SYM-AM-16-020

PROCEEDINGS
OF THE
THIRTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 4–5, 2016

Published April 30, 2016

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

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

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

SYM-AM-16-020

PROCEEDINGS
OF THE
THIRTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research: Creating Synergy for Informed Change

May 4–5, 2016

Published April 30, 2016

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

Keynote: The Honorable Frank Kendall III, Under Secretary of Defense for Acquisition, Technology, & Logistics

The Honorable Frank Kendall III—was Senate confirmed in May 2012 and currently serves as the Under Secretary of Defense for Acquisition, Technology, and Logistics (AT&L). In this capacity, he is responsible to the Secretary of Defense for all matters pertaining to acquisition; research and engineering; developmental testing; contract administration; logistics and materiel readiness; installations and environment; operational energy; chemical, biological, and nuclear weapons; the acquisition workforce; and the defense industrial base. He is the leader of the Department of Defense's efforts to increase the Department's buying power and improve the performance of the defense acquisition enterprise. Prior to this appointment, from March 2010–May 2012, he served as the Principal Deputy Under Secretary and also as the Acting Under Secretary.

Kendall has over 40 years of experience in engineering, management, defense acquisition, and national security affairs in private industry, government, and the military. He has been a consultant to defense industry firms, non-profit research organizations, and the Department of Defense in the areas of strategic planning, engineering management, and technology assessment. Kendall was Vice President of Engineering for Raytheon Company, where he was responsible for management direction to the engineering functions throughout the company and for internal research and development. Before assuming his current position, Kendall was a Managing Partner at Renaissance Strategic Advisors, a Virginia-based aerospace and defense sector consulting firm.

Within government, Kendall held the position of Director of Tactical Warfare Programs in the Office of the Secretary of Defense and the position of Assistant Deputy Under Secretary of Defense for Strategic Defense Systems. Kendall is a former member of the Army Science Board and the Defense Intelligence Agency Science and Technology Advisory Board, and he has been a consultant to the Defense Science Board and a Senior Advisor to the Center for Strategic and International Studies. Kendall also spent 10 years on active duty with the Army, serving in Germany, teaching engineering at West Point, and holding research and development positions.

Kendall is an attorney and has been active in the field of human rights, working primarily on a pro bono basis. He has worked with Amnesty International USA, where he served as a member of the Board of Directors; with Human Rights First, for which he was an observer at Guantanamo; and with the Tahirih Justice Center, where he was Chair of the Board of Directors.

Over the course of his career as a public servant, Kendall was awarded the following federal civilian awards: Defense Distinguished Civilian Service Medal, Secretary of Defense Meritorious Civilian Service Medal, Presidential Rank Award of Distinguished Executive (Senior Executive Service), Presidential Rank Award of Meritorious Executive (Senior Executive Service), and Army Commander's Award for Civilian Service. He also holds the following military awards for his service in the U.S. Army: Meritorious Service Medal with oak leaf cluster, Army Commendation Medal, and National Defense Service Medal.

Kendall is a distinguished graduate of the U.S. Military Academy at West Point, and he holds a master's degree in aerospace engineering from the California Institute of Technology, a Master of Business Administration degree from the C.W. Post Center of Long Island University, and a Juris Doctor degree from Georgetown University Law Center.

THIS PAGE INTENTIONALLY LEFT BLANK

Panel 11. Enabling Successful Outcomes in Performance Based Logistics

Thursday, May 5, 2016	
9:30 a.m. – 11:00 a.m.	**Chair: Stan Soloway,** President and CEO, Celero Strategies, LLC ***Performance-Based Logistics: Examining the Successes and Challenges When Operating in Stressful Environments*** William Lucyshyn, Senior Research Scholar, Center for Public Policy and Private Enterprise, UMD John Rigilano, Faculty Research Assistant, Center for Public Policy and Private Enterprise, UMD Darya Safai, Graduate Research Associate, School of Public Policy, UMD ***Effective PBLs Through Simultaneous Optimization and Simulation of Maintenance, Manpower, and Spare Parts*** Justin Woulfe, Executive Vice President, Technical Services, Systecon North America Samantha Alpert, Analyst, Systecon North America ***Future Contracting for Availability*** Lou Kratz, Vice President and Managing Director, Logistics & Sustainment, Lockheed Martin Bradd Buckingham, Senior Market Research Planner, Logistics & Sustainment, Lockheed Martin

Stan Soloway—is President and CEO of Celero Strategies, LLC, a Washington-based consultancy focused on the federal market for technology and professional services. From January 2001 to December 2015, Soloway served as the President and CEO of the Professional Services Council, the largest national association of government technology and professional services firms. Prior to joining PSC, Stan served as the Deputy Under Secretary of Defense during the Clinton administration. In that position, he was responsible for wide-ranging reforms to defense acquisition and technology policy and practices, and broader department-wide reengineering. In recognition of his leadership in the Defense department, Stan was awarded both the Secretary of Defense Medal for Exceptional Public Service and the Secretary of Defense Medal for Distinguished Public Service. Earlier in his career, he was a public policy and public affairs consultant for nearly 20 years. He also co-produced the acclaimed PBS series *Great Confrontations at the Oxford Union*.

Soloway was the recipient of the 2016 CES Government Technology Leadership Award and in 2015, was inducted into the Greater Washington Government Contractor Hall of Fame. He also was named the IT Industry Executive of the Year in 2013 by Government Computer News. He is a two-time winner of the Federal 100 Award.

Performance-Based Logistics: Examining the Successes and Challenges When Operating in Stressful Environments[1]

William Lucyshyn—is a Research Director at the Defense Advanced Research Projects Agency (DARPA) and a Visiting Senior Research Scholar at the Center for Public Policy and Private Enterprise in the School of Public Affairs at the University of Maryland. In this position, he conducts research into the public policy challenges posed by the increasing role information technologies play in improving government operations and their relationships with the private sector. Previously, Lucyshyn served as a program manager and the principal technical advisor to the Director, DARPA, on the identification, selection, research, development, and prototype production of advanced technology projects. Prior to this appointment, Lucyshyn completed a distinguished 25-year career in the U.S. Air Force serving various operations, staff, and acquisition positions. Lucyshyn received his bachelor degree in Engineering Science from the City University of New York in 1971. In 1985 he earned his master's degree in Nuclear Engineering from the Air Force Institute of Technology. He was certified Level III as an Acquisition Professional in Program Management in 1994. [lucyshyn@umd.edu]

John Rigilano—is a Faculty Research Assistant at the Center for Public Policy and Private Enterprise, School of Public Policy, University of Maryland. He earned his Master of Public Policy degree from University of Maryland, College Park, in 2011, and holds a Bachelor of Arts degree in anthropology from Pennsylvania State University. [jprig@umd.edu]

Darya Safai—is a Graduate Student in the School of Public Policy working toward her Master of Public Policy with a specialization in Federal Acquisition. As a Graduate Research Associate, Darya researches a variety of acquisition and logistics topics. Her latest research examines the effectiveness of performance-based sustainment strategies for weapon systems deployed in operational environments, including reduced costs and increased innovation. [dsafai09@gmail.com]

Introduction

Given current and anticipated budgetary constraints, the Department of Defense (DoD) must heighten its focus on affordability, especially with regard to operation and maintenance costs, which account for almost two thirds of the defense budget. At the same time, new and evolving threats demand superior technology that is highly-reliable. To a large extent, these twin objectives—reduced costs and better performance—can be achieved through the wider implementation of performance-based logistics (PBL) contracting, a proven strategy to obtain economical and innovative support solutions. Unfortunately, however, PBL contracting is not being aggressively pursued across the DoD.

Under traditional sustainment strategies, the government customer purchases spares, repairs, tools, and data in individual transactions. In contrast, PBL transfers inventory management, technical support, and the supply chain function to a provider, typically a private-sector contractor, who guarantees a specified level of performance, often at a fixed price to the government. In effect, this arrangement aligns both parties' objectives. The contractor strives to reduce a system's "downtime" through cost-efficient maintenance and technical improvements.

---

[1] This is a summary of the full report that will be released in June 2016.

Yet, despite their success, the number of PBL-supported systems has declined. In 2005, there were more than 200 PBL contracts in place within the DoD, with spending on PBL projects having more than tripled since their inception—from $1.4 billion in 2001 to $5 billion in 2009. Yet by 2013, the number of PBL contracts had dropped to 87, while total DoD sustainment costs continued to increase (Irwin, 2013). We believe that while PBL may appeal to users and program officials from a theoretical standpoint, some may be reluctant to embrace this strategy for fear that PBL arrangements may falter when supported systems are deployed in emergency and contingency operations.

This perception manifests itself in a number of ways. For instance, some question whether the PBL mechanism is flexible enough to adapt to rapidly-changing conditions, that perhaps PBL works—until it doesn't. There is also concern over whether contractors will be able to perform at the same high level during emergency and contingency operations, especially if these providers are deployed in theater. To address the validity of these claims, we examine the performance of four PBL sustainment programs that have operated in stressful environments.

A 2007 Congressional Budget Office (CBO) report noted that during the conflicts in Iraq and Afghanistan, major weapons systems operated at rates that exceeded—sometimes by factors of five or six—their average operating rates during peacetime. Chief among these were combat vehicles and helicopters, systems that required the highest levels of repair and reconditioning. We examine four such systems: the Stryker Armored Combat Vehicle, the High Mobility Artillery Rocket System (HIMARS), the Apache AH-64 helicopter, and the H-60 helicopter.

Not only are these among the most deployed PBL-supported systems, they represent a diversity of PBL arrangements with regard to contract type, terms, and length. For instance, the H-60 program encompasses a suite of PBLs that cover different subsystems, while the Apache PBL covers one component, the fire control system. In both cases, the required maintenance is performed by military personnel, with contractors responsible for supply chain operations and parts requisition. In contrast, both the HIMARS and Stryker programs relied on contractors working in theater with military personnel to perform maintenance and repair.

It is noteworthy that these two programs reverted to the use of a traditional support strategy for some functions; the Stryker program now relies on soldiers, not contractors, for maintenance and repair, both at home and in theater, while HIMARS shifted to organic inventory management. Needless to say, these developments add to the perception that the PBL mechanism may not be a practical support solution for highly-deployed systems. We address this concern through a detailed examination of the history, attributes, and, performance of these PBL programs.

Stryker

The DoD's first new vehicle since the early 1990s, the M1126 Stryker Combat vehicle is a rapidly-deployable wheeled armored vehicle, combining mobility, survivability, and versatility in combat environments with firepower and reduced logistics requirements (Boyer et al., 2015). Its lightweight design allows for easy transport by C-130 aircraft anywhere in the world within 96 hours, making it an extremely desirable commodity in unpredictable combat situations (GAO, 2006).

The Stryker vehicle's acquisition process was among the fastest of any major army system. The urgent need for an innovative and rapidly-deployable (anywhere within 96 hours) armored vehicle to meet demands in Iraq and Afghanistan, where improvised explosive devices (IEDs) were quickly becoming the number one threat to U.S. troops, played a major role in this expedited process (Coryell, 2007). The Stryker PBL was first implemented in 2002; in 2007 and 2012, follow-on contracts valued at $1.5 billion and $2.5 billion, respectively, extended support for an increasing number of vehicles (McLeary, 2014). Given the Army's heavy reliance on Stryker during these conflicts, the use of PBL, if successful, would go a long way in legitimizing PBL as a leading support strategy for deployed systems.

Under the PBL arrangement initiated in May 2002, General Dynamics assumed responsibility for the ordering, management, and distribution of all spare parts, as well as provision of any and all vehicle maintenance services (Coryell, 2007). Contractor personnel performed an array of functions, including wheeled vehicle mechanics, armament repairer, or automated logistics specialists. On account of Stryker's short operational history and unknown vulnerabilities, the PBL relied on a cost plus fixed fee contract (Coryell, 2007).

In short order, however, General Dynamics was able to implement a number of design and process innovations, including an ability to self-sustain operations for up to 72 hours, an array of on-system repair enablers, and logistics surge capabilities (Coryell, 2007). Ultimately, these innovations served to minimize the number of personnel and parts needed within each Stryker brigade while ensuring that the vehicles were prepared for sudden increases in operational tempo.

The PBL contract required a monthly readiness rate of 90% during deployments. Stateside, a 98% monthly rate was used during training exercises and a rate of 90% was used in garrison. Stryker continuously exceeded expectations, achieving, for example, 95% cumulative readiness during the height of the war in Iraq—a war in which Stryker vehicles were driven in excess of 6.5 million miles (Coryell, 2007).

From a cost perspective, however, contract performance is less clear. In 2012, DoD Inspector General asserted that the follow-on contract's continued use of a sole metric (system readiness) in combination with a high-ceiling, cost-plus contract unduly incentivized the contractor to accumulate significant excess inventory valued at $335.9 million (DoD IG, 2012). The Army responded that the excess inventory could be attributed, in part, to contractor improvements in reliability, and that the spare parts would be used eventually, albeit at a slower pace than anticipated (DoD IG, 2012).

Given the Army's heavy reliance on Stryker during the Iraq War, changing operational tempos, and the lack of historical cost data, the use of a cost-plus fixed fee contract (as opposed to a fixed-price contract) was well-founded. However, it appears that the Army could have implemented better cost controls, perhaps by tying the fixed fee to an agreed-upon cost-per-mile metric.

In November 2005, citing a need for increased flexibility in different combat environments, the Army determined that soldiers, as opposed to contractors, would perform unscheduled maintenance for all Stryker vehicles (GAO, 2006). The Army's plan called for replacing 45 Stryker vehicle maintenance contractors with 71 soldiers. This transition relied on the Army's ability to annually recruit or retain 497 additional soldiers with specific military specialties to support all seven Stryker brigades (GAO, 2006). The GAO questioned the Army's plan, asserting, ironically, that the larger logistics footprint could negatively impact Stryker's deployment flexibility. In 2006, the Army began the transition, which, at present, is still underway.

HIMARS

HIMARS is a wheeled, agile rocket and guided missile launcher. The Army awarded the first HIMARS PBL contract to Lockheed Martin in the amount of $96 million in February 2004 (Lockheed Martin, 2004). Given its increasing inventory of HIMARS, the existence of a successful partnership between the Army and Lockheed Martin, and the cost benefits that derive from economies of scale, the Marines sought to support its launchers through the same PBL upon completion of the initial contract.

HIMARS has been deployed extensively since PBL implementation in 2004, playing a significant role in operations in the Al Anbar province of Iraq. In January 2016, Lockheed Martin announced that HIMARS had reached one million operational hours with U.S. forces (Lockheed Martin, 2016).

By 2011, Lockheed Martin was supporting 620 Army and Marines fielded mobile launcher systems—396 HIMARS and 224 MLRS M270A1. A third PBL contract in the amount of $158 million extended HIMARS sustainment through December 2013 for services, and through December 2014 for hardware. The PBL strategy relied on firm-fixed price with incentive fee contracts for stateside operations and cost-plus fixed fee contracts for overseas contingency operations (Gardener, 2008). This strategy provided strong cost reduction incentives as well as the flexibility to meet overseas contingency requirements.

The PBL required that system readiness be maintained at or above 90%, and that response time fall within a specified range a certain percentage of the time, depending on the nature of the problem. For overseas operations, the response time ranges were extended to provide the flexibility necessary to meet fluctuations in demand that might arise in unpredictable operating environments (DoD, 2006).

The program consistently achieved these objectives at the required percentages, with the relative simplicity of the performance requirements facilitating straightforward monitoring and, thus, complete transparency. The HIMARS PBL program achieved success early on, reaching a 99% average system readiness rate, with no launcher out of service for more than 24 hours (DoD, 2006). Since the program's inception, the PBL has consistently exceeded performance requirements.

Lockheed Martin relies on a database that tracks the location of each launcher, including each spare part, and indicates whether the part is functional. There are 26 field service representatives (FSRs) that operate from 22 locations, eight of which are overseas (Hawkins, 2009). In-theater maintenance work is performed by soldiers, with the assistance of FSRs. Early on in the PBL program, Lockheed Martin reduced the number of diagnostics devices provided to each battalion from six to one in order to streamline the repair process.

In addition, early improvements in processes and design modularity allowed soldiers operating HIMARS in the field to remove and replace defective components quickly and easily. Perhaps one of the greatest benefits of the PBL is the provision of limited depot-level repair capability at each battalion, where repair work is provided by the FSR. Referred to as the capability to "Fix Forward," some 50% of HIMARS repairs are performed on location by the FSRs, eliminating wait times and significantly reducing costs (Hawkins, 2009). This in-the-field repair capability has also significantly improved deployed launcher availability. According to interviews with Lockheed Martin officials, FSRs voiced few concerns over their work environments, safety, or civilian status within the battalion, with several volunteering to return.

In 2015, the DoD transitioned inventory management from the contractor to the government in an effort to further reduce costs. It remains to be seen whether the DoD's decision will lead to lower costs and continued high performance.

AH-64 Apache

The AH-64 Apache was conceptualized as a high-powered, tank-killing attack helicopter, capable of repelling conventionally ground forces during a Soviet invasion of Europe. Still an essential part of the Army's fleet today, the primary mission of the Apache is to perform armed reconnaissance and conduct rear, close, and shaping missions, including deep precision strikes.

Since its inception, the Apache has accumulated over 3.9 million flight hours, with operational deployments during Desert Storm, Operation Iraqi Freedom, Operation Enduring Freedom, and Operation Inherent Resolve in Iraq. Central to the Apache's mission is the Modernized Target Acquisition Designation Sight/Pilot Night Vision Sensor (M-TADS/PNVS) system, nicknamed the "eye of the Apache." The system, consisting of two subsystems, enables Apache pilots to fly at low altitudes in total darkness and poor weather conditions, while also providing the capability for the co-pilot to identify and engage hostile targets (Cothran, 2012).

Since 2007, Lockheed Martin has provided sustainment for the AH-64 Apache Helicopter's M-TADS/PNVS system. The Apache sensors PBL relies on a firm-fixed price contract that is tied to the number of flight hours. This structure is ideally suited to heavily-deployed systems, such as the Apache, in that it provides the contractor with the traditional incentives associated with fixed-price contracts, translating to higher levels of innovation, reliability, and availability; at the same time, the contract is flexible, which ensures that the system is capable of supporting changes in operational tempo without unduly impacting tactics and strategy. The first four-year contract was valued at approximately $380 million; in 2012, a similar follow-on contract valued at $375 million was awarded (Lockheed Martin, 2012).

Lockheed has consistently achieved a supply availability rate of approximately 97%. The contract established a system of continuous improvements supporting the Apache sensors and covered complete post-production supply chain management, including inventory management, maintenance, modifications, procurement, repairs, and spares planning of fielded systems.

Under the initial contract, Lockheed successfully slashed sustainment costs for both sensor systems and improved supply availability primarily through improvements in supply chain and obsolescence management. Lockheed has lowered logistics and maintenance costs by leveraging data tracking for a number of health and maintenance indicators to improve demand forecasting, by determining appropriate inventory levels, and by ensuring the optimal locations of supply activities.

Other achievements include a monthly minimum supply availability rate of 96%, a 99% availability rate for depot repair parts, and material reliability improvements, leading to a 70% increase in Mean Time Between Failures. The PBL contract has also been credited with improving fleet readiness, reducing average flying hour cost and reducing the Army's long-term inventory investment. Over the course of the initial PBL contract, depot level repairable costs were reduced by 18%, supply inventory replenishment costs were reduced by 40%, and mean-time between maintenance actions was reduced by 9.6% (DoD, 2012).

Annual sustainment costs prior to the implementation of PBL totaled $218 million per year. In 2013, costs totaled $92 million, a drop of 58%. Other accomplishments include the mitigation of 759 obsolescence and diminishing manufacturing cases since 2007, resulting in $104.2 million in cost avoidance, the reduction of the maintenance support footprint, and a decrease of over 1,000 maintenance man hours per year through increased materiel reliability (DoD, 2012).

H-60

The H-60 is the U.S. Navy's family of multipurpose twin-engine, medium-lift helicopters—the legacy SH-60B, SH-60F, HH-60H, and the new MH-60R and MH-60S. These aircraft share upgraded mission systems, avionics, and components, including a common cockpit that allows pilots to shift from one aircraft to another with minimal retraining. The MH-60R, first deployed in 2009 to aid in Operation Iraqi Freedom, is a multi-purpose aircraft with many missions, including vertical replenishment, search and rescue, special operations support, and mine countermeasures, though its primary mission is anti-submarine and surface warfare.

The provision of traditional support for the H-60 was complicated by the number of versions that were in service, the length of their service, and the introduction of the two new models. The high operational tempo of the aircraft, combined with the unique challenges of maintaining rotary wing aircraft (e.g., the corrosive effects of maritime operations), led to increasing operating and support costs and lower availability (Heron, 2010). Obsolete parts represented an additional problem. Procurement necessitated the repair of small batches of custom-made parts, often at high cost, or undertaking expensive engineering changes to the aircraft to enable the use of newer parts.

A 1996 GAO report noted that one specific part "had a repair time of 232 hours, only 20 hours of which was spent actually repairing the item, [and that] the remaining 212 hours involved time to handle and move the part to different locations." In 2002, the Navy sought a new product support strategy. To this day, the Navy relies on a suite of fixed-price PBL contracts that, in effect, cover maintenance and repair of the entire aircraft. The largest of these, the Tip-to-Tail (T2T) PBL program, supports over 1,200 parts (avionics and airframe).

The original five-year T2T contract covered legacy models. It was awarded to a joint venture between Lockheed Martin Systems Integration (LMSI) and Sikorsky Aircraft Company in December 2003. Valued at approximately $417 million, the PBL provided requisition processing, requirements forecasting, inventory management, repair, overhaul, modification, packaging, handling, storage, transportation, configuration and obsolescence management, and reliability and technology improvement (Lockheed Martin, 2004). In order to capture contractor performance in the provision of these tasks, the PBL relied on a single metric, fill rate, which measures the percentage of requisitions filled on time.

Following the expiration of the initial contract, the Navy sought to use contractor cost data in order to develop the basis for the fixed-price follow-on contract. Following a series of challenging negotiations, the contractor supplied the data and in December 2010, the T2T was renewed for four years at an estimated five-year projected cost of $1.4 billion (DoD, 2010). As the price increase indicates, the contract was expanded to cover the newer models, the MH-60 R and S.

Since its implementation, the PBL has exceeded the established minimum fill rate (80%), averaging a rate of 88% (19% above the pre-PBL rate). Furthermore, the fill rate for special management items has also increased, from 80% to 99%. In addition, backorders were reduced by over 90% (Skotty, 2012).

Findings

In order to ensure the nation's continuing technological superiority and better prepare for the rapidly changing global environment, the DoD must strive to reduce life cycle costs, improve system availability, and incentivize innovation.

Based on our examination of the PBL mechanism, its proven applications, and four PBL-supported systems, we provide our findings.

1. **PBL-supported systems operating in stressful environments are capable of meeting or exceeding performance requirements, contributing to mission success.**

   In all four cases, the PBL programs met or exceeded performance requirements in operational availability and readiness. In light of new and emerging threats, a program's proven ability to consistently meet high performance standards in excess of 90, 95, or 99% availability/readiness cannot be overstated.

2. **PBL contractors have the proven ability to support weapons systems operating in stressful environments.**

   Over the last several years, there has been unprecedented contractor participation (in numerous roles) in the conflicts in Iraq and Afghanistan, with some voicing concern over the presence of "contractors on the battlefield." Needless to say, a line must be drawn between contractor support and direct participation in combat operations, but as these cases illustrate, this has not been an obstacle. The four cases also demonstrate that PBL contractors are willing and able to perform a critical supporting role, even in stressful environments.

3. **PBL provides sufficient flexibility and capacity to adapt to changing operational tempos.**

   The four cases suggest that PBL programs are adaptable and scalable, provided that they are structured appropriately. PBLs relying on cost-plus contracts provide inherent flexibility (to the government and the contractor) in the face of uncertainty, both technical and operational. Fixed-price PBLs also provide flexibility, especially when price is tied to operational tempo (number of flight hours, miles driven).

4. **All support contracts, including PBLs operating in theater, should apply stringent cost controls.**

   Owing in part to demonstrated success of PBL in meeting performance requirements, it may be that less attention is paid to contract specifics beyond readiness/availability metrics. Carefully-considered contract ceilings, cost-per-unit usage rates, and logistics footprint constraints should be included in cost-plus contracts. Without these features, contractors may be incentivized to accrue surplus inventory beyond what is necessary to meet the performance requirement.

Recommendations

Based on these findings, we provide the following recommendations to the DoD.

1. **Promote the use of PBL as a proven support strategy for weapons systems.**

   PBLs perform better than traditional support mechanisms, even in stressful environments. The DoD should renew its commitment to the expansion of PBL in order to improve weapons systems operation and reduce costs.

2. **Ensure proper alignment of government objectives with provider incentives.**

   Critics suggest, perhaps rightly, that PBL arrangements can be more challenging to develop and manage than other contract types. Just as an appropriate PBL program structure aligns the incentives of the customer (the government) and the support provider, leading to a win-win scenario, an inappropriate structure can create perverse incentives and result in undesired or unintended consequences.

3. **Structure PBL contracts appropriately.**

   In environments characterized by relatively low levels of uncertainty, both operational and technical, alignment of contractor and government objectives is optimized under fixed-price PBL contracts. These arrangements promote greater cost-reduction incentives, higher levels of innovation, and enhanced reliability. Often, these contacts rely on only one or two performance metrics, which ensures transparency and accountability. However, in stressful, unpredictable environments, cost-plus PBL contracts are often more suitable in that they provide greater flexibility to meet mission objectives. In these circumstances, however, programs may need to employ additional metrics beyond reliability and availability, including cost-per-unit usage rates and logistics footprint constraints, in order to strike the optimal balance between required availability and cost.

4. **Avoid distortions to the PBL paradigm.**

   From a theoretical standpoint, the power of PBL lies in affording the provider the discretion and flexibility to select the optimal mix of inventory levels, maintenance activities, and technology upgrades in order to meet performance requirements. Shifting one or more of these functions to the government customer distorts the PBL paradigm and may lead to reductions in performance and higher costs.

References

Boyer, M., Shurkin, M., Wong, J., Schwankhart, R., Albrich, A., Lewis, M., & Pernin, C. (2015). *Assessing conventional army demands and requirements for ultra-light tactical mobility.* Santa Monica, CA: RAND.

CBO. (2007). *Replacing and repairing equipment used in Iraq and Afghanistan: The Army's reset program*. Retrieved from https://www.cbo.gov/sites/default/files/110th-congress-2007-2008/reports/09-13-armyreset.pdf

Coryell, B. (2007). *Performance-based logistics, contractor logistics support, and Stryker* (Master's thesis). Fort Leavenworth, KS: U.S. Army Command and General Staff College.

Cothran, J. (2012, July). Lo*ckheed Martin PBL programs description and data for CAPE analysis.* Corporate Logistics and Sustainment, Lockheed Martin Corporation.

DoD. (2006). *The Secretary of Defense PBL Awards Program for Excellence in Performance-Based Logistics, summary of criteria accomplishments: HIMARS*.

DoD. (2012). *The Secretary of Defense PBL Awards Program for Excellence in Performance-Based Logistics, summary of criteria accomplishments: Apache M-TADS/PNVS.*

DoD Inspector General (DoD IG). (2012). *Better cost-control measures are needed on the Army's cost-reimbursable services contract for logistics support of Stryker Vehicles* (DoDIG-2012-102). Washington, DC: Author.

GAO. (2006). *Defense logistics: Changes to Stryker vehicle maintenance support should identify strategies for addressing implementation challenges* (GAO-06-928R). Washington, DC: Author.

Gardener, C. (2008). *Balancing government risks with contractor incentives in performance-based logistics contracts* (Master's thesis). Wright Patterson Air Force Base, OH: Air Force Institute of Technology.

Hawkins, K. (2009). HIMARS shoots high for award. Retrieved from U.S. Army website: http://www.army.mil/article/31909/HIMARS_Shoots_High_For_Award/

Heron, J. (2010, November 18). PowerPoint presentation at Defense Acquisition University.

Irwin, S. (2013). Industry and government butt heads over weapons maintenance contracts. *National Defense*.

Lockheed Martin. (2004, January 8). Sikorsky venture wins $417 million contract [Press release].

Lockheed Martin. (2012). U.S. Army awards Lockheed Martin $111 million Apache M-TADS/PNVS performance based logistics contract. [Press release]. Retrieved from http://www.lockheed.com/m/us/news/press-releases/2012/september/mfc-092412-usarmyawards.html

Lockheed Martin. (2016, January 12). HIMARS achieves one million operational hours milestone [Press release].

Skotty, R. (2012). PowerPoint presentation at Defense Acquisition University.

Effective PBLs Through Simultaneous Optimization and Simulation of Maintenance, Manpower, and Spare Parts

Justin Woulfe—is Executive Vice President, Technical Services, at Systecon North America. [jwoulfe@systecon.us]

Samantha Alpert—is an Analyst at Systecon North America. [salpert@systecon.us]

Abstract

The problem of determining the optimum repair strategy is sometimes called location-of-repair analysis, or LORA, while OPRAL is an analytical model for determining the optimal repair locations and spares allocations in a multi-level hierarchical support organization based on the spares optimization model OPUS10. The OPRAL optimization technique is based around the powerful concept of calculating the maximum function over all convex functions created from several different repair strategies in order to find the optimal one. The process of simultaneously optimizing LORA analysis, spare parts optimization and resource utilization is significant, hence why it is necessary to integrate the optimization techniques. These techniques can be instrumental in setting up and managing risk in Performance Based Logistics (PBL) contracts where a Product Support Manager (PSM) is responsible for a high level metric such as system availability, mission hours accomplished, etc., where optimizing all aspects of a system's operation is critical.

Introduction

Determining the optimal level of spares has been a use of the optimization software OPUS Suite (which consists of OPUS10, SIMLOX, and CATLOC) since the 1970s. Developed in Stockholm, Sweden, the software is now used to optimize spares in numerous countries for a wide span of projects that span across the commercial and defense sectors. However, the problem of determining the optimal repair strategy, called level-of-repair analysis or LORA, has become more prevalent, and as such, the ever-improving software had to accommodate.

The LORA calculation discussed in this paper is performed with the OPUS10 tool. OPUS10 contains an advanced LORA capability specifically created to optimize both spares and repair capabilities for Performance Based Logistics (PBL). This calculation is performed using an algorithm called OPRAL.

Background on OPRAL

The theory of the OPRAL optimization technique is based around the powerful concept of convexification. The convexification of a function f (x) is defined as the maximum over all convex functions g(x) such that g(x) ≤ f (x) for all x. For some values of x, it holds that f c (x) = f (x), that is, the function coincides with its convexification. We refer to these x as convex points (for the function f). In other words, convexification is the idea of finding the optimal curve from a group of curves. The OPRAL algorithm optimizes just like OPUS10, but instead of finding the optimal function from a large set of possible points, it takes the optimal curves of the different LORA candidates and finds a single optimal function.

As shown in Figure 1, the C/E-curves for different resource groups are combined to find the total C/E-curve. As in OPUS10, this curve represents maximum support system effectiveness when allocating a certain cost to the support system.

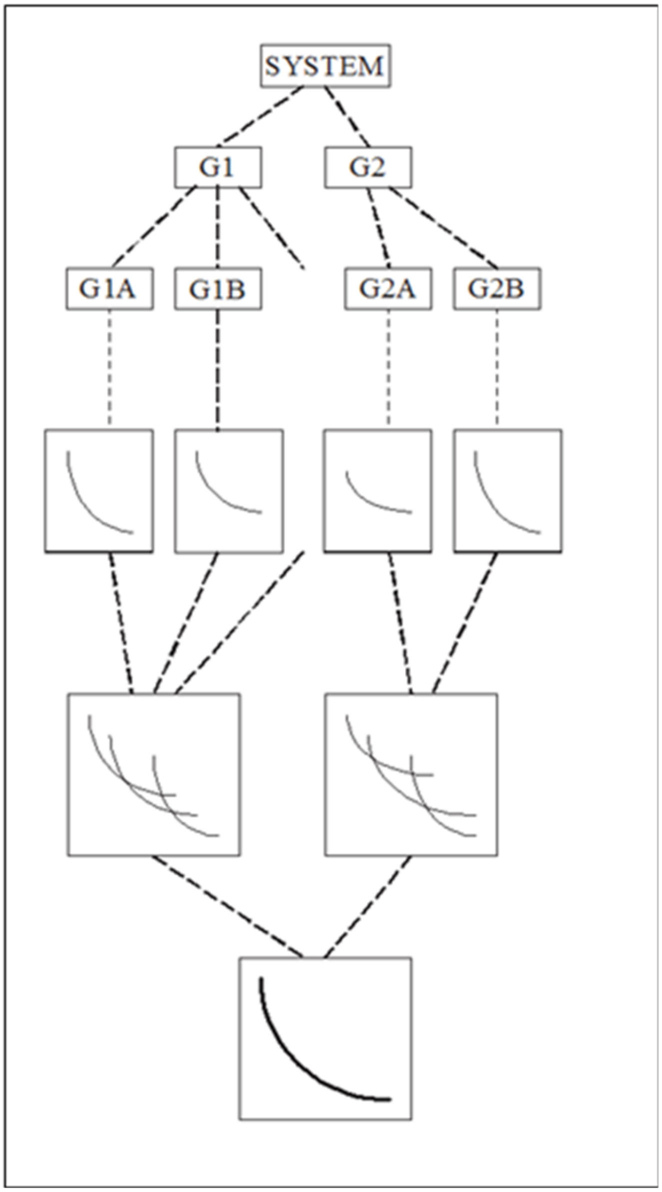


Figure 1. A Graphical Depiction of Convexification

Note. G1 and G2 represent subproblems for each resource group. A and B represent feasible resource allocations alternatives for each subproblem. The curves below represent the optimal curves for each allocation alternative, and the final curve is the optimal of all the previous curves.

