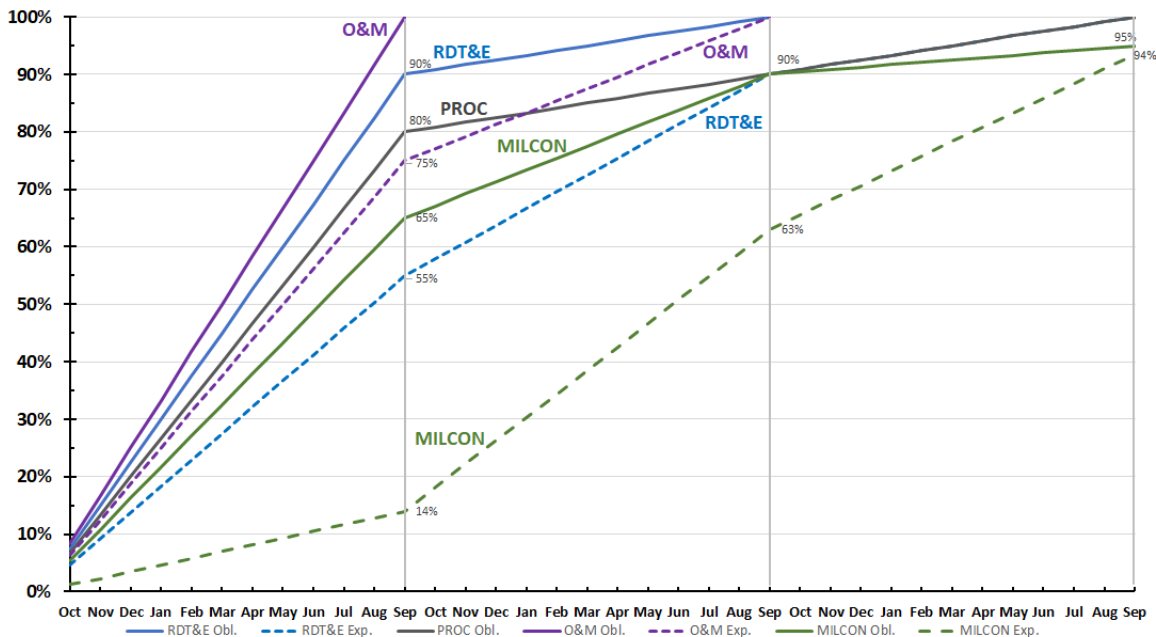
Obligation and Expenditure Benchmarks in the DoD

The DoD uses linear benchmarks for each category of funding (see Figure 1). Such benchmarks can help identify programs and activities that may have issues in spending funds within the year(s) of availability and thus may be candidates for further review to have portions of their budgets reprogrammed for critical priorities that emerged in the year of execution.

⁶⁴ See, for example, National Research Council (2005) for a very useful review of the theory and application of different types of metrics to achieve desired performance and outcomes.



Figure 42. Current Comptroller Obligations and Expenditures Rule-of-Thumb Benchmarks



SOURCE: Under Secretary of Defense (Comptroller), as reported in Tomasini (2017).

NOTES: The dashed lines are the obligation (Obl.) benchmarks over time, and the solid lines are the associated expenditure (Exp.) benchmarks over time. The O&M benchmark curves rise the fastest, followed by RDT&E and Procurement (PROC). Tomasini (2017) reports that Procurement expenditures are “N/A.” Exp. = expenditures; MILCON = Military Construction; O&M = Operation and Maintenance; Obl. = obligations; PROC = Procurement; RDT&E = Research, Development, Test, and Evaluation.

This paper assesses these benchmarks through quantitative analysis of DoD obligation and expenditures over time, observations from program managers, and a review of existing theory and qualitative data from experts. It also assesses the statistical effects of delayed full fiscal year (FY) appropriations associated with continuing resolutions (CRs), calendar-month effects (e.g., at the start and end of the FY), and time trends on DoD obligation rates. These analyses provide new insights into the realism of DoD obligation and expenditure benchmarks, leading to recommendations for improving these benchmarks.

Analysis of Obligation Rates: Effects of Continuing Resolutions and Other Events

We obtained data from the DoD’s Advana data environment on monthly obligated dollars for separate accounts (e.g., within military services or defense wide) and categories of funding (RDT&E, Procurement, O&M, MILPERS, and MILCON) going back to FY 2011. The data also included a range of other categories, such as MILCON and smaller accounts that are not analyzed in this paper.

We aggregated these data to obtain monthly obligation dollars by category across all accounts, then calculated the percentage obligated in each month compared to the total dollars obligated by the final month. For example, if the RDT&E obligated in month 2 was \$5,112,653, and the total obligated by month 24 was \$71,339,247, then the month 2 percentage is about 7.17% (= \$5,112,653 / \$71,339,247). This yielded a series of monthly obligation rates (percentages) for each FY’s authorization and appropriation out to the end of those obligations.

Using these data from Advana, we then conducted multivariate linear regressions on individual obligation categories (colors of money) to identify any variables that correlate with changes in the normal monthly obligation rates. Table 1 lists the variables tested for correlation (i.e., with a p-value no higher than 0.05). Visual examination of the monthly obligation rates



indicated that the first year of obligation behaved differently than any subsequent years, so for multi-year appropriations (RDT&E and PROC) we conducted separate regressions for the first and subsequent years.

Table 2 summarizes the statistical results with the following observations.

¹ The data are well modeled by a linear obligation rate (the constant base) with adjustments for the variables shown. For example, on average, the RDT&E rate in October of the first year of obligating a FY's appropriation was 5.9% – 2.3% = 3.6%. If March of the first spending year was also the third month after the budget was passed (3 MAB), then on average the obligation rate would be about 5.9% + 6.3% + 2.3% = 14.5% (which is close to the actual value of 15.1% for FY 2012, for example).

Table 19. Variables Tested for Effects on Monthly Obligation Rates

Type	Variables
FY calendar month	October April November May December June January July February August March September
Month after full budget passed	1 MAB 2 MAB 3 MAB 4 MAB 5 MAB
Time (month #)	Time

¹ Statistically, we note the following:

- While all have descent Adjusted R² values, the values for RDT&E (1st year), PROC, and O&M are the highest. Thus, the latter explain the variation in the data well.
- The models (the constant monthly linear contribution plus the contributions from the variables in the model) are fairly linear, with the RDT&E 1st-year for RDT&E and PROC (respectively) and O&M being very linear.



Table 20. Contributions of CR and Other Variables Affecting Obligation Rates (FY 2011–2023 Appropriations)

			RDT&E		PROC		O&M	MILPERS
			1 st Year	2 nd Year	1 st Year	2 nd –3 rd Years		
Average Base	Monthly Rate	<i>Base rate:</i>	5.9%	1.7%	5.0%	2.0%	7.5%	7.9%
CR Effects	1st MAB	<i>If true, add:</i>					1.2%	
	2nd MAB	<i>If true, add:</i>	4.3%		1.8%		2.4%	
	3rd MAB	<i>If true, add:</i>	6.3%		4.6%		2.4%	
	4th MAB	<i>If true, add:</i>	3.7%		2.2%		1.6%	
Calendar Month Effects	October	<i>If true, add:</i>	-2.3%		-3.6%		-1.1%	
	November	<i>If true, add:</i>		0.9%		0.6%	-0.8%	
	December	<i>If true, add:</i>				0.9%		
	March	<i>If true, add:</i>	2.3%		2.6%	0.4%		
	July	<i>If true, add:</i>					2.4%	
	August	<i>If true, add:</i>					-1.8%	
	September	<i>If true, add:</i>		1.0%	4.6%	1.4%	4.1%	0.4%
Time Trend	Time (month)	<i>If true, add:</i>		-0.1%		-0.1%		0.1%
Linearity	Multiple R		0.80	0.58	0.77	0.66	0.84	0.48
% of variation explained	Adjusted R²		63%	32%	57%	43%	68%	22%

The analysis above examined RDT&E obligations together. Subsequently, we obtained DoD monthly RDT&E obligations data broken down by Budget Activity (BA) for appropriation FYs 2013, 2014, and 2017–2022. Table 3 shows the results of the linear regressions on the RDT&E monthly obligation rates as percentages of appropriation-year dollars associated variables across all of RDT&E (original Advana data) and broken down by S&T, development, and management accounts.



Table 21. Contributions of CR and Other Variables Affecting RDT&E Obligation Rates (FY 2013–2014, 2017–2022 Appropriations)

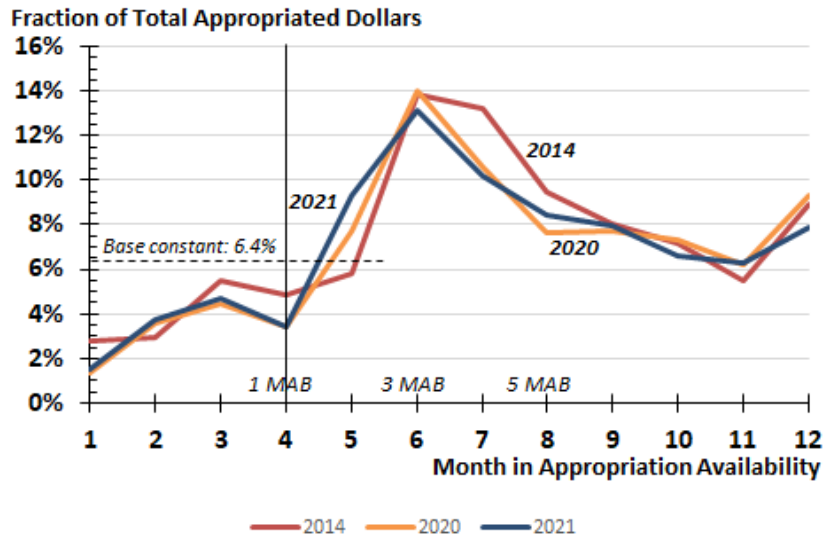
		1 st Year of Availability				2 nd Year of Availability			
		All	S&T	DEV	Mgt	All	S&T	DEV	Mgt
Avg. Base		5.9%	6.4%	6.0%	6.3%	1.7%	2.3%	1.6%	2.5%
Add CR Effects	CR		-1.0%		-1.4%				
	1 MAB		-1.6%						
	2 MAB	4.3%		3.0%	2.1%				
	3 MAB	6.3%	4.8%	7.1%	2.1%				
	4 MAB	3.7%	4.3%	4.8%					
	5 MAB		1.5%						
Add Calendar Month Effects	6 MAB		1.8%						
	Oct.	-2.3%	-3.5%	-1.9%				0.40%	-0.63%
	Nov.		-1.1%			0.91%		0.52%	
	Mar.	2.3%	1.6%	1.7%					0.68%
	Sept.		2.4%		2.9%	1.0%	1.3%	0.84%	1.3%
Time Trend	Time (mo.)					-0.10%	-0.15%	-0.10%	-0.10%
Adj. R2		63%	84%	60%	42%	32%	48%	50%	33%

MAB = month after budget is passed; CR = month under a continuing resolution (the months before 1 MAB); BA = Budget Activity; S&T = Science and Technology (BA-1, BA-2, and BA-3 combined); DEV = development (BA-4, BA-5, and BA-7 combined); Mgt = Management [Support] (BA-6); mo. = month; Oct. = October; Nov. = November; Jan. = January; Mar. = March; Aug. = August; Sept. = September.

In our data sample, appropriations from had their first MAB in January. Figure 2 illustrates the CR and month effects for three appropriation years (FYs 2014, 2020, and 2021) for which January was the first month after the budget was passed (the 1 MAB Here, months 1–3, October through December) operated under a CR and thus were below the base constant of 6.4%. The model showed that October was an additional 3.5 percentage points low and November an additional 1.1 percentage points lower than the baseline minus the CR effect. This can be seen in the lower left of the figure. The first MAB is 1.6 percentage points lower in the model, but rises significantly in MABs 3–8, then returning closer to the base constant with a final increase in September. The example shows that the expected cumulative obligations should be about 84% by September; the actuals for FYs 2014, 2020, and 2021 are 88.0%, 83.3%, and 83.1%, respectively (85% on average—close to 84% from the model).



Figure 43. Monthly S&T Obligation Rates for Appropriation FYs with Full Budget Passed in January (FY 2014, 2020, and 2021 Appropriations)



Comparison of Historical Data Against Current and Proposed DoD Obligation and Expenditure Benchmarks

The following figures graphically show the differences between historical obligations and expenditures against current and recommended benchmarks (e.g., with S-curves for RDT&E, Procurement, and O&M along with historically patterned benchmarks).

Figure 3 compares the current and recommended RDT&E benchmarks. The recommended obligation curves show the cumulative amounts, not counting the bumps that would be added in after the full FY budgets are passed. Figure 4 illustrate how the MAB obligation effect would add based on what month the full budget is passed. The first figure shows the effect when lowering the initial portion to further strengthen the s-curve effect while the second figure shows the effect based solely on the historical values from FY 2011–2022 actual obligations. These plots illustrate the strong effect on obligations of delayed final appropriations for the FY.

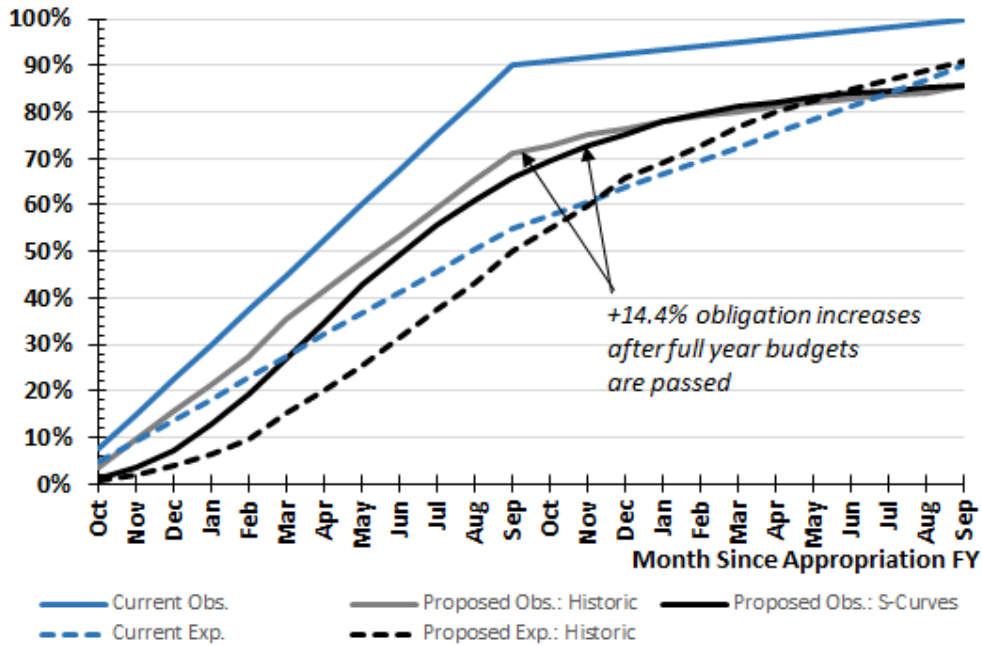
Figure 5 compares the current and recommended Procurement benchmarks. Again, the recommended obligation curves show the cumulative amounts not counting the bumps that would be added in after the full FY budgets are passed. There is no current expenditure benchmark for Procurement. The recommended benchmarks reflect insights from analyzing actual procurement benchmarks. Further analysis is needed to reflect the different lengths of different procurement accounts.

Figure 7 compares the current and recommended O&M benchmarks. Again, the recommended obligation curves show the cumulative amounts, not counting the bumps that would be added in after the full FY budgets are passed.

Finally, Figure 9 compares the current and recommended MILCON benchmarks.



Figure 44. Cumulative Fraction of RDT&E Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2022 Appropriations)



NOTE: See Figure 4 for how the 14.4% obligation increases are added depending on when the final FY appropriations are passed.

Figure 45. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of RDT&E Obligations by Month for (FY 2011–2022 Appropriations)

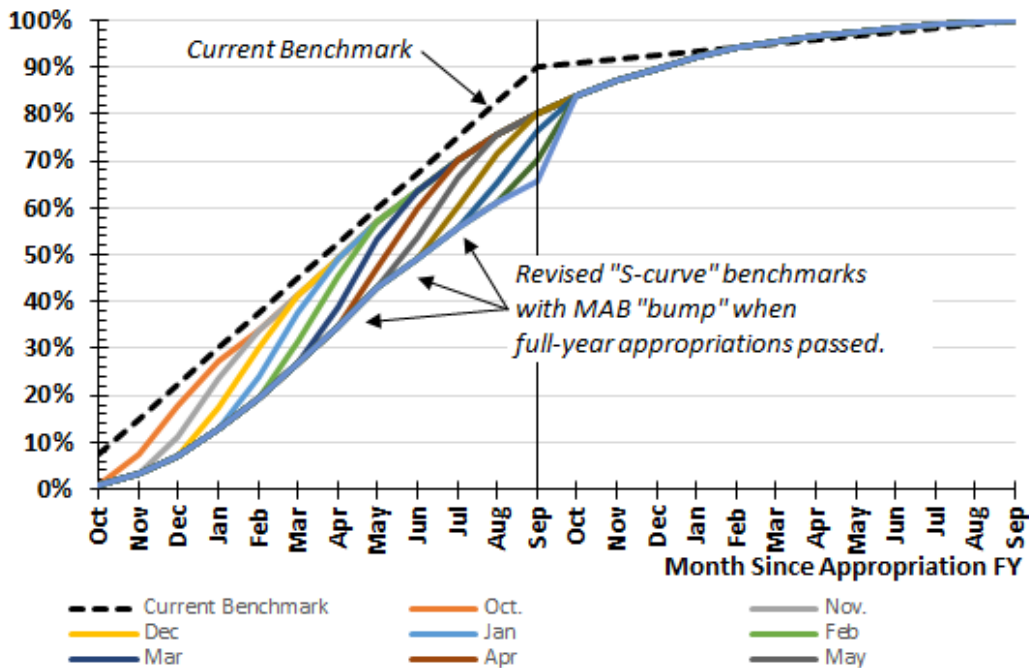
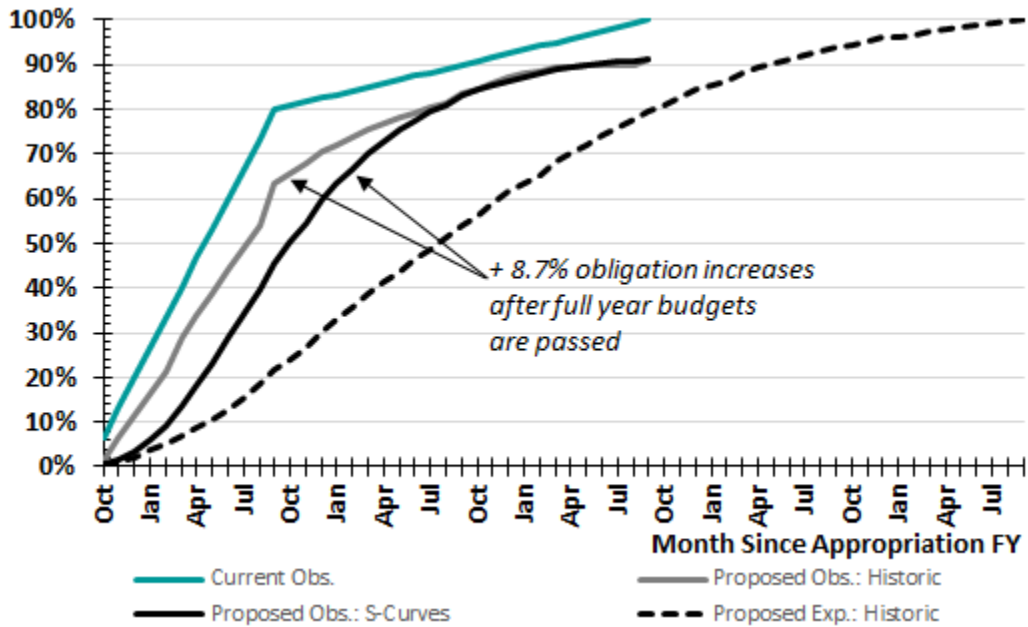


Figure 46. Cumulative Fraction of Procurement Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2021 Appropriations)



NOTE: See Figure 6 for how the 8.7% obligation increases are added depending on when the final FY appropriations are passed.