When thinking about a LORA model, the highest level questions are

- What repair strategy should be used for items of a given type?
- What sparing strategy should be used for items of a given type?

The choice of repair strategy concerns whether to discard or repair faulty items of a given type. Furthermore, if the item is to be repaired, it also concerns where the repair should take place.

One of the first techniques to approach this problem was the METRIC model from 1968 (Sherbrooke, 1968). Independently, yet concurrently, the first OPUS model was derived.

The step of marrying LORA and spare parts optimization is significant, which feeds into the next questions to be answered:

- If a given item fails, should it be repaired or discarded?
- If an item is to be repaired, where should the repair take place?

The correct answer to these questions depends on several things, for example, the cost of necessary resources to repair the item and the unit price of the item. Another issue that makes determining the optimal repair strategy difficult is the interrelation with sparing. The accessibility of spares will have an impact on how critical repair turn-around times will be. With large spare part stocks, we can allow longer repair turn-around times than with smaller stocks. Similarly, shorter repair turn-arounds will decrease the amount of spares required to reach a specified service performance.

The above-mentioned issues (that is, that repair decisions for different items) can be dependent on common expensive repair resources, and the relationship between sparing and repair complicates matters. Therefore, in the past, these issues have been ignored by traditional techniques and tools. However, a model aimed to accurately describe the real-world aspects of the problem must properly address them.

Performance Based Logistics

Because the OPUS Suite can simultaneously optimize manpower, spares and support and test equipment, and also simulate mission effectiveness of the optimal solution, the opportunity to effectively dimension and manage Performance Based Logistics, or PBL, contracts is significant.

Performance Based Logistics represents a potentially cost effective method for system sustainment. From the customer perspective, PBL means a shift away from buying parts to instead buying performance from the supplier. We can apply this concept at the system, subsystem, or major assembly level. A key element in PBL is the ability to measure the system performance in a well-defined way, either directly, like availability, or indirectly by measuring given logistic parameters, for example, backorders. Monitoring and following up logistic parameters in the supply chain can on its own be a driver for supply chain performance improvements. Applied correctly and tailored to the specific scenario, that potential is substantial. But as many Program Managers and Logisticians have experienced, setting up a PBL contract is a complex task. More importantly, if inadequately written, the outcome may be the opposite: increased costs and risks for government, contractor or both.

There are several success factors that can be realized through modeling and simulation as described by Olinger, Hell, and Wijk (2011):

- Success factor 1—A common pitfall in PBL contract design is that the supplier scope is not clearly defined and that the distinction between supplier and customer responsibilities is imprecise. A weak definition of this basic foundation of the contract can be detrimental and cause discussions and disagreements about what is included and not. It can also lead to the defined KPIs not corresponding to the actual interpretation of the contract scope.
- Success factor 2—Appropriate performance parameters (KPIs). The KPIs must be selected based on the nature and scope of the contract and give the customer performance, affordability and control. On the other side, KPIs must

give the contractor direction and incentive, but also maneuverability to build, adapt and manage the solution in the most cost effective way. To allow for the latter, a small number of well selected KPIs are preferable to many. It is a common mistake to try to compensate uncertainty with a long array of KPIs which are at best redundant and at worst conflicting and counterproductive.

- Success factor 3—Appropriate KPI target levels. It is crucial to understand the consequences of setting a certain target level in advance. For example, a target for average availability may seem acceptable if only considering a steady state situation, but can mean unacceptable sensitivity to changes or poor ability to handle peak loads. Meanwhile, a too high target typically escalates costs.
- Success factor 4—A clear and relevant incentive model. All involved should win when performance is on or above target, and, the very driving force of PBL, the revenue for the contractor must drop significantly when performing below target. The approach can be either penalties or rewards.
- Success factor 5—Performance measurement approach and intervals. The way performance is measured and calculated, and how often it is measured, can have a large impact on the outcome. Too long measuring intervals could for example mean that unsatisfactory performance over important periods can be averaged out by over-performing during the rest of the time. Too short intervals could mean that the contractor does not get enough time to adjust and correct deficiencies; hence the incentive to improve is lost.

Understanding the consequences of a PBL contract in advance, and the potential benefits, risks and costs involved, is equally important to customer and contractor. Design, evaluation and ultimately the negotiation of the terms in the contract should be based on thorough analysis by both parties.

Optimization in OPUS10 coupled with Monte Carlo simulation in a tool like SIMLOX can be used to design an effective incentive model and to set the performance levels and suitable measurement intervals, all based on proper decision support, mission understanding and consequence analyses.

A key element in PBL is the ability to measure the system performance in a well-defined way, either directly (e.g., availability) or indirectly by measuring given logistic parameters (e.g., backorders). Monitoring and following up logistic parameters in the supply chain can on its own be a driver for supply chain performance improvements.

The degree of PBL contract fulfillment has been shown to be able to be defined using a penalty function y(x), where y is the share (%) of the maximum penalty amount and x can be any logistics parameter of interest to mission capability. The parameter x is measured as an average over a time period T.

Using simulation, the time period T will influence the design of the penalty function y(x). In many cases, the backorder measure B is used for designing the penalty function Y(B), but the same approach can be used for any other logistics parameter. In fact, y(x) could be multidimensional (i.e., x being a vector of several types of logistics parameters).

Appropriate results collected from Monte Carlo simulations enable evaluation of alternative penalty (or reward) functions suggested in a PBL contract negotiation. Using these methods, a penalty function y(x) should be designed to meet the customer and supplier objectives in a satisfactory way for both parties.

It is important to consider different operational scenarios and the potential effects of the penalty function (e.g., mission success, mission readiness, and operational effectiveness) when designing the y(x) function. Typically, the penalty increases in steps if performance drops below target. Each step in y(x) should be simulated to demonstrate the capability impacts (positive and negative) of different outcomes. While requiring a thorough understanding of mission profiles, operational scenarios, and definitions of "success," this methodology allows both the customer and the supplier to make rational decisions and agree on a reward (or penalty) function commensurate with the relative impact of each y(x) step on overall mission readiness and mission capability. Simulation of mission readiness and capability instead of availability provide a y(x) function that aligns with operational realities and ensures cost effective capability to the warfighter.

LORA Example

When conducting a LORA, there are many factors to be considered. The measures of effectiveness are affected by the resources, spares, manpower, transportation, and much more. Using the tool OPUS10, we can measure how much each of your options for repair will change your results.

We connect each repair action at each possible location to a repair task. Those tasks can be broken into subtasks that can split the total repair time into specific steps. This is helpful when connecting the resources to those subtasks, causing the resources to only be used for the specific parts of the repair and not the entire repair turnaround time. The tasks are then divided by their complexity into task levels.

The task levels are used to create all of the different maintenance candidates. The maintenance candidates, or scenarios, are the possible combinations of places that will complete the task levels. All of these connections are shown graphically in Figure 2. OPUS10 optimizes each of the candidates as if it was its own model, then finds the optimal curve from the group. The result will show the most optimal candidate as well as the number of resources and spares required, where they should be, and how much they will all cost.

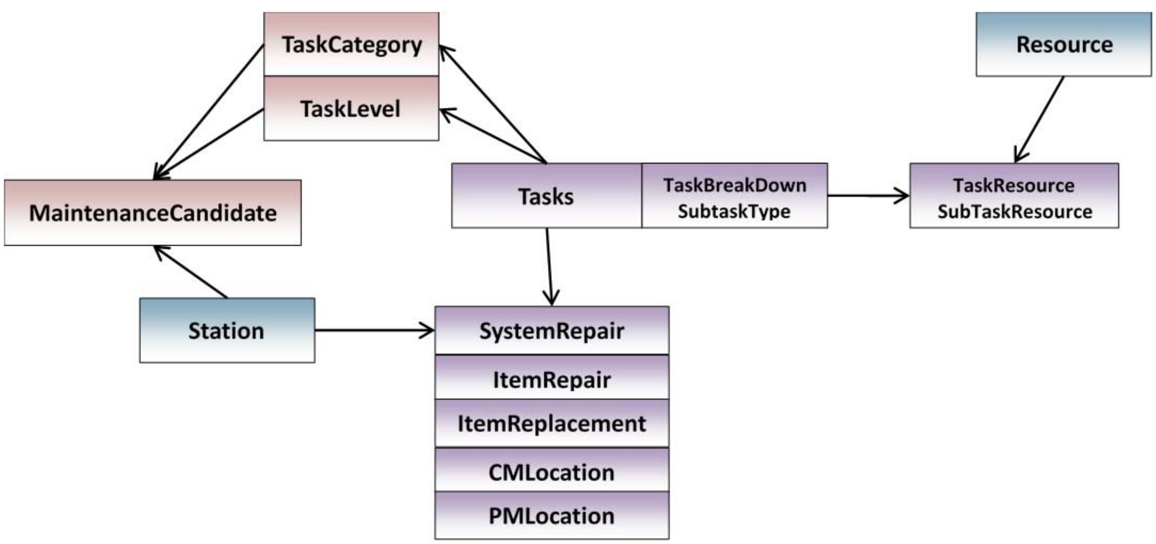


Figure 2. Connection Between Tables Built in OPUS10 to Perform LORA Calculation

Using the results of the OPUS10 calculation, we can make better decisions based on numerous metrics like availability of the system, total life support cost, investment costs, and more. Based on the different PBLs the analyst is optimizing for, there are many different reports that can be made from the results; some examples can be seen in Figure 3.

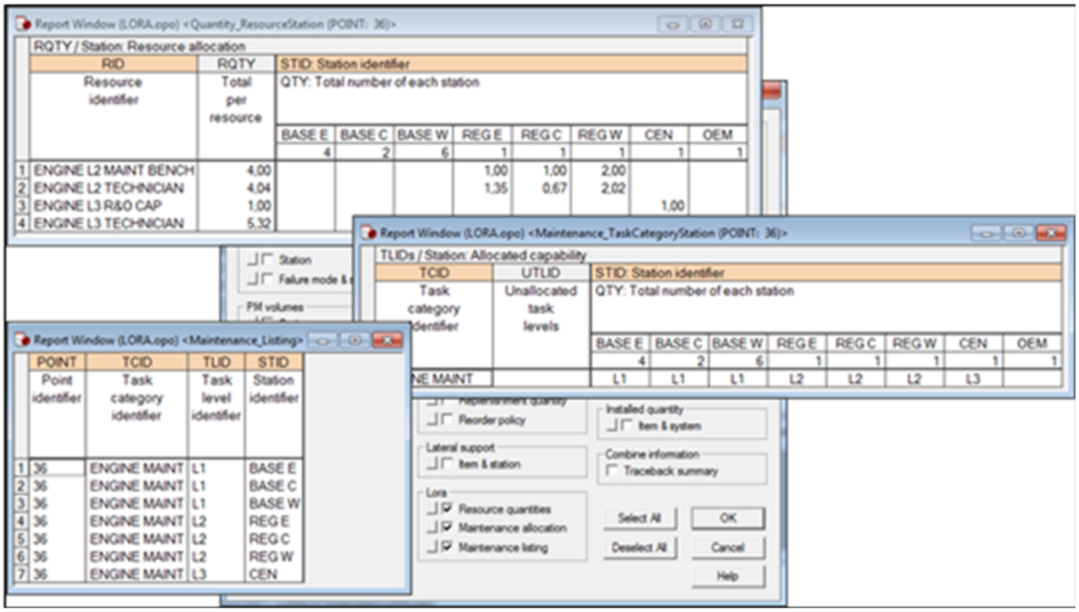


Figure 3. Examples of Reports Created From OPUS10 Calculations

Conclusion

To avoid a high cost with low effectiveness of a project, analytics must be performed. When a proper support structure has not been established, or the current structure has been shown to be suboptimal, a LORA must be conducted to assist with the decisions being made. In addition, the spares optimization cannot be ignored, nor should it be calculated separately. Combining these variables is key to truly optimal results.

References

Alfredsson. P. (1997). *On the optimization of support systems* (Doctoral thesis, KTH). Stockholm, Sweden.

Feeney, G. J., & Sherbrooke C. C. (1966). The (S-1,S) inventory policy under compound Poisson demand. *Management Science, 12,* 391–411.

Graves, S. C. (1985). A multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Science, 31*, 1247–1256.

Hadley, G., & Whitin, T. M. (1963). *Analysis of inventory systems*. Englewood Cliffs, NJ: Prentice-Hall.

Mahon, D. (2007). Performance-based logistics: Transforming sustainment. *Journal of Contract Management, 5*(1), 53–71.

Muckstadt, J. A. (1973). A model for a multi-item, multi-echelon, multi-indenture inventory system. *Management Science, 20*, 472–481.

OPUS10–Logistic support and spares optimization. (n.d.). Retrieved from http://www.systecon.se/

Palm, C. (1938). Analysis of the Erlang traffic formula for busy-signal arrangements. *Ericsson Technics, 5,* 39–58.

Sherbrooke, C. C. (1968). METRIC: A multi-echelon technique for recoverable item control. *Operations Research, 16,* 122–141.

Sherbrooke, C. C. (1986). VARI-METRIC: Improved approximation for multi-indenture, multi-echelon availability models. *Operations Research, 34*, 311–319.

Sherbrooke, C. C. (1992). *Optimal inventory modeling of systems: Multi-echelon techniques.* New York, NY: John Wiley & Sons.

SIMLOX—Simulation of logistics and operations. (n.d.). Retrieved from http://www.systecon.se/

Simon, R. M. (1971). Stationary properties of a two-echelon inventory model for low-demand items. *Operations Research, 19*, 761–773.

Systecon AB. (1998). *OPUS10 user's reference*. Stockholm, Sweden.

van Hoek, R. I. (1998). Measuring the unmeasurable—Measuring and improving performance in the supply chain. *Supply Chain Management, 3*(4), 187–191.

Wijk, O., Olinger, T., & Hell, R. (2011). Simulation as support for decision making in negotiations. *14th Annual System Engineering Conference.* San Diego, CA. Retrieved from www.systecon.se/documents/Simulation_as_support.pdf

Future Contracting for Availability

Lou Kratz—is the Vice President and Managing Director, Logistics and Sustainment, Lockheed Martin Corporation. Kratz is responsible for coordinating Lockheed Martin's logistics and weapon system sustainment efforts. Kratz leads Lockheed Martin's Automatic Identification Technology implementation, including RFID and UID. After successfully completing eight pilot projects, Kratz is now guiding Lockheed Martin's enterprise implementation of RFID. He also guides Lockheed Martin's logistics strategic planning, performance-based logistics efforts, logistics technology development, logistics human capital development, and cross-corporate logistics business initiatives. Previously, Kratz served as the Assistant Deputy Under Secretary of Defense (Logistics Plans and Programs), within the Office of the Deputy Under Secretary of Defense (Logistics and Materiel Readiness). As such, he was responsible for guiding the DoD's logistics transformation to meet the operational requirements of the 21st century. Kratz oversaw the DoD's long-range logistics planning to meet the requirements of the Quadrennial Defense Review (QDR) and Joint Vision 2020. [louis.kratz@lmco.com]

Bradd Buckingham—is the Senior Market Research Planner, Logistics & Sustainment, Lockheed Martin. Buckingham holds a Bachelor of Liberal Studies degree in Conflict, Politics, and National Policy from the University of Mary Washington. Buckingham currently provides research and analytic support in theater opening and sustainment, and in support of Army, USMC, and Defense Logistics Agency Depot management. His prior experience includes supporting the Army Manufacturing and Technology Program at Fort Belvoir, VA, as well as providing support to the Joint Defense Manufacturing and Technology Program's (JDMTP) Warfighter Brochure, and as a subject-matter expert for Army ManTech Small Business Innovative Research (SBIR) program. Buckingham was also the acting Army representative for the JDMTP SBIR Working Group. [bradd.a.buckingham@lmco.com]

Abstract

The United States faces unprecedented national security threats in an environment of continued federal budget limitations. The U.S. military must modernize its force to deter near-peer competitors and unstable states, while maintaining high readiness to deter and defeat extreme violent organizations. These factors put significant pressure on research, development, and procurement accounts to field critically needed capabilities in a time of overwhelming demands on resources.

These challenges are not unique to the United States. Many of our allies, faced with these same defense modernization and readiness issues, created new public–private partnerships through the implementation of Outcomes Based Service Contracting (OBSC). Under the outcomes based model, a customer (Defense) contracts and pays for business results delivered by a service provider (industry), rather than for defined activities, tasks, or assets. These types of contracts focus on the outcomes rather than piece parts or the manner in which the service is provided.

This paper explores the fundamental business decisions needed to identify opportunities that will allow the DoD to concentrate on its core competencies of deterrence and national defense. By buying outcomes versus equipment and services, the greater utilization of Outcomes Based Service Contracting will ensure readiness and modernization.

Introduction

The United States is facing significant economic challenges, as evidenced by its existing $18 trillion debt, rising entitlement expenditures, and increased national security needs. Despite recent calls by political leaders and industry to increase defense spending, the fundamental economic reality is that additional spending of any kind would merely add further to the national debt.

National Debt Interest Payments

The U.S. Federal Government debt is currently $18.1 trillion, with projected increases to the national debt for the foreseeable future. If interest rates go up, so does the cost of servicing both new debt and debt that is rolled over in the form of Treasury securities. With rising interest rates and expected increases in the Federal debt, at some point in the next 10 years annual interest payments are on pace to exceed the U.S. defense budget, as shown in Figure 1.

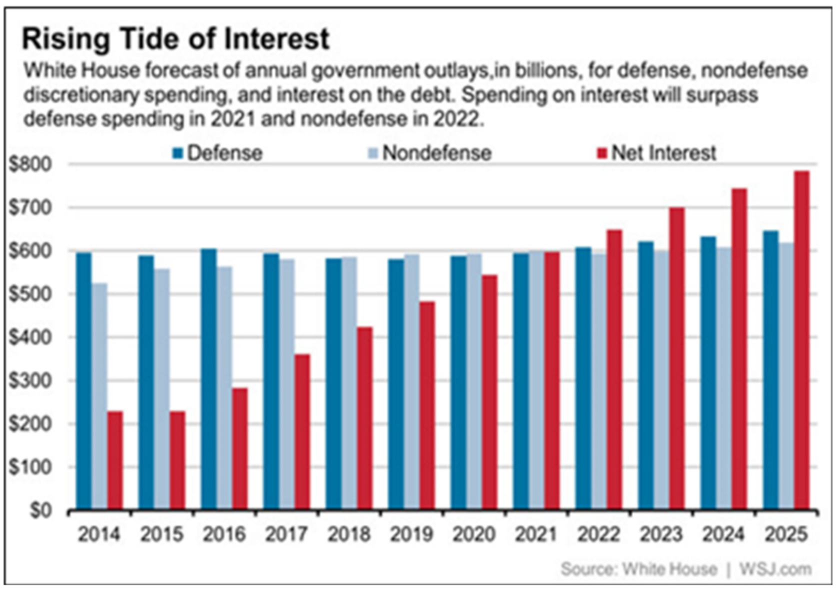


Figure 1. National Debt Payments vs. Defense Spending
(Zumbrun, 2015)

Entitlement Spending

The United States faces rising costs for its social welfare system as the population continues to age. Unless retirement and healthcare entitlement expenditures are reduced, these programs will generate enormous spending pressures, making it more difficult to support other national needs, as shown in Figure 2.

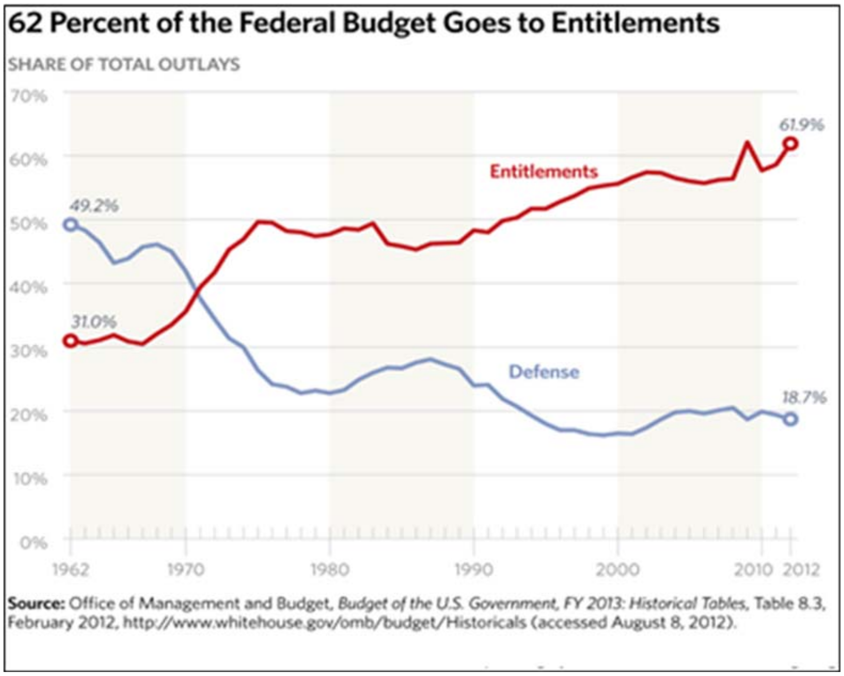


Figure 2. Entitlement Spending
(Office of Management and Budget, 2012)

The Department of Defense (DoD) is grappling with the drawdown from two wars and associated reset requirements, budget uncertainties, and program complexity.

One of the DoD's pressing concerns is how to get the most out of its sustainment funding while maintaining required weapons systems performance. Every dollar spent on Operations and Maintenance (O&M) and growing personnel costs reduces the resources available for required acquisition programs. As a result, the U.S. military must find innovative and practical solutions to modernize its force and maintain high readiness, as shown in Figure 3.

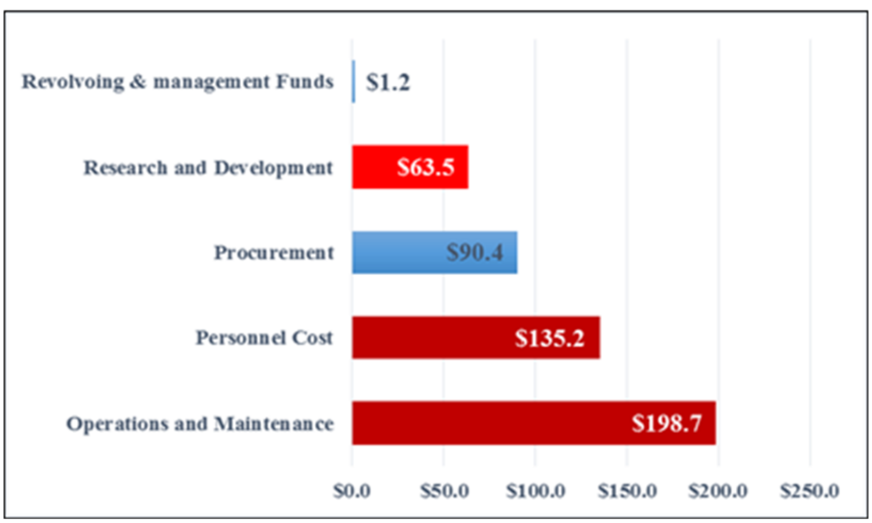


Figure 3. 2015 U.S. Military Budget, by Appropriations Title ($B)

To successfully address these challenges, the United States must re-assess fundamental business decisions to ensure readiness and modernization while maintaining force structure. More than a decade ago, those actions were taken by many of our allies when faced with similar challenges. These nations altered their military structure to concentrate on the core competencies (and responsibilities) of deterrence and conflict resolution. Key decisions made include

- Migration of uniformed personnel to combat/combat support functions
- Privatization of infrastructure
- Employment of public/private partnerships to buy outcomes (versus equipment and services)

These actions offer proven strategies for consideration by the DoD, particularly in the area of outcomes based service contracts.

Outcomes Based Services Contracting (OBSC)

Outcomes Based Service Contracting (OBSC) is a contracting mechanism that allows the customer to pay only when the contractor has delivered outcomes, rather than merely for activities and tasks. OBSC focuses on achieving required outcomes rather than performing to a set of prescribed specifications. In short, the buyer purchases the result of the product used (utilization of service or outcome) and not ownership of the product. The customer no longer directly manages or possibly even owns resources such as the inventory of spares. Suppliers find it in their interest to invest in designing more reliable products and more efficient repair and logistics capabilities to increase profitability.

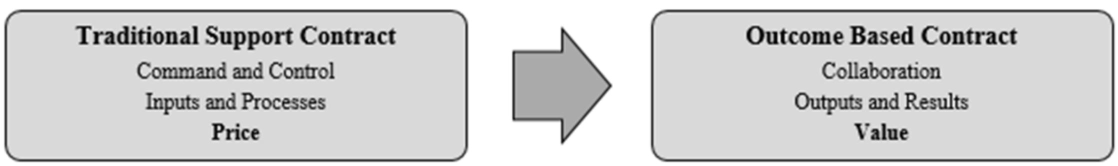


Figure 4. Traditional vs. Outcomes Based Model

OBSC has an ability to produce preferred performances arising from the incentives within the contract, consequently reducing the long-term cost of the contract for the customer. The added benefit of OBSC is that suppliers will be incentivized to think of innovative ways to sustain high operational availability rates. This new strategy is rapidly becoming a central component in the management of after-sales service supply chains, with implications that potentially reach beyond defense and aerospace contracting. A summarization of Outcomes Based Service Contracting benefits to both the DoD and industry are highlighted in Table 1.

Table 1. Outcomes Based Service Contracting Benefits

Outcomes Based Service Contracting	
DoD/Services Benefits	**Industry Service Provider Benefits**
Efficiency and predictability gains from paying for outcomes: By only paying for a measurable specified outcome that is predictable, DoD and the Services will able to make more accurate cost projections.	**Effectiveness, which breeds efficiency and security:** If payment of service providers is dependent on delivering measurable outcomes, there is greater motivation for service providers to perform high-quality work.
Lower servicing costs: OBSC lowers total contract costs as both the DoD and industry contribute complementary resources toward a joint outcome.	**Opportunities for greater control and efficiency:** OBSC involves a closer relationship between industry and the DoD. As a result the industry is more able to optimize and control outcome delivery, maximizing opportunities to reduce the cost of performance while still achieving acceptable outcomes.
Lower transaction and monitoring costs: Better alignment between the interests of the DoD/Services and industry and guaranteed outcomes means that less scrutiny of industry is required and internal DoD costs related to ensuring the outcome may be cut or reallocated.	**Opportunities for innovation:** Industry can use their first-hand experience of working alongside the DoD to drive innovation that meets DoD/Services' changing requirements. New processes required to deliver OBSC may also prompt internal innovations, such as the empowerment of cross functional service teams spanning multiple organizations.
Increased motivation of industry service providers to provide high quality outcomes: If payment to industry is dependent on delivering measurable outcomes, there is greater motivation for service providers to perform high-quality work.	**Sustainable competitive advantage**: Managing customers to optimize co-creation and co-production effectively is an integral part of successful OBSC. Service providers that become adept at customer management can develop a unique competitive advantage, providing more opportunities for the service provider to win contracts.

U.S. Allies and the Purchasing of Outcomes

Outcomes Based Service Contracting is a successful and proven strategy for many military allies of the United States that face significant fiscal constraints on defense spending. The United Kingdom, Australia, and Singapore offer examples of successful implementation of OBSC.

United Kingdom

The UK Ministry of Defence has executed outcomes based contracting for over a decade. After the UK reduced their defense budget by almost 30% in the late 1990s as a result of the Cold War peace dividend, their involvement in Afghanistan and Iraq conflicts from 2000 forward pushed defense spending upward. At the same time, rapidly escalating budget constraints created tremendous pressure to reengineer defense spending in order to deliver needed capability while improving cost and performance.

Establishing a goal to reduce cost by 20% by 2006, they transitioned to "availability contracting," paying industry for a given level of availability over long-term contracts with incentives to reduce support costs while making weapon systems more reliable and efficient. This shift from buying "inputs" (parts, labor, and services) to contracting for "outputs" (availability, capability) instituted a new approach based on partnering with industry and leveraging industry's capital infrastructure.

By 2008 this approach generated cumulative savings of about £1.4 billion while simultaneously achieving performance improvements. As a further benefit, this business model enabled the UK Ministry of Defense to focus on combat operations while utilizing industry partnerships and capabilities for weapons system sustainment.

Case studies have analyzed how this approach has been applied to major UK weapons systems, including Tornado and Harrier fast jets, and logistics activities related to aerial refueling. These cases outline the benefits achieved on these platforms—including savings in the billions of pounds.

Tornado and Harrier Aircraft

Under the Tornado and Harrier programs, the Royal Air Force paid industry to provide a given level of availability. The arrangement included incentives to reduce support chain costs and to make the weapons system more reliable and the support-maintenance processes more efficient. The cost-reduction goal was a key driver in the transformation of the maintenance, repair, and overhaul activity for the two jet aircraft. This approach has successfully reduced the cost of support and decreased manpower and maintenance times while maintaining operational availability. The success of this approach is due primarily to the redefined relationship between the Ministry of Defence and industry, with both sides taking responsibility for and having a stake in maintaining the aircraft.

Omega Aerial Refueling Services, Inc. (OARS)

Aerial tankers are essential when moving large numbers of men and materials long distances, or when stretching the range and length of fighter combat air patrols. Most tanker aircraft are government-owned, but a segment of semi-privatized services exist with their current military fleet counterparts. One such company is Omega Aerial Refueling Services, Inc.

Omega Aerial Refueling Services (OARS) has a very successful 15-year history as the only company in the world conducting commercial, fee-for-service, in-flight refueling services. Omega's service includes using Omega-owned K-707 and KDC-10 to refuel British Royal Air Force (RAF) GR-4A Tornadoes and Canadian Air Force CF-18s during training operations.

Over the past 14 years, Omega has flown over 5,000 missions and 15,000 hours, while off-loading 180 million pounds of fuel and 49,000 airborne aircraft refueling plugs, and while maintaining an exceptional 97% mission completion rate.

Australia Commercialization of Defense Support

Consistent with trends in the UK, Australia sought to maximize its Defense budget by contracting to industry the non-combat functions that support its fighting forces. This initiative has fundamentally changed the landscape of the Australian defense sector by greatly expanding the role of private industry in supporting the Australian Defense Force (ADF) and, conversely, increasing the dependence of the Defense Force on the sustainment of key capabilities from private industry. Two examples of procurement projects that reflect the use of OBSCs are the Hawk Lead-In Fighter and the Eurocopter Tiger Armed Reconnaissance Helicopter.

Hawk Lead-In Fighter

The Hawk was the first Defense aerospace acquisition in Australia that integrated acquisition and through-life-support into a single cradle-to-grave long-term contract that was outcomes driven. It is a performance-based contract that casts BAE Systems not only as the OEM prime in supplying the aircraft, but also as the support prime, or platform steward, for the aircraft where previously the Commonwealth acted as the prime in managing multiple support contracts for the support of an aircraft.

Eurocopter Tiger Armed Reconnaissance Helicopter

Building on the Hawk example above, the Eurocopter Tiger Armed Reconnaissance Helicopter contract with Australian Aerospace and the ADF contracted to acquire a comprehensive system that included the following:

- Sustaining the helicopter fleet by providing an ultramodern training system that included flight and ground-crew simulators
- Software support capabilities
- Ground-based mission planning and management system

The project was also novel for the way in which the final evaluation process was fast tracked to reduce costs to industry and Defense. It took only three months from receipt of proposals from the first four short-listed suppliers to select the tenderer to advance to the tender development stage. This sped up the process and saved tenderers money.

Singapore

Singapore Air Force Basic Wings Course

The Republic of Singapore Air Force (RSAF) Basic Wings Course is an outstanding example of outcomes based service contracting. The Singaporean Air Force is not focused on the reliability of the training airplane, the availability of classroom and simulator training, or even the training facilities and the base. They want the ultimate outcome—trained pilots. The Singaporean Air Force created a partnership with a Training System Integrator that designed the curriculum, procured and supports the equipment, delivers all round-based training, and provides aircraft availability for use by RSAF flight instructors. The training outcomes and cost savings are unmatched anywhere in the world. The success of the program led to its duplication by the Australian Defense Force to trains its next generation of pilots.

The adoption of outcomes based contracting by the Allies relied upon fundamental, strategic changes in their acquisition practices. Furthermore, success of their efforts was dependent upon several key enablers, including the following:

- Long-term contracts that enabled industry to amortize its capital investment.
- Government indemnification of the third party finance providers. This allowed leveraging of sovereign credit costs and provided a means to retain equipment while replacing the contractor if performance was lacking.
- Focus on delivered price and value for money (versus cost and profit). Customers focused on what they needed rather than what the contractor earned, which opened the door for incentive structures that greatly benefited the customer.
- Recognition of industry as full, committed partners—the industry partner only succeeds when the customer succeeds.

The above contracting examples highlight a developing theme—the commercialization of defense through the utilization of outcomes based contracting. The U.S. DoD partnered with industry should now build on these examples, enabling the DoD to better concentrate on its core competencies of deterrence and national defense by buying outcomes, versus equipment and services, to ensure readiness and modernization.

United States and Outcomes Based Service Contracting

The U.S. DoD employed similar procurement strategies and approaches in the late 1990s. Faced with post–Cold War budgets, crumbling infrastructure, and low material readiness, the DoD aggressively pursued third party modernization of base housing, private sector modernization of the DoD's energy infrastructure, and integrated, performance-based support to improve weapons system readiness. These initiatives enabled the DoD to secure modern facilities and enhanced readiness while minimizing pressure on procurement and Military Construction (MILCON) accounts.

The U.S. Navy also relies heavily on commercial merchant mariners for replenishment at sea and maritime force projection and distribution. Unfortunately, the adoption of many of these promising practices slowed as the DoD entered the Global War on Terror and budgets dramatically increased.

Targets of Opportunity

As summarized above, other countries have sought to maximize the effectiveness of their specific Defense budgets by contracting to industry the non-combat functions that support the fighting forces. The following are examples of how the United States could benefit from the increased utilization of OBSC.

Pilot Training

Both the United States Navy and Air Force are struggling to maintain and modernize their pilot training aircraft. The Navy is pursuing an outcomes based approach for rotary wing training while the U.S. Air Force is pursuing a more traditional approach for fixed wing pilot training.

Tanker Capability

The U.S. Air Force is focused on replacing the aging KC-35 with the KC-46. The capability provided by the 767 airframe based KC-46 greatly exceeds the range, endurance, and payload of the 707 based KC-135; however, the required capability is the delivery of fuel to receiver aircraft around the world, both in peacetime training and in wartime engagements. By examining capability-based service contracting options, there may be scenarios where contractors could deliver commercial fee-for-service in-flight refueling services similar to the UK experience.

Military Sealift Command

The Military Sealift Command (MSC) operates 19 Large, Medium-Speed, Roll-on/Roll-off ships or LMSRs. These ships have significantly expanded the nation's sealift capability in the 21st century. LMSRs can carry an entire U.S. Army Task Force, including 58 tanks and 48 other tracked vehicles, plus more than 900 trucks and other wheeled vehicles. The ships can carry vehicles and equipment to support humanitarian missions as well as combat missions. This significant capability is delivered to MSC by a contracted civilian crew of 26 mariners. With over 130 ships in the inventory, MSC should explore outcomes based service contracting opportunities.

Road Ahead

The United States will continue to face a chaotic threat environment and intense fiscal pressure, as shown in Figure 5.

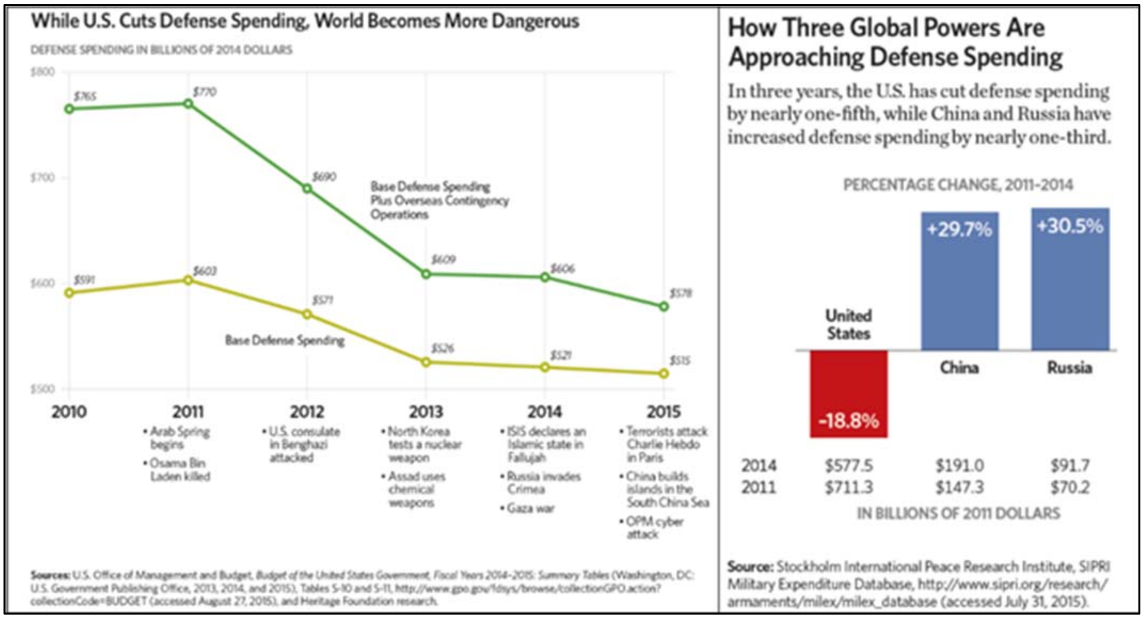


Figure 5. U.S. Defense Spending vs. Peer Competitors

A logical path forward would be to build upon our prior experience and the experience of our allies to employ outcomes based service contracts for non-combat support modernization. Specific areas for application may include the following:

- Operator and maintainer training
- Tactical distribution vehicles
- Air refueling
- Non-combat surface ship modernization
- Network operations
- Search and Rescue modernization
- Carrier On-board Delivery modernization

As an alternative to reduced force structure and combat capability, the DoD could make better use of available government and industry resources at the system, subsystem, and component levels. In that context, outsourcing should be considered in functions where robust capability already exists in the private sector. Outsourcing those functions could result in a 10–15% reduction in personnel, and a 20–30% cost savings.

Examples include maintenance and repair of commercial items, such as propulsion systems that are used and maintained in the private sector. In many instances, the DoD has established duplicate capabilities for maintenance, repair, and overhaul of commercial derivatives of these items with minor modifications for military use that could easily be supported by the private sector.

The DoD maintains 298,897 capital buildings and over 210,000 structures, valued at more than $772 billion. These capabilities were sized to support over 12 years of conflict, but in many cases trace their roots back to World War II. This infrastructure may be oversized for current needs.

The expanded use of Outcomes Based Service Contracts through public–private partnerships delivers increased real time capabilities to the DoD. These relationships provided concrete benefits to the government and would reduce the full life cycle cost.

- Private relationships reallocate risk and up front capital requirements, allowing the government to spread program cost over time. Freeing up the initial capital requirement affords the government the ability to acquire products and services with the limited resources provided in today's austere budget environment.
- Public–private partnerships provide the government with an increased infrastructure and technological capability without having to allocate current year dollars for additional property, plants, equipment, and unnecessary overhead.

When we buy capability as an outcome, the price point is no more than the operating cost of the legacy infrastructure, system, or platform.

Conclusion

The DoD is in a challenging environment characterized by budget deficits, economic uncertainties, and increased public scrutiny. These challenges are driving the need for a fundamental shift in the way the government acquires services, creating an opportunity to transition toward an outcomes-driven approach. Governments in the United Kingdom, Australia, and other countries have been successfully utilizing outcomes based service contracting for years. These same contracting principles can be readily applied across the DoD and its industry partners.

References

Office of Management and Budget. (2012, August 8). *Budget of the U.S. Government, FY 2013* [Table 8.3]. Retrieved from http://www.whitehouse.gov/omb/budget/historicals

Zumbrun, J. (2015, February 3). The legacy of debt: Interest costs poised to surpass defense and nondefense discretionary spending. *Wall Street Journal*. Retrieved from http://blogs.wsj.com/economics/2015/02/03/the-legacy-of-debt-interest-costs-poised-to-surpass-defense-and-nondefense-discretionary-spending/

Panel 12. Considerations for Focusing Development and Controlling Growth in MDAPs

Thursday, May 5, 2016	
9:30 a.m. – 11:00 a.m.	**Chair: Nancy Spruill,** Director, Acquisition Resources & Analysis, Office of the Under Secretary of Defense for Acquisition, Technology, & Logistics ***Blockmodeling and the Estimation of Evolutionary Architectural Growth in Major Defense Acquisition Programs*** LTC Matthew Dabkowski, U.S. Army, University of Arizona Ricardo Valerdi, Associate Professor, University of Arizona ***An Assessment of Early Competitive Prototyping for Major Defense Acquisition Programs*** William Fast, COL, U.S. Army (Ret.), Senior Lecturer, NPS ***The Sixth-Generation Quandary*** Raymond Franck, Professor Emeritus, U.S. Air Force Bernard Udis, Professor Emeritus, University of Colorado at Boulder

Nancy Spruill—is the Director of Acquisition Resources & Analysis at the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. Dr. Spruill received her Bachelor of Science degree in mathematics in 1971. From 1971 to 1983, she held a variety of positions with the Center for Naval Analyses, including technical staff analyst, professional staff analyst, and project director. She earned her Master of Arts in mathematical statistics in 1975, followed by her doctorate in 1980.

Dr. Spruill served on the staff of the Office of the Secretary of Defense from 1983 to 1993. Initially, she was the Senior Planning, Programming, and Budget Analyst in the Manpower, Reserve Affairs, and Logistics Secretariat. Later, she served as the Director for Support and Liaison for the Assistant Secretary of Defense for Force Management and Personnel. Then she served as the Senior Operations Research Analyst in the Office of the Assistant Secretary of Defense for Program Analysis and Evaluation.

In March 1995, she was selected as the Deputy Director for Acquisition Resources for the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). In February 1999, she was appointed the Director of Acquisition Resources and Analysis (ARA) for the USD(AT&L). In this capacity, she is responsible for all aspects of AT&L's participation in the Planning, Programming, and Budgeting System (PPBS); the congressional process; and the Defense Acquisition System. She serves as the Executive Secretary to the Defense Acquisition Board and is responsible for the timely and accurate submission to Congress of Selected Acquisition Reports and Unit Cost Reports for Major Defense Acquisition Programs. She manages the Defense Acquisition Execution Summary monthly review of programs; monitors cost and schedule status of high interest programs; and conducts analyses of contract and program cost performance, including analysis of the effective use of integrated program management principles through the use of earned value management. She leads the Department in developing plans to manage property, plants and equipment, inventory, operating materials, and supplies/deferred maintenance and environmental liabilities. She proposes modifications to, or acquisition of, new DoD feeder systems, in support of achieving an unqualified audit opinion on DoD financial statements as mandated by the Chief Financial Officers (CFO) Act.

She also manages the studies program for the OSD, oversees the USD(AT&L)'s office automation system, and manages its information system network. She serves as the focal point for DoD-wide software-intensive systems program initiatives to improve mechanisms for the management of defense acquisition programs; manages software intensive systems assessment initiatives; performs systemic analysis from independent expert program reviews to improve acquisition policy and education, and conducts special analyses for the under secretary.

Dr. Spruill has been a member of the Senior Executive Service since 1995. She is a certified acquisition professional and an active member of the American Statistical Association. Her many honors and awards include the Defense Medal for Exceptional Civilian Service, the Defense Medal for Meritorious Civilian Service, the Hammer Award, and the Presidential Rank Award. She has contributed papers in publications of the statistics and defense analyses communities and authored articles in the general press on how politicians use—and abuse—statistics.

Blockmodeling and the Estimation of Evolutionary Architectural Growth in Major Defense Acquisition Programs[1]

LTC Matthew Dabkowski, USA—is a Lieutenant Colonel in the U.S. Army, and a PhD candidate in Systems and Industrial Engineering at the University of Arizona (UA). He received his BS in operations research from the United States Military Academy (USMA) in 1997, and his MS in systems engineering from the UA in 2007. LTC Dabkowski is the winner of the Military Operations Research Society's (MORS) 2013 Wayne P. Hughes Junior Analyst Award, the MORS 2012 Rist Prize, the TRADOC Analysis Center's 2012 LTC Paul J. Finken Memorial Award, and the Army's 2009 Dr. Wilbur B. Payne Memorial Award. [mfd1@email.arizona.edu]

Ricardo Valerdi—is an Associate Professor in the Systems and Industrial Engineering Department at the UA, where he teaches courses in cost estimation and systems engineering. He received his BS/BA in electrical engineering from the University of San Diego in 1999, and his MS and PhD in systems architecting and engineering from the University of Southern California in 2002 and 2005, respectively. Dr. Valerdi is the Editor-in-Chief of the *Journal of Cost Analysis and Parametrics*, as well as a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE). [rvalerdi@arizona.edu]

Abstract

It is well-known that cost overruns in Major Defense Acquisition Programs (MDAPs) are endemic, and requirements volatility is at least partially to blame. In particular, when the desired capabilities of a system change during its life cycle, substantial reengineering can result, especially when a new subsystem must be incorporated into an existing architecture. Of course, the likelihood and specifics of such additions are rarely known ahead of time, and predicting integration costs is challenging. In this paper, we present a novel algorithm to address this issue. In particular, leveraging an integer programming implementation of the social network analysis technique blockmodeling, we optimally partition the subsystems represented in Department of Defense Architecture Framework (DoDAF) models into architectural positions. Using this abstracted structure, we subsequently grow the architecture according to its statistical properties, and we estimate this unforeseen cost of evolutionary architectural growth via the Constructive Systems Engineering Cost Model (COSYSMO). We illustrate this process with a real-world example, discuss limitations, and highlight areas for future research.

---

[1] The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Introduction[2]

Major Defense Acquisition Programs (MDAPs) are notoriously prone to excessive cost overruns (GAO, 2011), and requirements volatility is often partially to blame (e.g., Bolten et al., 2008; Peña & Valerdi, 2015). In fact, based on the GAO's most recent *Assessments of Selected Weapon Programs* (2015), 6 of the 14 largest increases in MDAP development costs were due to the addition of new capabilities, making it the most frequent cause of substantial post-Milestone B (MS B) cost growth. Given a general lack of system specification early in the system life cycle (Blanchard & Fabrycky, 1998), this is not surprising, as accurately estimating the cost of an unknown set of capabilities is difficult at best.

With this in mind, in 2009, Congress passed the Weapon Systems Acquisition Reform Act (WSARA), which implemented several initiatives to rein in cost growth, including shifting an MDAP's baseline cost estimate from MS B to MS A (WSARA, 2009). Acknowledging the need for detailed system information earlier in the life cycle, the DoD followed suit in 2013 by requiring the submission of a draft Capability Development Document (CDD) pre–MS A (USD[AT&L], 2013), replete with the DoD Architecture Framework (DoDAF) models required by the *Joint Capabilities Integration and Development System* (JCIDS; Chairman of the Joint Chiefs of Staff, 2012).

Given WSARA's call for accurate early life cycle cost estimates, this has favorable implications. Specifically, in Valerdi, Dabkowski, and Dixit (2015), we demonstrate that the DoDAF models required pre–MS A map to 14 of the 18 parameters of the Constructive Systems Engineering Cost Model (COSYSMO). Consisting of four size drivers (i.e., number of requirements, number of interfaces, number of algorithms, and number of operational scenarios) and 14 effort multipliers, COSYSMO has been used by a variety of organizations to estimate the amount of systems engineering effort required to bring a system to fruition (e.g.,Valerdi, 2008; Wang et al., 2012),[3] and industry has found this estimate to be a valuable proxy for total system cost (e.g., Honour, 2004; Cole, 2012).

Moreover, in Dabkowski, Valerdi, and Farr (2014), we develop an algorithm to estimate the cost of unforeseen architectural growth in MDAPs via the SV-3 (or Systems-Systems Matrix), providing a mechanism to assess the cost risk associated with alternative designs. Leveraging elements of network science and simulation, the algorithm exploits both the micro- and macrostructure of the SV-3 to connect a new subsystem to an MDAP's existing architecture, and it employs COSYSMO to estimate the cost of the associated growth. In 2016, we validated and further refined our approach using real-world SV-3s (Dabkowski & Valerdi, 2016). While the details of our most recent work are beyond the scope of this paper, one of our modeling considerations is not, namely, the detection and exploitation of architectural communities within the SV-3.

---

[2] The material in the Introduction and Identifying and Exploiting Architectural Communities sections is derived from our earlier Acquisition Research Symposium paper titled "The Budding SV3: Estimating the Cost of Architectural Growth Early in the Life Cycle" (Dabkowski & Valerdi, 2014). Copyright is retained by the authors.

[3] COSYSMO estimates systems engineering effort in person months (nominal schedule) or PM_{NS}

Identifying and Exploiting Architectural Communities

In order to facilitate the discussion that follows, consider the hypothetical SV-3 in Panel (a) of Figure 1, where cell (i, j) is shaded if subsystem i interfaces with subsystem j, and darker shades indicate greater interface complexity (i.e., light gray ⇒ easy, medium gray ⇒ nominal, black ⇒ difficult). Consisting of N=20 subsystems (labeled A through T) and E=47 undirected interfaces,[4] suppose we are interested in estimating the effort required to incorporate an additional subsystem (U) into the architecture *without knowing its purpose or function.* In light of COSYSMO's cost estimating relationship (CER), this ultimately forces us to estimate the number of interfaces (by complexity level) U will generate.

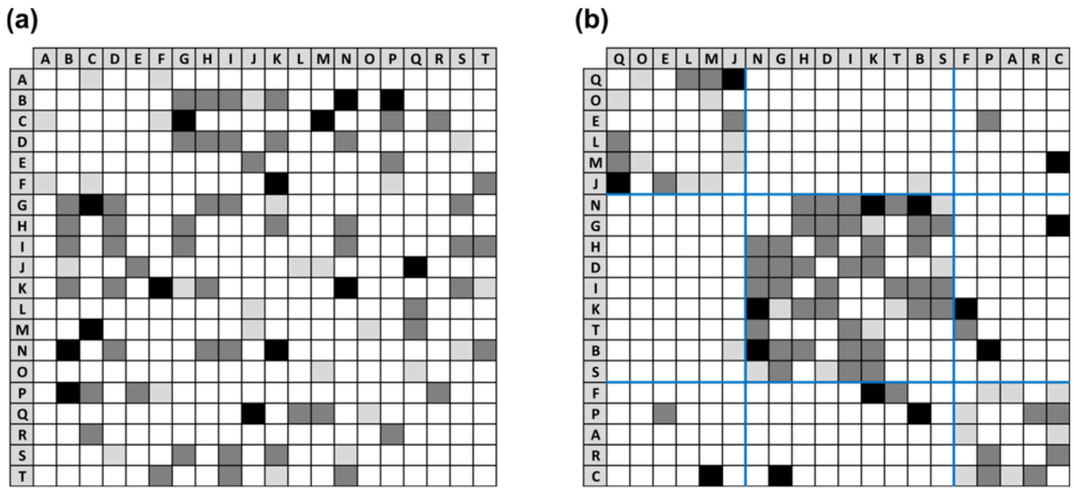


Figure 1. Hypothetical SV-3 in Its Original (Panel (a)) and Isomorphic (Panel (b)) Representations, Where Subsystems Have Been Permuted Into Architectural Communities
(Dabkowski et al., 2014)

More granularly, we need to answer three questions:

(Q1) How many subsystems should U connect to (degree, m)?;

(Q2) If U connects to m subsystems, which *m* subsystems should it connect to (adjacency)?; and

(Q3) If U connects to a specific set of *m* subsystems, what should the complexity of these interfaces be (weights)?

Under the scenario of evolutionary growth versus revolutionary change, we make the fundamental assumption that the current architecture foretells the future architecture. In other words, the existing patterns and characteristics of the subsystems' interfaces in Figure 1 provide us with useful evidence for predicting the pattern and characteristics of the

[4] In the parlance of network science, undirected interfaces are symmetric with respect to the SV-3's main diagonal. In other words, the interface from subsystem i to subsystem j implies the same interface from subsystem j to subsystem i. For directed interfaces, symmetry is not required, and the implication does not hold.

interfaces U will generate. As reported in our earlier Acquisition Research Symposium paper titled "The Budding SV3: Estimating the Cost of Architectural Growth Early in the Life Cycle" (Dabkowski & Valerdi, 2014), making this assumption allows us to address (Q1) through (Q3) as follows:[5]

(A1) Degree: To model a "rich-by-birth" effect, view the degree of U (M_U) as a random variable with a probability mass function (PMF) equal to the observed degree distribution of the existing system (Dorogovtsev & Mendes, 2003);

(A2) Adjacency: To incorporate a "rich-get-richer" effect, utilize the Barabási-Albert preferential attachment (PA) model from network science, where the probability subsystem i attaches to subsystem U is a linear function of its degree (d_i) or $p_i = d_i / \sum_{j=1}^{N} d_j$ (Barabási & Albert, 1999); and

(A3) Weights: To mimic the observed complexity in the existing architecture, cast the complexity of the interface between U and subsystem i (w_{iU}) as a conditional random variable, where the PMF for w_{iU} equates to the observed interface complexity distribution of subsystem i.

Furthermore, when searching for patterns in an MDAP's architecture, the manner in which systems engineers typically architect systems should be taken into account. For instance, in *The Art of Systems Architecting*, Maier and Rechtin (2000) note that the "most important aggregation and partitioning heuristics are to *minimize external coupling* and *maximize internal cohesion* [emphasis added]." Accordingly, looking for clusters or communities of subsystems where the density of intra- versus inter-community interfaces is high seems reasonable, and applying the Girvan-Newman community detection heuristic (Girvan & Newman, 2002) to the SV-3 in Panel (a) of Figure 1 identifies three *architectural communities*. As seen in Panel (b) of Figure 1, when the MDAP's subsystems are permuted by their community membership, the system's underlying macrostructure appears to abide Maier and Rechtin's (2000) heuristics. Exploiting these architectural communities in (A1) to (A3) yields the following mechanism for estimating the cost of connecting subsystem U to the existing architecture (Dabkowski et al., 2014):

For a specified, suitably large number of iterations (e.g., 10,000)[6]…

Preprocessing

1. Initialize the system as the current system,
2. Use Girvan-Newman (2002) to identify architectural communities,
3. Randomly assign U to community *j*,

---

[5] See Dabkowski et al. (2013) for additional details.

[6] When estimating the population mean of a random variable X (μ_X) using Monte Carlo simulation, the minimum number of iterations required is a function of (a) the researcher's desired accuracy for the estimate, which varies depending on the context, and (b) the population variance (σ_X^2), which is normally unknown. Accordingly, the researcher typically runs an initial set of iterations to generate unbiased estimates of μ_X and σ_X^2 from which the minimum number of iterations can be calculated (i.e., via Driels & Shin, 2004)

Intracommunity Growth

4. Generate a realization for $M_{U,intra}$ given U is assigned to community *j* (m_{intra}),
5. Connect U to m_{intra} subsystems inside community *j* using the PA model,
6. For each interface established in (5), assign complexity ($w_{iU,intra}$),

Intercommunity Growth

7. Generate a realization for $M_{U,inter}$ given U is assigned to community *j* (m_{inter}),
8. Connect U to m_{inter} communities using the PA model, and
9. For each interface established in (8), assign complexity ($w_{iU,inter}$),

Cost Estimation

10. Estimate the cost for the augmented system using COSYSMO (PM_{NS}*),
11. Calculate the additional cost of adding subsystem U (PM_{NS}* − PM_{NS}), and
12. Store results and return to (3).

Generalizing Beyond Architectural Communities via Blockmodeling

While the above algorithm has intuitive appeal, the SV-3 in Figure 1 is hypothetical, and this raises the following questions: "Do (A1) through (A3) adequately model the growth of real-world SV-3s, and do SV-3s actually harbor architectural communities?" In a recent paper, we address these questions using 24 different SV-3s from a wide variety of MDAPs (Dabkowski & Valerdi, 2016). First, with respect to (A1) and (A2), formal hypothesis testing suggested that using the observed degree distribution generated far too many interfaces and blindly applying the PA model was ill-advised. In fact, the PMF for an incoming subsystem's number of interfaces $(P(M = m))$ and the strength of preferential attachment β interact, which led us to identify and utilize an optimal set of $\{P(M = m), \beta\}$ pairs for each SV-3. Moving on to (A3), none of the real-world SV-3s we examined were valued; thus, the validity of using the observed interface complexity distribution to estimate future interface complexity could not be assessed. Finally, as regards architectural communities, less than 50% of the SV-3s exhibited community structure worth exploiting, suggesting a non-community version of the algorithm was necessary. Simply put, significant adjustments to our earlier algorithm were necessary, and these are documented in Dabkowski and Valerdi (2016).

Notwithstanding these refinements, restricting our attention to architectural communities may ignore other, more compelling macrostructures within the architecture. For example, consider the hypothetical SV-3 in Panel (a) of Figure 2.

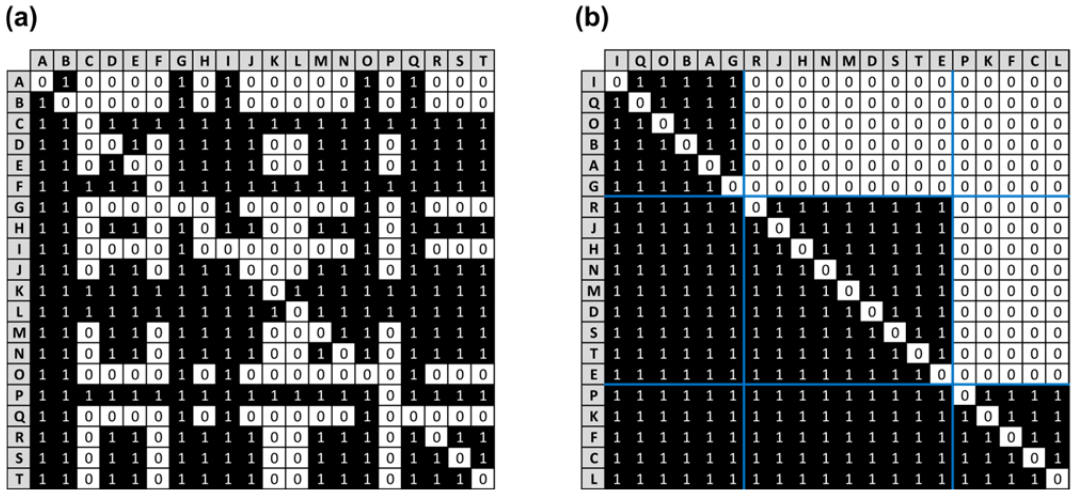


Figure 2. Hypothetical SV-3 With a Hierarchical Structure in Its Original (Panel (a)) and Isomorphic (Panel (b)) Representations, Where Subsystems Have Been Optimally Partitioned and Permuted

Consisting of $N = 20$ subsystems (labeled A through T) and $E = 251$ directed interfaces, the SV-3 is relatively dense, and, while the Girvan-Newman community detection heuristic identifies six architectural communities, the community structure is weak. Based on this result, we would invoke our non-community version of the algorithm. That said, the Girvan-Newman community detection heuristic was designed for sparse networks (Girvan & Newman, 2002), and the weak community structure may be spurious. Moreover, taking this approach would ignore the indisputable hierarchical structure of subsystems seen in Panel (b) of Figure 2, where subsystems in lower ranking clusters ({R, J, H, N, M, D, S, T, E} and {P, K, F, C, L}) not only have a high density of interfaces with subsystems inside their clusters but also have a high density of interfaces with subsystems inside higher ranking clusters.

To identify this and other hidden macrostructure, we can apply the network analysis technique known as blockmodeling, where a network consisting of $i = 1, \cdots, N$ objects (i.e., the SV-3 and its subsystems) is partitioned into $k = 1, \cdots, P$ nonoverlapping positions (or clusters) where the positions generally abide the structure represented in a $(P \times P)$ image matrix such that $P \ll N$ Conceived by computational sociologists at Harvard in the mid-1970s (e.g., White, Boorman, & Breiger, 1976; Boorman & White, 1976), blockmodeling methods have been an active area of research for over 40 years, and they have been integrated into popular network analysis software such as UCINET (Borgatti, Everett, & Freeman, 2002), R's *igraph* package (Csárdi & Nepusz, 2006), and Pajek (Mrvar & Batagelj, 2013).

Notable among these is Pajek's inclusion of Doreian, Batagelj, and Ferligoj's (2005) direct approach, which employs a simple object relocation routine that minimizes the number of inconsistencies between the permuted, partitioned $(N \times N)$ adjacency matrix (i.e., the SV-3) and a corresponding$(P \times P)$ image matrix. Invoked in Pajek via the commands `Network → Create Partition → Blockmodeling`, we ran Doreian et al.'s (2005) direct approach on the hypothetical SV-3 in Panel (a) of Figure 2, and this yielded the image matrix and reduced graph seen in Panels (a) and (b) of Figure 3, respectively. With zero inconsistences, the solution's partition matches Panel (b) of Figure 2, and it is the unique global optimum. As Figure 3 clearly demonstrates, unlike Girvan and Newman's (2002) community detection heuristic, Doreian et al.'s (2005) direct approach recovered the hierarchical clustering of subsystems.

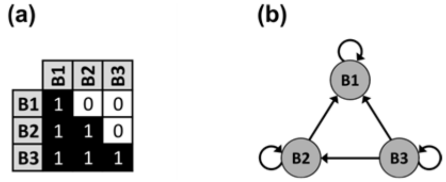


Figure 3. Globally Optimal Image Matrix (Panel (a)) and Reduced Graph (Panel (b)) for the Hypothetical SV-3 Seen in Panel (a) of Figure 2

In fact, blockmodeling can be seen as the natural generalization of community detection, as finding an optimal clustering of N objects into P communities is equivalent to finding the optimal partition of N objects for a P-position identity image matrix. For instance, consider the hypothetical SV-3 in Panels (a) and (b) of Figure 4.

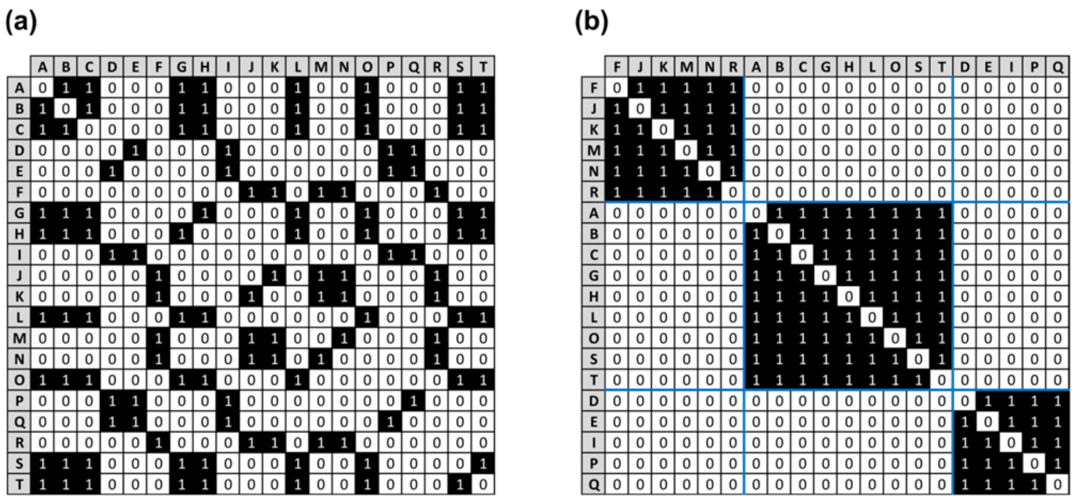


Figure 4. Hypothetical SV-3 With Community Structure in Its Original (Panel (a)) and Isomorphic (Panel (b)) Representations, Where Subsystems Have Been Optimally Partitioned and Permuted

With three isolated cliques and a sparse structure, we expect the Girvan-Newman community detection heuristic to identify the architectural communities, and it does. Similarly, Doreian et al.'s (2005) direct approach recovers the communities, yielding the globally optimal image matrix and reduced graph seen in Panels (a) and (b) of Figure 5, respectively.