Figure 47. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of PROC Obligations by Month for (FY 2011–2021 Appropriations)

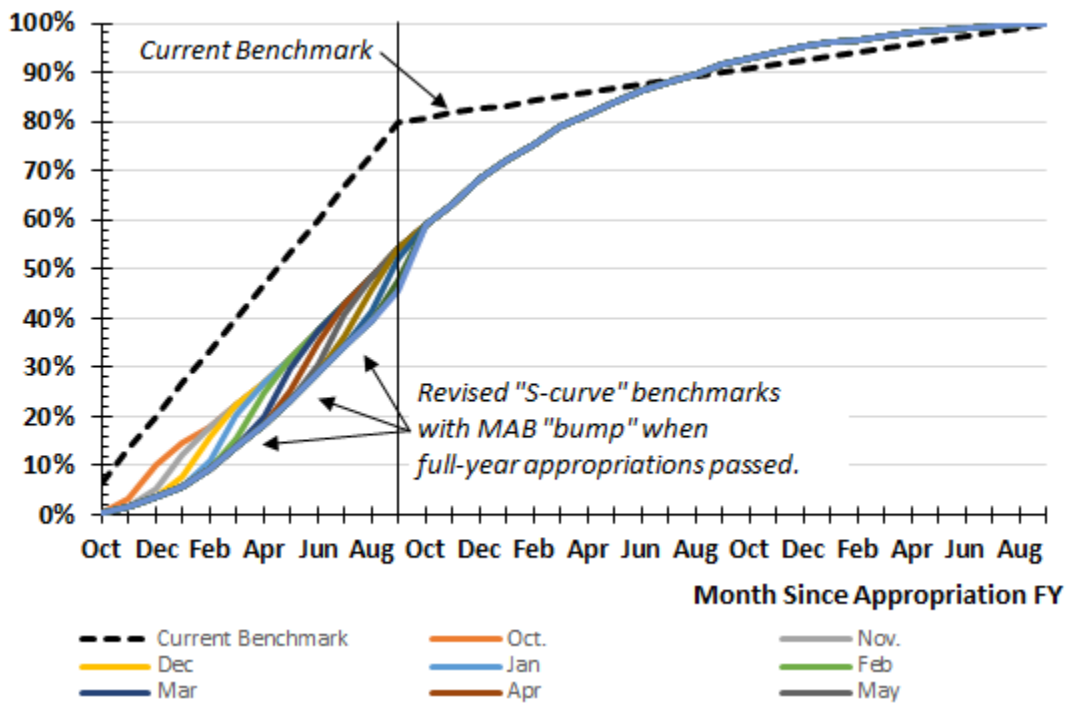
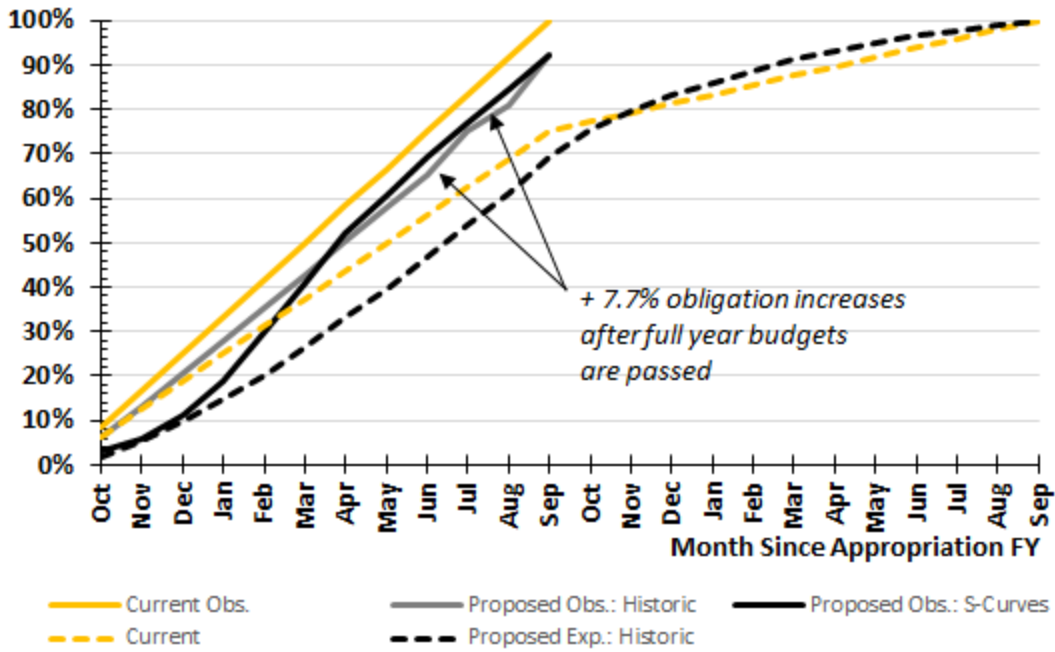


Figure 48. Cumulative Fraction of O&M Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2023 Appropriations)



NOTE: See Figure 8 for how the 7.7% obligation increases are added depending on when the final FY appropriations are passed.

Figure 49. Current and Proposed S-Curve Benchmarks: Cumulative Fraction of O&M Obligations by Month for (FY 2011–2023 Appropriations)

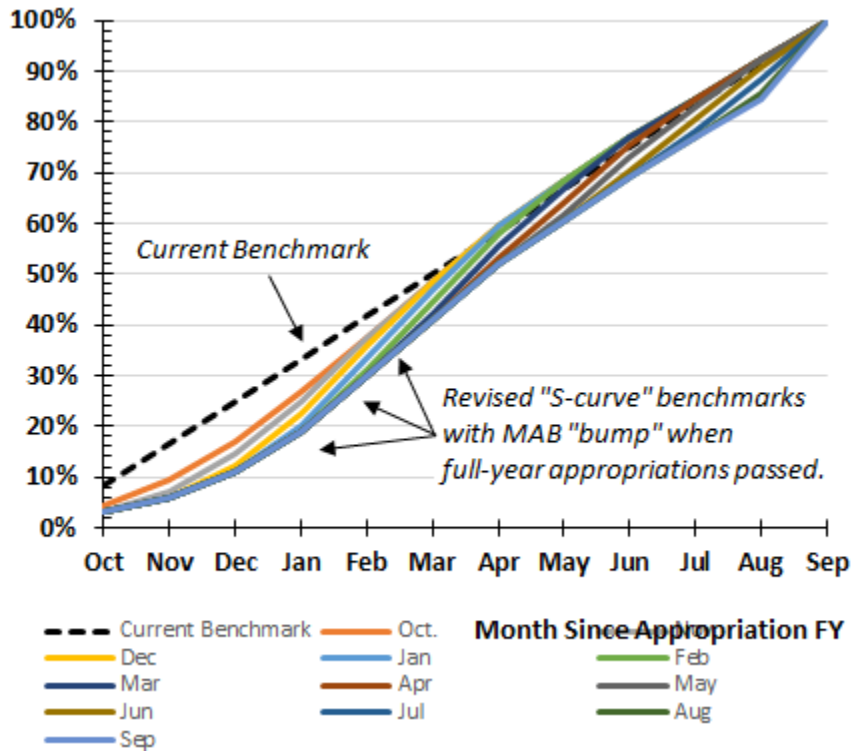
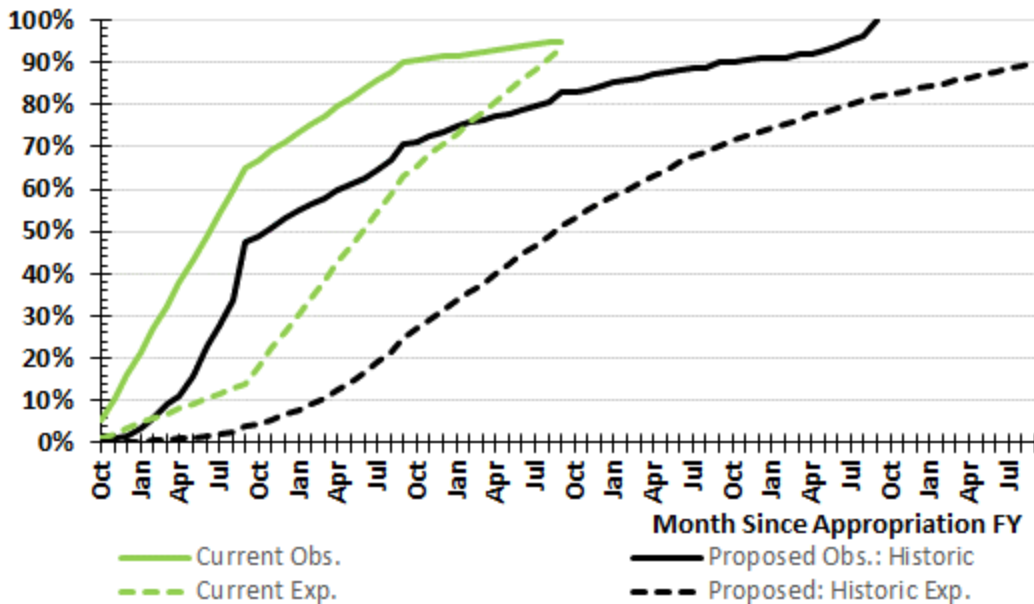


Figure 50. Cumulative Fraction of MILCON Obligations and Expenditure by Month for Current and Proposed Benchmarks (FY 2011–2021 Appropriations)



Observations on Variables Affecting Monthly Obligation Rates

Here are the significant effects uncovered by these analyses:

- Obligation rates are higher in the 2 to 6 months after the full budget is passed (MAB; i.e., once managers know their authorized spending). Thus, CRs delay a portion of funding into later in the FY.
- S&T and Management Support within RDT&E have a significantly lower obligation rate during CR months that other types of funding did not exhibit.
- Obligation rates are often lower the first October in the spending cycle, possibly reflecting assertions in the literature that it takes time to delegate spending authorization to program managers.
- Obligation rates are often higher the first March in the spending cycle (i.e., the month before the midyear spending reviews).
- Obligation rates for some types of funding are higher in September.
- While each category of funding has a general underlying linear trend, MILPERS obligations are linear with slight upward trend.
- RDT&E and Procurement dollars obligate the first year on a fairly linear basis but then inflect to a reduced, curved basis. Thus, obligations are modeled well by linear models with these variate effects.
- Military Construction (MILCON) shows a significant upward curve in the first year rather than the straight line in the benchmark but becomes fairly linear afterwards. Also, a significant fraction of MILCON obligations occurs after year 3, which is not in alignment with the benchmark targets.

These statistical models align somewhat with linear obligation rate targets set by the DoD Comptroller and are compatible with anecdotal assertions that when told to obligate, programs do. This does not account for any changes in DoD priorities given new threats or



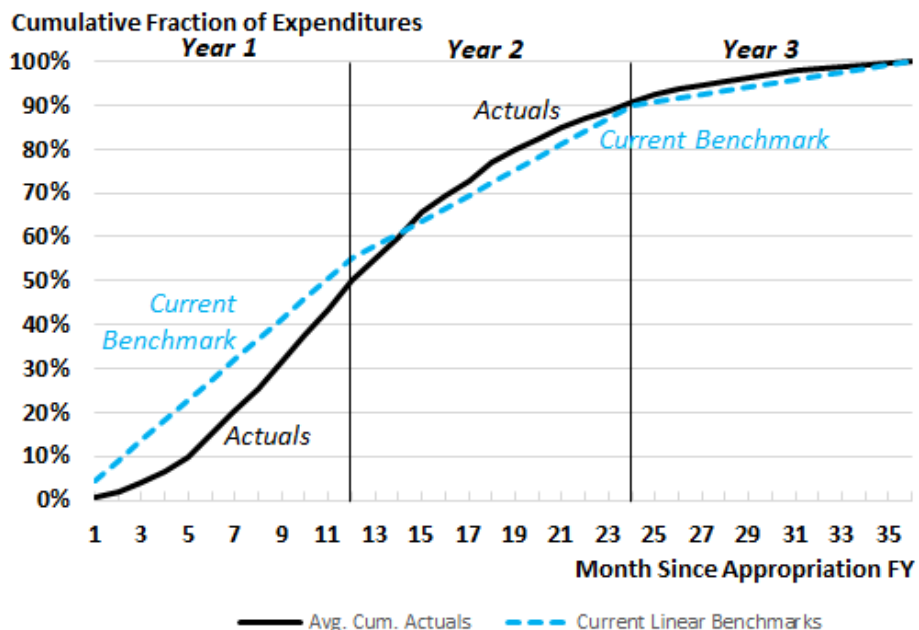
technological opportunities since the budgets were first drafted early in the PPBE process, but when told to spend or risk losing their funds, individuals across the DoD appear to do so to a large extent.

Expenditure Rates

Analysis of DoD expenditure data shows that RDT&E, Procurement, and O&M expenditures followed an S-curve shape rather than the linear profiles in the DoD’s benchmarks. This aligns with over 50 years of data and theory in the literature on program execution profiles.²

While the S-curve for RDT&E meets the Comptroller’s linear benchmarks at the 24-month point of 90%, but the average 6-month value of 15.5% is well below the benchmark of 27.5% and the 12-month value was also lower than the benchmark (see Figure 10). Thus, the DoD’s linear RDT&E benchmark poorly informs the midyear and first-year execution review for RDT&E. Similar profiles were seen for O&M and MILCON (see Figure 11 and Figure 12).

Figure 51. Average Cumulative RDT&E Expenditures Versus Benchmark as Percentage of Month 36 Obligations (FY 2011–2021 Appropriations)



NOTE: Month 1 is October of the FY in which the appropriations were made.

² See, for example, Behn, 2008; Burgess et al., 2014; Brown et al., 2015; Davis, 2008; Davis et al., 2009; Gallagher & Lee, 1996; Lee, Hogue, & Gallagher, 1993; Lee, Hogue, & Hoffman, 1993; Norden, 1970; Schiavoni, 2019; Watkins, 1982.



Figure 52. Average Cumulative O&M Expenditures Versus Benchmark as Percentage of Month 12 Obligations (FY 2011–2022 Appropriations)

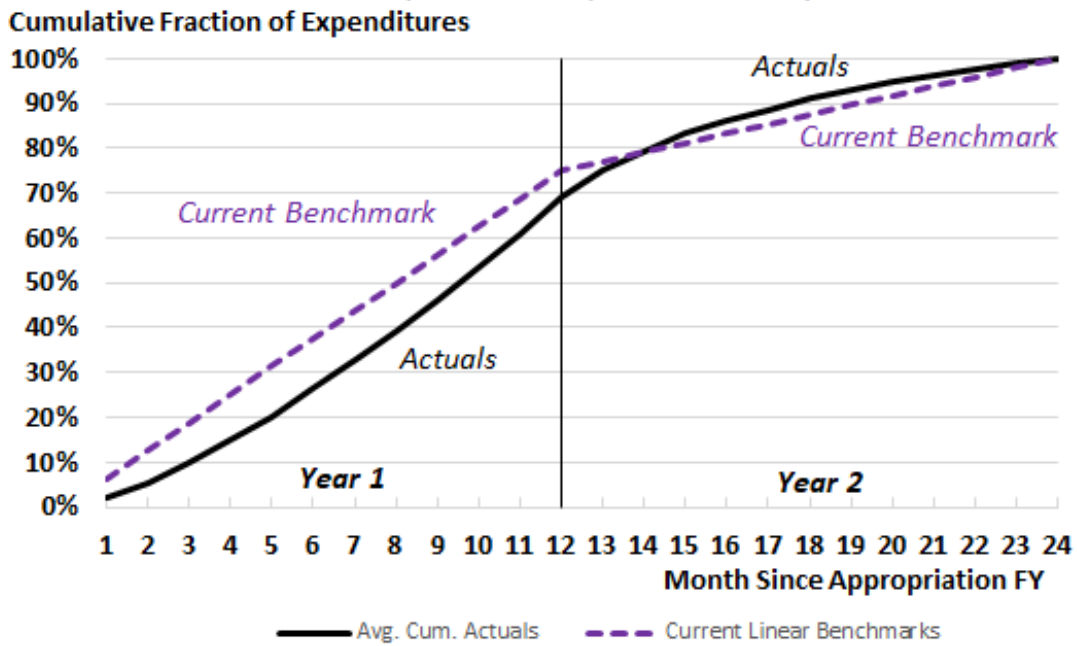
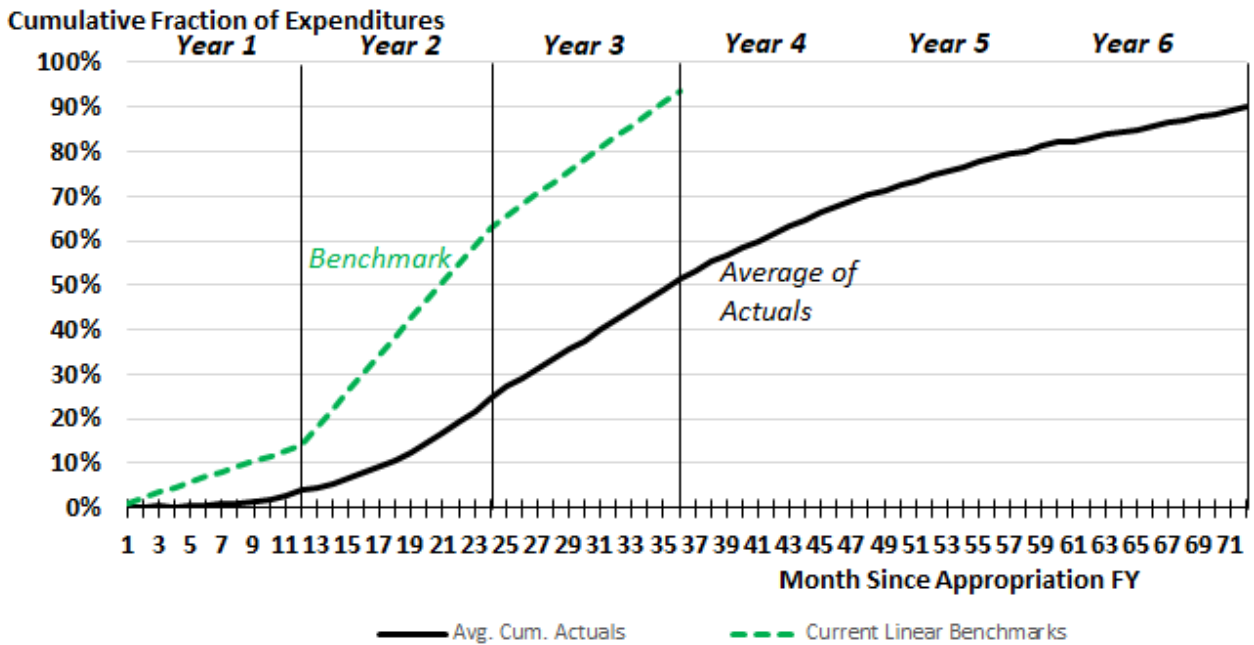