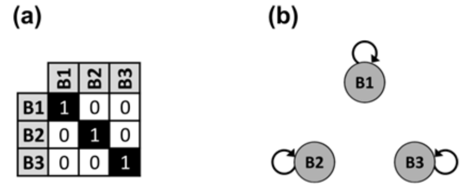


Figure 5. Globally Optimal Image Matrix (Panel (a)) and Reduced Graph (Panel (b)) for the Hypothetical SV-3 Seen in Panel (a) of Figure 4

Given these observations, the implication is that when it comes to identifying and exploiting the underlying macrostructure of a network, blockmodeling subsumes—and therefore trumps—community detection. Interestingly enough, however, this relationship has only recently been acknowledged by network scientists, as Newman and Leicht note in their 2007 paper extending earlier and more limited community detection methods:

> Here we describe a general technique for detecting structural features in large-scale network data that works by dividing the nodes of a network into classes such that the members of each class have similar patterns of connection to other nodes. … the idea is similar in philosophy to the block models proposed by White and others. (pp. 9564–9565)

Nonetheless, Doreian et al.'s (2005) direct approach is not a panacea, as it (1) generates locally optimal solutions and, thus, provides no guarantee that better fitting image matrices and partitions do not exist and (2) was designed to handle single one- or two-mode networks,[7] and, therefore, cannot readily accommodate multiple relations simultaneously. Unfortunately, both shortcomings are problematic. First, without a known optimality gap, we cannot definitively assess the quality of Pajek's solutions, and exact methods that generate global optima are necessary. Second, during our investigation of real-world SV-3s (Dabkowski & Valerdi, 2016), we discovered that 3 of the 24 SV-3s were actually mixed-mode networks. For example, consider the SV-3 in Figure 6, which consists of 10 internal subsystems and 7 external subsystems.

---

[7] One- and two-mode networks describe the connections that exist between a single set of objects and two distinct sets of objects, respectively. In the context of this paper, if an SV-3 is one-mode, the subsystems in its rows and columns are the same. If it is two-mode, they are different.

	Internal Subsystems										External Subsystems						
Internal Subsystems	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	E1	E2	E3	E4	E5	E6	E7
I1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
I2	1	0	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0
I3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
I4	1	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0
I5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
I6	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0
I7	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0
I8	0	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0
I9	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	1	1
I10	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	1

Figure 6. Multiple Relation Mixed-Mode SV-3 With 10 Internal Subsystems (Labeled I1 Through I10) and 7 External Subsystems (Labeled E1 Through E7)

In this SV-3, the 1-mode portion (located to the left of the vertical red line) shows the interfaces that exist between internal subsystems, where a 1 in cell (i, j) implies internal subsystem i interfaces with internal subsystem j. Similarly, the 2-mode portion (located to the right of the vertical red line) shows the interfaces that exist between internal and external subsystems, where a 1 in cell (i, m) implies internal subsystem i interfaces with external subsystem m. Clearly, each portion of the SV-3 contains valuable information for partitioning the internal subsystems, and we would like to include both in our analysis.

With this in mind, the first author embarked on a complementary line of research to develop an exact method for the blockmodeling of mixed-mode networks. Drawing on the integer programming approach of Brusco and Steinley (2009), this effort is chronicled in the "Exact Exploratory Blockmodeling of Multiple Relation, Mixed-Mode Networks Using Integer Programming" (Dabkowski, Fan, & Breiger, 2016), and it provides analysts with a reasonably efficient way to find globally optimal blockmodels for one-, two-, and mixed-mode SV-3s. Applying this method to the SV-3 in Figure 6 and capping the number of internal and external positons at three yields the results in Figure 7.

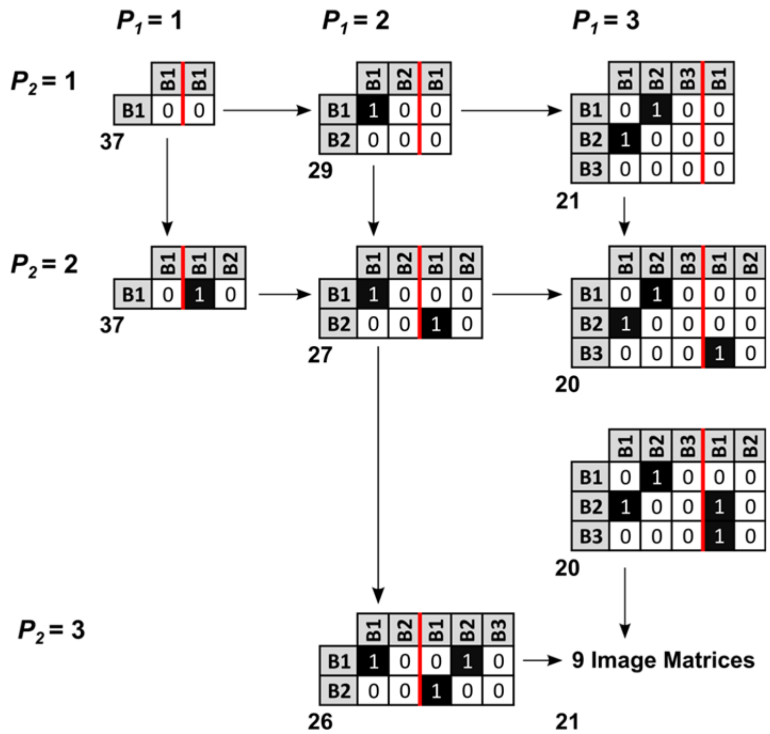


Figure 7. Globally Optimal Image Matrices for the Mixed-Mode SV-3 Seen in Figure 6, Where the Number of Inconsistencies Corresponding to the Globally Optimal $(P_1 \times P_1|P_1 \times P_2)$ Image Matrix Is Given at the Bottom Left of the Matrix

As Figure 7 shows, with the exception of the $(3 \times 3|3 \times 3)$ image matrix, the minimum number of inconsistencies decreases monotonically as the number of internal or external positions increases, eventually reaching a minimum of 20 for the two globally optimal $(3 \times 3|3 \times 2)$ image matrices. Moreover, for each of the two globally optimal $(3 \times 3|3 \times 2)$ image matrices in Figure 7, the clustering of the internal and external subsystems is the same, and the corresponding permuted, partitioned network is given in Figure 8.

		Internal Subsystems										External Subsystems						
		I1	I6	I7	I9	I10	I2	I8	I3	I4	I5	E2	E1	E3	E4	E5	E6	E7
Internal Subsystems	I1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0
	I6	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0
	I7	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0
	I9	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	1	1
	I10	0	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1	1
	I2	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	I8	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0
	I3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	I4	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0
	I5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0

Figure 8. Mixed-Mode SV-3, Where the Rows and Columns Have Been Permuted According to the Globally Optimal $(3 \times 3|3 \times 2)$ Image Matrices and Partition in Figure 7

Interestingly, the clustering of internal subsystems appears to be entirely driven by connections outside the clusters. As with the hypothetical SV-3 in Figure 2, traditional community detection algorithms cannot exploit this, and, as expected, Girvan and Newman's (2002) heuristic returned an insignificant, much different result using the one-mode portion of Figure 6. Nonetheless, as the number of positions increases the exact approach quickly becomes impractical, and mixed-mode blockmodeling heuristics are necessary. Accordingly, the first author built one in Pajek leveraging Doreian et al.'s (2005) direct approach, and its performance was outstanding, as it found the globally optimal solutions in a reasonable amount of time.

Integrating Results

Equipped with exact and heuristic methods for the blockmodeling of SV-3s, we can replace Step (2) in our earlier algorithm ("Use Girvan-Newman (2002) to identify architectural communities") with "Use Dabkowski-Fan-Breiger (2015; 2016) to identify an optimal P-position image matrix and partition of subsystems." If the optimal image matrix and partition suggest a compelling architectural structure, future evolutionary growth should abide it, and, similar to our earlier algorithm, we can randomly assign an incoming subsystem (X) to position k. However, unlike our earlier algorithm, the assignment of subsystem X's m interfaces to positions is no longer modeled via separate PMFs for each position (or community). It is the sum of m independent and identically distributed categorical random variables, where the probability interface j for $j = 1, \cdots, m$ links to a subsystem in position l for $l = 1, \cdots, P$ is given by:

$$\frac{\text{number of interfaces in block } (k,l) \text{ of the partitioned and permuted SV-3}}{\text{number of interfaces in row } k \text{ of the partitioned and permuted SV-3}} \quad (1)$$

[8]

As such, the collective assignment of subsystem X's m interfaces to positions can be modeled as a random $(1 \times P)$ vector $\boldsymbol{C}$, where $\boldsymbol{C}$ follows a Multinomial $m, \boldsymbol{p}$ distribution and $\boldsymbol{p}$ is the $(1 \times P)$ vector of multinomial probabilities defined in (Equation 1).

Of course, $\boldsymbol{C}$ could generate a realization ($\boldsymbol{c}$) where one or more of its elements ($\boldsymbol{c}_l$) exceeds the number of subsystems in its respective position (N_l). In this case, we can apply the following numerical recipe to generate a feasible realization for $\boldsymbol{C}$: (1) for all positions where $\boldsymbol{c}_l \geq N_l$,, aggregate the $\boldsymbol{c}_l - N_l$ excess interfaces into an accumulator variable, m', and set $\boldsymbol{c}_l$ as N_l; (2) remove these positions and their probability mass from $\boldsymbol{C}$; (3) renormalize the multinomial probabilities; and (4) redistribute the m' excess interfaces among the remaining positions, iterating as necessary.

Integrating these adjustments, as well as refinements from Dabkowski and Valerdi (2016), into our earlier algorithm yields the modified pseudocode below:

---

[8] As Kolaczyk and Csárdi (2014) note, in a nonstochastic blockmodel, "the edge probabilities π_{qr} [where q and r represent positions], and the maximum likelihood estimates—which are natural here—are simply the corresponding empirical frequencies" (p. 97).

For a specified, suitably large number of iterations …

Preprocessing

1. Initialize the system as the current system,
2. Build an optimal set of $\{P(M=m), \beta\}$ pairs,
3. Use Dabkowski-Fan-Breiger (2015; 2016) to identify an optimal P-position image matrix and partition of subsystems,

Growth

4. Randomly select a member from the optimal set of $\{P(M=m), \beta\}$ pairs,
5. Generate a realization for the incoming subsystem's (X's) number of interfaces using $P(M=m)$; if the optimal image matrix and partition suggest a compelling architectural structure, use *Connection Option A*; otherwise, use *Connection Option B*,

Connection Option A

6a. Randomly assign X to position k,

6b. Model the collective assignment of subsystem X's m interfaces to positions as a random $(1 \times P)$ vector $\boldsymbol{C}$, where $\boldsymbol{C}$ follows a Multinomial $(m, \boldsymbol{p})$ distribution and $\boldsymbol{p}$ is the $(1 \times P)$ vector of multinomial probabilities given by (Equation 1); generate a feasible realization for $\boldsymbol{C}$,

6c. For $l = 1, \cdots, P$, attach X to $\boldsymbol{c}_l$ subsystems inside position l using attachment probabilities $p_i = {d_i}^{\beta} / \sum_{j=1}^{N} {d_j}^{\beta}$,

6d. For each interface established in (6c), assign complexity (w_{iX}),

Connection Option B

6a. Attach X to m subsystems using attachment probabilities $p_i = {d_i}^{\beta} / \sum_{j=1}^{N} {d_j}^{\beta}$,

6b. For each interface established in (6a), assign complexity (w_{iX}),

Cost Estimation

7. Estimate the cost for the augmented system using COSYSMO (PM_{NS}*),
8. Calculate the additional cost of adding subsystem X ($PM_{NS}* - PM_{NS}$), and
9. Store results and return to (4).

As seen above, unlike our previous algorithm, *Connection Option B* provides an alternative, nonposition-based growth mechanism. Additionally, *Connection Option A* does not condition interface complexities based on the connected subsystems' positions of assignment (i.e., $w_{iX,l}$), as any patterns in intra- or interposition complexity could be due to chance. Specifically, the blockmodeling methods developed in Dabkowski et al. (2015; 2016) are for unvalued networks. Therefore, the statistical significance of apparent structure in the interface complexities must be assessed prior to leveraging them in the algorithm.

Using our improved algorithm, we can estimate the cost of unforeseen, internal architectural growth in mixed-mode SV-3s (as well as one- and two-mode SV-3s). For example, assume the system represented in Figure 6 has the following values for COSYSMO's parameters: $A = 0.25$; $E = 1.06$; $\prod_{j=1}^{14} EM_j = 0.89$; and 75 easy, 50 nominal, and 10 difficult requirements. Additionally, if we assume its interface complexities are portrayed in Figure 9, the system has 12 interfaces between internal subsystems (6 easy, 5 nominal, and 1 difficult) and 13 interfaces between external subsystems (6 easy, 6 nominal, and 1 difficult). Using COSYSMO's CER and weights from Valerdi (2008), we estimate that 59.24 PM_{NS} of systems engineering effort are required to successfully conceptualize, develop, and test the MDAP. At this point, we have initialized the system as the current system, and Step (1) of the algorithm is complete.

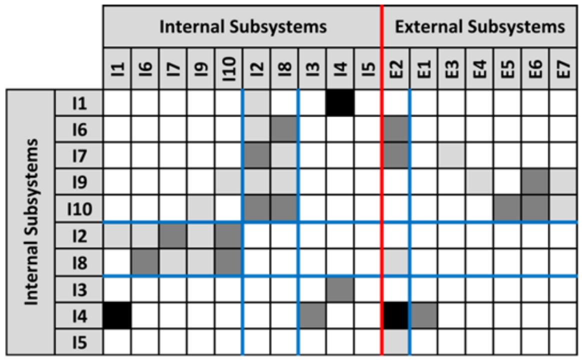
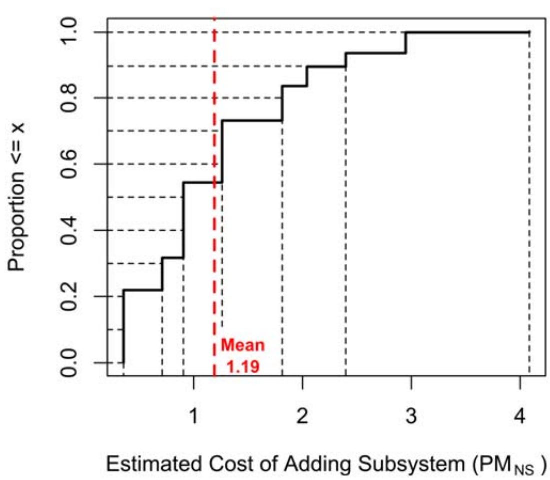


Figure 9. Hypothetical Interface Complexities for the System Represented in Figure 6, Where Cell (i, j) Is Shaded if Subsystem i Interfaces With Subsystem j, and Darker Shades Indicate Greater Interface Complexity (i.e., Light Gray ⇒ Easy, Medium Gray ⇒ Nominal, Black ⇒ Difficult)

Our next task is to build an optimal set of $\{P(M = m), \beta\}$ pairs. Using our approach in Dabkowski and Valerdi (2016), there are five feasible PMFs for m. Among these, the single optimum is $P(M = 2) = 0.5$ and $P(M = 1) = 0.5$, and the corresponding optimal set of β is {0, …, 0.4}.

With Step (2) complete, our last preprocessing step is to identify an optimal P-position image matrix and partition of subsystems, and the global optimal solution is given in Figure 9. This result, along with the optimal set of $\{P(M = m), \beta\}$ pairs, is then ingested into a Monte Carlo simulation, which performs Steps (4) through (9). Running the simulation for 10,000 iterations yields the results seen in Figure 10.



Percentile	PM_{NS}	Percentile	PM_{NS}
5%	0.36	55%	1.26
10%	0.36	60%	1.26
15%	0.36	65%	1.26
20%	0.36	70%	1.26
25%	0.71	75%	1.81
30%	0.71	80%	1.81
35%	0.91	85%	2.04
40%	0.91	90%	2.40
45%	0.91	95%	2.95
50%	0.91	100%	4.08

Figure 10. Empirical Cumulative Distribution Function and Percentiles for the Estimated Cost of Connecting an Additional Subsystem to the Internal Subsystems of Figure 9

As seen in Figure 10, the expected cost to connect an additional subsystem (X) to the internal subsystems of Figure 9 is 1.19 PM_{NS}, and the associated 95% confidence interval is (1.177, 1.206) PM_{NS}. Moreover, although the maximum cost to attach subsystem X should not exceed 4.08 PM_{NS}, there is only a 5% chance it will be more than 2.95 PM_{NS}. Finally, if we condition our estimate on X's position of assignment, the expected cost in person months (nominal schedule) is 1.00, 1.05, and 1.53 for positions 1, 2, and 3, respectively. In the absence of additional information, these estimates represent our "best guess" for the cost to attach a new subsystem to the existing architecture, and they help to quantify the likelihood of excessive cost growth.

Limitations and Future Work

Although our use of blockmodeling to identify and exploit an SV-3's globally optimal macrostructure provides a useful generalization, the algorithm and its supporting methods have several limitations, and these represent opportunities for future research. Starting with insufficient data, SV-3s are not currently weighted by interface complexity, and the validity of using the observed interface complexity distribution to estimate future interface complexity could not be assessed. Accordingly, sponsored research is required to generate the necessary data for statistical investigation.

Moving on to the algorithm's internal steps, Connection Option A assigns incoming subsystems to positions using a uniform distribution. If we assume unforeseen architectural growth is equally likely in all positions, this is appropriate. That said, other possibilities are worth exploring. For example, the probability subsystem X is assigned to position k could be modeled as either a function of position k's size or a function of subsystem X's number of interfaces. Additionally, although the algorithm is currently limited to estimating internal architectural growth, modifying it to address external architectural growth is natural, especially when we consider that its optimal macrostructure was obtained from the interfaces between its internal and external subsystems.

Finally, in a more general sense, mixed-mode blockmodeling remains a fruitful area for future research, as it suffers from scalability challenges, especially as the number of internal and external positions grow. Possible solutions to address this include improved integer programming formulations and the use of high throughput/high performance computing.

Conclusion

MDAPs are notoriously prone to cost overruns and schedule delays, and requirements volatility is at least partially to blame. In particular, when the desired capabilities of a system change during its life cycle, substantial reengineering and cost growth can result, especially when a new subsystem must be incorporated into an existing architecture. Of course, the likelihood and specifics of such additions are rarely known ahead of time, and predicting integration costs is challenging.

In this paper, we presented a novel algorithm to address this issue. Specifically, leveraging an integer programming implementation of the social network analysis technique blockmodeling, we optimally partitioned the subsystems represented in the SV-3 into architectural positions. Using this abstracted structure, we subsequently grew the architecture according to its statistical properties, and we estimated this unforeseen cost of evolutionary architectural growth via COSYSMO. Although our approach has limitations, the algorithm provides a useful prototype for pre–MS A cost risk analysis, and it continues to reinforce the potential of viewing DoDAF's models as computational objects.

References

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512. doi:10.1126/science.286.5439.509

Blanchard, B., & Fabrycky, W. (1998). *Systems engineering and analysis* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Bolten, J. G., Leonard, R. S., Arena, M. V., Younossi, O., & Sollinger, J. M. (2008). Sources of weapon system cost growth: Analysis of 35 major defense acquisition programs. Santa Monica, CA: RAND. Retrieved from http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG670.pdf

Boorman, S. A., & White, H. C. (1976). Social structure from multiple networks. II. Role structures. *American Journal of Sociology, 81*(6), 1384–1446. Retrieved from http://www.jstor.org/stable/2777009

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). Ucinet for Windows: Software for social network analysis (Ver. 6.488) [Software]. Harvard, MA: Analytic Technologies. Retrieved from http://www.analytictech.com/ucinet

Brusco, M., & Steinley, D. (2009). Integer programs for one- and two-mode blockmodeling based on prespecified image matrices for structural and regular equivalence. *Journal of Mathematical Psychology, 53*(6), 577–585. doi:10.1016/j.jmp.2009.08.003

CJCS. (2012, January 19). *Manual for the operation of the Joint Capabilities Integration and Development System.* Retrieved from https://dap.dau.mil/policy/Documents/2012/JCIDS%20Manual%2019%20Jan%202012.pdf

Csárdi, G., & Nepusz, T. (2006). *The igraph software package for complex network research* (Ver. 1.0.0) [Software]. Retrieved from http://igraph.org

Cole, R. (2012, October). *Proxy estimation costing for systems (PECS).* Paper presented at the 27th International Forum on COCOMO and Systems/Software Cost Modeling, Software Engineering Institute, Pittsburgh, PA. Retrieved from http://csse.usc.edu/csse/event/2012/COCOMO/presentations/CIIforum_2012_Reggie_Cole.pptx

Dabkowski, M., Fan, N., & Breiger, R. (2015). *Exploratory blockmodeling for one-mode, unsigned, deterministic networks using integer programming*. Manuscript submitted for publication.

Dabkowski, M., Fan, N., & Breiger, R. (2016). *Exact exploratory blockmodeling of multiple relation, mixed-mode networks using integer programming*. Manuscript in preparation.

Dabkowski, M., Reidy, B., Estrada, J., & Valerdi, R. (2013). Network science enabled cost estimation in support of model-based systems engineering. *Procedia Computer Science, 16*, 89–97. doi:10.1016/j.procs.2013.01.010

Dabkowski, M., & Valerdi, R. (2014, April). The budding SV3: Estimating the cost of architectural growth early in the life cycle. In D. A. Nussbaum (Chair of Panel 22—Enhancing Cost Estimating Techniques), *Proceedings of the 11th Annual Acquisition Research Symposium*. Retrieved from http://www.acquisitionresearch.net/files/FY2014/NPS-AM-14-C11P22R02-076.pdf

Dabkowski, M., & Valerdi, R. (2016). *Modeling evolutionary architectural growth in major defense acquisition programs.* Manuscript submitted for publication.

Dabkowski, M., Valerdi, R., & Farr, J. (2014). Exploiting architectural communities in early life cycle cost estimation. *Procedia Computer Science, 28*, 95–102. doi:10.1016/j.procs.2014.03.013

Doreian, P., Batagelj, V., & Ferligoj, A. (2005). *Generalized blockmodeling*. New York, NY: Cambridge University Press.

Dorogovtsev, S. N., & Mendes, J. F. F. (2003). Non-equilibrium networks. In *Evolution of networks: From biological nets to the Internet and WWW*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780198515906.003.0006

Driels, M. R., & Shin, Y. S. (2004, April). *Determining the number of iterations for Monte Carlo simulations of weapon effectiveness* (Technical Report No. NPS-MAE-04-005). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a423541.pdf

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences, 99*(12), 7821–7826. doi:10.1073/pnas.122653799

GAO. (2011). *DoD cost overruns: Trends in Nunn-McCurdy breaches and tools to manage weapon systems acquisition costs*. Retrieved from http://www.gao.gov/assets/130/125861.pdf

GAO. (2015). *Defense acquisitions: Assessments of selected weapon programs.* Retrieved from http://www.gao.gov/assets/670/668986.pdf

Honour, E. C. (2004). *Understanding the value of systems engineering.* Retrieved from http://www.seintelligence.fr/content/images/2015/12/ValueSE-INCOSE04.pdf

Kolaczyk, E. D., & Csárdi, G. (2014). *Statistical analysis of network data with R.* New York, NY: Springer.

Maier, M., & Rechtin, E. (2000). *The art of systems architecting* (2nd ed.). New York, NY: CRC Press.

Mrvar, A., & Batagelj, V. (2013). Pajek64: Program for analysis and visualization of very large networks (Ver. 3.12) [Software]. Retrieved from http://mrvar.fdv.uni-lj.si/pajek

Newman, M. E., & Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences, 104*(23), 9564–9569. doi:10.1073/pnas.0610537104

Peña, M., & Valerdi, R. (2015). Characterizing the impact of requirements volatility on systems engineering effort. *Systems Engineering, 18*(1), 59–70. doi:10.1111/sys.21288

Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). (2013, November 25). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

Valerdi, R. (2008). *The Constructive Systems Engineering Cost Model (COSYSMO): Quantifying the costs of systems engineering effort in complex systems.* Saarbrücken, Germany: VDM Verlag.

Valerdi, R., Dabkowski, M., & Dixit, I. (2015). Reliability improvement of major defense acquisition program cost estimates—Mapping DoDAF to COSYSMO. *Systems Engineering, 18*(5), 530–547. doi:10.1002/sys.21327

Wang, G., Valerdi, R., Roedler, G., Ankrum, A., & Gaffney, J. (2012). Harmonising software engineering and systems engineering cost estimation. International Journal of Computer *Integrated Manufacturing, 25*(4–5), 432–443. doi:10.1080/0951192X.2010.542182

Weapon Systems Acquisition Reform Act of 2009 (WASRA), Pub. L. No. 111-23. 123 Stat. 1704 (2009).

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology, 81*(4), 730–780. Retrieved from http://www.jstor.org/stable/2777596

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-13-1-0032, and the Office of the Secretary of Defense.

An Assessment of Early Competitive Prototyping for Major Defense Acquisition Programs

William R. Fast, COL, USA (Ret.)—has been a Senior Lecturer in Acquisition in the Graduate School of Business and Public Policy at the Naval Postgraduate School since September 2010. Prior to that, he taught acquisition and business, cost estimating, and financial management courses at the Defense Acquisition University. Fast is credentialed by the American Society of Comptrollers (ASMC) as a Certified Defense Financial Manager with Acquisition Specialty (CDFM-A). [wrfast@nps.edu]

Abstract

On May 22, 2009, the President signed into law the Weapon System Acquisition Reform Act of 2009 (WSARA). An important feature of WSARA is the requirement for all Major Defense Acquisition Programs (MDAPs) to conduct competitive prototyping prior to the Milestone B development decision. These prototypes must be demonstrated in a relevant environment to provide the milestone decision authority (MDA) with an assessment of their level of technology maturity. Competitive prototyping in this early phase can also identify program risk and help the MDA decide if there is a good match between the customer's needs and the available resource—technology, design, time, and funding. However, if the MDA determines that there is little or no benefit, competitive prototyping can be waived. The rationale behind such a waiver must be provided to the Government Accountability Office (GAO). The WSARA competitive prototyping requirement has now been in effect for nearly seven years. There is a considerable amount of data on the costs and benefits of early competitive prototyping efforts. In addition, the GAO has published numerous opinions regarding MDA waivers of competitive prototyping. This research analyzes MDAP data from the 2013, 2014, and 2015 Annual Reports on the Performance of the Defense Acquisition System, the Defense Acquisition Management Information Retrieval (DAMIR), Budget Exhibits that accompany the President's annual budget request, the Director of Operational Test and Evaluation (DOT&E), and the GAO. This research finds that early competitive prototyping has received only partial implementation for pre-MDAP and MDAP programs. However, when implemented, early competitive prototyping has reduced Program Acquisition Unit Cost (PAUC) and has reduced technology risk.

Research Questions

1. What MDAPs, entering the EMD phase after May 2009, have received a waiver from competitive prototyping prior to Milestone B? What were the most common reasons for these waivers? Did the GAO concur with these waivers and, if not, why not?
2. What MDAPs, entering the EMD phase after May 2009, have conducted competitive prototyping prior to Milestone B? For these MDAPs, what were the results of competitive prototyping in terms of technology maturity, design risk reduction, development cycle time, and cost estimate/funding stability, and system performance?
3. Based upon the answers to research questions 1 and 2, has the WSARA early competitive prototyping requirement helped reduce MDAP Program Acquisition Unit Cost (PAUC) and technology risk?

Background

Section 203 of the Weapon System Acquisition Reform Act (WSARA) of 2009 requires that the acquisition strategies for Major Defense Acquisition Programs (MDAPs) provide for competitive prototyping prior to the Milestone B approval. The Milestone Decision Authority (MDA) may waive the competitive prototyping requirement only on the basis that

> a) the cost of producing competitive prototypes exceeds the expected life-cycle benefits (in constant dollars) of producing such prototypes, including the benefits of improved performance and increased technological and design maturity that may be achieved through competitive prototyping; or b) on the basis that, but for such waiver, the Department would be unable to meet critical national security objectives. (WSARA, 2009, § 203)

The language of Section 203 was later modified to clarify that prototypes may be acquired from commercial, government, or academic sources and, if prototyping of the system is not feasible, prototypes may be acquired for critical subsystems (WSARA, 2009, § 813).

To better understand this statute, several terms must be defined. First, there are generally two types of prototyping incorporated into acquisition strategies. Early prototyping is done in the Technology Maturation and Risk Reduction (TMRR) phase of the program, prior to Milestone B. Prototyping does occur later, after Milestone B, when Engineering Design Models (EDM) are built in the Engineering and Manufacturing Development (EMD) phase of the program. However, the intent of the WSARA of 2009 is for MDAPs to have early prototypes, prior to approval of Milestone B.

Competitive prototyping involves two or more contractors developing prototypes and verifying (through testing or demonstration) that they meet contractual requirements. Ideally, after the contractor builds and demonstrates or tests the prototype, the government and contractor conduct a Preliminary Design Review (PDR). The purpose of the PDR is to assess the results of the prototyping effort and to conduct cost, schedule, and performance trades to meet user requirements, schedule, and resources. For example, some operational requirements may have to be deferred to a later increment of the program because certain critical technologies were determined as immature while testing a competitive prototype.

A Major Defense Acquisition Program (MDAP) is a program that expects to expend more than $480 million in Research, Development, Test, and Evaluation (RDT&E; FY2014 constant dollars) or $2.79 billion in procurement (FY2014 constant dollars) for all of its increments (10 U.S.C. § 2430). However, at the time that the WSARA of 2009 was enacted, the RDT&E expected expenditure threshold for an MDAP was $365 million and the procurement expected expenditure threshold was $2.19 billion (FY2000 constant dollars).

When a program is designated as an MDAP, the Milestone Decision Authority (MDA) is either the Under Secretary of Defense for Acquisition, Technology and Logistics (who is also the Defense Acquisition Executive, DAE) or the designated Component Acquisition Executive (CAE). The MDA determines when programs move from one acquisition phase to the next by reviewing program accomplishments at the various milestone reviews throughout the acquisition framework (DoDI, 2015a).

Since the WSARA of 2009 was enacted, DoD Instruction 5000.02, Operation of the Defense Acquisition System, has changed three times. To accommodate these changes, only the current names for acquisition phases, reviews, and capability documents are used throughout this paper.

Literature Review

The only published research on the implementation of early prototyping by MDAPs since 2009 are the Government Accountability Office (GAO) annual Assessments of Selected Weapon Programs. Recently, these assessments have been based upon an electronic questionnaire sent to selected program management offices. GAO pre-tests their questionnaire to ensure that the program offices of each Service understand the terminology used in the questionnaire. One set of questions in the survey asks if the program has attained knowledge that all of its critical technologies work in a relevant environment and in a realistic environment. Technologies that work in a relevant environment would be assessed at Technology Readiness Level (TRL) 6. Technologies proven to work in a realistic environment would be assessed at TRL 7. However, neither of these questions specifically ask if the technologies were demonstrated on early competitive prototypes. The program offices could have demonstrated the technology on components, subsystems, or even on a single prototype, rather than on competitive prototypes. Therefore, the results of the electronic questionnaire are inconclusive as to whether early competitive prototyping was actually used.

The program offices most certainly answer these two GAO questions based upon their interpretation of the early competitive prototyping requirement. For example, the F-22 Increment 3.2B Modernization (F-22 Inc 3.2B Mod) program office reported to the GAO that all critical technologies were demonstrated in a relevant environment prior to Milestone B. However, there is no evidence in any independent source documents that confirms that competitive prototypes were used for that technology assessment or that the critical technologies were tested on even a single prototype. Thus, based upon the GAO report, we can only assume that the program office tested the critical technologies on components or subsystems, but not using an integrated system prototype. Similar assumptions must be made for the K-46A Tanker Modernization program and the Global Positioning System (GPS) Next Generation Operational Control System (OCX)—no competitive prototypes or even a single prototype was developed and tested.

The other disconnect in the GAO questionnaire results involves programs that have had competitive and single prototype waivers. For example, the Combat Rescue Helicopter (CRH) program received a waiver from early prototyping. In the business case analysis for that waiver, the CRH program office reports that critical technologies are mature. Yet, in response to the GAO questionnaire, the CRH program office says that information is not available as to whether critical technologies have been demonstrated in a relevant or a realistic environment. A different kind of disconnect is seen in the Armored Multi-Purpose Vehicle (AMPV) and the VH-92A Presidential Helicopter Replacement (VH-92A) programs. In these two programs, the program offices report the demonstration of critical technologies in a relevant environment), yet both programs received prototyping waivers (and the GAO clearly states this in the assessment of the AMPV program). So, these two program offices must be referring to component or subsystem prototyping (GAO, 2015a).

Methodology

This research effort used only objective evidence, not questionnaires filled out by program offices that could be biased or open to interpretation. First, all new MDAPs entering the EMD phase since 2009 were identified by extracting new programs first reporting a development baseline in the Selected Acquisition Report (SAR) summary tables. SARs on all MDAPs are prepared by the Department of Defense (DoD) and sent to the Congress annually. Second, pre-MDAP programs that had early competitive prototyping efforts since 2009 were identified from the data reported in the Defense Acquisition Management Information Retrieval (DAMIR).

Several new MDAPs did not have an early prototyping effort because they were granted a waiver by the Milestone Decision Authority (MDA). In accordance with the WSARA of 2009, the MDA's rationale for granting a waiver from competitive prototyping must be reported to the Comptroller General of the United States. The GAO assesses this rationale and sends a written report of their findings to the congressional defense committees. Copies of these waiver assessments are publicly available on the GAO website. All GAO assessments of competitive prototyping waivers since 2009 were reviewed, to include the rationale for the waiver. For example, was the waiver based upon National Security urgency? Or, was the waiver based upon a cost benefit analysis (CBA) where the costs of prototyping outweighed the dollarized benefits in constant dollars? If the waiver was based upon a CBA, then answers were sought for these additional questions:

- What were the costs and benefits cited in the CBA? Were the costs and benefits properly determined and compared?
- What type of contract was subsequently used in the EMD phase? If a firm fixed price-type contract was used in the EMD phase, it may be that there was little to no technical risk. That would validate the decision to waive early prototyping.
- What was the baseline cost estimate for Program Acquisition Cost (in constant dollars), and what was the quantity established at the development baseline (i.e., program initiation and the first SAR to Congress)? What is the current Program Acquisition Cost estimate (in constant dollars) and what is the quantity reported in the 2015 SAR? Is there evidence of cost growth or cost reduction between the Program Acquisition Unit Cost (PAUC) at the development baseline and the current PAUC as of the 2015 SAR?

For each MDAP that did not have a competitive prototyping waiver, the type of early prototyping done prior to the Engineering and Manufacturing Development (EMD) phase was identified (i.e., subsystem/component prototypes, a single system prototype, and competitive prototypes). Then, these questions about those prototyping efforts were answered:

- What early prototypes were built, and how were they tested? Whenever available, independent reports of test efforts by the DoD Director of Systems Engineering, the DoD Director of Developmental Testing, and the DoD Director of Operational Testing were used.
- What technology readiness levels (TRL) and manufacturing readiness levels (MRL) resulted from the prototyping effort? Whenever available, independent TRL/MRL assessments by the DoD Director of Systems Engineering were used.
- What type of contract was used in the TMRR phase for the prototyping effort, and what type of contract was subsequently used in the EMD phase? When available, contract types were taken from Form R-3 Budget Exhibits. The point of asking this question is to understand the level of technical risk, especially after early prototyping. For example, if a Cost Plus Fixed Fee (CPFF) contract was used in the EMD phase, one might question whether the early prototyping effort revealed the technical maturity of the design.
- What was the baseline cost estimate for Program Acquisition Cost (in constant dollars), and what was the quantity established at the development baseline (i.e., program initiation and the first SAR to Congress)? What is the current Program Acquisition Cost estimate (in constant dollars) and what is the quantity reported in the 2015 SAR? Is there evidence of cost growth or cost reduction between the Program Acquisition Unit Cost (PAUC) at the development baseline and the current PAUC as of the 2015 SAR?

Findings

Since the enactment of the WSARA of 2009, there have been about 28 pre-MDAP or MDAP programs subject to statutory competitive prototyping requirements. These programs can be placed into three categories as shown in Table 1.

Table 1. Pre-MDAPs and MDAPs Subject to Statutory Prototyping Since 2009

Waiver	No Prototyping	Prototyping
• AMPV	• B61 Mod 12 LEP TKA	• AMDR
• B-2 DMS Mod*	• F-22 Inc 3.2B Mod	• Chem Demil-ACWA
• CRH	• FAB-T FET	• EPS
• EPS CAPS	• GPS OCX	• F-35 Aircraft
• F-15 EPAWSS*	• ICBM Fuze Mod	• F-35 Engine
• IFPC Inc 2 Blk 1	• KC-46A	• JAGM
• T-AO (X)*	• LCS MM	• JLTV
• VH-92A		• LCS
		• NGJ*
		• SDB II
		• SF
		• SSC
*pre-MDAP		• 3DELRR*

The next three sections summarize the common characteristics of the programs with prototyping waivers, no prototyping, or prototyping efforts.

Waivers

Since the enactment of the WSARA of 2009, 8 pre-MDAP or MDAP programs received a waiver from early prototyping (see Table 1). A description of each of these programs and their cost benefit analyses are at Appendix A. The findings follow:

- **Rationale Behind Waivers**. Seven (7) of these waivers were based upon the rationale that the cost of producing the competitive prototypes outweighed the life-cycle benefits of the prototyping effort. The waiver for EPS CAPS was based upon the rationale that the cost of prototyping outweighed the benefits and that the delay caused by prototyping would jeopardize National Security.
- **GAO's Assessments**. In most cases, the GAO's assessments found that the rationale behind the prototyping waivers was consistent with the intent of the WSARA of 2009. However, in assessing the waivers for AMPV, EPS CAPS, B-2 DMS-M, EPS CAPS, and VH-92A, the GOA questioned whether a sufficient number of prototyping alternatives had been analyzed or criticized the effort (or lack of effort) to dollarize prototyping benefits.
- **Some Waivers Supported by Acquisition Strategy**. Acquisition strategies to reduce technology risk by using only mature technologies and/or by reducing user requirements often obviated the need for early risk reduction prototyping. This was the case with the waivers for AMPV, CRH, F-15 EPAWSS, IFPC Inc 2 Blk 1, T-AO(X), and VH-92A.
- **Cost Reimbursement Contract for EMD Questioned.** The GAO's assessments of the prototyping waivers always questioned whenever a cost reimbursement-type contract was used or proposed for the EMD phase. Their rationale is that early prototyping may reduce technology risk, permitting the

use of a fixed price-type contract for the EMD phase. Two areas of risk often identified as justifying the use of cost reimbursement-type contracts were software development and integration. This was the case with the waivers for AMPV, B-2 DMS Mod, and EPS CAPS.

- **Schedule Impacts Not Quantified in Dollars.** Schedule impacts due to prototyping were discussed in 6 of the 8 waivers. Depending on the number of competing contractors, the delays due to prototyping ranged from 6 months to 60 months. However, none of the schedule impacts appear to have been added into the cost of the prototyping effort for direct comparison to the benefits in dollars.
- **BCA Not Posted With Waiver Assessment.** The business case analysis (BCA) done by the program office was not posted along with the GAO assessment of the waiver. Had the BCA been made available, it might have been useful in answering some of these research questions. More importantly, a copy of the BCA posted with the GAO assessment of that BCA could be a good teaching tool for the Defense Acquisition Workforce.
- **ASD Acquisition Comments on GAO Assessments**. The Assistant Secretary of Defense for Acquisition commented on three of the GAO's waiver assessment reports that *"competitive prototyping needs to be tailored to the needs and risks of each specific program, balanced with any potential adverse cost and schedule impact."* This comment may indicate that the ASD Acquisition has a different interpretation of the early prototyping requirement than does the GAO.

No Prototyping

Since the enactment of the WSARA of 2009, seven (7) pre-MDAP or MDAP programs did not conduct any early prototyping (see Table 1). A description of each of these programs is at Appendix B. The findings follow:

- **Programs Started in EMD Phase.** All 7 of these programs may have assumed that just because they bypassed the TMRR phase and started in the EMD phase, they didn't have to do competitive prototyping. However, that assumption is clearly not the intent of the WSARA of 2009 which states that competitive prototyping is required before Milestone B (per Section 203 of the WSARA of 2009). All of these programs had to go through Milestone B before entering the EMD phase. In addition, prior to the Milestone B decision, the Milestone Decision Authority must certify, to the Congress, that the technology in the program has been demonstrated in a relevant environment (TRL 6) and that the program complies with all relevant policies, regulations, and directives of the Department of Defense (per 10 U.S.C. § 2366b). DoDI 5000.02, *Operation of the Defense Acquisition System*, has required competitive prototyping since 2008.
- **Reductions in PAUC.** When the Program Acquisition Unit Cost (PAUC) in constant dollars, based upon the development (EMD) baseline, is compared to the current PAUC in constant dollars, based upon the 2015 Annual SAR, four (4) of the programs have shown a reduction in PAUC. The B61 Mod 12 LEP TKA program has the largest reduction in PAUC at -9.6%. The LCS MM program has the smallest reduction in PAUC at -0.1%.
- **Increases in PAUC.** When the Program Acquisition Unit Cost (PAUC) in constant dollars, based upon the development (EMD) baseline, is compared

to the current PAUC in constant dollars, based upon the 2015 Annual SAR, three (3) of the programs have shown an increase in PAUC. The GPS OCX program has the largest increase in PAUC at +21.2%. The ICBM Fuze Mod program has the smallest increase in PAUC at +1.0%.

- **Cost Reimbursement Development (EMD) Contracts.** Five (5) of the programs awarded a cost reimbursement-type contract for the development (EMD) phase. Cost Plus Fixed Fee (CPFF) contracts were awarded for ICBM Fuze Mod and LCS MM. Cost Plus Award Fee (CPAF) contracts were awarded for GPS OCX and LCS MM. Cost Plus Incentive Fee (CPIF) contracts were awarded for B61 Mod 12 LEP TKA and F-22 Inc 3.2B Mod. The determination of the contract type for the development (EMD) phase of an MDAP is the responsibility of the Milestone Decision Authority (MDA). If the MDA selects a cost reimbursement–type contract for the development (EMD) phase, this decision has to be justified to the Congress. In the justification, the MDA is to explain why the technology is so immature that a cost reimbursement contract is necessary (DoD, 2015a).
- **Identification of Critical Technology Elements.** Software development and system integration were often not considered as critical technology elements that could be assessed and matured through early prototyping. This was the case with F-22 Inc 3.2B Mod, GPS OCX, and LCS MM.

Prototyping

Since the enactment of the WSARA of 2009, thirteen (13) pre-MDAP or MDAP programs did conduct early prototyping (see Table 1). A description of each of these programs is at Appendix C. The findings follow:

- **Types of Prototyping Used.** All 13 of the programs used some type of prototyping permitted under the WSARA of 2009. One (1) program (Chem Demil-ACWA) prototyped two separate components. One (1) program (SSC) used a subsystem prototype. Eleven (11) programs (AMDR, EPS, F-35 Aircraft, F-35 Engine, JAGM, JLTV, LCS, NGJ, SDB II, SF, and 3DELRR) used 2 or more subsystem or competitive prototypes.
- **Reductions in PAUC.** When the Program Acquisition Unit Cost (PAUC) in constant dollars, based upon the development (EMD) baseline, is compared to the current PAUC in constant dollars, based upon the 2015 Annual SAR, eight (8) of the programs have shown a reduction in PAUC. The SDB II program has the largest reduction in PAUC at -21.3%. The EPS program has the smallest reduction in PAUC at -0.6%.
- **Increases in PAUC.** When the Program Acquisition Unit Cost (PAUC) in constant dollars, based upon the development (EMD) baseline, is compared to the current PAUC in constant dollars, based upon the 2015 Annual SAR, three (3) of the programs have shown an increase in PAUC. The Chem Demil-AWCA program has the largest increase in PAUC at +3.0%. The JAGM program has the smallest increase in PAUC at +0.2%.
- Two (2) pre-MDAP programs (NGJ and 3DELRR) did not have development (EMD) baselines from which to determine any increase or decrease in PAUC.
- **Fixed Price Development (EMD) Contracts.** Five (5) of the programs awarded a fixed price–type contract for the development (EMD) phase. Fixed Price Incentive (Firm Target; FPIF) contracts were awarded for LCS, SDB II, SF, and SSC. A Firm Fixed Price (FFP) contract was awarded for JLTV. Use

of a fixed price-type contract is an indication that technology is mature and cost risk has been reduced.

- **Cost Reimbursement Development (EMD) Contracts.** Five (5) of the programs were awarded a cost reimbursement–type contract for the development (EMD) phase. Cost Plus Award Fee (CPAF) contracts were awarded for Chem Demil-ACWA, F-35 Aircraft, and F-35 Engine. Cost Plus Incentive Fee (CPIF) contracts were awarded for AMDR, Chem Demil-ACWA, EPS, F-35 Aircraft, and F-35 Engine. The determination of the contract type for the development (EMD) phase of an MDAP is the responsibility of the Milestone Decision Authority (MDA). If the MDA selects a cost reimbursement–type contract for the development (EMD) phase, this decision has to be justified to the Congress. In the justification, the MDA is to explain why the technology is so immature that a cost reimbursement contract is necessary (DoD, 2015a).
- Two (2) programs have yet to award contracts for the EMD phase. These are both pre-MDAP programs: NGJ and 3DELRR.

Conclusion

This research finds that early competitive prototyping has received only partial implementation for pre-MDAP and MDAP programs. However, when implemented, early competitive prototyping in the TMRR phase does reduce technology and cost risk going into the EMD phase and does reduce Program Acquisition Unit Cost (PAUC) as measured from the development (EMD) baseline estimate.

Eight (8) pre-MDAP or MDAP programs entering EMD after May 2009 have received waivers from competitive prototyping prior to Milestone B. The most common reason for these waivers was that the cost of early prototyping exceeded the life-cycle benefits in constant dollars. While the GAO assessed that the rationale behind each of these waivers was consistent with the WSARA of 2009, deficiencies in the supporting business case analyses were exposed. The Defense Acquisition Workforce would benefit from training in the identification of viable prototyping scenarios and how to estimate the life-cycle benefits (in dollars) that can accrue from prototyping.

Thirteen (13) pre-MDAP or MDAP programs entering EMD after May 2009 conducted competitive prototyping prior to Milestone B. These programs are showing more reduction in PAUC compared with the programs for which prototyping was waived or the programs that did not use early prototyping. In addition, the fact that many of these programs were able to use a fixed price–type contract in their EMD phase points to a reduction in technology and cost risk as a result of competitive prototyping.

References

DoD. (2008, December 8). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: U.S. Government Printing Office.

DoD. (2013, November 26). *Operation of the defense acquisition system* (Interim DoDI 5000.02). Washington, DC: U.S. Government Printing Office.

DoD. (2015a, January 7). *Operation of the defense acquisition system* (DoDI 5000.02). Washington, DC: U.S. Government Printing Office.

DoD. (2015b, March 19). *Selected acquisition reports*, December 2014. Retrieved from http://www.defense.gov/News/News-Releases/News-Release-

View/Article/605423/department-of-defense-selected-acquisition-reports-sars-as-of-december-31-2014

DoD. (2016a, March 21). *Deputy Assistant Secretary of Defense for Developmental Test & Evaluation annual reports, FY2009–FY2014*. Retrieved from http://www.acq.osd.mil/dte-trmc/Publications.html

DoD. (2016b, March 21). *Deputy Assistant Secretary of Defense for Systems Engineering annual reports, FY2009–FY2014.* Retrieved from http://www.acq.osd.mil/se/docs/

DoD. (2016c, March 21). *Director, Operational Test & Evaluation annual reports, 1999–2015.* Retrieved from http://www.dote.osd.mil/annual-report/index.html

DoD. (2016d, March 24). *Selected acquisition reports, December 2015*. Retrieved from http://www.defense.gov/News/News-Releases/News-Release-View/Article/703529/department-of-defense-selected-acquisition-reports-sars

GAO. (2010, March). Defen*se acquisitions: Assessments of selected weapon programs* (GAO-10-388SP). Retrieved from http://www.gao.gov

GAO. (2011, March). *Defense acquisitions: Assessments of selected weapon programs* (GAO-11-233SP). Retrieved from http://www.gao.gov

GAO. (2012a, March). *Defense acquisitions: Assessments of selected weapon programs* (GAO-12-400SP). Retrieved from http://www.gao.gov

GAO. (2012b, August 23). *Prototyping waiver* (GAO-12-983R EPS). Retrieved from http://www.gao.gov

GAO. (2013a, March). *Defense acquisitions: Assessments of selected weapon programs* (GAO-13-294SP). Retrieved from http://www.gao.gov

GAO. (2013b, March 7). *CRH prototyping waiver* (GAO-13-313R). Retrieved from http://www.gao.gov

GAO. (2013c, September 6). *VXX prototyping waiver* (GAO-13-826R). Retrieved from http://www.gao.gov

GAO. (2014a, March). *Defense acquisitions: Assessments of selected weapon programs* (GAO-14-340SP). Retrieved from http://www.gao.gov

GAO. (2014b, April 22*). B-2 DMS-M prototyping waiver* (GAO-14-522R). Retrieved from http://www.gao.gov

GAO. (2014c, April 25). *AMPV prototyping waiver* (GAO-14-521R). Retrieved from http://www.gao.gov

GAO. (2014d, June 11). *IFPC INC 2, block 1 prototyping waiver* (GAO-14-643R). Retrieved from http://www.gao.gov

GAO. (2014e, October 8). *T-AO(X) prototyping waiver* (GAO-15-57R). Retrieved from http://www.gao.gov

GAO. (2015a, March). *Defense acquisitions: Assessments of selected weapon programs* (GAO-15-342SP). Retrieved from http://www.gao.gov

GAO. (2015b, September 21). *F-15 EPAWSS prototyping waiver* (GAO-15-800R). Retrieved from http://www.gao.gov

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (n.d.). Defense acquisition management information retrieval (DAMIR), large active contract database, fiscal years 2001–2016. Retrieved March 21, 2016, from http://www.acq.osd.mil/damir/

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2013, June 28). *Performance of the defense acquisition system, 2013 annual report*. Washington, DC: Author.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2014, June 13). *Performance of the defense acquisition system, 2014 annual report.* Washington, DC: Author.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2015, September 16*). Performance of the defense acquisition system, 2015 annual report.* Washington, DC: Author.

Weapon System Acquisition Reform Act of 2009, Pub. L. No. 111-23, § 203 Stat. 1704 (2009).

Appendix A. Waivers

1. **Armored Multi-Purpose Vehicle (AMPV) Program.** The Army's AMPV fleet is to replace the M113 family of vehicles. The AMPV will come in five variants: general purpose, medical evacuation, medical treatment, mortar carrier and mission command. The AMPV program acquisition strategy is based on modifying an existing and operationally proven military vehicle and plans to bypass technology development and begin in system development. The Army has also modified or eliminated some AMPV requirements to ensure that no technology development is needed for the AMPV (GAO, 2014c).
    - Costs Estimated for Prototyping Alternatives. (1) One contractor producing 5 variants ($198M); (2) Two contractors producing 5 variants each and 1 Live Fire Test & Evaluation vehicle each ($341M).
    - Benefits Estimated. (1) None ($0); (2) None ($0).
    - Schedule Impact. (1) 19 months; (2) 31 months.
    - Type Contract for TMRR phase. None.
    - Type Contract for EMD phase. Cost Plus Incentive Fee (CPIF) with 70/30 share ratio.
    - Percent change in PAUC from development baseline. +0.2%.
2. **B-2 Defensive Management System Modernization (B-2 DMS Mod) Program.** The B-2 DMS Mod will detect, identify, and locate enemy radar systems and provide real-time threat avoidance, threat warning, and threat situational awareness information to the aircrew. The program is an analog to digital upgrade that consists of three subsystems (antenna, electronic support measures, and avionics and graphics processors) integrated onto existing B-2 aircraft. Entry into the EMD phase was in late fiscal year 2015 (GAO, 2014b). The prime contractor has conducted competitions for three key subsystems—antennas, processors, and electronic support measures. Competitive prototyping was also conducted for the antennas (GAO, 2015a).
    - Costs Estimated for Prototyping Alternatives. (1) System prototyping by a separate contractor from Northrup Grumman ($524.8M); (2) Subsystem prototyping ($28.5M).
    - Benefits Estimated. (1) $6.3M; (2) $1.3M.
    - Schedule Impact. (1) Unknown; (2) Unknown.
    - Type Contract for TMRR phase. None.

- Type Contract for EMD phase. Cost Reimbursement.
- Percent change in PAUC from development baseline. Unknown.

3. **Combat Rescue Helicopter (CRH) Program.** The Air Force's CRH program will replace legacy HH-60G Pave Hawk helicopters. The CRH will recover personnel from hostile territory, conduct civil search and rescue, provide disaster relief, and evacuate non-combatants. Learning from the Combat Search and Rescue Replacement Vehicle program that was canceled in 2009, the Air Force reduced requirements for the CRH to lower the cost and ensure that no technology development is needed (GAO, 2013b).
    - Costs Estimated for Prototyping Alternatives. 2 competing contractor prototypes during the EMD phase ($725M).
    - Benefits Estimated. Reduction of software and integration risk ($12M).
    - Schedule Impact. Unknown.
    - Type Contract for TMRR phase. None.
    - Type Contract for EMD phase. Fixed Price Incentive (Firm Target) with 50/50 share ratio.
    - Percent change in PAUC from development baseline. +1.2%.
4. **Enhanced Polar System (EPS) Program Control and Planning Segment (CAPS).** EPS will provide extremely high frequency, jam-resistant, and secure satellite communications to forces operating in the North Pole region. EPS consists of two payloads hosted on classified satellites, a gateway to connect user terminals to other communication systems, and a control and planning segment (CAPS) to control the payloads and manage communications (GAO, 2012b).
    - Costs Estimated for Prototyping Alternatives. One additional prototype and funding for a second contractor through Preliminary Design Review ($49M).
    - Benefits Estimated. Negligible ($0).
    - Schedule Impact. 6 to 24 months.
    - Type Contract for TMRR phase. None.
    - Type Contract for EMD phase. Cost Plus Incentive Fee (CPIF) with 50/50 share ratio.
    - Percent change in PAUC from development baseline. -0.7%.
5. **F-15 Eagle Passive/Active Warning and Survivability System (F-15 EPAWSS) Program.** The program will replace and upgrade internal self-protection electronic warfare systems on fielded F-15C/E aircraft. EPAWSS consists of four major subsystems: radar warning receiver, electronic countermeasure processer, countermeasure dispenser system, and fiber-optic towed decoy. These subsystem capabilities will be fielded in two increments. The program entered the TMRR phase in August 2015 (GAO, 2015b).
    - Costs Estimated for Prototyping Alternatives. (1) Competitive prototypes by 2 contractors ($116.3M); (2) Single prototype by one contractor ($38.3M); Critical subsystem prototyping by one contractor ($36.3M).
    - Benefits Estimated. (1) $7.2M; (2) 7.2M; (3) 6.5M.

- Schedule Impact. (1) 41 months; (2) 26 months; (3) 21 months
- Type Contract for TMRR phase. Unknown.
- Type Contract for EMD phase. Unknown.
- Percent change in PAUC from development baseline. Unknown.

6. **Indirect Fire Protection Capability Increment 2, Block 1 (IFPC Inc 2 Blk 1) Program.** IFPC Increment 2 will detect, assess, and defend against threats from rockets, artillery, mortars, cruise missiles, and unmanned aircraft. Block 1 counters cruise missiles and unmanned aircraft (GAO, 2014d).
    - Costs Estimated for Prototyping Alternatives. (1) Two competing contractor prototypes ($208.5M); (2) Two in-house prototypes ($219.6M).
    - Benefits Estimated. (1) $9.8M; (2) $9.8M.
    - Schedule Impact. (1) 24 months; (2) 24 months.
    - Type Contract for TMRR phase. Unknown.
    - Type Contract for EMD phase. Unknown.
    - Percent change in PAUC from development baseline. Unknown.
7. **Fleet Replenishment Oiler (T-AO[X]) Program.** Fleet oilers transfer fuel to Navy surface ships that are operating at sea. This non-developmental and commercial-based program replaces the Navy's 15 legacy T-AO 187 Class Fleet Replenishment Oilers. The acquisition strategy includes competitive contract awards for industry studies, detailed design and construction, and follow-on ship procurement (GAO, 2014e). On 5 April 2013, the MDA signed the program's Milestone B ADM citing no technology development required and low programmatic risk (Compendium, 2015).
    - Costs Estimated for Prototyping Alternatives. Two prototypes ($1,350M); One prototype ($742M).
    - Benefits Estimated. None ($370M); None ($370M).
    - Schedule Impact. (1) 60 months; (2) 60 months.
    - Type Contract for TMRR phase. None.
    - Type Contract for EMD phase. Unknown.
    - Percent change in PAUC from development baseline. Unknown.
8. **Presidential Helicopter Replacement (VH-92A) Program.** The Navy's VH-92A helicopter will transport the President, Vice President, and heads of state. The program provides replacements for legacy VH-3D and VH-60N helicopters. Compared with the VH-71 program that was canceled in 2009, the VH-92A has reduced performance requirements. The acquisition approach is to integrate mature communications and mission systems into an existing commercial or military helicopter. Since programmatic and technology risks have been reduced, there is no TMRR phase (GAO, 2013c).
    - Costs Estimated for Prototyping Alternatives. (1) One contractor producing system and subsystem prototypes ($782M); (2) Two contractors producing system and subsystem prototypes ($3,380M).
    - Benefits Estimated. (1) $0; (2) $542M.
    - Schedule Impact. (1) 16 months; (2) 16 months.
    - Type Contract for TMRR phase. None.

- Type Contract for EMD phase. Fixed Price Incentive (Firm Target) with 50/50 share ratio.
- Percent change in PAUC from development baseline. -0.5%.

Appendix B. No Prototyping

1. **B61 Mod 12 LEP TKA.** The objective of the Air Force's B61 Modification 12 Life Extension Program (LEP) is to upgrade and improve the accuracy of the tactical nuclear weapons stockpile. The National Nuclear Security Administration (Department of Energy) is responsible for the B61 Mod 12 LEP warhead design, manufacture and assembly. The Air Force is responsible for the tail kit assembly (TKA) for use on aircraft, including the F-35 JSF and the B-21 Long-Range Strike Bomber. The MDA directed that the program move directly into the EMD phase, bypassing the TMRR phase and any early competitive prototyping. The first 35-month phase of a total 74-month EMD contract was awarded to Boeing in November 2012. Preliminary Design Review (PDR) was planned for January 2014, with Critical Design Review (CDR) in the summer of 2015. The development contract includes the requirement to build 77 engineering design models for testing and evaluation.
    - Competitive Prototyping Tests. None (TMRR phase bypassed).
    - TRL/MRL achieved through competitive prototyping. Not applicable.
    - Type Contract for TMRR phase. No TMRR phase or contract.
    - Type Contract for EMD phase. CPIF with 30/70 share ratio.
    - Percent change in PAUC from development baseline. -9.6%.
2. **F-22 Inc 3.2B Mod.** With this modification, the Air Force is improving the F-22 Raptor hardware and software with air-to-air missile upgrades (AIM-9X and AIM-120D), Geolocation 2, and other Electronic Protection capabilities. The modernization effort started as an engineering change proposal (ECP). However, early development and production cost estimates revealed that the ECP should become a separate ACAT 1D acquisition program (DoD SE FY12 Annual Report). The Materiel Development Decision (MDD) for the F-22 Inc 3.2B Mod was conducted in October 2011. The February 2012 R-2 Budget Exhibits for the program mention the use of an Avionics Integration Lab (AIL), but no prototyping. The program achieved Milestone B in June 2013.
    - Competitive Prototyping Tests. None. No prototypes were built.
    - TRL/MRL achieved through competitive prototyping. Not applicable.
    - Type Contract for TMRR phase. IDIQ
    - Type Contract for EMD phase. CPIF with 20/80 share ratio for EMD; CPFF for test execution.
    - Percent change in PAUC from development baseline. -2.3%.
3. **FAB-T–FET.** The Air Force's Family of Advanced Beyond Line-of-Sight Terminals (FAB-T) provides for survivable terminals for communicating strategic nuclear execution orders via jam-resistant, low probability of intercept waveforms through the Milstar and Advanced Extremely High Frequency (AEHF) satellite constellations. FAB-T includes both Command Post Terminals (CPT) and Force Element Terminals (FET). On 30 July 2015, Congress was notified by the USD(AT&L) that the FAB-T program had been

split into two subprograms for more effective management: CPT and FET. The FAB-T-CPT subprogram entered the Production and Deployment phase in October 2015. However, execution of the FAB-T-FET subprogram has been deferred, in part due to bomber aircraft requirements still under development by Air Force Global Strike Command and approval of the Acquisition Program Baseline (APB; DAES, 2016a).

- Competitive Prototyping Tests. None. No prototypes have been built.
- TRL/MRL achieved through competitive prototyping. Not applicable.
- Type Contract for TMRR phase. Unknown.
- Type Contract for EMD phase. Not awarded yet.
- Percent change in PAUC from development baseline. +4.1%.

4. **GPS OCX.** The Air Force's Global Positioning System (GPS) OCX program provides for a modernized satellite command and control (C2) system which replaces the current ground control system for legacy and new GPS satellites. OCX implements a modern flexible architecture with information assurance built in to address emerging cyber threats. The Air Force is taking a block approach to develop OCX, with each block delivering upgrades as they become available (DAES, 2016b).
    - Competitive Prototyping Tests. None. No prototypes have been built.
    - TRL/MRL achieved through competitive prototyping. Not applicable.
    - Type Contract for TMRR phase.
    - Type Contract for EMD phase. Cost Plus Award Fee (CPAF)
    - Percent change in PAUC from development baseline. +21.2%.
5. **ICBM Fuze Mod.** The Air Force's Intercontinental Ballistic Missile (ICBM) Fuze Modernization (Fuze Mod) program entails the design and development of a form, fit, and function replacement for the current Mk21 fuze used on the W87 warhead. This is a cooperative effort between the Department of Energy's National Nuclear Security Administration (NNSA), the U.S. Navy, and the United Kingdom. The ICBM Fuze Mod program will leverage technologies, components, and development/production capabilities from previous work performed by the U.S. Navy and NNSA on the Mk5 fuze for Submarine Launched Ballistic Missile warheads. The Mk 5 fuze entered EMD in August 2012. The ICBM Fuze Mod program entered the EMD phase in August 2013 after 6.3 Development Engineering entry approval by the Nuclear Weapons Council (chaired by the USD[AT&L]). Entry was in accordance with DOE Instruction 5030.55. The equivalent milestone review in DoDI 5000.02 is Milestone B (R-2 Budget Exhibit, 2014).
    - Competitive Prototyping Tests. None. No prototypes have been built.
    - TRL/MRL achieved through competitive prototyping. Not applicable.
    - Type Contract for TMRR phase. None. The TMRR phase was bypassed.
    - Type Contract for EMD phase. Sandia National Labs, Albuquerque, NM, MIPR; Northrup Grumman, Clearfield, CT, CPAF
    - Percent change in PAUC from development baseline. +1.0%.
6. **KC-46A.** The Air Force's Tanker Modernization (KC-46A) program will replace the aging fleet of air refueling tankers. Since the program is based

upon a commercial aircraft modified for air refueling and military avionics, the program bypassed the TMRR phase and any competitive prototyping. An EMD phase development contract was awarded to Boeing in February 2011. A successful PDR was conducted in April 2013. A Milestone C decision is expected in April 2016 (DAES, 2016a).

- Competitive Prototyping Tests. None. No competitive prototypes were built.
- TRL/MRL achieved through competitive prototyping. Not applicable.
- Type Contract for TMRR phase. None. TMRR phase was bypassed.
- Type Contract for EMD phase. Fixed Price Incentive (Firm Target; FPIF) with a 40/60 share ratio.
- Percent change in PAUC from development baseline. -8.2%.

7. **LCS MM.** The Mission Modules (MM) for the Navy's Littoral Combat Ship (LCS) provide Mine Countermeasures (MCM), Surface Warfare (SUW), and Anti-Submarine Warfare (ASW) capabilities. While EMD phase prototypes are currently in development and testing, early MM prototyping in the TMRR phase is unknown. This could be because the MM became reportable as an MDAP only when all three MM were combined into the single LCS MM program. However, the LCS MM program office feels that it is complying with DoD guidance and regulations because all MM were demonstrated in a relevant environment prior to MM integration with the LCS (GAO, 2015a).
   - Competitive Prototyping Tests. None. No early LCS MM prototypes were tested on the actual LCS.
   - TRL/MRL achieved through competitive prototyping. Not applicable.
   - Type contract for TMRR phase. Unknown.
   - Type Contract for EMD phase. CPFF and CPAF
   - Percent change in PAUC from development baseline. -0.1%.