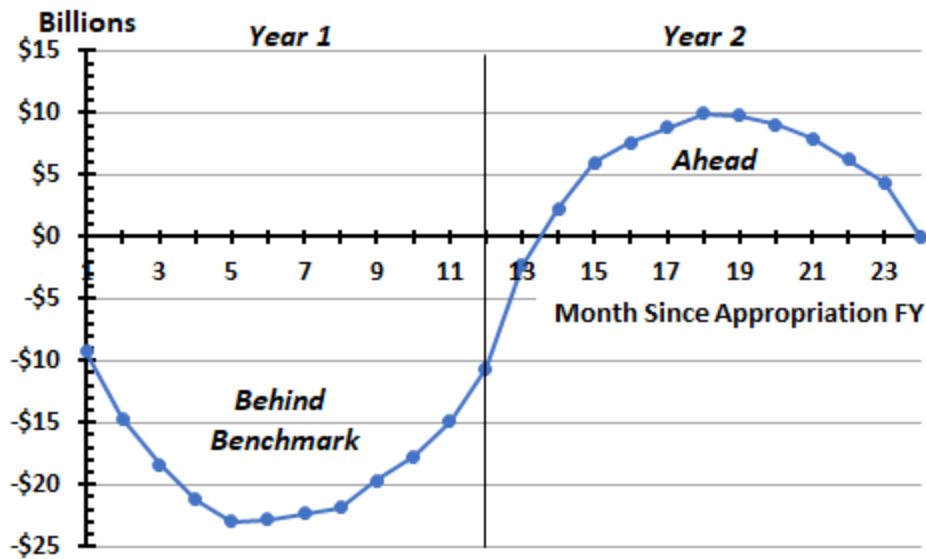
Figure 53. Average Cumulative MILCON Expenditures Versus Benchmark as Percentage of Month 36 Obligations (FY 2011–2018 Appropriations)



Overall, RDT&E, O&M, and MILCON expenditure differences between actuals over the last decade and the current linear benchmarks can be as large as \$10 billion, \$23 billion, and \$3 billion, respectively (see, for example, Figure 13).



Figure 54. Dollar Difference Between Average Cumulative O&M Expenditures and Current Benchmark (FY 2011–2022 Appropriations)



Aligning Obligation and Expenditure Benchmarks with Theory and Data

This paper reached the following conclusions based on the review of theory and analysis of available data (see also Anton and Buettner, forthcoming, for further details).

At the least, benchmarks should be adjusted to reflect realities evident in recent years. DoD obligation and expenditure data consistently show statistically significant differences between average actuals and simple linear benchmarks. If benchmarks are not adjusted, then benchmarks are less effective at identifying potential issues. When average (normal) actuals are behind the benchmark, then too many programs may be undergoing subsequent deep-dive performance reviews. Likewise, when average actuals are above the benchmarks, then too few programs may be undergoing subsequent deep-dive performance reviews. Thus, these are indicators that updating benchmarks may improve the effectiveness and efficiency of performance reviews by helping to focus on programs that may be behind. For example, Figure 13 shows that O&M expenditures are, on average, as much as \$23 billion below benchmarks in months 5–6 (right before mid-year reviews) and as much as \$10 billion over benchmarks by month 18. This indicates potentially significant inefficiencies given limited oversight resources.

The best shape of obligation benchmark curves ultimately comes down to intent and theory. While our analysis shows that managers in the DoD have tended to obligate at rates that generally align with current linear obligation benchmarks, there are good reasons to reconsider these profiles. First, even with pressures to obligate on a straight line, actual data show startup delays as well as reductions due to CR effects. Also, RDT&E inherently involves engineering uncertainty and surprises, so it may be more effective for the DoD and the country to target more obligations in the second year than in the first. In addition, shifting more obligation targets for RDT&E and Procurement into the second year would give DoD managers more time to make investments when needed (earlier or later), negotiate better deals (e.g., prices, intellectual property rights, and deliverables), and fully assess contractors' execution, subcontracting, and supply-chain plans and risks.

Benchmarks should be adjusted for CR and financial-management realities. Regardless of the basic shape of the benchmarks, the statistical analysis in this paper shows real-world effects that should be considered for RDT&E, Procurement, and O&M. CRs result in obligation bumps after full budgets are passed as well as reductions during CRs for S&T and



Management Support. Obligation rates in the first month (October of the first year) are lower than the current benchmarks (probably from the time it takes for the financial management system to allocate spending authority to program managers). These CR effects introduce some level of S-curve patterns into actual obligation rates.

S-curves for obligation benchmarks may be beneficial for RDT&E, PROC, and O&M. While actual obligations have underlying linear bases, shifting to an S-curve profile for obligations would allow more time for improved performance and deals, addressing the points above.

Benchmarks can be useful but require additional due diligence. When combined with further due diligence, benchmarks can help the DoD and Congress identify funds that could be reprogrammed to address higher-priority threats and needs that emerge during the spending periods. The combined effects of these benefits can improve DoD mission outcomes by identifying badly needed resources. However, the emphasis here is on proper use and due diligence to ensure a balance between the benefits and issues. The use of benchmarks alone does not provide insight into the practical realities and issues in execution. Anecdotes indicate that DoD and Congressional leadership do not rely solely on benchmarks to identify from whom to take money for new urgent priorities that arise during the year of execution. However, other anecdotal evidence indicates that program managers believe otherwise, adding to the concern that these managers may prioritize spending to benchmarks over more prudent uses of financial resources, leading to undesirable or unforeseen negative side effects.

Avoid unforeseen negative consequences from managing to benchmarks. Finally, metrics drive behavior. This concern is well documented in the literature³ and can be seen in the increased obligation rates in March immediately before the midyear reviews that identify programs spending below the benchmark rates for potential budget reprogramming to other programs and needs. While management metrics can be useful tools for insight, management pressures will drive behavior to the exclusion of other factors. Forcing people to spend to a curve will get spending to that curve whether or not that spending results in the best use of taxpayer dollars and the best results for national security. This axiom also applies to other potential uses of these benchmarks, such as adjusting Office of Management and Budget (OMB) apportionments based on changes in benchmarks.

Suggestions for Future Research

Future research in the following areas may lead to additional recommendations:

- Piloting modified benchmarks.
- Identifying expenditure benchmark profiles for Procurement.
- Assessing obligation and expenditure rates at the account level within each category.
- Assessing sources of obligation and expenditure data errors.

Recommendations

In this paper we provided the bulk of our theoretical work from Anton and Buettner (forthcoming) and Stalebrink (forthcoming). Based on these observations, we recommended that the DoD Comptroller consider modifying their benchmarks. Four optional variants are discussed in the report and are provided in Table 4. The preferred option includes adding additional S-curve ramp-up elements on top of historical obligation behaviors and recommends replacing linear expenditure profiles with historical S-curve profiles. Table 5 summarizes our

³ See, for example, National Research Council, 2005; Behn, 2008.



recommendations. Anton and Buettner (forthcoming) provide candidate benchmark tables that better reflect recent history as well as the correlative effects of month, CR, and time.

In addition to aligning expenditure benchmarks to actual data and theoretical objectives, such changes could help eliminate the negative side effects cited in theory and the literature that program managers may seek expenditures prematurely just to meet comptroller benchmarks at the expense of other program and department objectives of prudent use of the resources (see, for example, Commission on PPBE Reform, 2023, p. 33; Marsalis, 2022). Slight delays in switching to S-curves with their lower initial expenditure benchmarks should give program managers more time to get good deals for the program, the DoD, and taxpayers rather than having to rush negotiations and contracting to meet somewhat arbitrary benchmarks or risk losing their funding.

There would be some cultural and process adjustments for both Congress and the DoD (and industry) to adjusting the obligation and expenditure benchmark profiles over time, but the benefits could be improved performance given the financial resources provided by Congress and the taxpayers to the DoD. In the end, keep in mind the following insightful quote.

Tell me how you measure me, and I will tell you how I will behave. If you measure me in an illogical way ... do not complain about illogical behavior.

Eliyahu Moshe Goldratt

Table 22. Benchmarks Options: Elements and Ranking

	Obligations		Expenditures	RDT&E	Rank
	Base Shape	Variables			
Option 1	S-curves on historic	CR, MAB, Calendar, and Time Effects	Historic (S-curved)	Separate S&T, DEV, Mgt	1 (Preferred)
Option 2	S-curves on historic			Combined	2
Option 3	Historic (linear base)			Separate S&T, DEV, Mgt	2
Option 4	Historic (linear base)			Combined	3
Option 5	As-is (arbitrary lines)	None	As-is (arbitrary lines)	Combined	4



Table 23. Recommendations for Improving Obligation and Expenditure Benchmarks

Obligations	Expenditures
<ul style="list-style-type: none"> • Reduce obligation benchmarks for the first 1–2 months for RDT&E, PROC, and O&M to reflect process delays in allocating spending authorities. • Consider changing benchmarks to S-curves instead of straight lines. • Consider allowing more time in benchmarks for later spending to give time to get better negotiated deals and address surprises. 	<ul style="list-style-type: none"> • Change benchmarks to S-curves for RDT&E, PROC, and O&M. • At a minimum, if the benchmarks are not changed to S-curves, consider: <ul style="list-style-type: none"> ○ Reducing expenditure benchmarks for the first 3 months. ○ Changing benchmark shapes to straight lines across <u>all</u> years for multi-year funds rather than front-loading in the first year. • Add predictive metrics to identify more likely spending shortfalls.
<ul style="list-style-type: none"> • Explore switching to plan-based benchmarks instead of fixed benchmark curves, using Advana to collect plans from program offices. • Ensure proper due diligence along with spending relative to benchmarks before taking program funds. • Use needs, plans, and priorities for budgeting—not just spending. • Avoid overly enforcing benchmarks and other metrics. Keep these as information tools. • To avoid slowing down DoD acquisition, do not use obligation and expenditure benchmarks as a guide to OMB apportionments—instead inform apportionments based on the distribution data of recent actual obligations and expenditures. • Pilot these changes before pursuing more aggressive shifts to lower benchmarks in earlier years to understand better the effects (if any) on changes in unobligated and unexpended funds at the end of normal availability. 	

Acknowledgements

We would like to thank the Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform for their sponsorship and support throughout this research. In particular, the questions, discussions, and feedback from the Commission chairs and staff were extremely helpful: the Hon. Robert Hale, the Hon. Ellen Lord, as well as the commission staff, especially Lara Sayer (our sponsor of record), Annie Crum, Elizabeth Bieri, Caroline Bledsoe, Kelle McCluskey, Brooks Minnick, and Soleil Sykes. We would also like to thank John Stedge in the Chief Digital and Artificial Intelligence Office (CDAO), for guiding us on how to pull the Obligations and Expenditures data from Advana. Thanks to Scott Lucero and Nathan Self from Virginia Tech for support with extracting USASpending.com data from their data repository. Finally, we would like to acknowledge and thank our Stevens staff: Technical Editor, Nancy Méndez-Booth for her editorial review comments and suggestions, and Tara Kelly for her project management skills, edits, and overall support.

The views, findings, conclusions, and recommendations expressed in this material are solely those of the authors and do not necessarily reflect the views or positions of the United States Government (including the Department of Defense (DoD) and any government personnel) or the Stevens Institute of Technology.

References

Anton, P. S., & Buettner, D. J. (Forthcoming). *DoD obligation and expenditure rates: More-realistic benchmarks and the effects of continuing resolutions and other events on*



obligation rates. Acquisition Innovation Research Center, Systems Engineering Research Center UARC, Stevens Institute of Technology.

- Behn, R. D. (2008). On why, to improve performance, measurement is rarely enough. *Bob Behn's Public Management Report*, 5(9).
- Brown, G. E., White, E. D., Ritschel, J. D., & Seibel, M. J. (2015). Time phasing aircraft R&D using the Weibull and beta distributions. *Journal of Cost Analysis and Parametrics*, 8(3), 150–164. <https://doi.org/10.1080/1941658X.2015.1096219>
- Burgess, E., Smirnoff, J., & Wong, B. (2014, June). *Weibull analysis method* [Briefing]. ICEAA Annual Symposium, Denver, CO.
- Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform. (2023, August). *Interim report*. <https://ppbereform.senate.gov/wp-content/uploads/2023/08/PPBE-Commission-Interim-Report-Final.pdf>
- Davis, D. (2008, April 15). Early warning model for acquisition program cost and schedule growth. *2008 SCEA-ISPA Joint Annual Conference and Training Workshop*. <https://www.iceaaonline.com/wp-content/uploads/2017/09/EVM04.pdf>
- Davis, D., Christle, G., & Abba, W. (2009, January). *Using the Rayleigh model to assess future acquisition contract performance and overall contract risk: The future isn't what it used to be. (Yogi Berra) (Volume I)* (CRM D0019289.A2/Final). Center for Naval Analysis. https://www.cna.org/archive/CNA_Files/pdf/d0019289.a2.pdf
- Department of the Air Force. (2022, February). *Planning, Programming, Budgeting, and Execution System training program: Reference manual*. <https://afacpo.com/AQDocs/PPBE.pdf>
- Gallagher, M. A., & Lee, D. A. (1996, June). Final-cost estimates for research & development programs conditioned on realized costs. *Military Operations Research*, 2(2), 51–65. https://www.researchgate.net/profile/Mark-Gallagher-5/publication/233620264_Final-Cost_Estimates_for_Research_Development_Programs_Conditioned_on_Realized_Costs/links/55956b8808ae5d8f3930f1ff/Final-Cost-Estimates-for-Research-Development-Programs-Conditioned-on-Realized-Costs.pdf
- GAO. (2005). *A glossary of terms used in the federal budget process* (GAO-05-734SP). <https://www.gao.gov/assets/gao-05-734sp.pdf>
- Lee, D. A., Hogue, M. R., & Gallagher, M. A. (1993, September 1). *Determining a budget profile from a development cost estimate*. <https://apps.dtic.mil/sti/pdfs/ADA275864.pdf>
- Lee, D., Hogue, M., & Hoffman, D. (1993). *Time histories of expenditures for defense acquisition programs in the development phase*. Annual Meeting of the International Society for Parametric Analysis.
- Marsalis, J. R. (2022). *The negative impacts of the benchmarks*. Senior Acquisition Course, National Defense University. https://nps.edu/documents/105938399/0/Negative+Impacts+of+Benchmarks_Marsalis.pdf/fc216616-7b4c-45e6-d4fa-c4bc48a6b71f?t=1652994212754
- National Research Council. (2005). *Thinking strategically: The appropriate use of metrics for the climate change science program*. The National Academies Press. <https://doi.org/10.17226/11292>
- Norden, P. V. (1970). Useful tools for project management. In W. Voris & M. K. Starr (eds.), *Management of Production* (pp. 71–101). Penguin.



- Schiavoni, M. L. (2019, May 14–17). *Adaptive curve fitting: An algorithm in a sea of models* [Briefing]. ICEAA Professional Development & Training Workshop, Tampa FL.
<https://www.iceaaonline.com/wp-content/uploads/2019/06/AM05-Adaptive-Curve-Fitting-Schiavoni.pdf>
- Stalebrink, O. J. (Forthcoming). *A budget-theoretical perspective on DoD PPBE reform*. Acquisition Innovation Research Center, Stevens Institute of Technology.
- Tomasini, R. (2017). *OUSDC rule-of-thumb acquisition obligation and expenditure rates*. Defense Acquisition University.
<https://www.dau.edu/sites/default/files/Migrated/ToolAttachments/OSD%20%28C%29%20Color%20Rule-of-Thumb%20Acq%20Obligation%20and%20Expenditure%20Rates.pdf> and
<https://www.dau.edu/tools/ousdc-rule-thumb-acquisition-obligation-and-expenditure-rates>.
- Watkins, H. (1982). *An application of Rayleigh curve theory to contract cost estimation and control* [Master's thesis, Naval Postgraduate School].
<https://apps.dtic.mil/sti/pdfs/ADA118213.pdf>



Assessing the Impact of Continuing Resolutions (CRs) on DoD Contract Obligation Rates

Philip S. Antón, PhD—is the Chief Scientist of the Acquisition Innovation Research Center (AIRC) in the Systems Engineering Research Center (SERC) led by the Stevens Institute of Technology. He assesses DoD needs, envisions and develops innovative research, and facilitates application to DoD practices, policies, and workforce development. For 23 years, Dr. Antón was a Senior Information Scientist at the RAND Corporation. From 2011–2016, Dr. Antón directed Acquisition Policy Analysis for USD(AT&L). His PhD and MS are in Information and Computer Science from UC Irvine, specializing in Computational Neuroscience and Artificial Intelligence. His BS in engineering is from UCLA. [panton@stevens.edu]

Douglas J. Buettner, PhD—is the AIRC Deputy Chief Scientist at the Stevens Institute of Technology and is an Adjunct Assistant Professor for the University of Utah’s Systems Engineering Program. His doctoral research at the University of Southern California’s Viterbi School of Engineering examined issues with DoD software-acquisition strategies, identifying acquisition optimizations. His research is found in Acquisition Games and Software Defects: Dynamics of the Software Developer’s Dilemma. Dr. Buettner has BS and MS degrees in physics, with a minor in atmospheric science from Oregon State University, and his PhD from USC is in astronautical engineering. [dbuettne@stevens.edu]

Abstract

One concern often raised with the Department of Defense (DoD) Planning, Programming, Budgeting, and Execution (PPBE) process is the potential effect of continuing resolutions (CRs) on spending. Using DoD data on monthly total contract obligations from fiscal years (FYs) 2011–2023, this research examines whether CRs or other variables had a statistically significant effect on the monthly rate at which the DoD obligates funds on contracts. Linear regression models of obligation data find that budgetary CRs from Congress generally correlated with higher obligation rates in the few months after the full budget is passed (i.e., a bump in spending once managers know their authorized spending). The months affected varied by category (“color”) of money. Also, in some Research, Development, Test, and Evaluation (RDT&E) budget activities, spending was also lower during CR months.