Appendix C. Prototyping

1. **AMDR.** The Navy's next generation radar system for air defense and ballistic missile defense is the Air and Missile Defense Radar (AMDR). Using S-band radar, C-band radar, and a Radar Suite Controller, the AMDR will be deployed on Guided Missile Destroyer (DDG) 51 Flight III. TMRR phase prototypes of these AMDR components were developed by Lockheed Martin, Northrup Grumman, and Raytheon based upon contracts awarded in September 2010. An initial PDR was conducted prior to MS B with each of the three TMRR phase solutions. However, a system-level Preliminary Design Review (PDR) for the AMDR was not conducted until 27 August 2014 under the EMD contract awarded to Raytheon. The MDA had waived the requirement to conduct a system-level PDR prior to Milestone B (AMDR DAES, 2016).
   - Competitive Prototyping Tests. Hardware components were prototyped by three contractors. Software was not prototyped (DoD SE FY2013 Annual Report).
   - TRL/MRL achieved through competitive prototyping. The TMRR phase demonstrated the critical technologies necessary for a scalable AMDR (DoD SE FY2013 Annual Report).

- Type Contract for TMRR phase. 3 FPIF contracts for $85.392M each and 3 FFP contracts for $10M each to LM, NG, and Raytheon (R-3 Budget Exhibit, 2011).
- Type Contract for EMD phase. CPIF with 50/50 share ratio
- Percent change in PAUC from development baseline. -10.5%.

2. **Chem-Demil ACWA.** DoD's Chemical Demilitarization–Assembled Chemical Weapons Alternatives (ACWA) program includes two chemical demilitarization plants designed to safely dispose of the chemical weapons stored at Blue Grass and Pueblo Army Depots. Instead of incineration, both plants use safe and environmentally sound neutralization alternatives. The program suffered a critical unit cost breach in 2010 and was rebaselined to cover risk and to prove out first-of-a-kind equipment development (Annual SAR, 2010). Since then, event-based technical reviews of a rocket cutting machine and a rocket shearing machine were conducted prior to the March 2012 Milestone B decision. Both machines were found ready for implementation. In addition, each plant was modeled to assist in understanding facility control system issues (DoD SE, FY12 Annual Report).
    - Competitive Prototyping Tests. Component testing only as discussed above.
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. Bluegrass, CPIF (awarded 13 June 2003); Pueblo, CPAF (awarded 27 September 2002)
    - Type Contract for EMD phase. Contracts from TMRR phase have been extended.
    - Percent change in PAUC from development baseline. +3.0%.
3. **EPS.** The Air Force's Enhanced Polar System (EPS) program will provide assured communications over the North Polar Region. The system consists of two communications payloads on host satellites in highly elliptical orbits. The EPS is composed of and managed in four segments: an extended data rate payload, user terminals (acquired separately by users), a fixed installation Gateway, and a fixed installation Control and Planning Segment (OUSD[AT&L], n.d.). The EPS payloads are a simplification of the Advanced Extremely High Frequency (AEHF) payloads, and the EPS segments use the mature extended data rate waveform with common Global Information Grid interfaces. Since technology was determined to be mature, the EPS program was directed in 2007 to proceed to Milestone B. Before that review, the EPS program office and contractor did analysis and modeling and conducted a Preliminary Design Review (PDR) at which an allocated baseline was established. The DoD Director of Systems Engineering assessed that allocated baseline for software development and determined that it lacked integration detail (DoD SE, FY13 Annual Report). Regardless, a successful Milestone B decision review was conducted on 2 April 2014. Today, the prime contractor, Northrup Grumman, is experiencing some unfavorable cost variance due to unplanned software and systems engineering, integration, and testing; however, there are no significant software development issues (DAES, 2016a).

- Competitive Prototyping Tests. Early prototyping included flight equivalent payloads, a gateway engineering development model, and prototype control and planning software (GAO, 2015a).
- TRL/MRL achieved through competitive prototyping. Not available.
- Type Contract for TMRR phase. Cost Plus Incentive Fee (CPIF)
- Type Contract for EMD phase. CPIF with a 50/50 share ratio.
- Percent change in PAUC from development baseline. -0.6%.

4. **F-35 Aircraft.** The Joint Strike Fighter (F-35 Aircraft) program develops, produces, and fields the next generation multi-role tactical aircraft. The three variants are F-35A Conventional Takeoff and Landing (CTOL) for the Air Force, F-35B Short Takeoff and Vertical Landing (STOVL) for the Marine Corps, and F-35C Aircraft Carrier suitable Variant (CV) for the Navy (DAES, 2016b).
    - Competitive Prototyping Tests. Two competitive prototypes were built.
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. CPAF/CPIF
    - Type Contract for EMD phase. CPAF/CPIF
    - Percent change in PAUC from development baseline. -5.3%.
5. **F-35 Engine.** The Joint Strike Fighter Engine (F-35 Engine) program involved competition between two contractors: General Electric and Pratt and Whitney. Due to cost, the DoD decided to go with just one contractor, Pratt and Whitney.
    - Competitive Prototyping Tests. Prototypes were built by 2 different contractors (Pratt and Whitney and General Electric).
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. CPAF/CPIF
    - Type Contract for EMD phase. CPAF/CPIF
    - Percent change in PAUC from development baseline. -4.7%.
6. **JAGM.** The Joint Air Ground Missile (JAGM) program is led by the Army with joint interest with the U.S. Marine Corps and U.S. Navy. The JAGM represents the next generation of aviation launched fire and forget missiles to replace the HELLFIRE laser and Longbow radar missiles. JAGM will be used by joint service aircraft for destruction of high value stationary, moving, and relocatable land and maritime targets from standoff range in day, night, adverse weather, and obscured battlefield conditions (DAES, 2016b).
    - Competitive Prototyping Tests. Two contractors built prototypes to mature technologies and designs. Each contractor completed a flight test and a Preliminary Design Review (PDR). The program’s four critical technologies were all approaching maturity before the EMD phase (GAO, 2015a).
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. Unknown.
    - Type Contract for EMD phase. Not yet awarded (FFP contract planned).
    - Percent change in PAUC from development baseline. +0.2%.

7. **JLTV.** The DoD's Joint Lightweight Tactical Vehicle (JLTV) Program was one of the first programs to comply with competitive prototyping in the TMRR phase (at the time, competitive prototyping was a USD[AT&L] policy, not a statutory requirement). In October 2008, three separate Cost Plus Fixed Fee (CPFF) contracts, totaling $239.8 million, were awarded for the design, development, modeling, simulation, fabrication, test, and test support of 24 JLTVs and companion trailers (JLTV DAES ExSum, 2016). The three vendor prototypes were subjected to endurance testing to demonstrate reliability and maintainability, user assessments of suitability, and ballistic testing to assess force protection requirements (DOT&E, 2011).
    - Competitive Prototyping Tests. During TMRR phase testing, the DOT&E reported that all three vendor prototypes had problems with mobility in soft soil and integrating government furnished mission equipment into the prototypes. Reliability was also an issue for all prototypes, falling well short of the threshold reliability of 2,400 mean miles between operational mission failure. In addition, during the TMRR phase, the JLTV underbody threat requirement was increased to that of the Mine Resistant Ambush Protected (MRAP) vehicle. As a result, this increased force protection requirement was not tested on the JLTV competitive prototypes (DOT&E, 2011; DOT&E, 2012).
    - TRL/MRL achieved through Competitive Prototyping. Not available.
    - Type Contract for TMRR phase. Cost Plus Fixed Fee (CPFF)/Cost Share
    - Type Contract for EMD phase. Three Firm Fixed Price (FFP) contracts, with a 14-month period of performance, for 22 vehicles and six trailers from each vendor (JLTV DAES ExSum, 2016).
    - Percent change in PAUC from development baseline. -16.2%.
8. **LCS.** The Navy's Littoral Combat Ship (LCS) is optimized for flexibility in the littorals as a system of systems, both manned and unmanned. The LCS will be reconfigurable with three mission packages: surface warfare, mine warfare, and littoral anti-submarine warfare (DAES, 2016b).
    - Competitive Prototyping Tests. Two competitive prototypes were built, each of a different design, by separate contractors.
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. FPIF
    - Type Contract for EMD phase. FPIF
    - Percent change in PAUC from development baseline. +2.7%.
9. **NGJ.** The Navy's Next Generation Jammer (NGJ) is an electronic attack system that will provide significantly improved Airborne Electronic Attack capabilities against advanced threats through enhanced agility and precision jamming, increased interoperability, and greater coverage against a wide variety of radio frequency emitters. The NGJ system will be integrated on the EA-18G tactical aircraft to replace aging AN/ALQ-99 Tactical Jamming System (TJS) pods (DAES, 2016b).
    - Competitive Prototyping Tests. Four prototypes have been built and tested: flight demonstration pod, 8% scale model in a high speed wind tunnel, and component-level prototypes for testing of the Mid-Band 1 and Mid-Band 2 array subsystem.

- TRL/MRL achieved through competitive prototyping. Not available.
- Type Contract for TMRR phase. Unknown.
- Type Contract for EMD phase. Pre-MDAP, not awarded yet.
- Percent change in PAUC from development baseline. Unknown. The NGJ development baseline cost estimate is not known yet.

10. **SDB II.** The Small Diameter Bomb II (SDB II) program is a joint program with the Air Force as the lead service. The SDB II is a 250 pound, air-launched, precision-glide weapon that can attack both stationary and moving targets in degraded weather conditions at standoff ranges. The SDB II uses millimeter-wave radar, infrared sensor, and semi-active laser technologies, in addition to GPS and inertial navigation. The SDB II also has a weapon datalink network for in-flight target updates, retargeting, weapon tracking, and weapon abort (DOT&E, FY15 Annual Report). The SDB II program entered the TMRR phase in 2006. Competitive prototype contracts (CPFF) were awarded to Boeing and Raytheon in May 2006. By October 2009, SDB II had completed a 42-month competitive TMRR phase. The program had a successful MS B review and entered the EMD phase on 29 July 2010 (R-2 Budget Exhibits, 2007, 2010, 2011).
    - Competitive Prototyping Tests: Both contractor prototypes were subjected to free flight demonstrations, captive carriage on F-15s, and seeker testing at the component level. Each contractor's warhead was also tested for lethality (Director, OT&E FY09 Annual Report).
    - TRL/MRL achieved through competitive prototyping. Competitive prototyping lead to a near-CDR level of maturity prior to the down-select and MS B decision. One or both contractors were on track to achieve TRL 6 or greater and MRL 6 or greater by MS B (DoD SE FY09 Annual Report).
    - Type Contract for TMRR phase. Cost Plus Fixed Fee (CPFF)
    - Type Contract for EMD phase. Fixed Price Incentive (Firm Target; FPIF) with priced options for the first 5 production lots (FPIF, lots 1-3 and FP[EPA], lots 4 and 5).
    - Percent change in PAUC from development baseline. -21.3%.
11. **SF Inc 1.** The Air Force's Space Fence (SF) Ground Based Radar System Increment 1 (Inc 1) will provide space situational awareness by detecting and reporting objects in Low Earth and Medium Earth Orbits (LEO/MEO). The system will have an operations center and two radar sites operating at S-band frequencies. SF Inc 1 consists of the operations center at Reagan Test Site Operations Center Huntsville, AL, and one radar site located at Kwajalein Atoll, Republic of the Marshall Islands (OUSD[AT&L], n.d.). The program had a successful MS B in 3rd Qtr FY2014 and is now in the EMD phase. Negotiations are still underway with the potential host nation for the location for the second radar site (R-2, PE 0604425F, Project 65A009, February 2016).
    - Competitive Prototyping Tests. Two working prototypes were built in the TMRR phase to reduce risk. Also used prototype assembly to validate production planning for IOC (DoD SE FY2012 and 2013 Annual Reports).

- TRL/MRL achieved through competitive prototyping. An Independent Program Assessment (IPA) in FY2012 assessed the two contractor designs as technically mature and validated by working prototypes (DoD SE FY2012 Annual Report).
- Type Contract for TMRR phase. Firm Fixed Price (FFP)
- Type Contract for EMD phase. Fixed Price Incentive (Firm Target) with a 20/80 share ratio.
- Percent change in PAUC from development baseline. -4.9%.

12. **SSC.** The Ship to Shore Connector (SSC) is the Navy's replacement for the Landing Craft, Air Cushion (LCAC) class of ships. The SSC will project, sustain, and retrograde combat power from the sea, independent of tides, water depth, underwater obstacles, or beach gradient. The lead craft (funded with RDT&E, Navy) will be maintained as a test and training craft to test fixes to problems that arise during fleet introduction (R-2 Budget Exhibit, 2010). The SSC received Milestone B approval in June 2012. In July 2015, the program was granted approval to enter the Production and Deployment phase.
    - Competitive Prototyping Tests. The Command Module (a subsystem) was prototyped (DoD SE FY2014 Annual Report).
    - TRL/MRL achieved through competitive prototyping. Not available.
    - Type Contract for TMRR phase. Cost Plus Fixed Fee (CPFF) and Firm Fixed Price (FFP)
    - Type Contract for EMD phase. Fixed Price Incentive, Firm Target (FPI [F]) to Textron, Inc. for detailed design and construction of lead test and training craft (Craft 100) and technical manuals.
    - Percent change in PAUC from development baseline. -6.7%.
13. **3DELRR.** The Air Force's Three Dimensional Long-Range Radar (3DELRR) is a replacement for the current legacy AN/TPS-75 Radar. The 3DELRR is a software intensive program that uses new semiconductor technology (gallium-nitride-based transmit/receive modules that have lower power requirements).
    - Competitive Prototyping Tests. Three competing contractors built and demonstrated critical software elements and internal system integration.
    - TRL/MRL achieved through competitive prototyping. Results presented at the Preliminary Design Reviews from Technical Performance Measures indicate that the program is on track to meet its six KPPs and seven KSAs (DoD SE FY2013 Annual Report).
    - Type Contract for TMRR phase. Cost Plus Incentive Fee (CPIF; 1 contractor) and Firm Fixed Price (FFP; 2 contractors).
    - Type Contract for EMD phase. Fixed Price Incentive Firm Target (FPIF) planned.
    - Percent change in PAUC from development baseline. Unknown. The 3DELRR development baseline cost estimate is not known yet.

The Sixth-Generation Quandary

Raymond E. Franck—PhD, retired from the faculty of the Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School (NPS) in 2012. He retired from the Air Force in 2000. His active duty career included a number of operational tours, staff positions, and service as head of the Department of Economics and Geography, USAF Academy. His responsibilities at NPS included chairman, Systems Engineering Department, and associate dean, GSBPP. His research has emphasized defense acquisition and military innovation. His published work includes a number of journal articles and research reports in those areas. [cfranck215@aol.com]

Bernard Udis—PhD, is a Professor Emeritus of Economics at the University of Colorado at Boulder. He has also served as Distinguished Visiting Professor of Economics at the U.S. Air Force Academy and as a William C. Foster Fellow at the U.S. Arms Control & Disarmament Agency. His NATO research fellowship examined the costs and benefits of offsets in defense trade. A recognized authority on the economics of defense, particularly offsets, his published work includes three books, plus a number of book chapters and journal articles. [bernard.udis@colorado.edu]

Abstract

During the Cold War and its aftermath, technical superiority was a core competency of the U.S. military, which relied on platforms that were high-performance, multi-role, expensive, and with long development times. This approach generally worked because adversaries couldn't easily counter those capabilities. However, the "unipolar moment" featuring the U.S. as the sole superpower may well be ending, and a number of capable rivals have emerged.

In this changed world, a well-considered, timely response is therefore strongly indicated. But U.S. acquisition programs are taking ever longer to field combat capability. At the same time, adversaries are becoming more sophisticated and agile.

Accordingly our paper addresses the following questions concerning 6th-gen air combat. First, what are the lessons learned from 5th-generation fighter programs, especially the F-35? Second, how many new 6th-generation fighter aircraft should the U.S. develop and field? Two, one, or none? Third, what are the likely building blocks of the kinetic component of the next generation of air combat forces? Fourth, what might all this mean for acquisition professionals?

Introduction

Based on open sources, a 6th-Generation Fighter(s) with an Initial Operational Capability (IOC) of 2030 is taken as a commitment, if not a requirement. However, Figure 2 strongly suggests that this is not an attainable goal within the current state of the art for defense acquisition management. Moreover, our adversaries (real and potential) are becoming increasingly sophisticated, agile, and capable. And the combination of those developments is central to the 6th-gen quandary.

Accordingly our paper addresses the following questions.

First, what are the lessons learned from 5th-generation fighter programs, especially the F-35?

Second, how many new 6th-generation fighter aircraft should the U.S. develop and field? Two, one, or none? We do not intend to offer a definite answer, but these alternatives should be (a) described and (b) provided with a rationale. We think doing this will be a useful addition to the ongoing discussion of 6th-gen air combat capabilities.

Third, what are the likely building blocks of the kinetic component of the next generation of air combat forces? The kinetic component is, of course, not the entire force, but we regard this as a prudent limitation for this paper.[1]

Fourth, what might all this mean for acquisition professionals? It's natural to expect that with changes in military affairs, there would also be changes in defense acquisition. In fact, there is good reason to believe that "we can't keep doing things the way we did before," as one authority on military aviation put it.[2] It also indicates that the operating environment for defense acquisition ("small A") is increasingly shaped by the imperatives of network-centric warfare, and the requirements process ("big A"), as depicted in Figure 1. And as we'll discuss below, there's good reason to believe that acquisition of systems will be supplanted by acquisition of systems of systems—at least to some extent.

During the Cold War and its aftermath, the United States could rely on technical superiority as a core competency, relying on "highly capable, multi-function platforms," which were expensive and had long development times. However, their "sophisticated military technology" could not be quickly or easily countered (Shaw, 2016). However, the "unipolar moment" featuring the United States as the sole superpower may well be ending.

Within this context, a number of the threats to U.S. air superiority are in place or developing, as part of access-denial military complexes, to include long-range, stealthy tactical fighters; ballistic and cruise-missile weapons capable of targeting U.S. air bases (land and sea). In particular, modern integrated air defenses pose a particularly acute threat to U.S. air superiority and therefore to global precision strike warfare. Threat systems include advanced surface-to-air missile (SAM) systems, highly networked command and control, improved ground and airborne radar systems and advanced airborne interceptors—all enabled by modern information technology.

Given these circumstances, a well-considered, timely response (offset strategy) is strongly indicated. However, U.S. acquisition programs are taking ever longer to field combat capability. The situation for tactical fighters is discussed below.

With the change in military affairs will likely come a change in acquisition needs and acquisition management practices, discussed below.

---

[1] Based on space constraints, limitations on our areas of expertise, and (most important) on our current levels of clearance.
[2] An opinion offered not for attribution in May 2015.

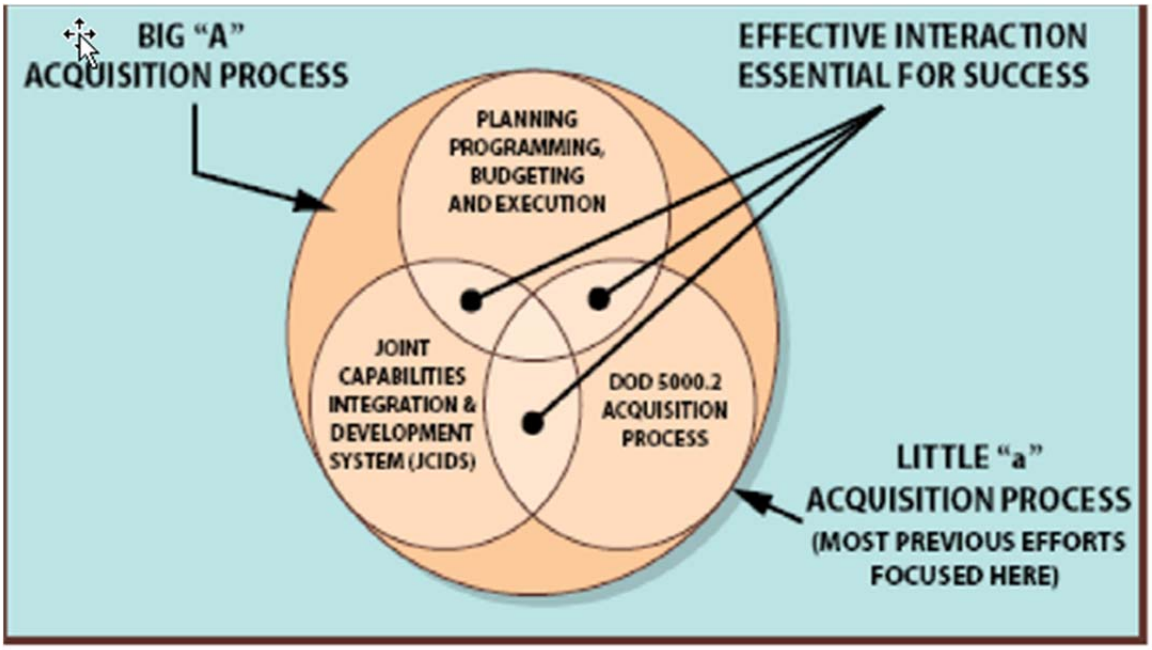


Figure 1. Depiction of the Defense Acquisition System
(Schwartz, 2014)

Air Dominance and Related Initiatives

One aspect of the U.S. response to the new international environment is a multi-faceted effort to study air combat needs for 2030 and beyond. A new start on next-generation air combat capabilities is underway with modest resource levels (e.g., LaGrone, 2015; DoD, 2014). It is, we think appropriately, a wide-ranging, and decentralized effort.

The multiple initiatives include the following:

- DARPA's air dominance initiative is charged to study means "for maintaining air dominance beyond the next decade" (Under Secretary of Defense for Acquisition, Technology & Logistics (USD[AT&L]), 2014). Its tasks include "exploring systems-of-systems concepts in which networks … interact to succeed in a contested battlespace" (Senate Appropriations Committee, Defense Subcommittee [SAC-D]), 2014, p. 4);
- An Air Force study of air dominance in 2030 and beyond, which was expected to issue a report in March this year (Tirpak, 2015);
- Navy initiatives which study replacements for the F-18 Super Hornet, to include new technologies and a joint analysis of alternatives, working with the Air Force (LaGrone, 2015);
- The Air Force Capability Collaboration Team, charged with identifying relevant technologies, drafting a course of action (road map) to field them. The team is expected to issue a final report in 2018 (Mehta, 2015).

1. Lessons From the F-35 Program

The F-35 experience has produced a number of lessons for future acquisition efforts. And there have been serious efforts to understand that experience and glean those lessons (some of which are cited in this section).

Cost Growth Was a Result of Acquisition Strategy

The F-35 emerged from Milestone B with a highly optimistic, success-oriented acquisition strategy: The "Milestone B program schedule, driven by the need to develop an affordable aircraft to replace aging combat aircraft, was aggressive and highly concurrent" (Blickstein et al., 2011, p. 37). Moreover, F-35 design requirements posed difficult design choices (Blickstein et al., 2011, p. 49, Table 4.6). Notably, an independent DoD cost estimate in 2001 rated the F-35 as high risk for both technical and schedule reasons (Blickstein et al., 2011, p. 37). Nonetheless, this fragile plan was adopted and pursued.

When unexpected difficulties (or problems that were assumed away) emerged during the SDD process, there were cost and schedule difficulties directly related to that problem. There were also "spillover" problems because of effects on other parts of the design. The result was a significant increase in cost and also significant delays (Arnold et al., 2010, esp. pp. 6–9; Blickstein et al., 2011, esp. pp. 39–41, 55).

Cost Growth Events Also Had Schedule Effects

F-35 experience suggests that platform density has also been a cost driver for aircraft. RAND's Root Cause Analysis of F-35 cost overruns contains some interesting observations. Requirements for stealth, supersonic flight (all models), STOVL[3] capability (B), and carrier landings (C) (Blickstein et al., 2011, p. 49, esp. Table 4.6). These requirements were frequently conflicting (Blickstein et al., p. 36); that is, the F-35 entered development with its engineering "trade space" considerably truncated. This design problem caused the cascading design issues that arose from a more powerful engine[4] (F-135 vice F-119): "the increase in thrust also lead to an increase in the engine size by a reported 1.5 inches in diameter. This small change in the engine generated a need to redesign the airframe, which in turn changed everything from aerodynamics to stealth signature, all of which needed to be re-baselined" (Blickstein et al., 2011, p. 53). While RAND's RCA focused on cost implications, there were schedule effects as well. These schedule slippages were reported (as of 2009) in Table 4.3 of Blickstein et al. (2011, p. 43)

Requirements Growth Was a Key Factor in Cost Growth

> *"Sometimes stakeholders despite their best intentions can derail your program."* —Maj Gen Christopher Bodgan, 2012

The Joint Strike Fighter began with timeliness and affordability as key program considerations. This was in its CALF (Common Affordable Lightweight Fighter) and JAST (Joint Attack Strike Technology) incarnations (Aboulafia, 2015, p. 8; Arnold et al., 2010, p. 2; Blickstein et al., 2011, p. 35). The design strategy then evolved to making the new fighter something described as a revolution in air combat (e.g., Laird et al., 2015). The acquisition strategy might then be described as devolving to the acquisition of that aircraft no matter how long it took or how much it cost.

Schedule Delays Had Wide-Reaching Effects

The IOC in the F-35 was scheduled for an IOC (Initial Operational Capability) in June 2011, 117 months after contract award (Blickstein et al., 2011, esp. p. 48). The F-35B

---

[3] Short takeoff, vertical landing

[4] That, in turn, arose from actual weight exceeding planned weight.

(Marine Corps) was declared operational in July 2015 (166 months), with software limitations to operational capability. F-35A IOC is expected in late 2016 (~179 months; DOT&E, 2015). F-35C IOC is expected in late 2018 (~202 months; DoD, 2013).

The effects of F-35 delays were not limited to the JSF program itself. Delays meant deferred production, which, in turn led to shortages in fighter aircraft. Addressing these shortages, entailed keeping "legacy" fighters in service longer than planned, with associated O&S expenditures. It also meant new programs, and associated expenses to extend airframe life, and upgrades to lessen degree of obsolescence against improving threat (Tirpak, 2011). As one observer put it, "the failure of … fifth-generation fighters … to arrive on time and on cost is having cascading effects throughout U.S. and allied fighter forces" (Sweetman, 2012).

Joint Programs Don't Save Money

We believe a definitive answer to the joint-program cost question comes from a recent RAND report (Lorell et al., 2013). Its basic conclusion is that the practical disadvantages of joint programs outweigh their theoretical advantages. The putative advantages for joint systems programs are

1. Lower total R&D costs for one joint system vs. multiple single-service systems
2. One production line offers economies of scale and greater learning curve effects
3. Lower O&S costs for highly common models vs. total O&S for multiple types (Lorell et al., 2013, esp. pp. 12–14)

One practical disadvantage of joint systems is that there is inevitably a compromise between individual service needs and preserving design commonality, which practically guarantees lower system performance and less commonality than originally planned (Lorell et al., 2013, p. 20). And instead of cost savings, the RAND research identified "a joint cost growth premium" (pp. 10–11).

The result has been cost increases from joint programs, relative to single-service programs. The R&D cost savings have been more than offset by relative cost increases later in the life cycle. (This finding is summarized particularly well by Figure 3.2 [p. 27] and Figure 3.4 [p. 29] in Lorell, 2013).

The bottom line seems plain. Joint programs deliver less performance at higher cost. And this conclusion appears to have been taken as a lesson learned throughout the DoD (e.g., Seligman & Swarts, 2016).

How Many New Fighters?

A significant part of the ongoing discussion concerns manned fighters. It seems there's an emerging consensus for two fighters. However, that approach, while sensible, should not be chosen without full consideration of alternatives that appear discredited, or have been neglected.

Two New Fighters?

The case for two fighters appears to be highly credible with the two services most affected: the Navy and the Air Force. It draws much support from two sources. First is the F-35 experience, analyzed in a RAND study (Lorell et al. 2013) whose results are discussed above.

Second is the common-sense (and empirically-supported) view that Navy- and Air Force–designed operational roles (and design requirements) are sufficiently different that pursuing a common airframe is probably not indicated (e.g., Lt Gen James Holmes, quoted in Seligman and Swarts, 2016). These conclusions were strongly stated in a 2014 RAND research brief: "Unless the participating services have identical, stable requirements, DoD should avoid future joint fighter and other complex joint aircraft programs" (RAND, 2014).

A variant of this alternative is two aircraft with some common subsystems—engines and software being most commonly mentioned. The principal author of the 2013 RAND study, Mark Lorell, noted, "Initial analysis suggests jointly developed engines, avionics, and subsystems can lead to significant savings, even if these common elements are installed in completely different airframes optimized for different service requirement" (Lorell, 2015). And indeed, the two services are planning to collaborate on studies of new aircraft designs (LaGrone, 2015).

Based on publicly available sources, two aircraft with significant commonality is the most widely accepted view of the best approach to a 6th-generation fighter. The case for two aircraft summarized above for two aircraft (as opposed to one) makes persuasive points and appears to be what will emerge if nothing changes the current discussion.

One New Fighter: Apparently Discredited

However, it's possible to make a case that inadvisability of a multi-service program is based on a lesson overlearned. There are historical cases of one fighter aircraft being successfully developed for multiple users with different needs—both domestic and international. These include the F/A-18 (discussed below), in which one aircraft was adapted to multiple customers' needs and situations. Our take is that the F/A-18's success in service with multiple air services was due to good management, a user community governance structure, and transfer of technology sufficient to adapt the aircraft to customer-specific needs.

Among other things, having a clearly defined lead customer (U.S. Navy) provided a clear demarcation of responsibility that was not present in the JSF program. The Air Force's C-17 program was similarly successful (Franck et al., 2012, esp. pp. 20–21). Perhaps an even better example is the F-4 Phantom II, which was used by all three U.S. services (Navy, Air Force, Marines), plus a number of allied nations.

We consider the F/A-18 case in more detail below.

The RAND conclusions notwithstanding (Lorell et al., 2013), it might be premature to dismiss a joint program. But in any case, it would appear to be wise and emulate the F/A-18 and C-17 (Franck et al., 2012a, esp. pp. 20–31) approaches, rather than the F-111 and F-35. In short, joint-use systems can be successful if they do not start as jointly-managed acquisitions.

No New Fighters: Largely Neglected

Although advocating 6th-generation air combat as a system-of-systems problem (e.g., Prabhakar, 2015, p. 4), there has been no strong advocacy for no new fighters, with the possible exception of Admiral Mullen (quoted in McQuain, 2009). Until recently, that is. Rob Weiss, head of Lockheed-Martin's Skunk Works, recently stated that fielding a 6th-gen fighter should wait until 2045 (Tirpak, 2016).

It's easy to interpret this statement cynically; in effect, Weiss said that the Lockheed-Martin monopoly on new fighters should stay in place for at least another quarter century. One could also note that his first suggested priority is to complete acquisition of all the planned F-35s (and maybe more). His second priority is to "accelerate modernization of the F-22 and F-35" (Tirpak, 2016). Also, Weiss recommends minimized expenditures on 4th-gen fighters (Drew, 2016). This can translate readily to spending lots of money with Lockheed-Martin and minimizing how much goes to Boeing.

Nonetheless, there is a coherent case for delaying new fighters for a very long time. We'll try to summarize that case in this section. However, a variation of this alternative could include selections from the following list:

- a weapons truck (F-15SE variant perhaps),
- new models of the F-35 (perhaps optimized for air combat [Majumdar, 2014]),
- a variant of the B-21 as a long-range, multi-role aircraft,
- arsenal platforms (like B-52s [Harper, 2015a]), and
- UAV mother ships (like C-130s [Atherton, 2014]).

We'll essay a summary of that case in this section. It rests on the improbability of a 6th-generation fighter by 2030, or even 2035. It's also founded on a strategy emphasizing opportunities that are arguably more promising.

<u>The Time Curve Argues Against Timely Fielding of a New Aircraft</u>

First is the Time Curve. The time it takes to get new weapons system in service continues to grow, as shown in Figure 2. Lacking a serious bending of the "time curve" below, we can expect a new fighter no sooner than the late 2030s, even with a forced march from here to Milestone B.