Introduction

In the Federal government, program and contract spending involve two basic steps:

- *Obligations*, which commit funds from the U.S. Treasury for payment of goods and services (e.g., on a contract).¹
- *Expenditures* (also called *outlays* or *disbursements*), which are the actual financial payments (money) from the U.S. Treasury to liquidate an obligation (e.g., upon receipt of goods or services under a contract).²

If Congress does not pass a Department of Defense (DoD) budget by the beginning of the fiscal year (FY) starting on October 1, Congress often passes one or more continuing resolutions (CRs) to provide interim funding until the final FY appropriations act is passed. Through quantitative analysis of actual DoD obligations, we examine whether, when, and to what extent

¹ The Government Accountability Office (GAO; 2005) defines an *obligation* as “A definite commitment that creates a legal liability of the government for the payment of goods and services ordered or received, or a legal duty on the part of the United States that could mature into a legal liability by virtue of actions on the part of the other party beyond the control of the United States. Payment may be made immediately or in the future. An agency incurs an obligation, for example, when it places an order, signs a contract, awards a grant, purchases a service, or takes other actions that require the government to make payments to the public or from one government account to another” (p. 70).

² The GAO (2005) defines an *outlay* (i.e., expenditure or disbursement) as “The issuance of checks, disbursement of cash, or electronic transfer of funds made to liquidate a federal obligation. Outlays also occur when interest on the Treasury debt held by the public accrues and when the government issues bonds, notes, debentures, monetary credits, or other cash-equivalent instruments in order to liquidate obligations” (pp. 73–74).



monthly obligation rates are affected by operating under or after a CR rather than a final budget for the FY.

Obligation Rates and the Effects of Continuing Resolutions and Other Events

CRs often provide only partial funding—say, spending up to 80 or 90% of last year’s appropriations—and usually prevent new program starts unless explicitly authorized in the CR. Also, if the final appropriations are lower than the requested budgets, then the program will likely need to replan and spend differently to reoptimize; this uncertainty makes it hard to know what to obligate during a CR. Even with CRs, final funding delays could potentially affect obligation rates on acquisition programs, procurement, sustainment, or operations. As a result, one might expect DoD obligations to be slower during a CR and then increase at a higher rate after a full budget is passed (i.e., showing an inflection point after the CR is passed).

In response to a request from the Commission on PPBE Reform, we conducted statistical analyses on available data to see if periods of CRs correlate with changes in the rate at which the DoD obligates funds.

Below, we first present summary data on CRs, followed by analysis of correlates that affect monthly obligation rates following the authorization and appropriation of funds for a FY’s budget—including CRs and other variables. This includes analysis of the following major categories of funding: Research, Development, Test, and Evaluation (RDT&E), Procurement (PROC), Operation and Maintenance (O&M), Military Personnel (MILPERS), and Military Construction (MILCON).

Continuing Resolutions in Fiscal Years 2010–2022

When the U.S. Congress is unable to authorize and appropriate spending by the beginning of a FY, which starts on October 1, then Congress often passes one or more CRs. These resolutions provide interim funding to avoid a partial government shutdown until a full year-long budget (authorized and appropriations) can be enacted (for example, GAO, 2022). As shown in Figure 1, there were CRs in every FY from 2011–2022, as well as three lapses in appropriations that resulted in government shutdowns.

Table 1 lists the dates when final authorization and appropriation acts were passed by Congress and became law (usually when signed by the president of the United States). For our analysis, we coded a variable for each month after the full-year budget was passed to test for consistent changes in obligation rates after CRs were over. These months after budget (MABs) are also shown in Table 1.



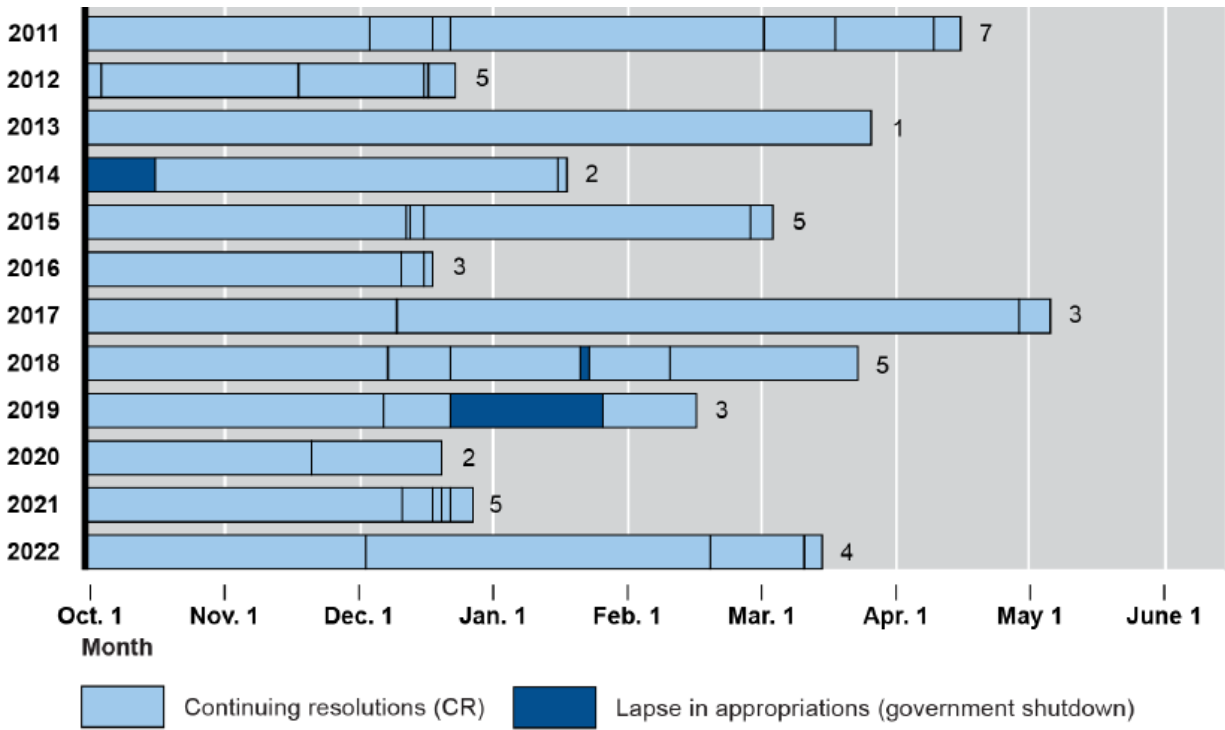


Figure 55. Number and Duration of Federal-Wide CRs and Lapses in Appropriations (FY 2011–2022) (GAO, 2022)

NOTE: These are federal-wide CRs. In FY 2019, the DoD actually received its authorization and appropriation acts before the start of the FY.



Table 24. Dates of Final DoD Authorization and Appropriation Acts and MABs (FY 2011–2023)

(Commission on PPBE Reform, 2023)

FY	Authorizations	Appropriations	1 MAB	2 MAB	3 MAB	4 MAB
2011	1/7/2011	4/15/2011	April	May	June	July
2012	12/31/2011	12/23/2011	January	February	March	April
2013	1/2/2013	3/26/2013	April	May	June	July
2014	12/26/2013	1/17/2014	January	February	March	April
2015	12/19/2014	12/16/2014	January	February	March	April
2016	11/25/2015	12/18/2015	January	February	March	April
2017	12/23/2016	5/5/2017	May	June	July	August
2018	12/12/2017	3/3/2018	March	April	May	June
2019	8/13/2018	9/28/2018	October	November	December	January
2020	12/20/2019	12/20/2019	January	February	March	April
2021	1/1/2021	12/27/2020	January	February	March	April
2022	3/15/2022	3/15/2022	March	April	May	June
2023	12/23/2022	12/29/2022	January	February	March	April

NOTE: If there were at least 10 working days in the month that the final budget (appropriation) was passed, then the 1st month after budget (MAB) is the month of passage, else the following calendar month is the 1st MAB (1 MAB). For example, for FY 2014, passage was on 1/17/2014 with at least 10 working days in January, so 1 MAB was January. However, in FY 2015, passage was on 12/16/2014, so with the end-of-year holidays, we used January instead of December as the 1 MAB.

Monthly Obligation Rates by Categories of Funding

We obtained data from the DoD’s Advana data environment on monthly obligated dollars for separate accounts (e.g., within military services or defense-wide) and categories of funding (RDT&E, Procurement, O&M, MILPERS, and MILCON) going back to FY 2011. The data also included a range of other categories, such as MILCON and smaller accounts, that are not analyzed in this paper.

We aggregated these data to obtain monthly obligation dollars by category across all accounts, then calculated the percentage obligated in each month compared to the total dollars obligated by the final month. For example, if the RDT&E obligated in month 2 (November of the first year) was \$5,112,653, and the total obligated by month 24 (September of the second year) was \$71,339,247, then the month 2 percentage is about 7.17% (= \$5,112,653/\$71,339,247). Table 2 lists the number of nominal months available to obligate by category of funding.

Table 25. Month Used for Final Obligation Values (DoD, 2008)

Category	Years to Obligate	Final Month
RDT&E	2	24
Procurement	3	36
O&M	1	12
MILPERS	1	12
MILCON	5	60

Note: See Section 1.7.2.25 (years for new obligations unexpired). MILCON final month was based on when the reported obligations in the DoD data began to flatten on average.

This yielded a series of monthly obligation rates (percentages) for each FY’s authorization and appropriation out to the end of those obligations. Thus, we have a 24-month series for FY 2011 RDT&E obligations. Likewise, we have 36-month, 12-month, and 12-month



series of obligation rates for FY 2011 Proc, O&M, and MILPERS, respectively. This allows us to analyze the obligation rates for each FY's authorized and appropriated dollars separately, even when they extend into subsequent years (e.g., analyze the FY 2011 RDT&E dollars obligated in FY 2012 separately from the FY 2012 RDT&E dollars obligated in FY 2012).

RDT&E Monthly Obligation Rates

Figure 2 shows the cumulative RDT&E obligations as a percentage of the total by month 24 of the allotted time to obligate RDT&E. Note that 85.8% of the RDT&E is obligated by the end of the first year. Figure 3 shows these data as monthly obligation rates for the first year.

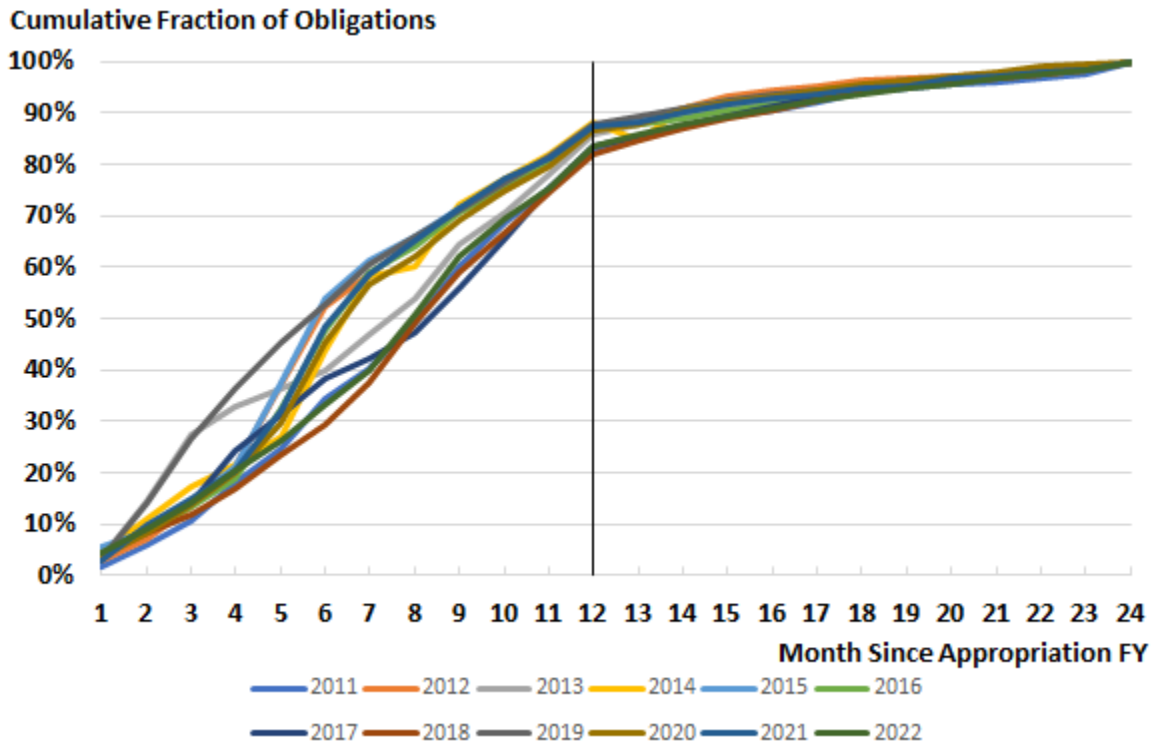


Figure 56. Fraction of Cumulative RDT&E Obligations by Month (FY 2011–2022 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.



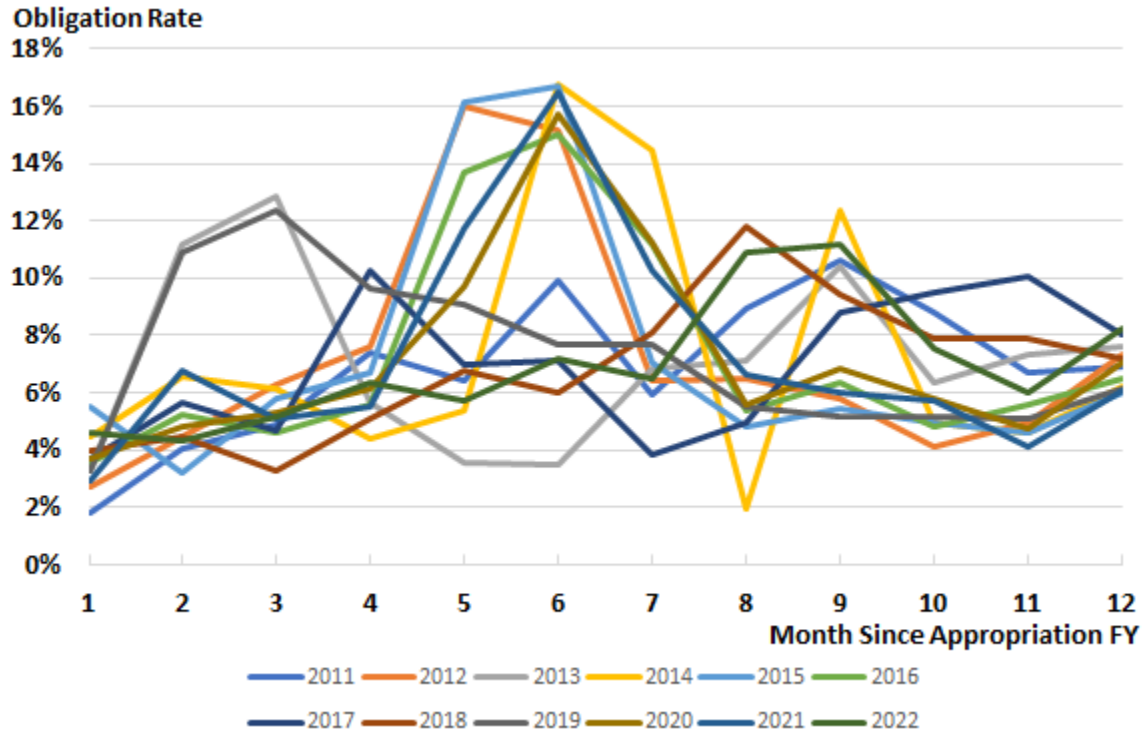


Figure 57. RDT&E Monthly Obligation Fractional Rates: Year 1 (FY 2011–2022 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

Procurement Monthly Obligation Rates

Figure 4 shows the cumulative Procurement obligations as a percentage of the total by month 36. Note that on average, 72.3% of Procurement dollars are obligated by the end of the first year. Figure 5 shows these data as monthly obligation rates for the first year.



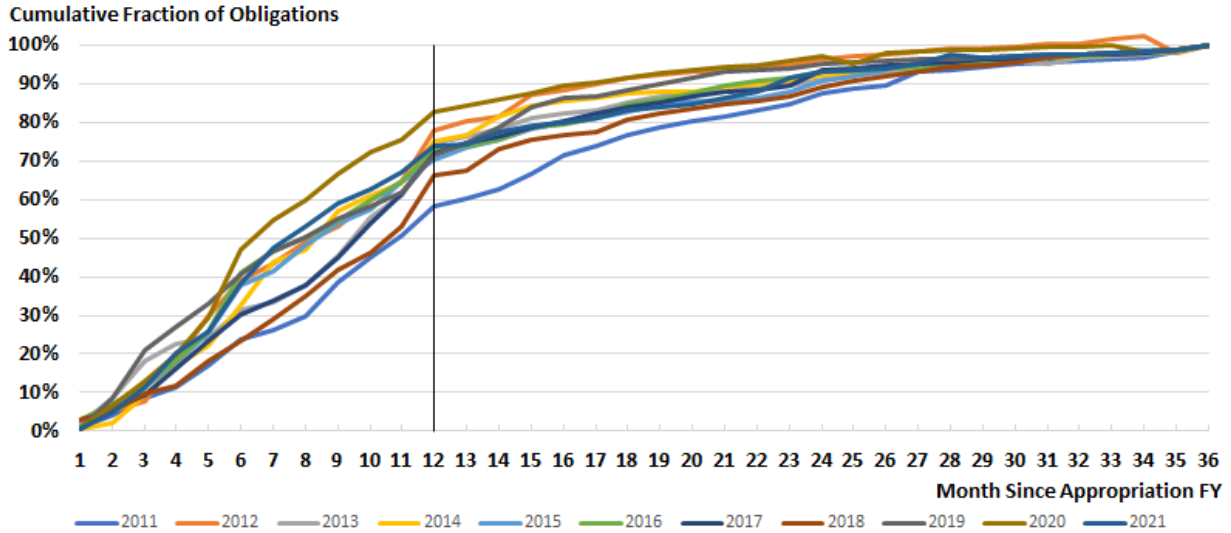


Figure 58. Fraction of Cumulative Procurement Obligations by Month (FY 2011–2021 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

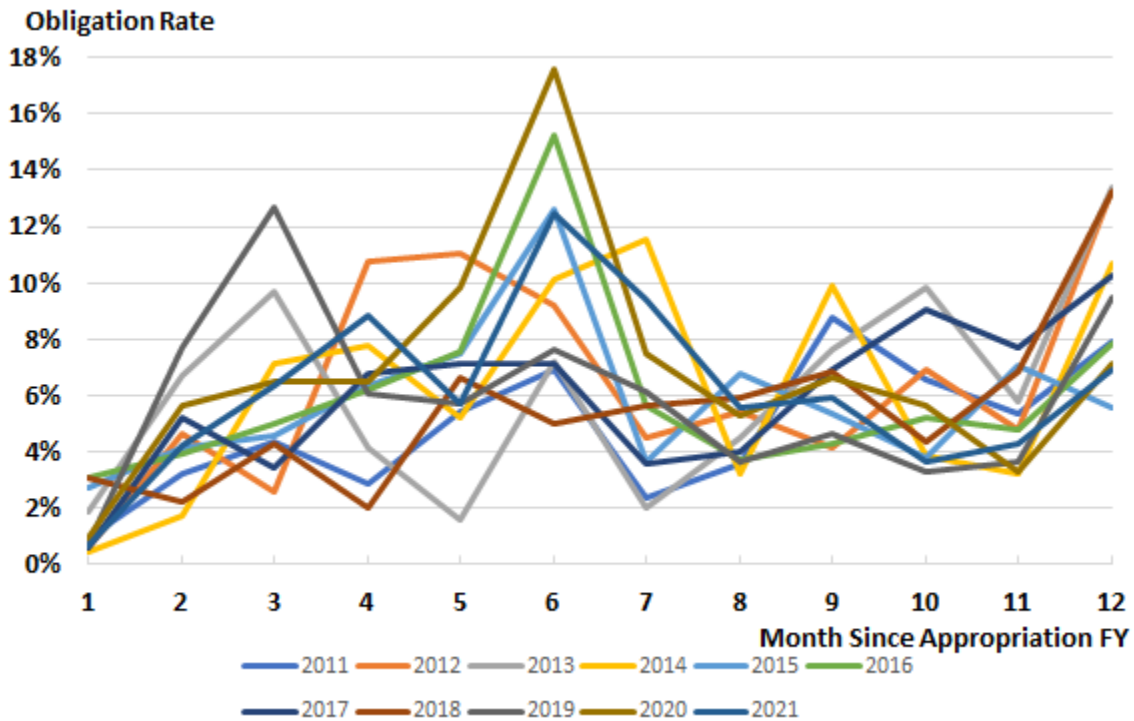


Figure 59. Procurement Monthly Obligation Fractional Rates: Year 1 (FY 2011–2021 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

O&M Monthly Obligation Rates

Figure 6 shows the cumulative O&M obligations as a percentage of the total by the end of the FY. Figure 7 shows the monthly obligation rates for the year.



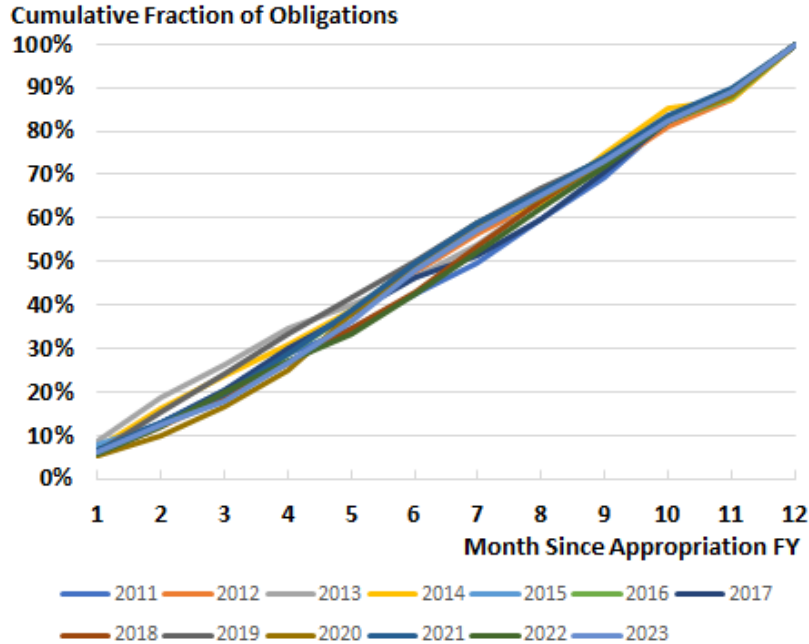


Figure 60. Fraction of Cumulative O&M Obligations by Month (FY 2011–2023 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

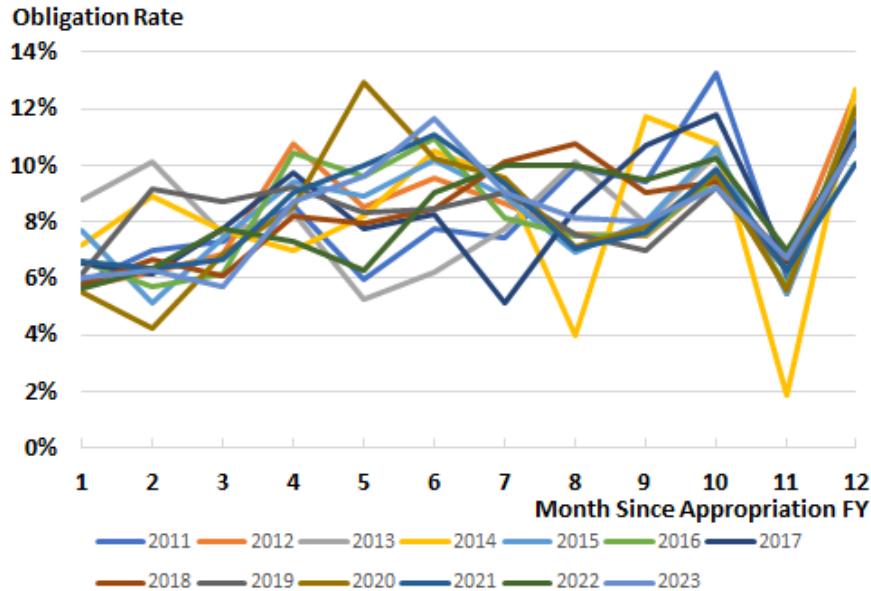


Figure 61. O&M Monthly Obligation Fractional Rates (FY 2011–2023 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

Military Personnel Monthly Obligation Rates

Figure 8 shows the cumulative MILPERS obligations as a percentage of the total by month 36. Figure 9 shows these data as monthly obligation rates for the year available for obligation.



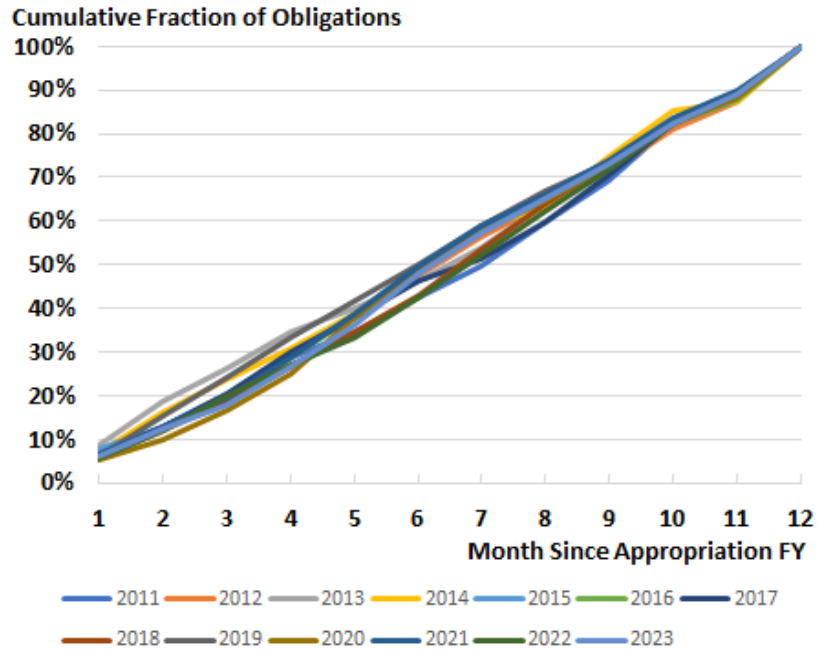


Figure 62. Fraction of Cumulative MILPERS Obligations by Month (FY 2011–2023 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

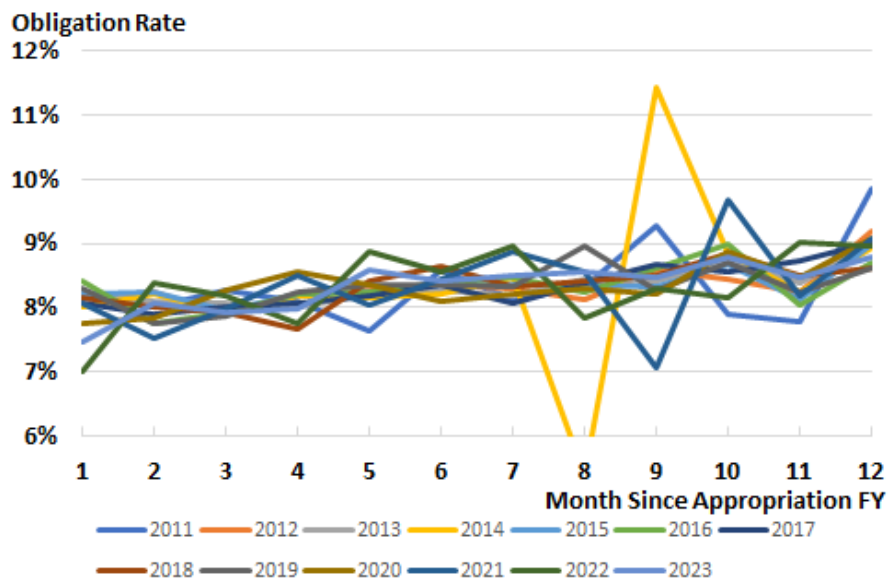


Figure 63. MILPERS Monthly Obligation Fractional Rates (FY 2011–2023 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.



MILCON Monthly Obligation Rates

Figure 10 shows the cumulative dollars obligated by month for MILCON in appropriation years FY 2011–2023. The larger obvious data issues were cleaned, especially through month 60, but what appear to be drops in data reporting can still be seen. Also, there continue to be significant obligations after month 36 (despite what the benchmarks seek).

Figure 11 shows the cumulative MILCON obligations as a percentage of the total by month 60. Because it is unclear what the final obligation values are (see Anton & Buettner, n.d., Figure 2-12), we used the value at month 60 as the final total for the actuals when calculating percentages of total. This skews the tail end of this chart and may not align with how the Comptroller applies the benchmarks against the total obligation authority values that they have. Thus, the values towards the end of these figures are less accurate.

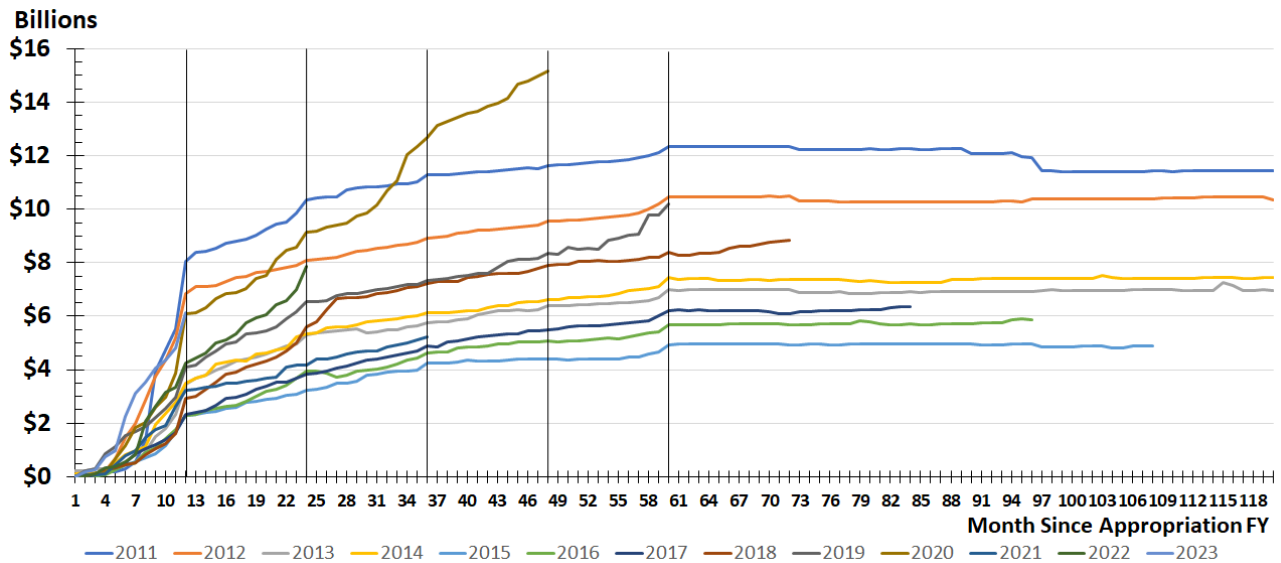


Figure 64. Cumulative MILCON Obligation Dollars by Month (FY 2011–2023 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.



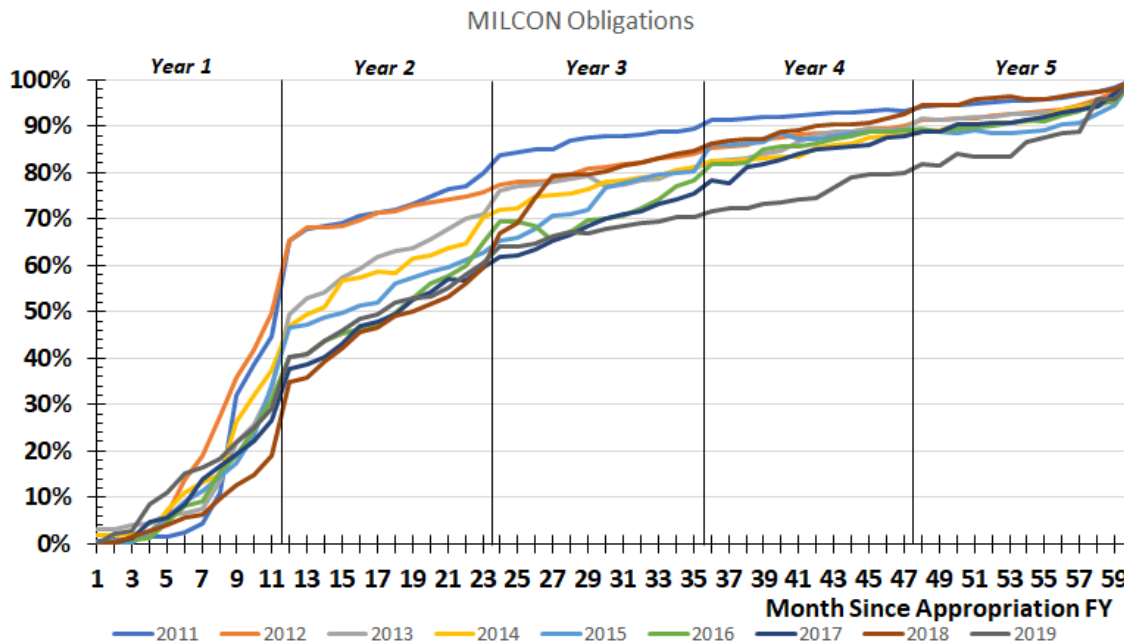


Figure 65. Fraction of Cumulative MILCON Obligations by Month (FY 2011–2019 Appropriations)

NOTE: Month 1 is October of the FY in which the appropriations were made.

CR and Other Effects on DoD Obligation Monthly Rates

Using the data from Advana, we conducted multivariate linear regressions on individual obligation categories (colors of money) to identify any variables that correlate with changes in the normal monthly obligation rates. Table 3 lists the variables tested for correlation (i.e., with a p-value no higher than 0.05). Visual examination of the monthly obligation rates indicated that the first year of obligation behaved differently than any subsequent years, so for multi-year appropriations (RDT&E and PROC), we conducted separate regressions for the first and subsequent years.

Table 4 summarizes the statistical results with the following observations.³ The data are well modeled by a linear obligation rate (the constant base) with adjustments for the variables shown. For example, on average, the RDT&E rate in October of the first year of obligating a FY’s appropriation was $5.9\% - 2.3\% = 3.6\%$. If March of the first spending year was also the third month after the budget was passed (3 MAB), then on average, the obligation rate would be about $5.9\% + 6.3\% + 2.3\% = 14.5\%$ (which is close to the actual value of 15.1% for FY 2012, for example).

³ Statistically, we note the following:

- While all have descent Adjusted R² values, the values for RDT&E (1st year), PROC, and O&M are the highest. Thus, the latter explain the variation in the data well.
- The models (the constant monthly linear contribution plus the contributions from the variables in the model) are fairly linear, with the RDT&E 1st-year for RDT&E and PROC (respectively) and O&M being very linear.



Table 26. Variables Tested for Effects on Monthly Obligation Rates

Type	Variables
FY calendar month	October April November May December June January July February August March September
Month after full budget passed	1 MAB 2 MAB 3 MAB 4 MAB 5 MAB
Time (month #)	Time

Table 27. Contributions of CR and Other Variables Affecting Obligation Rates (FY 2011–2023 Appropriations)

			RDT&E		PROC		O&M	MILPERS
			1 st Year	2 nd Year	1 st Year	2 nd – 3 rd Years		
Average Base	Monthly Rate	<i>Base rate:</i>	5.9%	1.7%	5.0%	2.0%	7.5%	7.9%
CR Effects	1st MAB	<i>If true, add:</i>					1.2%	
	2nd MAB	<i>If true, add:</i>	4.3%		1.8%		2.4%	
	3rd MAB	<i>If true, add:</i>	6.3%		4.6%		2.4%	
	4th MAB	<i>If true, add:</i>	3.7%		2.2%		1.6%	
Calendar Month Effects	October	<i>If true, add:</i>	-2.3 %		-3.6%		-1.1%	
	November	<i>If true, add:</i>		0.9%		0.6%	-0.8%	
	December	<i>If true, add:</i>				0.9%		
	March	<i>If true, add:</i>	2.3%		2.6%	0.4%		
	July	<i>If true, add:</i>					2.4%	
	August	<i>If true, add:</i>					-1.8%	
September	<i>If true, add:</i>		1.0%	4.6%	1.4%	4.1%	0.4%	
Time Trend	Time (month)	<i>If true, add:</i>		-0.1 %		-0.1%		0.1%
Linearity	Multiple R		0.80	0.58	0.77	0.66	0.84	0.48
% of variation explained	Adjusted R²		63%	32%	57%	43%	68%	22%

CRs affect obligation rates as a 2–3-month increase after the budget is passed. CRs affect obligations for RDT&E, PROC, and O&M but not MILPERS. These effects are seen predominantly as increased obligations in the second through fourth months after the budgets are passed but do not extend into subsequent months. The effect is strongest for RDT&E, wherein about a seventh (14.4%) of all RDT&E appears to be withheld until after the budget is passed, contributing to a bump of 14.4% spread over the second through fourth months after passage, about doubling the obligation rates in months 2–3 MAB and six-tenths higher in 4 MAB. This is true also for PROC and O&M, but the bump is smaller at 8.7% and 7.7%,



respectively. With PROC rates about double in 3 MAB and O&M rates about a third higher in 2–3 MAB. Interestingly, this effect is a short-lived bump rather than an inflection in the rates. MILPERS shows no significant post-CR effects.