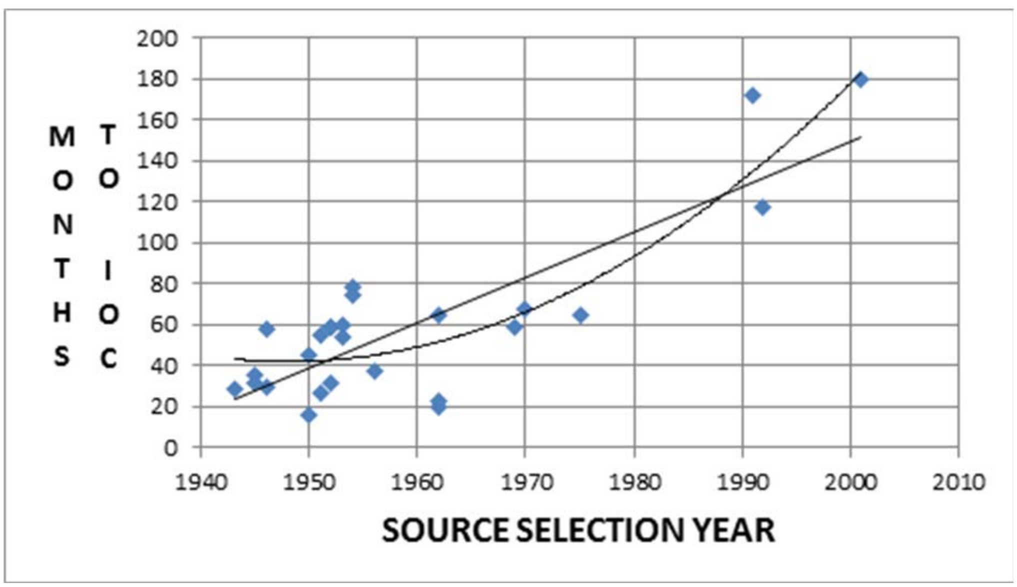


Figure 2. The Time Curve
(adapted from Blickstein, 2011, p. 48, Table 4.5)

Extrapolating from this curve gets us an IOC in the late 2030s, assuming a source selection and start of SSD in the early 2020s (which seems optimistic). This raises two critical questions for those advocating new airframes. First, how important is a 2030 IOC? Much public discussion of U.S. air combat capabilities (e.g., DSB, 2013) uses 2030 as a reference point. At one point, the commander of the Air Force Air Combat Command held the 2030 IOC to be a "requirement" (General Hostage, quoted in Mehta, 2012). Similarly, RADM Manazir (Navy Director of Air Warfare) has noted a need for a replacement for the Super Hornet fleet starting in 2030 (LaGrone, 2015).

Second, if an IOC sooner rather than later is preferred, how will we bend the Time Curve downward? The F-35 program set out to do just that (e.g., Blickstein et al., 2011, p. 43) but has turned out to be (at best) just an extension of the overall trend.

Exploit the Weapons Revolution

There have also been serious efforts to upgrade existing weapons and develop new-technology munitions. These include the following:

- Upgrade initiatives for older weapons, such as improved seekers (Tomkins, 2016) and warheads (Defense Industry Daily Staff, 2016) for Tomahawk cruise missiles; and new seekers for bombs (Tucker, 2016).
- New-technology developments such as directed energy (Wilson, 2015) and hypersonic missiles (Seligman, 2016);
- Unmanned, expendable UAVs (Tucker, 2016), including swarms.

These are in various stages of development, but all have attracted both interest, with the Secretary of Defense's stated willingness to fund their acquisition by cutting back on other programs (Harper, 2016a). In short, there's a case for emphasizing weapons now and letting the 6th-generation manned aircraft wait for a good while, perhaps until 2045. There's also a good case for upgrades of existing aircraft being a better exploitation of the weapons revolution than developing a new fighter aircraft.

Finish the Nail Soup

An old folktale is about starting with a nail in boiling water, and adding various ingredients to make an excellent meal.[5] The nail soup analogy applies here. An operational F-35 fleet, taken alone, looks a lot like that nail in boiling water; it's merely a start.

The F-35 has only four weapons stations in stealthy configuration. To contribute significantly to the fight, it needs to collect sensor data, fuse data, and bring other weapons to bear. One British commentator put it this way:

> If seamless interoperability is reached, the F-35 will allow … legacy assets to operate against targets and in areas which otherwise would be too heavily defended—either by providing targeting data in real time for stand-off munitions or by suppressing key defensive nodes to provide a window for the main force. (Bronk, 2016)

The same observation applies to the U.S. force. In addition, U.S. planners want the F-35 to direct stand-off strikes from non-stealthy platforms: "In practice, the arsenal plane will function as a very large airborne magazine, networked to 5th-generation aircraft that act as forward sensor and targeting nodes—essentially combining different systems already in our inventory to create wholly new capabilities" (Secretary Carter, quoted in Harper, 2016a).

---

[5] One version of the story is available at http://wayback.archive.org/web/20020316195723/http://hem.fyristorg.com/kulturkemi/net/soup.htm.

That implies seamless interoperability and excellent networking, with a decentralized command and control focus. Or as ACC commander Gen Hawk Carlisle put it, "The centralized hub-and-spoke architecture becomes a decentralized many-to-many network" (Laird et al., 2015).

However," networked operations" is much easier to say than to do. Just a year ago, for example, the Air Force reported difficulties in sharing operational awareness even within F-35 formations (Butler & Norris, 2015). Likewise reported was an unsatisfactory degree of connectivity between F-22s and F-35s, and other Air Force assets. Networking with other services was even worse: "USAF can't buy a solution unless it's compatible with an inter-service interoperability standard … and there may not be one, yet" (Tirpak, 2015). More recent reports likewise don't indicate a quick or easy solution. According to RADM Mike Manzir, director of Navy air warfare,

> I would *hope* … that when that aircraft in the mid '20s comes off the flight deck doing an ISR and tanking role, we can connect it through a waveform *still to be determined* to an F-35 or an E-2 or a Super Hornet and be able to give that aircraft commands. (emphasis added; Harper, 2016b)

In short, without networked operations, the F-35 doesn't add all that much; and, judging from the open literature, we aren't even close to achieving the networking that can "deliver the operational situational awareness critical to joint forces" (Wynne, 2012). Those problems are undoubtedly solvable, but doing that will take time, effort and resources. And reasonable people could decide that this capability is more important that what's offered by a new fighter platform.

By the way, we do not take any position on the "how many fighters" question. We do, however, believe that all three approaches presented here have serious rationales. Our emphasis on "no fighters" reflects our perception that this alternative has received much less attention than is warranted.

Two Useful Questions

We close this part of the discussion posing two useful questions. There appears to be a consensus that stealth is a necessary condition for air operations in contemporary high-threat environments. This gets to our first question: is stealth sufficient? Second, is the F-35 platform sufficiently "persistent" to stay effective over a long operational life against improving threats?

<u>Is stealth sufficient for successful operations in high-threat environments?</u>

While the United States has emphasized stealthy designs, all concerned parties (including the United States) have been developing countermeasures to the stealth threat. The list includes the following:

- Advanced lower-frequency radars, which can cause a resonant return from a stealthy airframe (McGarry, 2014). New developments include phased-array radars operating in the VHF frequency band (Sweetman, 2015b);
- New-generation, higher-frequency, airborne radars such as SAAB's ErieEye, which provide improved detection of lower-RCS targets and better tracking through improved interpretation algorithms (Sweetman, 2016);
- Bistatic (and multi-static) radar networks featuring passive receivers and rapid analysis of sensor information (Westra, 2009);
- Detection and tracking using non-radar emissions, such as heat (Sweetman, 2015a) and sound (Smith, n.d.);

- Far-forward airborne interceptors, such as the J-20, which could cause major problems for, *inter alia,* U.S. refueling orbits for stealth aircraft, particularly if China can improve its engines relative to those that are currently planned (Erickson & Collins, 2012).

Many of these programs are still in development, and it's safe to say not all will succeed. Nonetheless, these stealth countermeasures constitute a rich and promising menu. As time goes on, we can expect at least some of them to be operational and effective. While there are certainly counter-countermeasures possible, it may be difficult to keep pace (Sternstein, 2015). Or as an Air Force flag officer put it: "Emerging threats' timelines are decreasing. (Our) acquisition times are increasing."[6]

<u>Second, is the F-35 platform sufficiently persistent to stay ahead of the threat environment over its very long operational life?</u>

"Persistent platform" means a system that is sufficiently adaptable (with respect to both technical upgrades and new tactics) to remain effective despite changes in operational environment and mission. Lewis (2015) operationalized this idea in the context of the DDG-51 destroyer class, "(which) features the expandability (growth margin) and open systems characteristic that continues in ... service for a greater period of time than … originally ... contemplated." Franck et al. (2012, pp. 101–106) similarly narrates the persistent-platform aspect of the B-52's service.

While the F-35s may indeed age as well as DDG-51s and B-52s, and fit into Weiss' program of accelerated modernization (Tirpak, 2016), that's not self-evident. The set of requirements for the aircraft (supersonic flight, vertical landing, carrier operations, stealth) did much to reduce trade space (Blickstein, et al., p. 36, Table 4.1), and perhaps limit the expandability so important to platform persistence. (The "flying blivet"[7] epithet for the F-35 is unnecessarily pejorative, but not unfounded.)

The F/A-18 Case

This section is intended to amplify the rationale for one new fighter. While F-35 experience has rightly cast some doubt on the advisability of one fighter aircraft for many customers, the F/A-18 is a successful example of such a program. It has had multiple users, not only across services but across nations. This might have seemed unlikely since the F/A-18 was designed, to operate from catapult-launch carriers, most of which are in the U.S. Navy. Aircraft flying such missions require special features such as very strong airframes and undercarriages as well as hook mechanisms to facilitate carrier landings. Somewhat longer wings are also necessary to permit slower approach speeds.

<u>Export Sales of the F/A-18</u>

Deliveries of F/A-18 aircraft between 1980 and 2000 totaled 1,480, with over 400 exported (Powell & Renko, 2010). These countries purchased an already existing aircraft currently in use by the U.S. Navy, unlike the F-35 acquisition strategy. Hence, while some production modifications were possible, their role essentially is that of customer rather than a partner.

---

[6] Observation offered at a 2015 symposium, not for attribution.
[7] A "blivet" is basically a large amount of stuff put into a small sack.

Many issues were resolved between the customer and the principal contractors. These include location of the assembly facilities and the identity of the organization performing that function, as well as subsequent maintenance and modification. This takes the U.S. government outside of the loop dealing with industrial participation in the buyer country.

Throughout the entire post–World War II period, countries buying foreign military aircraft and other advanced technology products have attempted to acquire the underlying technologies in order to lessen their dependence on foreign sources. Frequently, this goal also reflected a belief that advanced technologies were the key to modern economic growth and a higher standard of living. These demands for industrial participation often were a major factor in selecting the winner of contract competitions (Udis, 2009).

With the exception of Kuwait and Malaysia, all of the export buyers participated in the assembly of their aircraft, and in mid-life upgrades. Without exception, those nations claimed significant industrial benefits and technological advances from their experiences with the aircraft.[8] This was matched with a high level of satisfaction with the performance of their aircraft and their working relationship with U.S. Navy and industry personnel (Powell & Renko, 2010).

Worth noting is that carrier-specific design features (strengthened undercarriage and tail hook) did prove useful for some customers. For example, the Finnish and Swiss Air Forces operating concepts included launching and recovering their F/A-18s from selective sections of their highway systems (Embassy of Finland, personal communication, December 16, 2009; Embassy of Switzerland, personal communication, November 15, 2011), which greatly benefited from those features.

F/A-18 International Governance

Most disputes dealt with rather mundane issues like transfer of spare parts, and test and repair capabilities between countries that had already been certified as members of the F-18 user community. However, the F/A-18 international community had a governance structure that worked well in resolving many of these problems.

A very active user community discussed common problems with U.S. Navy and Boeing representatives. One important example deals with efforts by the Navy to have the State Department standardize and clarify the application of U.S. export control regulations to the activities of the F/A-18 Community (Powell & Renko, 2010). The work of the Community has been divided into several interest groups as follows:

- HISC: Hornet International Steering Committee;
- HIRG: Hornet International Requirements Group, now called CCC;
- THRILL: Logistics;
- AV TCM: Logistics and Engineering;
- LPIT: Logistics Process Improvement Team;
- FISIF: F/A-18 International Structural Integrity Forum;

[8] This and related information were obtained in a series of confidential interviews held with representatives of these five countries in June 2010 and February 2011

- CREDP: F/A-18: Composite Repair Engineering Development Program; and
- NDTWG: Nondestructive Testing Working Group (Embassy of Finland, personal communications, February 8 and 22, 2016).

Very close relations are maintained between members of the User Community and representatives of the U.S. Navy and Boeing (F/A-18 Users Group Meeting discussions, 2010).

Export Control and Technology Transfer Issues

Occasionally, there were conflicts with the United States over a need to protect technologies deemed crucial to national security, against an interest in making the most of a national (albeit imported) military asset. In the F/A-18 case, these conflicting goals were resolved in the context of U.S. export control and technology transfer regulations. However, relations were not trouble-free, particularly with respect to the application of U.S. export control and technology transfer regulations. Over time it became clear that authority was also necessary to allow users to coordinate joint development efforts. Finally, a memorandum of understanding was obtained in 2005 to address this issue. It allowed multinational exchange of information and initiation, conduct and management of cooperative efforts [and also permitted] cooperation in acquisition arrangements and research, development, testing, evaluation, and production (including follow-on support) efforts (Powell & Renko, 2010).

Despite such efforts at clarification and simplification, minor problems seemed to appear without limit, requiring creative attention. According to ITAR regulations, the export of components and spare parts required separate approval, even when they are to be used in support of previously approved and exported end products. The U.S. Navy played an important role in resolving such issues (Powell & Renko, 2010). Of particular significance was its role in dealing with customer concerns about the continued access to U.S. supplied parts and other essential components for foreign inventories as the Navy moved to retire its use of the F/A-18, series A-D. NAVAIR's International Programs group conducted a major effort to alleviate that potential problem through careful advanced planning (Powell, 2010).

Lessons Learned From the F/A-18 Experience

In multinational weapon systems projects, somewhat different results have emerged regarding the problem of dealing with administrative disputes, especially the U.S. export control and technology transfer regimen. There are several factors that may explain such different experiences.

- In the case of the F/A-18, there was one clear lead service, which had undisputed responsibility for program success. This was not a joint acquisition program.
- The nature of the purchase agreement may influence access to information. Foreign Military Sales (FMS) arrangements are more likely to be associated with liberal information sharing than direct commercial sales since the military service whose weapon system is involved in the transaction serves as something of an intermediary between the buyer and the U.S. government.
- An active interest community with a defined governance structure and robust communications channels can do much to resolve issue among the participating natures.

- Having a lead service willing to act as a champion for the other participating services is potentially useful in dealing with issues related to export control and technology transfer.

The F/A-18 community was generally successful in solving problems while satisfying the needs of the participating services. By way of contrast, we note the following statement from the Right Honourable James Arbuthnot, Chair of the UK House of Parliament Defence Committee concerning the F-35 experience: “In all candour, I would encourage UK industry to design around the ITAR and produce ITAR-free items” (Moore et al., 2011, p. 86).

A Notional Force for 2035 Air Combat

Starting with a concept of operations, we offer some thoughts on a future air combat force structure, with emphasis on the “kinetic” component. Based on open discussions of the topic, we think it’s reasonable to suppose the following constitutes a reasonable, if sketchy, concept of operations for an air offensive against a near-peer competitor in the late 2030s.

First Phase: targets include military command & control (especially for air defense and space), long-range strike assets, political control nodes, power projection forces (air, sea, amphibious). Objectives are to degrade enemy’s ability to exercise political and economic control; reduce force projection capabilities (particularly air and missile), and opponent’s control of airspace outside his frontiers.

Strike sorties involve stealthy aircraft operating forward to find targets and direct strikes, generally by air-breathing missiles fired from a distance). That is, weapons would be mostly standoff: hypersonic missiles, and subsonic cruise missiles.

Objectives are to degrade enemy’s ability to exercise political and economic control; minimize force projection capabilities (particularly air and missile), and reduce enemy’s air control outside his frontiers.

Second Phase is intended to roll back enemy air control to permit operations by aircraft carriers and arsenal ships against forces operating (or preparing to operate) outside of the enemy homeland. Among other things, it’s intended to enable operations of non-stealthy aircraft, mated with stealthy aircraft, in previously contested air space.[9] It also aims to degrade the opponent’s energy production and distribution capabilities. There would undoubtedly be further phases for the campaign.

The forces implied by this concept of operations include the following building blocks, and could be a useful benchmark for force planning and acquisition:[10]

- Stealthy aircraft intended primarily to obtain and share situational awareness with other forces;
- Non-stealthy, “legacy,” weapons carriers;
- Specialized “weapons trucks” (e.g., upgraded F-15s);
- Arsenal aircraft (large weapons trucks such as B-52s);
- C4ISR assets, both airborne and in space;

---

[9] A concept called the “Wolfpack” (Wynne, 2012).
[10] An approach recommended by Jumper et al. (2009) and others.

- Naval forces to include arsenal ships;
- Land combat forces capable of taking and holding island bases; and
- Lots of tankers.

Getting from where we are to that force involves substantial efforts at networking, upgrades to existing aircraft, continue aerial tanker modernization, and lots of new non-nuclear weapons—plus a new fighter, maybe. There's no guarantee that the entire package would be affordable.

What This Could Mean for Acquisition Professionals

The 6th-generation air combat is clearly something of a quandary for warfighters (and therefore to the requirements community). It also is something of a quandary for the acquisition community.

First, there are at least three reasonable answers to the question of how many new fighters should be in the DoD's portfolio of 6th-generation air combat capabilities, depending in part on views of the requirements. There is accordingly good reason to pay more attention to what the requirements community specifies. As the JCIDS instruction puts it, "Close collaboration between requirements and acquisition communities is a key aspect of ensuring that knowledge gained early in the acquisition process is leveraged to enable the setting of achievable risk-informed capability requirements, and the making of effective cost, performance, schedule, and quantity trade-offs" (CJCS, 2015). That's always been sound guidance, but it's becoming even more important. It's critical to understand not only the requirements pertaining to a particular platform type, but how it performs within a larger system ("ecology" or "complex"). How to do it is a difficult problem.

Second, the 6th-generation quandary might well be a watershed event beginning a new era in defense acquisition management. The DoD appears to be doing less acquisition of platforms (systems) and more acquisition of systems of systems. An interesting representation of this idea is shown in Figure 3.

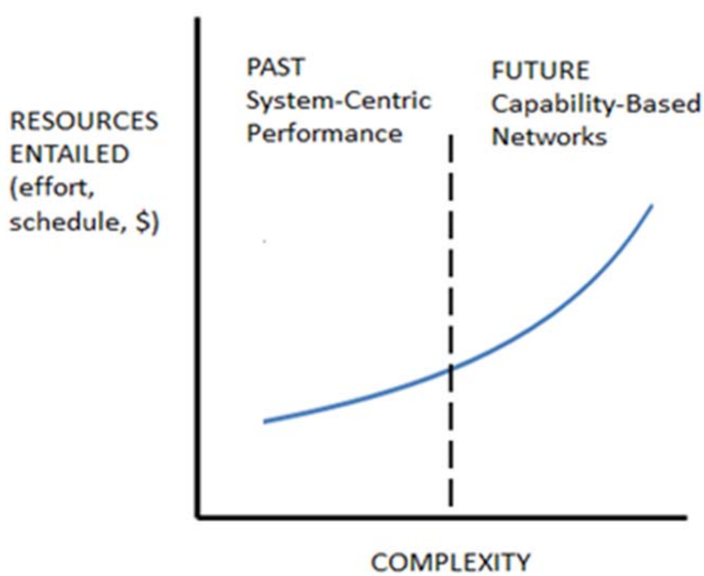


Figure 3. Changing Context of Defense Acquisition
(adapted from Angelis et al., 2008, p. 2)

Or as one DARPA official put it,

> The globalization of technology has made (previous practices) increasingly unsustainable. Potential adversaries are now able to access advanced technologies with relative ease and incorporate them quickly into military systems—sometimes accomplishing multiple upgrades during a U.S. weapon system's development and acquisition period. (Shaw, 2016)

Third, all of this likely means that Systems of Systems (SoS) Engineering will become a more important management method in the future acquisition enterprise. While we have no particular expertise in System of Systems Engineering, we think the following items are the key takeaways from System of Systems literature:

- SoS acquisition management is hard to do. Extant research (e.g., Angelis et al., 2008, esp. pp. 25, 29–30) strongly suggests that system of system acquisition programs are more likely to encounter cost and schedule difficulties.
- SoS research has identified causes for these difficulties (e.g., DeLaurentis & Ghose, 2008, p. 188; Huynh et al., 2011, p. 237). We take the central themes as being related to problems with coordination, organization, persistence of platform-centric management practices, and complexity (with emergent behavior of the system of systems).
- There have also been serious efforts to formulate methods and tools for dealing with SoS difficulties (e.g., Huynh et al., 2011; Shaw, 2016).
- Open-source reports indicate that those methods are not fully developed yet, or have yet to be fully applied (as discussed in the no-fighters rationale above).

In short, significant, ongoing changes in contemporary military affairs are driving the United States and its allies to networked, system-of-systems solutions to ever more difficult threats. However, acquiring systems is hard, and it's not clear that the tools available are up to the task of achieving good outcomes in such acquisition programs.

And it seems that the 6th-generation quandary poses significant problems for the operational, planning, and acquisition communities.

References

Aboulafia, R. (2015). *Lockheed-Martin F-35 Joint Strike Fighter*. Washington, DC: The Teal Group.

Angelis, D., Dillard, J., Franck, R., Melese, F., Brown, M., & Flowe, R M. (2008). *Exploring the implications of transaction cost economics on joint and system-of-systems programs.* Monterey, CA: Naval Postgraduate School. Retrieved from Acquisition Research Program website: http://www.acquisitionresearch.net/files/FY2008/NPS-AM-08-124.pdf

Arnold, S. A., et al. (2010). WSARA 2009: *Joint Strike Fighter root cause analysis* (Paper P-4612). Alexandria, VA: Institute for Defense Analysis.

Atherton, K. D. (2014). DARPA wants airborne aircraft carriers. *Popular Science*. Retrieved from http://www.popsci.com/article/technology/darpa-wants-airborne-aircraft-carriers

Blickstein, I., et al. (2011). *Root cause analyses of Nunn-McCurdy breaches, Volume 1: Zumwalt-Class Destroyer, Joint Strike Fighter, Longbow Apache, and Wideband Global Satellite*. Santa Monica, CA: RAND.

Bogdan, C. (2012, September 17). Deputy PPO, Joint Strike Fighter Program, Air Force Association, Air & Space Conference. Retrieved from http://www.af.mil/shared/media/document/AFD-120919-046.pdf

Bronk, J. (2016, February 23). Maximum value from the F-35: Harnessing transformational fifth-generation capabilities for the UK military. *RUSI Papers*. Retrieved from https://rusi.org/publication/whitehall-reports/maximum-value-f-35-harnessing-transformational-fifth-generation

Butler, A., & Norris, G. (2015, March 20). Patching the F-35's data fusion gap. *Aviation Week.* Retrieved from http://aviationweek.com/defense/patching-f-35-s-data-fusion-gap

Charles, S. (2006, April 6). Buy lines: 'Big A' acquisition is here to stay. *Washington Technology.*

CJCS. (2015, January 23). *Joint Capabilities Integration and Development System (JCIDS)* (CJCSI 3170.01I). Washington, DC: Author. Retrieved from https://dap.dau.mil/policy/Documents/2015/CJCSI_3170_01I.pdf

Defense Acquisition University (DAU). (2016). *Defense acquisition guidebook*, Section 4.2.1.2, Systems of Systems. Retrieved March 28, 2016, from https://acc.dau.mil/CommunityBrowser.aspx?id=638308#4.2.1.2

Defense Industry Daily Staff. (2016, January 19). Dev program seeks to increase Tomahawks payload. *Defense Industry Daily*. Retrieved from http://www.defenseindustrydaily.com/pratt-whitney-to-produce-engines-for-f-35-dev-program-seeks-to-increase-tomahawks-payload-afghanistan-gets-4-super-tucanos-032484/

DeLaurentis, D., & Ghose, S. (2008). Defense acquisition management for systems-of-systems. In *Proceedings of the Fifth Annual Acquisition Research Symposium* (pp. 171–189). Retrieved from http://www.acquisitionresearch.net/files/FY2008/NPS-AM-08-033.pdf

DoD. (2013). *F-35 IOC report*. Retrieved from https://www.f35.com/assets/uploads/downloads/12994/f-35_ioc_joint_report_final.pdf

Erickson, A., & Collins, G. (2012, December 9). The "long pole in the tent": China's military jet engines. *The Diplomat*. Retrieved from http://thediplomat.com/2012/12/the-long-pole-in-the-tent-chinas-military-jet-engines/

F/A-18 Users Group Meeting discussions. (2010, June 17). Patuxent Naval Air Station, MD.

Franck, R. E., Lewis, I., & Udis, B. (2012a). *Impact of export control and technology transfer regimes: International perspectives* (NPS-AM-12-214). Monterey, CA: Naval Postgraduate School. Retrieved from Acquisition Research Program website: http://www.acquisitionresearch.net/files/FY2012/NPS-AM-12-001.pdf

Harper, J. (2015a, February 3). 2017 Pentagon budget targets high end capabilities to counter Russia, China. *National Defense*. Retrieved from http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=2074

Harper, J. (2015b, March 22). Admiral pushing plan to network aerial fleet. *National Defense.* Retrieved from http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=2133

Huynh, T., Osmundson, J. & Rendon, R. (2011). System-of-systems acquisition: Alignment and collaboration (NPS-AM-11-C8P19R02-067). In *Proceedings of the Eighth Annual Acquisition Research Symposium* (pp. 236–248). Retrieved from Acquisition Research Program website: http://www.acquisitionresearch.net/publications/detail/822/

Jumper, J. P., Deptula, D. A., & Adams, H. B. (2009). Integrating CONOPS into the acquisition process. *Joint Forces Quarterly, 55,* pp. 66–68. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a515171.pdf

LaGrone, S. (2015, March 27). Navy and Air Force planning joint exploration of next generation fighter follow-ons to F-22 and F/A-18E/F. *USNI News*. Retrieved from http://news.usni.org/2015/03/27/navy-and-air-force-planning-joint-exploration-of-next-generation-fighter-follow-ons-to-f-22-and-fa-18

Laird, R., Timperlake, E., & Delaporte, M. (2015, December 29). F-22, Typhoon, Rafale: Lessons from the trilateral wargame. Retrieved from http://breakingdefense.com/2015/12/f-22-typhoon-rafale-lessons-from-the-trilateral-wargame/

Lewis, I. (2015). *Persistent platforms*—The DDG-51 case (NPS-AM-15-124) Monterey, CA: Naval Postgraduate School. Retrieved from the Acquisition Research Program website: http://www.acquisitionresearch.net/files/FY2015/NPS-AM-15-124.pdf

Lorell, M. A. (2015, March 2). Opinion: Where commonality can work in a sixth-gen fighter. *Aviation Week*. Retrieved from http://aviationweek.com/defense/opinion-where-commonality-can-work-sixth-gen-fighter

Lorell, M. A., et al. (2013). *Do joint fighter programs save money?* Santa Monica, CA: RAND.

Majumdar, D. (2014, December 19). A new, "super" F-35 to rule the U.S. military? *National Interest*. Retrieved from http://nationalinterest.org/feature/new-super-f-35-rule-the-us-military-11892

McGarry, B. (2014, July 31). Chinese radar may pierce F-35 stealth armor report. Retrieved from http://www.defensetech.org/2014/07/31/chinese-radar-may-pierce-f-35-stealth-armor-report/

McQuain, B. (2009, July 11). F-35: Last manned fighter? Retrieved from http://www.qando.net/?p=3490

Mehta, A. (2012, November 30). ACC chief hints at 6th gen fighter. *Air Force Times*. Retrieved from http://www.airforcetimes.com/news/2012/11/dn-air-combat-command-hints-6th-gen-fighter-113012/

Mehta, A. (2015, March 7). Planning begins for USAF next-gen air dominance. *Defense News.* Retrieved from http://www.defensenews.com/story/defense/air-space/strike/2015/03/07/next-gen-air-dominance-6th-gen-fighter/24393673/

Moore, D., Ito, P., Young, S., Burgess, K., & Antill, P. (2011). The impact of U.S. export control and technology transfer regime on the Joint Strike Fighter (JSF) project: A UK perspective. In *Proceedings of the Eighth Annual Acquisition Research Symposium* (pp. 72–91). Retrieved from the Acquisition Research Program website: http://www.acquisitionresearch.net/files/FY2011/NPS-AM-11-C8P03R01-025.pdf

Perrett, B. (2015, August 21). Japan working on anti-stealth missile guidance. *Aviation Week*. Retrieved from http://aviationweek.com/defense/japan-working-anti-stealth-missile-guidance

Powell, R. (2010). *Sundown/sunrise plan.* Patuxent River Naval Air Station, MD: U.S. Navy Air Systems Command (NAVAIR).

Powell, R., & Renko, K. (2010). *F/A-18A-D Hornet international cooperation: A case study on the role of export control*. Patuxent River Naval Air Station, MD: U.S. Navy Air Systems Command (NAVAIR).

Prabhakar, A. (2015, May 14). Statement to the Senate Appropriations Committee, Defense. Retrieved from http://www.appropriations.senate.gov/sites/default/files/hearings/Dr%20Prabhakar%20Statement.pdf

RAND. (2014). *The Department of Defense should avoid a joint acquisition approach to the sixth-generation fighter*. Santa Monica, CA: Author.

Schwartz, M. (2014). *How DoD acquires weapon systems and recent efforts to reform the process.* Washington, DC: Congressional Research Service.

Seligman, L. (2016, March 16). Lockheed's Marillyn Hewson touts breakthroughs in hypersonic weapons. *Defense News*. Retrieved from http://www.defensenews.com/story/defense/air-space/2016/03/16/lockheeds-marilyn-hewson-touts-breakthroughs-hypersonic-weapons/81836070/

Seligman, L., & Swarts, P. (2016, February 12). Sixth-gen fighter likely won't be common across U.S. services, Air Force general says. *Military Times*. Retrieved from http://www.militarytimes.com/story/military/2016/02/12/sixth-gen-fighter-likely-wont-common-across-us-services-air-force-general-says/80308582/

Senate Appropriations Committee, Defense Subcommittee (SAC-D). (2014, May 14). Statement of Dr. Arati Prabhakar, Director, Defense Advance Projects Research Agency. Retrieved from http://www.appropriations.senate.gov/sites/default/files/hearings/Dr%20Prabhakar%20Statement.pdf

Shaw, J. (2016). System of Systems Integration Technology and Experimentation (SoSITE), DARPA. Retrieved from http://www.darpa.mil/program/system-of-systems-integration-technology-and-experimentation

Smith, S. J. (n.d.). Stealth countermeasures. Retrieved February 13, 2016, from http://www.whale.to/b/stealth_countermeasures.html

Sternstein, A. (2015, April 27). Here's how you hack a military drone. *Next Gov*. Retrieved from http://www.nextgov.com/defense/2015/04/heres-how-you-hack-drone/111229/

Sweetman, B. (2012, November 26). Generation gap. *Aviation Week*, pp. 22–23.

Sweetman, B. (2015a, May 7). Selex talks details on IRST technology. *Aviation Week*. Retrieved from http://aviationweek.com/defense/selex-talks-details-irst-technology

Sweetman, B. (2015b, September 11). The development of stealth and counterstealth. *Aviation Week*. Retrieved from http://awin.aviationweek.com/ArticlesStory.aspx?id=3bf19237-89b4-4a6a-ac84-fe95f321496a

Sweetman, B. (2016. February 17). Counter-stealth radar key to growing AEW market. *Aviation Week*. Retrieved from http://aviationweek.com/defense/counter-stealth-radar-key-growing-aew-market

Tirpak, J. (2011). New life for old fighters. *Air Force Magazine, 94*(2). Retrieved from http://www.airforce-magazine.com/MagazineArchive/Pages/2011/February%202011/0211fighters.aspx

Tirpak, J. A. (2015, May 22). And no fix on fifth gen comms. *Air Force Magazine Daily Report*. Retrieved from http://www.airforcemag.com/DRArchive/Pages/2015/May%202015/May%2022%202015/And-No-Fix-On-Fifth-Gen-Comms.aspx

Tirpak, J. A. (2016, March 16). A long wait for sixth gen. *Air Force Magazine Daily Report.* Retrieved from http://www.airforcemag.com/DRArchive/Pages/2016/March%202016/March%2016%202016/A-Long-Wait-for-Sixth-Gen.aspx

Tomkins, R. (2016, January 14). Raytheon tests new seeker for Tomahawk cruise missiles. Retrieved from http://www.upi.com/Business_News/Security-Industry/2016/01/14/Raytheon-tests-new-seeker-for-Tomahawk-cruise-missiles/5861452789391/

Tucker, P. (2016, February 2). These are the new weapons the Pentagon chief wants for tomorrow's wars. Retrieved from http://www.defenseone.com/technology/2016/02/New-weapons-pentagon-wants-tomorrows-wars/125611/

Udis, B. (2009). Offsets and international industrial participation. In Richard A. Bitzinger (Ed.), *The modern defense industry: Political, economic, and technological issues* (pp. 257–271). Denver, CO: Praeger Security International.