Obligation rates are higher the month before the midyear appropriation review. For RDT&E and PROC, there is an increase in March of about 40–50%. This may be due to proactive increases in anticipation of the midyear budget reviews in the April and May timeframe. These reviews compare planned to actual obligation and expenditure rates and are “the beginning of major reprogramming activity within the Defense Department, to ensure all appropriated funds are used” (Department of the Air Force, 2022, p. 54).

CR and Other Effects on DoD Obligation Monthly Rates: RDT&E by Budget Activity

The analysis above examined RDT&E obligations together. Subsequently, we obtained DoD monthly RDT&E obligations data broken down by Budget Activity (BA) for appropriation FYs 2013, 2014, and 2017–2022. Table 5 shows the results of the linear regressions on the RDT&E monthly obligation rates as percentages of appropriation-year dollars associated variables across all of RDT&E (original Advana data) and broken down by S&T, development, and management accounts. Table 6 further shows the breakout of S&T by Budget Activity (BA) 1–3. RDT&E obligations for each FY extend for two years. The monthly obligation rates show a major inflection after the first year, so separate regressions were made for each obligation FY. These tables show the average percentage rates (the “average base”) plus the additional percentage-points associated with the variables shown.

When breaking S&T and BAs out of the earlier aggregated RDT&E model uncovered a direct continuing resolution (CR) effect. BAs 2–3 (in S&T) and BA-6 (Management Support [Mgt]) had statistically significant reductions in monthly obligation rates in the first year of about 1.0 to 1.4%. That is in addition to the prior short-term increases in the few months after the full FY budget was passed (MAB). Both the increase in obligation rates in the first few MABs as well as the CR effects for RDT&E and BA-6 can be seen in Figure 12, which plots monthly obligation rates for different appropriation FYs with the budget passage (MABs) aligned instead of by calendar month. This figure also shows that the MAB result is larger for non-S&T but not BA-6 (which has the lowest and more inconsistent MAB effect). In addition, the CR effect also can be seen clearly in the S&T plot in Figure 13.



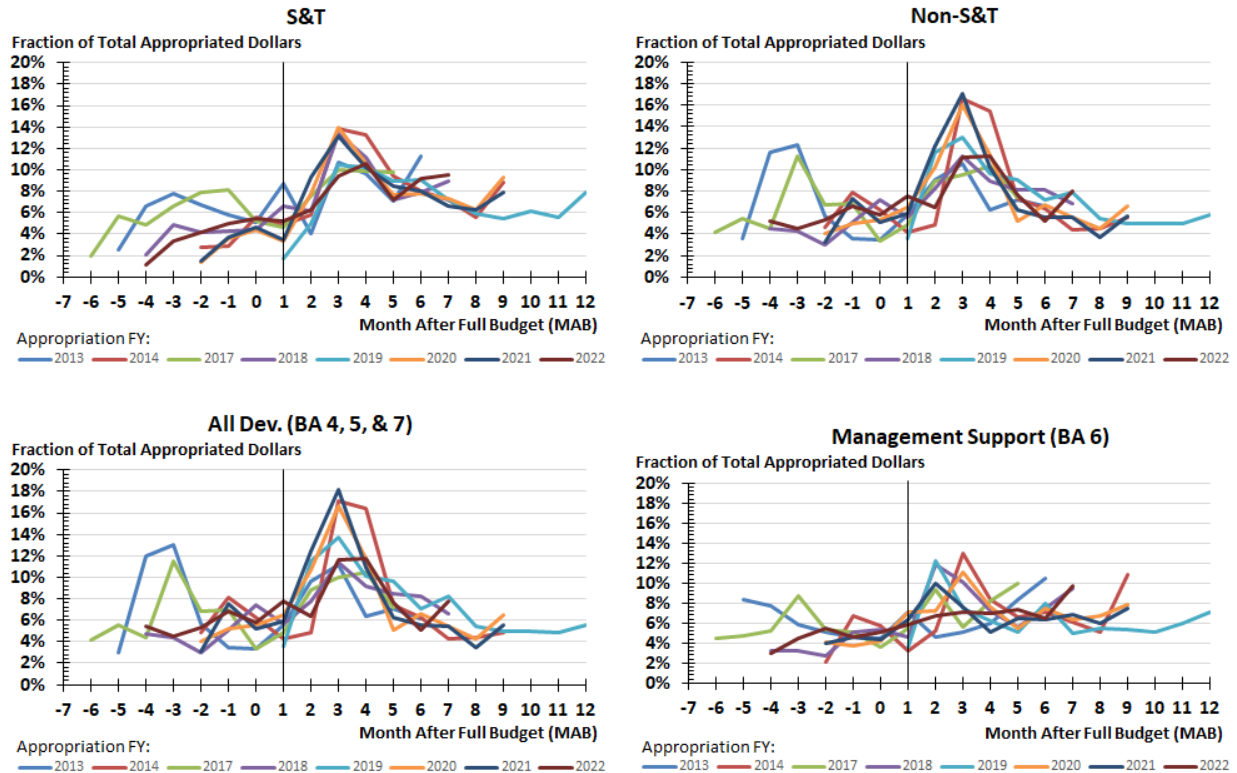


Figure 66. Monthly S&T Obligation Rates for Appropriation FYs Aligned to When the Full Budget Is Passed (FY 2013–2014, 2017–2022 Appropriations)

As an example of how to read the model results in Table 5 and Table 6, if January was the first month after the budget was passed (the “1 MAB”) for an appropriation FY, then the model would predict the monthly appropriation rates shown in Table 7. In our data sample, appropriations from FYs 2014, 2020, and 2021 had their first MAB in January. Figure 13 shows that their incremental monthly obligation rates align to each other and to the model shown in Table 7 for FYs with January 1 MABs. Here, months 1–3 (October through December) operated under a CR and thus were below the base constant of 6.4%. The model showed that October was an additional 3.5 percentage points low and November an additional 1.1 percentage points lower than the baseline minus the CR effect. This can be seen in the lower-left of the figure. The first MAB is 1.6 percentage points lower in the model but rises significantly in MABs 3–8, then returns closer to the base constant with a final increase in September. The example shows that the expected cumulative obligations should be about 84% by September; the actuals for FYs 2014, 2020, and 2021 are 88.0%, 83.3%, and 83.1%, respectively (85% on average—close to 84% from the model).

Not only are there differences between S&T and non-S&T BAs, but the BAs within S&T show differences on which variables correlate with the monthly incremental obligation rates. For example, rates for BA-2 and BA-3 now show a significant drop in all months when there is a CR (-1.1 and -1.2 percentage points below the average base rate). There is neither such CR correlation in RDT&E overall nor any of the other account categories (Procurement, O&M, and MILPERS).



Table 28. Contributions of CR and Other Variables Affecting RDT&E Obligation Rates (FY 2013–2014, 2017–2022 Appropriations)

		1 st Year of Availability				2 nd Year of Availability			
		All	S&T	DEV	Mgt	All	S&T	DEV	Mgt
Avg. Base		5.9%	6.4%	6.0%	6.3%	1.7%	2.3%	1.6%	2.5%
Add CR Effects	CR		-1.0%		-1.4%				
	1 MAB		-1.6%						
	2 MAB	4.3%		3.0%	2.1%				
	3 MAB	6.3%	4.8%	7.1%	2.1%				
	4 MAB	3.7%	4.3%	4.8%					
	5 MAB		1.5%						
	6 MAB		1.8%						
Add Calendar Month Effects	Oct.	-2.3%	-3.5%	-1.9%				0.40%	-0.63%
	Nov.		-1.1%			0.91%		0.52%	
	Mar.	2.3%	1.6%	1.7%					0.68%
	Sept.		2.4%		2.9%	1.0%	1.3%	0.84%	1.3%
Time Trend	Time (mo.)					-0.10%	-0.15%	-0.10%	-0.10%

Adj. R2		63%	84%	60%	42%	32%	48%	50%	33%
----------------	--	-----	-----	-----	-----	-----	-----	-----	-----

MAB = month after budget is passed; CR = month under a continuing resolution (the months before 1 MAB); BA = Budget Activity; S&T = Science and Technology (BA-1, BA-2, and BA-3 combined); DEV = development (BA-4, BA-5, and BA-7 combined); Mgt = Management [Support] (BA-6); mo. = month; Oct. = October; Nov. = November; Jan. = January; Mar. = March; Aug. = August; Sept. = September.



Table 29. Contributions of CR and Other Variables Affecting RDT&E Obligation Rates (FY 2013–2014, 2017–2022 Appropriations)

		1 st Year of Availability								2 nd Year of Availability							
		All	BA-1	BA-2	BA-3	S&T	ACDP	BA-5,7	Mgt	All	BA-1	BA-2	BA-3	S&T	ACDP	BA-5,7	Mgt
Avg. Base		5.9	7.5	6.3	6.2	6.4	5.7	5.9	6.3	1.7	1.8	2.4	2.4	2.3	1.6	1.7	2.5
Add CR Effects	CR			-1.1	-1.2	-1.0			-1.4								
	1 MAB			-1.2	-1.6	-1.6											
	2 MAB	4.3					4.2	2.6	2.1								
	3 MAB	6.3	5.1	4.8	4.8	4.8	7.5	7.7	2.1								
	4 MAB	3.7	3.4	4.0	4.9	4.3	3.3	5.8									
	5 MAB			1.4	1.7	1.5		2.0									
	6 MAB			2.0	1.9	1.8											
Add Calendar Month Effects	Oct.	-2.3	-5.4		-3.6	-3.5		-2.5				-0.42				0.53	-0.63
	Nov.		-3.1		-0.95	-1.1	1.7			0.91						0.59	
	Jan.		-3.5														
	Mar.	2.3		1.5	2.0	1.6	2.5										0.68
	May														-0.48		
	July														0.51		
	Sept.			2.9	2.4	2.4		2.9		1.0	1.2	1.3	1.5	1.3	0.82	0.91	1.3
Time Trend	Time (mo.)									-0.10	-0.11	-0.17	-0.15	-0.15	-0.10	-0.11	-0.10

Adj. R2		63%	54%	77%	87%	84%	56%	57%	42%	32%	21%	49%	39%	48%	22%	51%	33%
----------------	--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

MAB = month after budget is passed; CR = month under a continuing resolution (the months before 1 MAB); BA = Budget Activity; S&T = Science and Technology (BA-1, BA-2, and BA-3 combined); ACDP = Advanced Component Development and Prototypes (BA-4); Mgt = Management [Support] (BA-6); mo. = month; Oct. = October; Nov. = November; Jan. = January; Mar. = March; Aug. = August; Sept. = September.

NOTE: The regression models for BA-1 and ACDP (highlighted in yellow) in the second year of availability are rather low, only explaining 21% and 22% of the variation in the data, respectively. Also, the data for appropriations from FY 2015 and 2016 were not available.



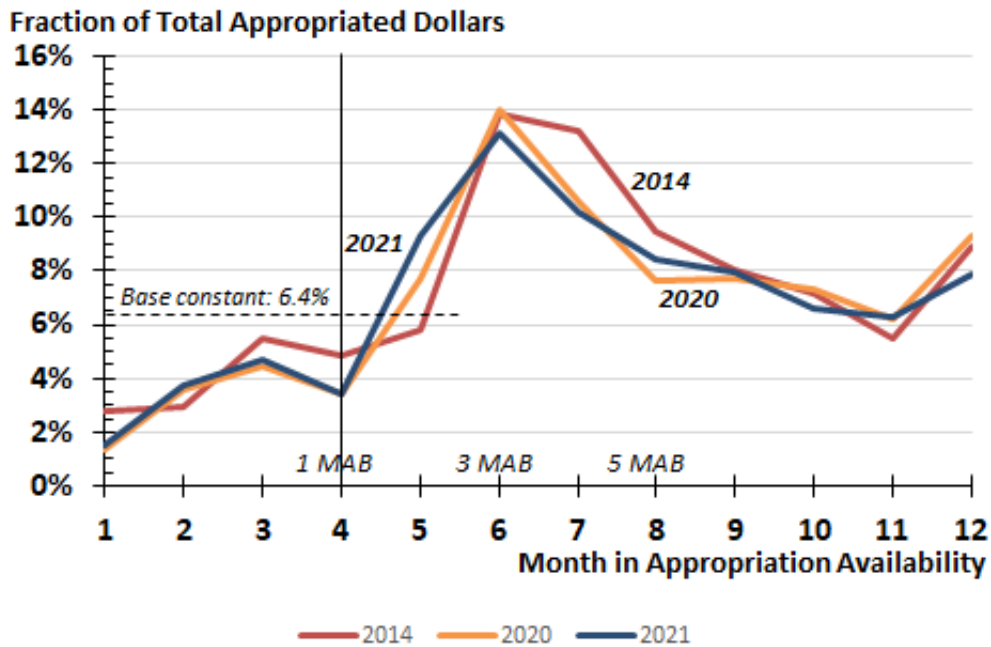
Table 30. Example of How to Read the Obligation Models

Monthly S&T Obligation Rates From the Model for a FY With Full Budget Passed in January

Mo.	#	Variables Active in that month	Avg. Base	+ Variable Effects (CR effects + Oct.–Sept. + MAB effects)	= Expected Obligation Rate
Oct.	1	CR & Oct.	6.4%	-1.0% - 3.5% (the CR and Oct. constants)	= 1.9%
Nov.	2	CR & Nov.	6.4%	-1.0% -1.1% (the CR and Nov. constants)	= 4.3%
Dec.	3	CR	6.4%	-1.0% (the CR constant; no Dec. constant)	= 5.4%
Jan.	4	1 MAB	6.4%	-1.6% (the 1 MAB constant)	= 4.8%
Feb.	5	2 MAB	6.4%	(no 2 MAB constant in the model)	= 6.4%
March	6	3 MAB & March	6.4%	+ 4.8% + 1.6% (the 3 MAB and March constants)	= 12.8%
April	7	4 MAB	6.4%	+ 4.3% (the 4 MAB constant)	= 10.7%
May	8	5 MAB	6.4%	+ 1.5% (the 5 MAB constant)	= 7.9%
June	9	6 MAB	6.4%	+ 1.8% (the 6 MAB constant)	= 8.2%
July	10	none	6.4%	(no July constant in the model)	= 6.4%
Aug.	11	none	6.4%	(no Aug. constant in the model)	= 6.4%
Sept.	12	Sept.	6.4%	+ 2.4% (the Sept. constant)	= 8.8%
Subtotal for 1st Year					84%

NOTE: The constants from the S&T obligation model are taken from the second column in Table 5, which is the same as the fifth column in Table 6.

Figure 67. Monthly S&T Obligation Rates for Appropriation FYs With Full Budget Passed in January (FY 2014, 2020, and 2021 Appropriations)



Conclusions on Obligation Benchmarks

Theory, data, and analysis identified several complications that affect the realism of the current Comptroller benchmark profiles.



Operating under CRs can also delay obligations. As mentioned earlier, while there is no general slowing of actual obligations during a CR (except for S&T and BA-6 within RDT&E), statistical analysis found statistically significant increases in several months after final appropriations are passed (see the MAB variable results in the regressions above. Tremaine and Seligman (2013) also published a survey of 229 DoD personnel. The survey indicated that CRs can delay obligations. Thus, there is quantitative and qualitative evidence supporting an adjustment in obligation benchmarks due to CRs. The three-month surge after budget passage may be due to new starts or to other program authorization and appropriation effects on obligations.

There is also a concern that the general restrictions on new starts during a CR drives the DoD to spend on older priorities rather than providing an ability to respond to new immediate threats and their associated reprioritizations on spending needs in the DoD.

Monthly DoD Expenditure Rates by Categories of Funding: Actual DoD Data

As with obligations, we obtained expenditure data from the DoD’s Advana data environment in the form of monthly expended dollars (also called disbursements or outlays) for military services and defense-wide funding accounts for categories of funding (e.g., Army RDT&E, Navy Aircraft Procurement, Defense-Wide O&M, and Air Force Military Personnel) from FY 2011–2023. We aggregated these data to obtain monthly expenditure dollars by category across all accounts, then calculated the percentage expended in each month compared to the total dollars expended by the final month. Table 8 lists the time periods used to observe monthly expenditure rates and trends as a percentage of the final month. Although some trends will extend further in time, this provides a good window into the bulk of the trends for analysis.

Table 31. Month Used for Final Expenditure Values

Category	Years to Expend	Final Month
RDT&E	3	36
Procurement	3*	36*
O&M	2	24
MILPERS	1	12
MILCON	6**	72**

* We used 36 months as the reference time to expend PROC dollars, although some procurements have longer to obligate and disburse funds.

** In the DoD data, MILCON reached about 90% on average by the end of year 6.

NOTE: These periods provide us with a simple way to measure and observe expenditure trends and rates.

Conclusions and Recommendations

In this paper, we provided the bulk of our analytical work from our report to the PPBE Reform Commission (Anton & Buettner, n.d.). Based on these observations, we recommended that the DoD Comptroller consider modifying their benchmarks. Four optional variants were discussed in our report. The preferred option includes adding additional S-curve ramp-up elements on top of historical obligation behaviors and recommends replacing linear expenditure profiles with historical S-curve profiles. Table S-3 in our report summarizes our recommendations.⁴ Anton and Buettner (n.d.) also provide example benchmark tables that better reflect recent history as well as the correlative effects of month, CR, and time.

⁴ Some of these conclusions and recommendations are also reflected in our companion NPS ARS 2024 paper, *Improving Comptroller Benchmarks on Program Spending* (Anton, Stalebrink, and Buettner, forthcoming), and that paper includes the tables from the forthcoming report.



In addition to aligning expenditure benchmarks to actual data and theoretical objectives, such changes could help eliminate the negative side effects cited in theory and the literature that program managers may seek expenditures prematurely just to meet comptroller benchmarks at the expense of other program and department objectives of prudent use of the resources (see, for example, Marsalis, 2022; Commission on PPBE Reform, 2023, p. 33). Slight delays in switching to S-curves with their lower initial expenditure benchmarks should give program managers more time to get good deals for the program, the DoD, and taxpayers rather than having to rush negotiations and contracting to meet somewhat arbitrary benchmarks or risk losing their funding.

There would be some cultural and process adjustments for both Congress and the DoD (and industry) to adjusting the obligation and expenditure benchmark profiles over time, but the benefits could be improved performance given the financial resources provided by Congress and the taxpayers to the DoD. In the end, keep in mind the following insightful quote.

Suggested Future Research

These analyses have highlighted areas for future research that could lead to a deeper understanding of the effects from the recommendations made in Anton and Buettner (n.d.).

Assess obligation and expenditure rates at the account level within each category. For expedience, this research aggregated all the accounts within each funding category (RDT&E, PROC, O&M, MILPERS, and MILCON). There may be additional insights from statistical analysis at the account level. For example, there may be significant differences among the military departments or insights into the variation by account type (e.g., aircraft versus missiles).

Assess sources of obligation and expenditure data errors. As noted, there are some obvious errors in the reported data. Further research could identify these sources and help the Comptroller improve data reported to Advana. For example, some errors may just be that programs that have completed obligations or expenditures stopped reporting for a FY's appropriations, resulting in drops in the totals reported at the account level. Correcting such data issues will become increasingly important as these data are made more available for this type of longitudinal trend analysis. Also, further data cleaning of the MILCON data after the third year is warranted; this may enlighten exactly how long into the future MILCON obligation and expenditures extend—and by what fractions of the total.

Acknowledgements

We would like to thank the Commission on Planning, Programming, Budgeting, and Execution (PPBE) Reform for their sponsorship and support throughout this research. In particular, the questions, discussions, and feedback from the Commission chairs and staff were extremely helpful: the HON. Robert Hale, the HON. Ellen Lord, as well as the commission staff, especially Lara Sayer (our sponsor of record), Annie Crum, Elizabeth Bieri, Caroline Bledsoe, Kelle McCluskey, Brooks Minnick, and Soleil Sykes. We would also like to thank John Stedge in the Chief Digital and Artificial Intelligence Office (CDAO), for guiding us on how to pull the Obligations and Expenditures data from Advana. Finally, we would like to acknowledge and thank our Stevens staff: Technical Editor Nancy Méndez-Booth for her editorial review comments and suggestions, and Tara Kelly for her project management skills, edits, and overall support.



References

- Anton, P. S., & Buettner, D. J. (n.d.). *DoD obligation and expenditure rates: More-realistic benchmarks and the effects of continuing resolutions and other events on obligation rates* [Forthcoming]. Acquisition Innovation Research Center, Stevens Institute of Technology.
- Anton, P. S., Stalebrink, O. J., & Buettner, D. J. (n.d.). *Improving comptroller benchmarks on program spending* [Forthcoming]. 2024 Acquisition Research Symposium, Naval Postgraduate School, Monterey, CA.
- Commission on Planning, Programming, Budgeting, and Execution Reform. (2023, August). *Interim report*. <https://ppbereform.senate.gov/wp-content/uploads/2023/08/PPBE-Commission-Interim-Report-Final.pdf>
- Department of the Air Force. (2022, February). *Planning, programming, budgeting, and execution system training program: Reference manual*. <https://afacpo.com/AQDOcs/PPBE.pdf>
- DoD. (2008, October). *Financial management regulation* (DoD 7000.14R, Vol. 2A). https://comptroller.defense.gov/Portals/45/documents/fmr/current/02a/02a_01.pdf
- GAO. (2005, September). *A glossary of terms used in the federal budget process* (GAO-05-734SP). <https://www.gao.gov/assets/gao-05-734sp.pdf>
- GAO. (2022, June). *Federal budget: Selected agencies and programs used strategies to manage constraints of continuing resolutions* (GAO-22-104701). <https://www.gao.gov/assets/gao-22-104701.pdf>
- Marsalis, J. R. (2022). *The negative impacts of the benchmarks* [Senior acquisition course]. National Defense University. [fc216616-7b4c-45e6-d4fa-c4bc48a6b71f \(nps.edu\)](https://www.ndu.edu/learning-center/online-courses/216616-7b4c-45e6-d4fa-c4bc48a6b71f)
- Office of the Secretary of Defense. (2012, September 10). Department of Defense management of unobligated funds; obligation rate tenets [Memorandum by the USD(Comptroller) and USD(Acquisition, Technology, and Logistics)]. DoD. https://www.acq.osd.mil/dpap/policy/policyvault/DoD_Mgt_of_Unobligated_Funds_Obligation_Rate_Tenets.pdf
- Tremaine, R. L., & Seligman, D. J. (2013, July–August). OSD’s obligation and expenditure rate goals. *Defense AT&L*, 14–21. <https://apps.dtic.mil/sti/pdfs/ADA608878.pdf>



Unlock the Hidden Secrets of AI Transformation on the Workforce

Symantha “Sam” Loflin—has an MS in program management and a certification in advanced acquisition studies from the Naval Postgraduate School (NPS), where she was a contributor to the 18th, 19th, and 20th Annual NPS Acquisition Research Symposium. She also holds a BS in finance from the University of Houston. Loflin is a Case Manager and Researcher at Tanner and Associates. She has more than 20 years of acquisition experience supporting DCMA, the DoD, NASA, and the military services. She recently served as an acquisition program manager on the Coronavirus Task Force that focused on building the industrial base for personal protective equipment. Loflin’s career began at NASA, supporting the Space Shuttle, ISS, and the Constellation Programs in Houston. [symanthaloflin@gmail.com]

Abstract

The author has written this paper to defend and strengthen the use of governmental risk mitigation measures that prevent divergence, ensure safety, and highlight the possibilities of future growth in workforce skills. The purpose of the investigation is to discover and disclose of the impact of artificial intelligence (AI) and cyberspace on the nation and global human population and workforce. The analysis of the research revealed the key results related to the current and future measures that the United States government, military services, the country, and the world, endorse to secure and protect consumers and the workforce, while promoting innovation from the use of safe AI and cyberspace. Each year, the federal government increases the funding of developmental contracts as a measure to “Protect Sea, Air, and Space” (White House, 2022b). These efforts aim to protect U.S. interests in developing technologies, creating economic opportunities, and enabling climate surveillance, and to responsibly oversee the space environment. The government is working with allies and industrial base partners in advancing and developing new technologies with trusted artificial intelligence and secure internet to create prosperity and economic security. Industry and the United States government’s ability to engage the right people, processes, and tools at the right time is essential to effective program management policy and control (Hite, 2010. p. 23).

Research Issues

How will the United States defend, protect, and safely navigate the workforce throughout the emerging technology of artificial intelligence and the risk of cyber-attacks? Are the AI secrets really hidden or is there so much information on the web that unless it touches us we do not see it?

Research Results Statement

The results are clear and convincing that it takes a whole-of-nation and global approach to defend and protect the workforce to achieve and enhance economic growth and protect national security. The conformality of legal and regulatory procedures is necessary to protect human rights and safely use AI and protect against cyber vulnerabilities and cyber-attacks is essential to strengthen the future of workforce jobs. Investing in our workforce effectively and efficiently increases economic growth in the production of good, services, and materials in the United States that meet Federal procurement needs. On January 29, 2024, the White House published *Fact Sheet: Biden-Harris Administration Announces Key AI Actions Following President Biden’s Landmark Executive Order* (White House, 2024).

Recommendations

1. Provide public service announcements on the measures the U.S. government has taken to educate the American public and the workforce on the safe and secure AI and cybersecurity (White House, 2024). Research shows that subject matter knowledge increases



the likelihood of successful problem resolutions “experiential learning offers a way to ensure we are imparting not just rote learning and certifications but providing our people the knowledge, skills, and experience to effectively control the efforts we charge them to lead” (Pickar, 2020). Human capabilities lead to improved preventative measures. 2. Enhance government oversight of statutory and regulatory policies, standards, and procedures to safeguard and preserve the rights of the workforce. 3. Create a consortium for enhanced collaboration. (See the *Evidence-Based Policymaking Act of 2018 (Evidence Act)* (The Evidence Act, 2019).)¹ See also learning agendas:

According to Office of Management and Budget (OMB) guidance for implementing the Evidence Act, a learning agenda is to define and prioritize relevant questions and identify strategies for building evidence to answer them. In developing a learning agenda, an agency should involve key leaders and stakeholders, to help (1) meet their evidence needs for decision-making and (2) coordinate evidence-building activities across the agency. (GAO, 2019)

Introduction

The United States government, humans, companies, the nation, and the world’s ability to develop the workforce and engage the right people, processes, and tools at the right time is challenging. Education and communication for humans to safely use AI technology and prevent cyber-attacks and guard against vulnerabilities are essential to be effective and efficient. To ensure the rights of humans there must be accountability through government oversight of AI and cybersecurity laws, regulations, standards, guidance, and policy. These measures will provide the capabilities required to maximize the innovations of the workforce and support proper stewardship of taxpayer dollars.

Biden-Harris Administration’s National Security Strategy,

By enhancing our industrial capacity, investing in our people, and strengthening our democracy, we will have strengthened the foundation of our economy, bolstered our national resilience, enhanced our credibility on the world stage, and ensured our competitive advantages. (White House, 2022)

In the March 2017 issue of the *ScienceDirect Engineering Journal*, the Research Intelligent Manufacturing-Review issued the publication, “Intelligent Manufacturing in the Context of Industry 4.0: A Review Industry,” and provided the research results that provide an understanding of intelligent manufacturing in Industry 4.0. “Our next generation of Industry—Industry 4.0—holds the promise of increased flexibility in manufacturing, along with mass customization, better quality, and improved productivity” (Zhong et al., 2017). Additionally, the results showed that some technologies have AI, including cyber-physical systems that allows learning with manufacturing systems. In Industry 4.0, the expected collaborations between humans and machines will enable humans to be more efficient and effective with decision

¹ The Evidence Act adopts as its definition of evidence “information produced as a result of statistical activities conducted for a statistical purpose.” It adopts as its definition of statistical purpose “the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups and includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support” those actions. Pub. L. No. 115-435, § 101(a)(1); 44 U.S.C. § 3561(6), (12). OMB’s June 2019 update to Cir. No. A-11 contains these definitions. The guidance also states that in the context of improving organizational and agency performance, “evidence” can be viewed more broadly, in line with OMB’s definition.



making (Zhong et al., 2017). Even though in Industry 4.0, AI and an IT infrastructure provide the workforce with improved efficiency, reliability, and human involvement, there are inherent risks of cyber attacks and cybersecurity must be a top priority for humans, companies, the nation, and the world.

June 23, 2009—United States Cyber Command (USCYBERCOM)



It is important to note that Dr. Robert M. Gates, in his book *Duty* (2014, p. 300), emphasized the urgency to direct the establishment of the United States Cyber Command (USCYBERCOM). On June 23, 2009 Gates's memorandum also stipulated that the Under Secretary for Policy would develop a new national cybersecurity strategy (p. 1). The nation and military operations dependence on cyberspace continues to increase, along with the threats and exposures. The cyber risk of exposing required technical warfighting capabilities across the globe is a matter of national security. Gates's highest mission priority was realized with the establishment and successful accomplishments of the USCYBERCOM. The command is now almost 14 years old and the nation's unified combatant command operates globally to ensure the United States and its allies have freedom of action in cyberspace while defending against the same adversaries. Additionally, as shown on the command website, "This is our code. As the nation's first line of defense in cyberspace, we operate at the speed, relevance, and scale necessary to win" (United States Cyber Command, n.d.). Mr. Gates notable book cover and quote about the importance of knowledge are shown in Figure 1.

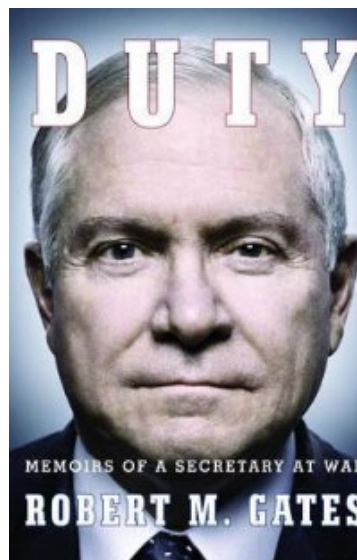


Figure 1. Cover of Robert M. Gates Book,
Duty: Memoirs of a Secretary at War
(Gates, 2014)

"What I know concerns me. What I don't know concerns me even more.
What people aren't telling me worries me the most" (Gates, 2014).



National Institute of Standards and Technology (NIST)—Industry 4.0—Cybersecurity

On October 26, 2023, Katie Rapp a writer/editor for the National Institute of Standards and Technology's (NIST's) Manufacturing Extension Partnership, published "Infographic—Integrating Cybersecurity with Industry 4.0: What it Means for Manufacturing" (Rapp, 2023, p. 1). She asked the questions and provides her research answers,



Did you know that manufacturing is now the most targeted industry for cyber attacks? The average cost of a data breach for a small business is \$105,000. Can your firm absorb that cost? Can you risk the down time and the damage that a data breach would cause your business? (Rapp, 2023)

The NIST has a blog that is a series on cybersecurity and Industry 4.0. On May 11, 2022, Pat Toth, a 30 year cybersecurity guidance document professional at the NIST, posted "Cybersecurity and Industry 4.0—What You Need to Know" (infographic is shown in Figure 2 (Toth, 2022). She provided a thorough background of Industry 4.0 and the relationship of the *NIST Special Publication 1500-201, Framework for Cyber-Physical Systems: Volume 1* (Giffor, 2017) that is intended to document external viewpoints/artifacts (use cases with activities and artifacts for certification and regulatory compliance testing) in the application of the CPS Framework from industry, academia, and government. Note: The reports do not represent official NIST positions.

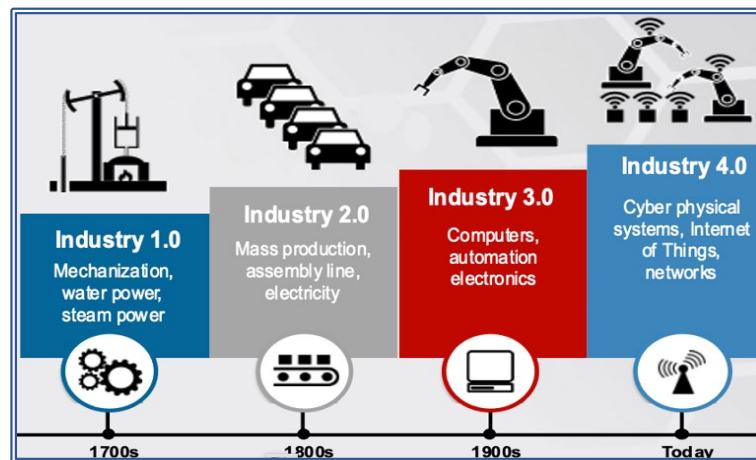


Figure 2. NIST Industrial Revolutions—Cybersecurity and Industry 4.0—What You Need to Know (Toth, 2022)

The cybersecurity and Industry 4.0 blog provides the "Need to Know" on Cyber-Physical Systems (CPS). Cobots are engineered interacting networks between cyber-physical systems and the networks of physical and computational components. The systems are providing new functionalities in many domains including our quality of life, smart manufacturing, defense, and homeland security. Cobots (used to perform some manufacturing tasks) are rapidly being adopted by manufacturers to assist humans in a shared share workspace and requires ongoing safety risk assessments.

Additionally, Internet of Things (IoT) and Big Data, as well as Cloud Manufacturing have transformed resources and capabilities that improve production and provide a safer and higher quality manufacturing life cycle. However, hackers, cybercriminals, and industry competitors pose extreme cyber risk. Cyber risk assessments are essential prior to the introduction of new



technology. Industry 4.0 brings robotic and automation efficiencies to manufacturing, as well as the risk cyber vulnerabilities.

January 26, 2023, AI Risk Management Framework (NIST AI 100-1)

The world must be vigilant in creating safe and trustworthy AI systems with ongoing surveillance to ensure compliance. Safe AI systems contribute to human efficiencies when used in manufacturing through automation and robotics. The NIST published examples of the documented harms of AI systems in Figure 3 (NIST, 2023a).



Figure 3. NIST AI 100-1, Harms from AI systems (NIST, 2023a)

January 26, 2023, AI Risk Management Framework (RMF)

Ideally, the life cycle stages of the AI area teams include actions from a diverse team in AI risk management, operations, plan and design function, as well as all stakeholders that exists through the AI system life cycle. The teams share ideas, develop, deploy, test, and evaluate to identify, analyze, track, and manage-prioritize current and emerging risk. The risk are mitigated and preventive measures are integrated. The AI RMF life cycle shows the stages of the AI system actors in planning and design as an integral part of the risk management efforts as shown in Figure 4.



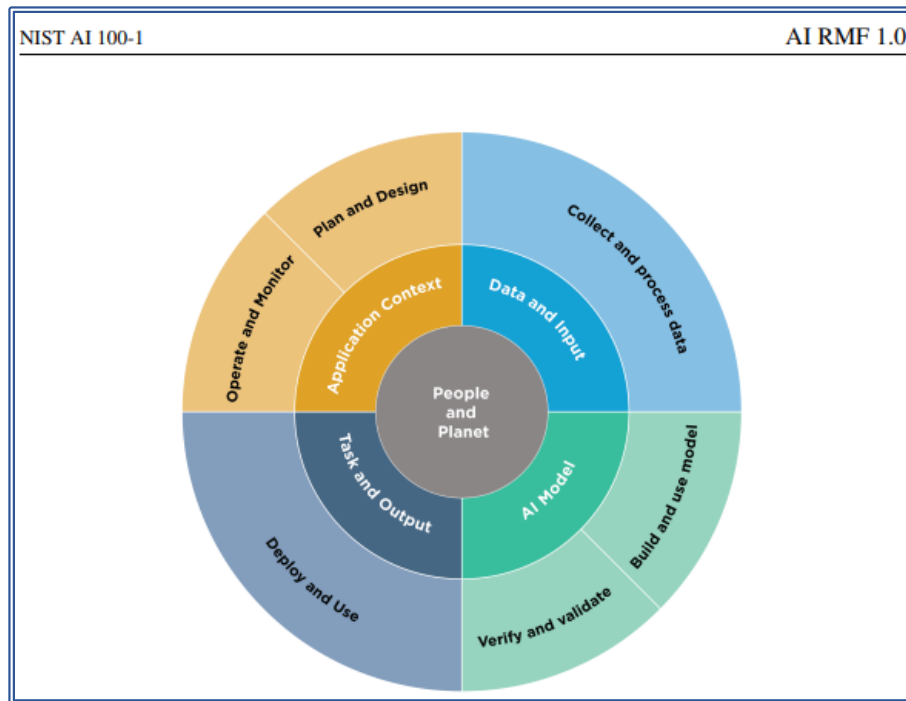


Figure 4. Life Cycle and Key Dimensions of an AI System (Organization for the Economic Cooperation and Development [OCED], 2022)

March 30, 2023, Trustworthy & Responsible AI Resource Center

The center was established by the NIST Trustworthy & Responsible Artificial Intelligence Center (AIRC; NIST n.d.). The AIRC was developed to provide AI actors in the “development and deployment of trustworthy and responsible AI technologies” (NIST, 2023). The center offers numerous resources (e.g., NIST AI RMF 1.0, Playbook, Training, etc.). The AI RMF Timeline and Engagements shows a that the establishment of the AI RMF from its July 29, 2021 inception to publication of the AI RMF 1.0 and the Playbook on July 26, 2023, took more than 30 months (Figure 5).

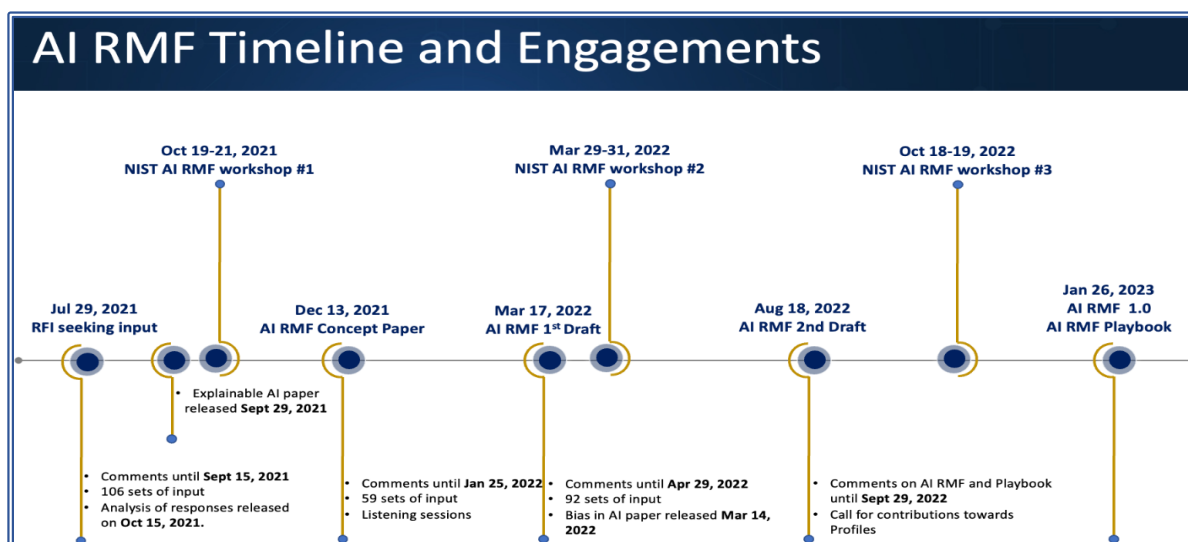


Figure 5. AI Risk Management Framework Timeline and Engagements (NIST, 2023)



Table 1 Shows examples of some of the notable government measures related to AI.

Table 1. Notable Government Measures Related to AI

United States Government Measures ~Artificial Intelligence (AI)~
 <p>United States Department of State</p>
<p>January 1, 2021, <i>National Artificial Intelligence Initiative Act of 2020 (NAIIA)</i>, became law. H.R.6216 - 116th Congress (2019-2020): National Artificial Intelligence Initiative Act of 2020 Congress.gov Library of Congress</p>
<p>April 28, 2022, Declaration for the Future of the Internet.</p>
 <p>CHIEF DIGITAL AND ARTIFICIAL INTELLIGENCE OFFICE OSD Bias Bounty Presented by  January 29–March 11, 2024 (closed), AI Bias Testing Office of Secretary of Defense. (2024, January 29–March 11). OSD Bias Bounty. https://osdbiasbounty.com/sign-in?callbackUrl=https%3A%2F%2Fosdbiasbounty.com%2Fsign-in</p>

<p>October 4, 2022, What is the Blueprint for an AI Bill of Rights? Relationship to Existing Law and Policy, E.O. 13960, 13985, and Fair Information Practice Principles (FIIPs) https://www.whitehouse.gov/ostp/ai-bill-of-rights/what-is-the-blueprint-for-an-ai-bill-of-rights/ and https://www.whitehouse.gov/ostp/ai-bill-of-rights/relationship-to-existing-law-and-policy/</p>
<p>October 30, 2023, AI Bill of Rights on Safe, Secure, and Trustworthy Artificial Intelligence Executive Order 14110 Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/</p>

May 2023, World Economic Forum, Future of Jobs Report 2023—Industry 4.0

The World Economic Forum was established in 2014 and its first edition report was published in 2016 (World Economic Forum, 2023). The forum serves a significant purpose by collaborating with governments from around the globe and public-private sectors as an international organization to create surveys that are distributed to employees. The survey results provide insight over the next five years for analysis and employer planning into the future of jobs and skills under the **Forth Industrial Revolution—Industry 4.0**.

The 2023–2027 survey included 44 survey questions in 12 languages with data collection between November 2022 and February 2023. The survey was administrated by



participating companies for future of jobs and skills projections in the November 2023 through February 2027 timeframe. The data was collected by 830 companies with a collective 11.3 million workers in 45 world region economies and 22 country acknowledgements. The results provide the private- and public-sector leaders with insight into the future of jobs and skills needed to be successful in providing a better work future for all. The World Economic Forum, Future of Jobs Insight Report 2023 cover is shown in Figure 6.



Figure 6. World Economic Forum, Future of Jobs Report (World Economic Forum, 2023)

United States of America Job Creation/Displacer Outlook 2023–2027

The World Economic Forum determined that age 25 and older is the working age population. The major focus was on the economy and global five-year business transformation for trends and technologies and the impact as a job creator, displacer, net effect, and global net effect. The Future of Jobs 2023 Insight Report calculated 226 million people of the working age in the United States. The economic profile for the United States of America shows the key impacts in job creation in Figure 7. The largest impact globally is in skills (cognitive and self-efficacy), reskilling (AI and big data analytical and creative thinking), and training funding (org. and free training) as shown in Figure 8.



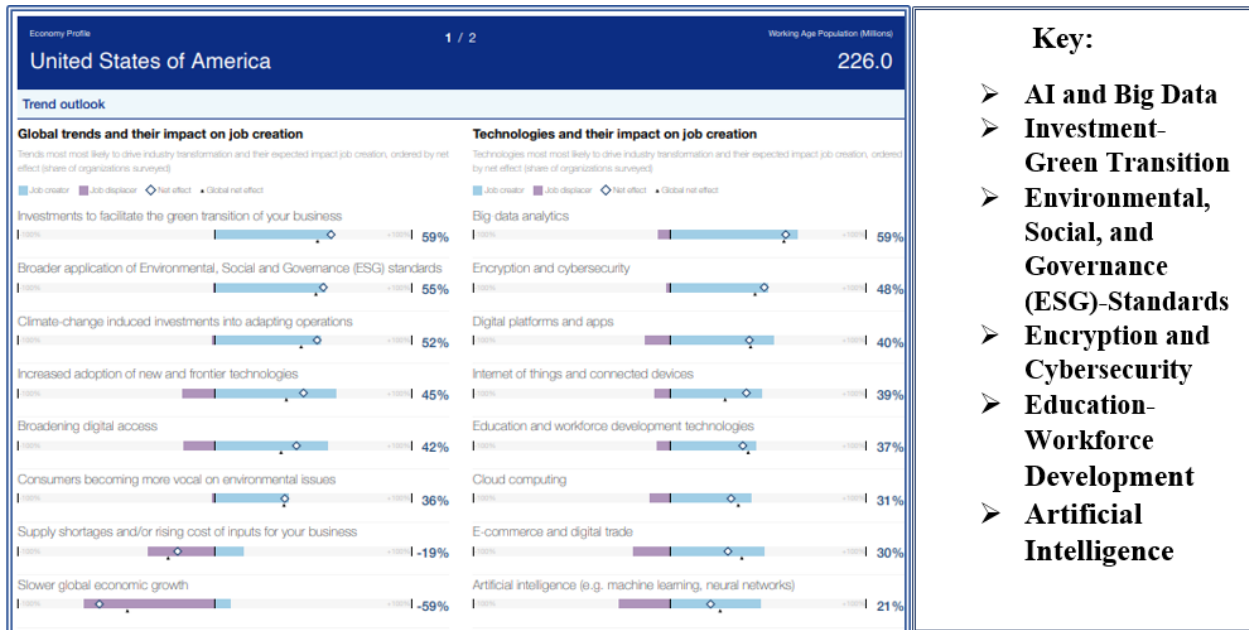


Figure 7. Economic Profile for the United States of America



Figure 8. Economy and Global Skills-Reskilling-Training





ARTIFICIAL INTELLIGENCE

Addendum to the NPS Strategic Plan 2018–2023



“Humanity is at the edge of a revolution driven by artificial intelligence ... This revolution is unstoppable. Attempts to halt it would cede the future to that element of humanity more courageous in facing the implications of its own inventiveness. Instead, we should accept that AI is bound to become increasingly sophisticated and ubiquitous, and ask ourselves: How will its evolution affect human perception, cognition and interaction? What will be its impact on our culture and, in the end, our history?”

-Henry Kissinger, Eric Schmidt and Daniel Huttenlocher
The Atlantic • August 2019

Figure 9. *Artificial Intelligence. An Addendum to the Strategic Plan* (Naval Postgraduate School, 2019, p. 2)

References

Bias Bounty. (n.d.). <https://osdbiasbounty.com/>

Bureau of Labor Statistics. (2024 January 23). *News release*. U.S. Department of Labor.
<https://www.bls.gov/news.release/pdf/union2.pdf>

COPILOT. (n.d.). Microsoft. <https://copilot.microsoft.com/>



- The Evidence Act, Pub. L. No. 115-435, 132 Stat. 5529 (Jan. 14, 2019).
<https://www.congress.gov/bill/115th-congress/house-bill/4174/text>
- Federal Register. (2020, December 3). *Promoting the use of trustworthy artificial intelligence in the federal government* (Executive order 13960).
- Gates, R. M. (2009, June 23). *Establishment of a subordinate unified U.S. cyber command under U.S. strategic command for military cyberspace operations* [Memorandum]. Department of Defense. <https://nsarchive.gwu.edu/document/21425-document-29>
- Gates, R. M. (2014). *Duty: Memoirs of a secretary at war*. Alfred A. Knopf.
<https://www.google.com/books/edition/Duty/IYzZAAAAQBAJ?hl=en&gbpv=1&bsq=What%20I%20know%20concerns%20me>
- Giffor, E. (2017). *Framework for cyber-physical systems* [Special Publication 1500-201]. Department of Commerce. National Institute of Standards and Technology (NIST).
<https://doi.org/10.6028/NIST.SP.1500-201>
- Government Accountability Office. (2019, December). *Report to congressional requestors. Evidence-based policymaking. Selected agencies coordinate activities, but could enhance collaboration*. <https://www.gao.gov/products/gao-20-119>
- Government Accountability Office. (2023, December). *Artificial intelligence agencies have begun implementation but need to complete key requirements* (GAO-24-105980).
<https://www.gao.gov/assets/d24105980.pdf>
- Hite, R. C. (2010, August). *Organizational transformation: A framework for assessing and improving enterprise architecture management (version 2.0)* (GAO-10-846G). Government Accountability Office. <https://www.gao.gov/assets/gao-10-846g.pdf>
- Homeland Security. (2008, December 29). *Fair information practice principles (FIPPs)*.
https://www.dhs.gov/sites/default/files/2024-01/Fair%20Information%20Principles_12_2008.pdf
- Foundations for Evidence-Based Policymaking Act of 2018, Pub. L. No. 115-435 (2018).
<https://www.congress.gov/bill/115th-congress/house-bill/4174/text>
- International Association of Machinists and Aerospace Workers (IAMAW). (1888).
<https://www.goiam.org/>
- International Brotherhood of Teamsters (IBT). (1903). <https://teamster.org/>
- Kehoe, A. (2022, May 21). *Navy ships swarmed by drones, not UFOs, defense officials confirm*. The Warzone. <https://www.twz.com/navy-ships-swarmed-by-drones-not-ufos-defense-officials-confirm>
- Kennedy, B., Tyson, A., & Saks, E. (2023, February 15). *Public awareness of artificial intelligence in everyday activities*. Pew Research Center.
<https://www.pewresearch.org/science/2023/02/15/public-awareness-of-artificial-intelligence-in-everyday-activities/>
- Kobielus, J. (2020, July 10). *Robotics: How crisis is accelerating automation*. Information Week.
<https://www.leadershipreview.net/robotics-how-crisis-is-accelerating-automation/>
- Livescience. (2014, August 25). *Timeline of artificial intelligence*.
<https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html>
- Lynch, S. (2023, April 3). *2023 State of AI in 14 charts*. Stanford University.
<https://hai.stanford.edu/news/2023-state-ai-14-charts>



- National Artificial Intelligence Initiative Act of 2020*, H.R. 6216, 116th Cong. (2020). <https://www.congress.gov/bill/116th-congress/house-bill/6216/text>
- National Institute of Standards and Technology (NIST). (2023a, January 26). *AI risk management framework (RFM)*. <https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf>
- National Institute Standards and Technology (NIST). (2023b, March 30). Trustworthy & responsible AI resource center. <https://airc.nist.gov/Home>
- National Institute of Standards and Technology (NIST). (2023c, March 30). *AI RMF timeline*. <https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development>
- National Institute of Standards and Technology (NIST). (n.d.). *Crosswalk document list*. <https://airc.nist.gov/home>
- Naval Postgraduate School. (2019, August 15). *Artificial intelligence. An addendum to the strategic plan*. https://nps.edu/documents/115153495/115155926/Al_StratPlan_Addendum_WEB-FINAL.pdf/ed7ddfbf-49c1-4a3e-b3ca-67d04775e799?t=1568929349000
- Office of Secretary of Defense. (2024). *OSD bias bounty*. <https://osdbiasbounty.com/sign-in?callbackUrl=https%3A%2F%2Fosdbiasbounty.com%2Fsign-in>
- Organization for the Economic Cooperation and Development (OCED). (2022, February 22). *OECD framework for the classification of AI systems*. https://www.oecd-ilibrary.org/science-and-technology/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en
- Pickar, C. (2020). *Learning from experience: Acquisition professional education for this century* (SYM-AM-20-070). Naval Postgraduate School, Acquisition Research Symposium. nps.edu
- Rapp, K. (2023, October 26). Infographic—Integrating cybersecurity with industry 4.0: What it means for manufacturing. National Institute of Standards and Technology. <https://www.nist.gov/blogs/manufacturing-innovation-blog/infographic-integrating-cybersecurity-industry-40-what-it-means>
- Raymond, N. (2024, January 29). *Lawyers voice opposition to 5th Circuit's proposed AI rule*. <https://www.reuters.com/legal/transactional/lawyers-voice-opposition-5th-circuits-proposed-ai-rule-2024-01-29/>
- Rogoway, T. (2024a, February 5). Drone warfare's terrifying AI-enabled next step is imminent. *The Warzone*. <https://www.twz.com/news-features/drone-warfares-terrifying-ai-enabled-next-step-is-imminent>
- Rogoway, T. (2024b, April 4). The compelling case for arming U.S. navy warships with drone swarms. *The Warzone*. <https://www.twz.com/sea/the-compelling-case-for-arming-u-s-navy-warships-with-drone-swarms>
- Service Employees International Union (SEIU). (1921). <https://www.seiu.org/about?fdonav>
- Sheffi, Y. (2023, September 15). The UAW and other unions must focus more on AI and automation in their negotiations. *Harvard Business Review*. <https://hbr.org/2023/09/the-uaw-and-other-unions-must-focus-more-on-ai-and-automation-in-their-negotiations>
- The Federal Privacy Council. (n.d.). *Fair information practice principles (FIPPs), roles in federal privacy, the Privacy Act of 1974, privacy impact assessments*. <https://www.fpc.gov/learn-about-federal-privacy-program/>



- The Internet in Real–Time. (2024, February 15). <https://visual.ly/community/Infographics/how/internet-real-time>
- Thomson Reuters. (n.d.-a). AI infographic. <https://tax.thomsonreuters.com/content/dam/ewp-m/documents/tax/en/pdf/infographics/7-things-professionals-need-to-know-about-ai-tr840725.pdf>
- Thomson Reuters. (n.d.-b). AI infographic. <https://www.thomsonreuters.com/en/artificial-intelligence.html>
- Toth, P. (2022, May 11). *Cybersecurity and industry 4.0—What you need to know*. National Institute of Standards and Technology (NIST). <https://www.nist.gov/blogs/manufacturing-innovation-blog/cybersecurity-and-industry-40-what-you-need-know>
- United Auto Workers (UAW). (1935). <https://uaw.org/about/>
- United Nations. (n.d.). *Member States*. <https://www.un.org/en/about-us/member-states>.
- United States Cyber Command (USCYBERCOM). (n.d.). *This is our code*. <https://www.cybercom.mil/>
- United States Cyber Command (USCYBERCOM). (2009, June 23). <https://www.cybercom.mil/About/History/>
- United States Department of Justice Overview of the Privacy Act of 1974, 2020 Edition. Overview of the Privacy Act of 1974. https://www.justice.gov/Overview_2020/dl?inline
- United States Department of Labor. (2023, September 1). *Worker organizing resource and knowledge center*. <https://www.workcenter.gov/>
- United States Department of Labor Blog. (2023, November 14). *Worker organizing resource and knowledge center*. <https://blog.dol.gov/2023/11/14/find-collective-bargaining-resources-at-the-work-center>
- United States Department of State. (2021, January 1). *National artificial intelligence initiative act of 2020 (NAIIA)*. <https://www.state.gov/artificial-intelligence/>
- United States Department of State. (2022, April 28). *Declaration for the future of the internet*. <https://www.state.gov/wp-content/uploads/2022/04/Declaration-for-the-Future-for-the-Internet.pdf>
- United Steelworkers (USW). (1942, May 22). <https://www.usw.org/union>
- White House. (2021, January 20). *Advancing racial equity and support for underserved* (Executive order 13985). <https://www.govinfo.gov/content/pkg/DCPD-202100054/pdf/DCPD-202100054.pdf>
- White House. (2021, January 25). *Executive order 14005, section 4(a): Ensuring the future is made in all of America by all of America’s workers*. <https://www.whitehouse.gov/briefingroom/presidential-actions/2021/01/25/executive-order-on-ensuring-the-future-is-made-in-all-of-america-by-all-of-americas-workers/>
- White House. (2022a, October 4). *Blueprint for an AI bill of rights*. Office of Science and Technology Policy. <https://www.whitehouse.gov/ostp/>
- White House. (2022b, October 12). *Biden-Harris administration’s national security strategy*. <https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government>



- White House. (2023a, February 16). *Executive order on further advancing racial equity and support for underserved communities through the federal government*. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/02/16/executive-order-on-further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>
- White House. (2023b, September 1). *Fact sheet: Ahead of labor day, Biden-Harris administration announces new actions to empower workers—Building on the president’s historic support for workers and unions*. <https://www.whitehouse.gov/briefing-room/statements-releases/2023/09/01/fact-sheet-ahead-of-labor-day-biden-harris-administration-announces-new-actions-to-empower-workers-building-on-the-presidents-historic-support-for-workers-and-unions/>
- White House. (2023c, October 30). *Blueprint for an AI bill of rights*. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- White House. (2023d, October 30). *Fact sheet: President Biden issues executive order on safe, secure, and trustworthy artificial intelligence* (Executive order 14110). <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>
- White House. (2024, January 29). *Fact Sheet: Biden-Harris administration announces key AI actions following President Biden’s landmark executive order*. <https://www.whitehouse.gov/briefing-room/statements-releases/2024/01/29/fact-sheet-biden-harris-administration-announces-key-ai-actions-following-president-bidens-landmark-executive-order/>
- World Economic Forum. (2023, May). *Future of jobs report 2023*. https://www3.weforum.org/docs/WEF_Future_of_Jobs_2023.pdf
- Young, S. D. (2024, March 28). *Memorandum for the heads of executive departments and agencies: Advancing governance, innovation, and risk management for agency use of artificial intelligence* (M-24-10). Office of Management and Budget. [M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf](https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf) (whitehouse.gov)
- Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, 3(5), 616–630. <https://www.sciencedirect.com/science/article/pii/S2095809917307130>





ACQUISITION RESEARCH PROGRAM
DEPARTMENT OF DEFENSE MANAGEMENT
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

WWW.ACQUISITIONRESEARCH.NET