U.S. Air Force. (2015). F-15E Strike Eagle [Fact sheet]. Retrieved from http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104499/f-15e-strike-eagle.aspx

Under Secretary of Defense for Acquisition, Technology, & Logistics (USD[AT&L]). (2014, October 9). *Terms of reference—Defense Science Board Task Force on Air Dominance.* Retrieved from http://www.acq.osd.mil/dsb/tors/TOR-2014-10-09-Air_Dominance.pdf

Westra, A. G. (2009, Winter). Radar versus stealth: Passive radar and the future of U.S. military power. *Joint Forces Quarterly, 55,* 136–143. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a515171.pdf

Wilson, J. R. (2015, November 18). Laser weapons at the crossroads*. Military & Aerospace Electronics.* Retrieved from http://www.militaryaerospace.com/articles/print/volume-26/issue-11/special-report/laser-weapons-at-the-crossroads.html

Wynne, M. (2012, September 14). Shaping the Wolfpack: Leveraging the 5th-generation revolution. SLDinfo.com. Retrieved from http://www.sldinfo.com/shaping-the-wolfpack-leveraging-the-5th-generation-revolution/

Panel 13. Setting Requirements and Managing Risk in Complex, Networked Projects

Thursday, May 5, 2016	
9:30 a.m. – 11:00 a.m.	**Joseph Yakovac, Lieutenant General, USA (Ret.)**– Former Principal Military Deputy, Assistant Secretary of the Army for Acquisition, Logistics and Technology ***Acquisition in a World of Joint Capabilities: Methods for Understanding Cross-Organizational Network Performance*** Mary Brown, Professor, UNCC Zachary Mohr, Assistant Professor, UNCC ***Modeling Uncertainty and Its Implications in Complex Interdependent Networks*** Anita Raja, Professor, The Cooper Union Mohammad Rashedul Hasan, Assistant Professor of Practice, UNL Robert Flowe, Office of Acquisition Resources & Analysis, OUSD (AT&L) Brendan Fernes, Student, The Cooper Union ***An Optimization-Based Approach to Determine System Requirements Under Multiple Domain-Specific Uncertainties*** Parithi Govindaraju, Graduate Research Assistant, Purdue University Navindran Davendralingam, Research Scientist, Purdue University William Crossley, Professor, Purdue University

Joseph L. Yakovac, Jr., LTG, U.S. Army (Ret.)—retired from the United States Army in 2007, concluding 30 years of military service. His last assignment was as director of the Army Acquisition Corps and military deputy to the assistant secretary of defense for acquisition, logistics, and technology. In those roles, Lt. Gen. Yakovac managed a dedicated team of military and civilian acquisition experts to make sure America's soldiers received state-of-the-art critical systems and support across a full spectrum of Army operations. He also provided critical military insight to the Department of Defense senior civilian leadership on acquisition management, technological infrastructure development, and systems management.

Previously, Lt. Gen. Yakovac worked in systems acquisition, U.S. Army Tank-Automotive Command (TACOM), and in systems management and horizontal technology integration for the Office of the Assistant Secretary of the Army for Acquisition, Logistics, and Technology. He has also served as executive officer and branch chief for the Bradley Fighting Vehicle and as a brigade operations officer and battalion executive officer, U.S. Army Europe and TACOM.

Lt. Gen. Yakovac was commissioned in the infantry upon his graduation from the U.S. Military Academy at West Point. He served as a platoon leader, executive officer, and company commander in mechanized infantry units. He earned a Master of Science in mechanical engineering from the University of Colorado at Boulder before returning to West Point as an assistant professor.

Lt. Gen. Yakovac is a graduate of the Armor Officer Advanced Course, the Army Command and General Staff College, the Defense Systems Management College, and the Industrial College of the Armed Forces. He has earned the Expert Infantry Badge, the Ranger Tab, the Parachutist Badge, and for his service has received the Distinguished Service Medal, the Legion of Merit three times, and the Army Meritorious Service Medal seven times.

Acquisition in a World of Joint Capabilities: Methods for Understanding Cross-Organizational Network Performance

Mary M. Brown—is a Professor at University of North Carolina at Charlotte. [marbrown@uncc.edu]

Zachary Mohr—is an Assistant Professor at University of North Carolina at Charlotte. [zmohr@uncc.edu]

Abstract

Increasingly, government managers are turning to cross-organizational networks for the acquisition and delivery of services. The use of networks is lauded as a means to eliminate service gaps, achieve synergistic benefits, and provide better buying power. Cross-organizational networks now support a large number of local, state, and federal level activities (i.e., health care, social services, emergency management, and transportation). It has long been recognized that organizations are susceptible to the vagaries of their environment and that performance is often a function of how well organizations adapt to environmental fluctuations (Ashby, 1954; Holland, 1975). Despite the popularity of networks, little is known about the unique risks they encounter and the susceptibility of cascades. The objectives of this research are to (1) identify the exposure and vulnerability mechanisms that relate to cross-organizational network risk, contagion, and performance; (2) provide managerial recommendations on cross-organizational networks as a form of service delivery; and (3) provide a theoretical framework for conceptualizing cross-organizational networks as a service delivery option. This research models the Major Defense Acquisition Programs (MPADs) as a network of interconnecting programs and employs Contagion Modeling (mixed effects linear regression with a modularity maximization algorithm) as a method for understanding MDAP performance. The presentation will provide the statistical results gained from the contagion modeling and provide insights on risk susceptibility. Understanding the nature of how exposure triggers state changes across networks levels is likely to yield new strategies on how to manage network risk.

Introduction

Whether explicitly pronounced or implicitly performed, “jointness” has become a dominant means for modern warfare acquisition. For this research, jointness, interdependency, exchange, and partnerships all refer to a similar concept: the notion that autonomous organizations build relationships to obtain resources to provide capabilities that, when looked at in totality, form network structures. While it is true that at the individual pair-wise level, these exchanges exist as explicit transactions for the transfer of data, labor, capital, or materials, it is also true that the totality of the various dimensions, coupled with the turbulence of perturbations, influences the cost, schedule, and performance of the acquisition effort.

Organizations in the past sought to limit interdependencies to maintain control over the environment. Concerned about environmental instabilities, organizations either limited the scope of their activities or sought to expand their domain by bringing mission critical activities internally. More recently, however, organizations have found that the costs and limitations of environmental control behaviors are both impractical and infeasible.

Typically, jointness appears in the context of shared resources, supply chains, or shared requirements. The benefits of joint activities can be great. Jointness can eliminate redundancy, streamline activities, and lead to “Better Buying Power.” Jointness can also make possible what was previously improbable. Jointness has been known to result in critical synergistic opportunities, that is, battlespace awareness.

But jointness does not come without risk. Collaborative efforts are known to experience the problems of suboptimization and moral hazard and principal-agent issues (Pfeffer & Salancik, 1978). In ideal terms, the decision calculus to engage in a relationship would involve weighing the costs of lost opportunities (e.g., in terms of response time, flexibility, etc.) against the benefits of the relationship (e.g., synergy, shared resources, and economies of scale and scope). In the world of transaction costs, collaborative efforts are rarely free. Uncertainties regarding a partner's ability to commitment to a relationship for the duration of the initiative can influence the decision to engage. Transaction risk, or the probability that a loss might occur due to a partner default, is a concern for many public managers. Recognizing that the environment of a given organization can exert powerful and unintended consequences on the relationship, collaboration, or jointness, is often avoided (Wilson, 1994).

For this research, jointness, interdependency, exchange, and partnerships all refer to a similar concept: the notion that autonomous organizations build relationships to obtain resources to provide capabilities that, when looked at in totality, form network structures. While it is true that at the individual pair-wise level, these exchanges exist as explicit transactions for the transfer of data, labor, capital, or materials, it is also true that the totality of the various dimensions, coupled with the turbulence of perturbations, influences the cost, schedule, and performance of the acquisition effort.

Unfortunately, by and large, the literature on interdependent activities is steeped in contradictory findings. For example, some argue that tight-knit arrangements are more likely to have the social traction needed to overcome environmental difficulties (Sosa, 2011), whereas others argue that loose coupling, or weak ties, may be a better solution (Granovetter, 1973). Some claim that more information is the key to benefit attainment (Comfort, 1994), whereas others claim that more information leads to a false sense of security (Hall, Ariss, & Todorov, 2007). Yet, despite the absence of consistent sage advice, resource limitations and a demand for comprehensive solutions continue to push organizations toward complex structures for the delivery of products and services.

As discussed, jointness does not occur without some degree of risk. This research examines one particular form of risk: contagion. The discussion below examines the funding interdependencies that arise from shared program elements and begs the question, are neighborhood programs contagious when it comes to cost variance? The study examines MDAP performance in light of the cost variance reports in the annual SARs over a period of six years.

Methods

As alluded to above, MDAP programs often share program elements. Shared resources, that is, program elements, are a common form of jointness. The analysis below tests for the presence of contagion as it relates to the cost variances of neighbor programs.

To test for the presence of contagion, mixed effects linear regression with a modularity maximization algorithm was employed. The modularity maximization algorithm allowed us to divide the network into groups and the mixed effects linear regression allowed us to obtain coefficients to test for the presence of contagion. With mixed effects we are able to model the random effect of the network community (j) by employing a modularity maximization algorithm.

The modularity maximization algorithm splits the network into a number of communities or groups. In other words, it tells us which MDAP programs belong together in a single cluster and which do not. Put simply, employing iteration methods modularity is the fraction of the edges that fall within the given groups minus the expected such fraction if edges were distributed at random. The benefit of using the modularity algorithm is that no single program can be identified in two groups. Hence, the groups are orthogonal.

Because we were testing the individual variance of each MDAP within each of the groups, a mixed effects model was needed (Raudenbush & Bryk, 2002). The mixed effects models that were estimated are linear regressions that account for the total cost variance of all network partners, B5 Model 1, and component cost variances of schedule, estimation, economic, and engineering that correspond with B6, B7, B8, and B9 in Model 2 respectively. The other predictors of interest in both models are β1, which models the effect of the number of network partners that are directly connected to the MDAP program yi. The β2 estimator is the diversity of network partners based upon the rank abundance curve. The β3 is the percent of network partners that are considered joint programs. The β4 is the percent of network partners that are classified as in production. The δk is a vector of year dummies to account for the years 2010–2014; therefore, the baseline year is 2009. The network community is the random effects term (j) in the model. The αj is the varying intercept based upon the network community upon which the MDAP program is classified.

Model 1: $y_i = \alpha_{ji} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \delta_k X_k + \varepsilon_i$

$\alpha_j = \mu_\alpha + \eta_j$

Model 2: $y_i = \alpha_{ji} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_6 X_{i6} + \beta_7 X_{i7} + \beta_7 X_{i7} + \beta_7 X_{i7} + \delta_k X_k + \varepsilon_i$

$\alpha_j = \mu_\alpha + \eta_j$

Due to the leptokurtic nature of the untransformed y_i, the y_i was transformed using the cube root, $y_i^{1/3}$, to make the error distribution further reflect the Gaussian assumptions of the linear mixed effects model. Because the cube root equally reduces the variance of large positive and negative values, this transformation was found to be the simplest transformation possible but other transformations are also possible. The nature of the transformation does not influence the estimation of the relationship between the linear predictors. The major influence that this has upon the model is to shrink the variance of the untransformed y_i to make the model better fit the data. The interpretation of this transformation is discussed below.

Measures

As mentioned above, the goal was to test the cost variance of neighborhood partners and contagion to other programs. Consequently, the cost variances reported in the annual SARs were collected. Additionally, several control variables were employed. The first was a complexity metric that measures the number of programs that share a program element. The second was a diversity measure. Diversity was measured by the slope of the rank abundance curve. The percent of the partners that were explicitly joint as well as the percent of the partners that were in production were included in the models as controls.

Findings

The two best-fitting models are presented in Table 1, and they reveal that both complexity and the cost variances of the network partners influence the cost variances of the MDAP programs. Of the two theoretical classifications of variables, we find that the complexity variable is the better predictor of cost variances in the network. First, we describe the results of the first model of the total cost variance of the network partners, which does not seem to support the hypothesis that network partners' cost variances should influence the MDAP program cost variance. Next, we describe the second model, which shows when we look at the component cost variances of the network partners we see modest support for the network partner to MDAP program cost variance connection, at least for estimation cost variance. Throughout all of the models, the complexity and diversity measures are significant (p<.1) and of the sign predicted by theory.

Table 1. Models of Network Partner Cost Variance Effects on the MDAP Program Total Cost Variance in the MDAP Financial Network, 2009–2014

	Model 1—Total cost variance of network partners*			Model 2—Component cost variances of network partners *		
Parameter	Est.	Std. Error	Sig.	Est.	Std. Error	Sig.
Number of network partners	0.0714	0.0394	0.071	0.1134	0.0449	0.015
Diversity of network partner services	5.9373	3.0053	0.049	6.3388	3.0339	0.038
Percent of network partners that are joint	-0.0955	0.6853	0.889	-0.2419	0.6976	0.729
Percent production of network partners	0.6047	0.6460	0.350	0.5017	0.6475	0.439
Network partner total cost variance	0.1847	0.1615	0.253	-	-	-
Network partner schedule cost variance	-	-	-	-0.0041	0.0029	0.162
Network partner estimation cost variance	-	-	-	0.0003	0.0002	0.090
Network partner economic cost variance	-	-	-	0.0025	0.0030	0.418
Network partner engineering cost variance	-	-	-	-0.0006	0.0006	0.295
Intercept	0.0240	1.0674	0.982	0.0225	1.0897	0.984
Network community (variance est.)	0.2734	0.3881	0.481	0.0760	0.1890	0.688
-2loglik	1723.35			1766.17		
BIC	1734.95			1777.75		

*MDAP program total cost variance is estimated in the model based upon the cube root of the MDAP program total cost variance. Year fixed effects are not shown in the table.

The first model shows that the network partner total cost variance is not a significant predictor of the MDAP program cost variance when we account for complexity, year, and network community. It is of the correct theoretical sign, which would indicate that when the network partners have greater cost variance, the MDAP programs also have greater cost variance. The fact that network partners' total cost variance is not a significant predictor of the MDAP program total cost variance may be due to the fact that they are unrelated, but it may also be because there are simply too many cost variances being added together in the total network partner cost variance, which creates noise in the analysis and supports the analysis of the components of cost variance as we do in the second model.

The complexity and diversity variables that were included in the model were significant predictors of cost variance in the model as well. The complexity variable number of network partners was significant (p<.1) and of the direction predicted by theory. The weak significance of this variable strengthens when we look at the second model, but it is substantively significant in terms of its effect on the cost of the MDAP program. One thing to remember is that these models are based on the cube root of the total MDAP program cost variance, due to the leptokurtic nature of the distribution. Therefore, the effect of all of these variables is nonlinear and is dependent on the current level of cost variance. Because of this, we observe that a unit change in the number of network partners is associated with a change in the cost variance of 0.214 times the square of the cube root of the estimated cost variance.[1] Given that the average cost variance of the programs in the dataset is $38 million, this means that a one-unit change in the number of network partners for the average program would result in a $2.42 million increase in the cost of the program.

Likewise, the diversity of network partners services based upon the rank abundance curve is very strong. A one-unit change that takes us from no network partner diversity to most theoretical network partners diversity has a significant impact on the cost variance. The change, therefore, from least possible diversity to most possible diversity of network partners leads to an increase in the cost of the program of $201.34 million in the first model.

Overall, the first model fit better (BIC = 1734.95) than the second (BIC = 1777.75). The network community variance estimate is 0.27 but is not significant. This variable is included in the model because preliminary data analysis suggested that the network community was associated with the MDAP program cost variance. Therefore, the random effects or hierarchical model of cost variances in the network is theoretically warranted but may not be needed given the other variables included in the model. In the conclusion, we provide suggested research approaches to further test if network communities have an influence on the cost of programs.

Interpreting the significant coefficients from the second model, we see that both the complexity and diversity variables are now both significant at the $p<.05$ level and the substantive effect of the variables increases. The increase in the cost to a program based upon the regression coefficients in the second for complexity and diversity are $3.84 million and $214.94 million, respectively. In the second model, the sum of the network partners estimation variances is now associated with the MDAP program cost variance ($p<.1$). This effect, like the complexity and diversity variables, is non-linear based upon the underlying cost variance; however, unlike the diversity and complexity variables this effect is not nearly as strong in practice. For example, if network partners estimation variance increased by a million dollars, then the cost variance of the average MDAP program is predicted to increase by $10,172. In conclusion, this variable provides only weak evidence that network partners cost variances are associated with the MDAP program’s cost variances once the models account for the year of the cost variance, the complexity of the network partners, and the diversity of the network partners.

Many of the variables in the model were not significant, including the total MDAP program cost variance in the first model and the component cost variances, with the exception of estimation cost variance. This suggests that much of the cost variance is strongly attributable to the complexity and diversity of the programs that are being developed.

---

[1] Because the linear model estimates the effect of the independent variable on the dependent variable as dY/dX and Y is to the 1/3 power, estimates of the effect must apply the chain rule of $Y = (b_0 + b_iX_i)^3$, where x is the vector of regressors. The chain rule tells us that a unit change in any of the x_i is associated with a change of Y such that $dY/d_{xi} = 3b_i\,(b_0+b_iX_i)^2 = 3b_iY^{2/3}$. If we concentrate on just the second form of the equation, we are be able to interpret the b_i effect of a unit change on x_i given a particular level of cost variance, which we do in terms of the mean cost variance in the dataset of $38 million.

In sum, none of the neighbor cost variance measures (neither the production nor percent joint) proved instrumental in predicting individual program cost variance. However, both the diversity and the number of neighbors did prove instrumental and do appear correlated with cost variance growth.

As discussed, jointness does not occur without some degree of risk. This research examined one particular form of risk—contagion—employing one particular statistical technique mixed effects linear regression with a modularity maximization algorithm. The results did yield interesting findings in terms of size of neighborhood and diversity. Further research will test a number of different algorithms for their strengths and weaknesses in providing insights on joint activities.

References

Comfort, L. (1994). Self-organization in complex systems. *Journal of Public Administration Research and Theory, 4*(3), 393–410.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Hall, C.C., Ariss, L., & Todorov A. (2007). The illusion of knowledge: When more information reduces accuracy and increases confidence. *Organizational Behavior and Human Decision Processes, 103*, 277–290

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods (Advanced quantitative techniques in the social sciences)* (2nd ed.). SAGE.

Sosa, M. (2011). Where do creative interactions come from? The role of tie content and social networks. *Organization Science, 22*(1), 1–21.

Wilson, J. Q. (1989). *Bureaucracy: What agencies do and why they do it.* Basic Books.

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-15-1-0062.

Modeling Uncertainty and Its Implications in Complex Interdependent Networks

Anita Raja—is the Associate Dean of Research and Graduate Programs and Professor of Computer Science in the Albert Nerken School of Engineering at The Cooper Union. She directs the Distributed Intelligent Agents Lab. She received her PhD in computer science from the University of Massachusetts at Amherst in 1998 and 2003, respectively. Raja's research focus is in the field of artificial intelligence, specifically as it relates to the study of decentralized control and reasoning in software agent systems operating in the context of uncertainty and limited computational resources. [araja@cooper.edu]

Mohammad Rashedul Hasan—is an Assistant Professor of Practice in the Department of Computer Science and Engineering at the University of Nebraska–Lincoln. He received his PhD in Computing and Informatics from the University of North Carolina at Charlotte in 2014. His PhD adviser was Anita Raja. He obtained a Master of Applied Information Technology degree from Monash University (Australia) under the Australian Government Development Scholarship. He was the entrepreneur of a non-profit IT start-up company named Global Village.com that provided telecommunication services to the rural people in Bangladesh. Hasan's research interest revolves around artificial intelligence, network science and complex systems. [hasan@unl.edu]

Robert Flowe—is a Senior Program Analyst supporting the Deputy Director Office of the Secretary of Defense Studies (OSD) and Federally Funded Research and Development Center (FFRDC) Management, conducting policy development and oversight of DoD FFRDCs and University Affiliated Research Centers (UARCs). He facilitates the proper utilization and management of DoD FFRDCs and UARCS in accordance with relevant laws, regulations, and policies, and enhances the effectiveness of these strategic resources in support of DoD goals. [robert.m.flowe.civ@mail.mil]

Brendan Fernes—is an undergraduate student in engineering at The Cooper Union. He is in his sophomore year. [fernes@cooper.edu]

Abstract

The overall goal of this paper is to continue our efforts to forge new ground in identifying the effects of interdependencies in large complex networked applications and, if needed, uncovering early indicators of interdependency risk so that appropriate risk mitigation actions may be taken. Specifically, we seek to study and quantify the impact of network characteristics on cascading risk. Cascading risk is defined as the propagation of programmatic issues across networked programs due to the interdependency of one program upon the other. Harnessing the extensive data that has been collected over the years in the form of Defense Acquisition Execution Summary (DAES) and Selected Acquisition Reports (SARs) documents for Major Defense Acquisition Programs (MDAPS), we will present our intermediate results in our ongoing efforts on leveraging network structure and sequential data to study cascading risks. We will also identify the challenges to data acquisition.

Introduction

Our work is motivated by the need for "what-if" analysis in large complex interdependent and networked applications such as the critical infrastructure network (electric, water, gas grids). The research goal is to develop methodologies and algorithms to proactively model and reason about non-linear cascading risks to facilitate this analysis. Networked applications often operate under uncertainty in environmental response and the temporal state and action choices of the nodes are captured in the form of structured and unstructured text data as well as image data.

We build on our previous work (Raja, Hasan, & Brown, 2012; Raja et al., 2013, 2014), where we used state-of-the-art extraction technologies including Latent Dirichlet

Allocation (LDA) topic modeling algorithms to develop automated text and image extraction techniques to extract features from various types of structured and unstructured text and image data. In addition to this automated data extraction module, we developed two executable modules: one to identify the relationships in a network (Network Identifier module) and the other to compute the weight of the links among the neighboring nodes (Interdependency Index Determiner). We tested and evaluated these algorithms on a small network and showed that the performance of the automatic extraction algorithms was comparable to the performance of manual extraction.

We use the MDAP network as a case study to study cascading risks and develop methodologies and algorithms that can be generalizable to similar networks. Individual MDAP performance across months and years has been captured by a combination of structured and unstructured temporal data, including Selected Acquisition Reports (SARs), Defense Acquisition Execution Summary (DAES) reports, and milestone reviews are evaluated from an individual program point of view without emphasizing the dynamics of joint space. The question of modeling cascading risk across programs with funding or data relationships is important since we conjecture that poor performance of the MDAPs (various breach conditions) can be attributed to local (individual MDAP) as well as non-local (related MDAPs) sources that result due to interdependencies among the MDAPs.

In this paper, we present a network-centric approach that has the dual goal of contributing to advances in reasoning about uncertainty, large-scale text and image data analysis, as well as understanding of complex networks. This project breaks ground in the areas of (a) defining a metric to quantify the influence of network characteristics on performance and (b) identifying the type of data required to formulate appropriate mathematical models for understanding the dynamics of complex networks.

Network Performance Study From an Interdependent Hierarchical Network Perspective

The joint space of major defense acquisition programs (MDAPs) creates interdependencies among MDAPs. These interdependencies contain the characteristics of a complex network (Brown, 2014). Programs in the MDAP network share diverse relationships. Mainly, there are two types of ties that exist among the MDAPs: (1) programmatic ties (also called programmatic interdependencies) are defined by the program managers in terms of inbound and outbound connections to support hardware/software requirement of the programs, and (2) funding ties that identify the programs as funding neighbors if they draw funding support from the same “program element” (PE) account. These two types of ties result in two types of network relationships among the MDAPs, namely, programmatic network and funding network.

A systemic understanding of the performance of the MDAPs requires the understanding of these two types of networks. Therefore, the system of MDAPs can be considered as a multiplex network that is a superposition of both programmatic and funding networks defined on the same set of programs (Szell, Lambiotte, & Thurner, 2010).

Interdependencies among the program influence the performance of the MDAPs (Brown, 2014; Raja et al., 2012). However, the multiplex nature of the MDAP network has not been considered to examine the performance of the programs. Moreover, the effect of interdependency on the programs was not quantified previously. Our goal is to investigate the joint space of the MDAP multiplex network as it influences program performance and to define a metric (the risk parameter) that quantifies this influence. The values of this risk parameter for each program in the multiplex network would be useful to forecast a potential cascading effect. Moreover, the program managers would be able to identify critical programs using this parameter and take necessary measures to improve programs' performance. The risk parameter is formally defined in the following section based on the Probabilistic Risk Analysis (PRA) methodology for networked systems.

Probabilistic Risk Analysis (PRA)

Probabilistic risk analysis (PRA) is a methodology (Lewis, 2009) to evaluate risks associated with a complex engineering entity. It systematically looks at how the pieces of a system work together to ensure safety. PRA allows analysts to quantify risk and identify what could have the most impact on safety (Lewis, 2009). Therefore, we use the risk parameter from PRA methodology to quantify the influence of interdependency in a complex network, specifically the MDAP network.

The PRA equations for risk in a system use the notion of vulnerability and consequence. Although the concept of vulnerability, risk, and consequence in non-network systems share standard definitions in financial and engineering communities, these terms are not well understood for networked systems. This is because network science is a new field, and it is not very clear how to understand the failure of the assets in networks.

According to the standard definitions (Lewis, 2009) in non-networked systems, vulnerability V is the probability that a component or asset will be compromised after successful attacks. Risk R measures the expected loss due to the failure of an asset. Threat T is the probability that an attack will be attempted. Consequence C is the outcome of a successful attack. Therefore, standard risk is defined as the product *R=TVC.*

These definitions, however, do not provide an appropriate measure for risk in networked systems. In a network, system failure is a function of the interdependence of the nodes. These definitions do not incorporate the interdependency of the various components of a system. Therefore, it is important to consider the connectivity among the nodes in a network for computing risk.

Lewis (2009) extended these standard definitions to networks containing many components or assets (nodes and links). Threat (t), vulnerability (v), consequence (c), and risk (R) in a networked system are an aggregation of individual component or asset threat, vulnerability, and consequences. Network risk is defined in the following PRA equation as an expected value by taking the sum over all nodes (n) and links (m) of the individual components: $\boldsymbol{R} = \sum_{i=1}^{n+m} \boldsymbol{t_i v_i c_i} = \sum_{i=1}^{n+m} \boldsymbol{v_i c_i},$ assuming $t_i = 1$. Here, threat and vulnerability are a priori estimates of the probability of failure. Consequence is typically measured in dollars or lives. This PRA equation for risk is applicable for any system where a priori approximations of the probability of failure can be reasonably estimated. For reducing risk in a networked system, Lewis (2009) argues that it is important to identify the critical nodes that have higher risk values.

Earlier works by Albert, Jeong, and Barabasi (2000) and others explored why highly-connected nodes were more critical nodes than others. However, these studies were done in the context of a single-plex network. Al-Mannai and Lewis (2007) proposed a static technique for critical node analysis in a **multiplex network** where criticality of a node not only depends on the number of connections, but also on other measures. They use a degree-weighted model of network risk to identify the most critical nodes in a network. Intuitively, critical nodes either have many connections or have larger target values. Based on this observation, Al-Mannai and Lewis (2007) extended the simple PRA definition of risk to define the target value of a node as g_iC_i , where g_i is the degree of the node and C_i is the consequence associated with the node's intrinsic value. Therefore, according to their model, extended risk r for an n-node network is related to network topology as follows:

$$\boldsymbol{r_{ext}} = \sum_{i=1}^{n} g_i \, V_i C_i \qquad (1)$$

where g is the degree of node i, while V and C are its vulnerability and consequence, respectively.

Example: PRA for a Small Synthetic Network

As an illustration of the above-mentioned extended PRA technique, let's consider the following network (Figure 1) of four nodes (A, B, C, and D). Connectivity among the nodes is shown for three years. We will use fictitious values for vulnerability and consequence of the nodes in this network in order to understand how the above-mentioned model helps to identify nodes that are most critical for the operation of the network. Also, it will facilitate in understanding the various factors that contribute towards the criticality measure.

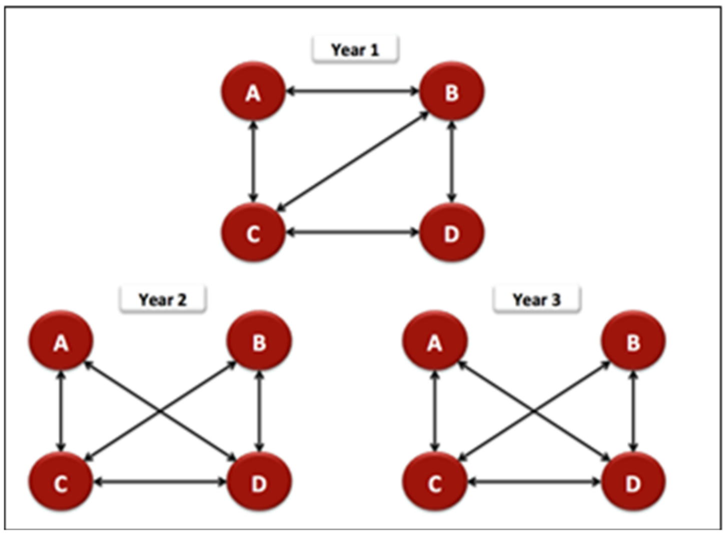


Figure 1. Critical Node Analysis for a Synthetic Network

Table 1 shows the results for extended PRA. In Year 1, we notice that node C is the most critical. Although it has the smallest consequence value in the network, its high connectivity and largest value for the vulnerability are responsible for its critical condition.

In Year 2, however, node C is not the most critical node anymore. This is due to the reduction in its consequence measure. Node D appears to be the most critical because of the increase in its degree. Its vulnerability and consequence values did not increase from the previous year.

In Year 3, Node A becomes the most critical node because of the increase in its vulnerability and consequence values. However, its degree did not increase.

Table 1. PRA of the Synthetic Network for Three Years

Year 1

	g	V	C	r = gVC
A	2	0.01	15	0.3
B	3	0.02	20	1.2
C	*3*	*0.6*	*10*	***18***
D	2	0.3	15	9

Year 2

	g	V	C	r = gVC
A	1	0.07	10	0.7
B	2	0.01	30	0.6
C	3	0.5	2	3
D	3	0.3	15	***13.5***

Year 3

	g	V	C	r = gVC
A	1	0.6	20	***12***
B	2	0.01	30	0.6
C	3	0.5	7	10.5
D	3	0.1	10	3

Figure 2 shows the change in extended risk values for the nodes that indicate node criticality during the three-year time-span. This simple illustration helps us to understand the significance of incorporating a node's degree (g) for the computation of its risk along with its vulnerability and consequence (Al-Mannai & Lewis, 2007).

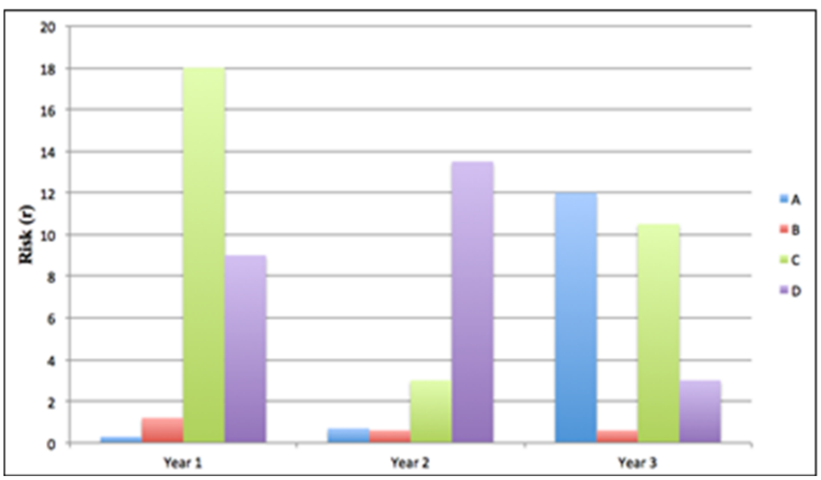


Figure 2. Critical Node Analysis for the Synthetic Network

Critical Node Analysis for a Small MDAP Network

Implementing the above-mentioned technique of extended PRA is a non-trivial task for MDAP networks. PRA requires a reasonable estimation of a priori approximations of vulnerability and consequence of the network assets. However, there is no guideline to do such estimation for MDAPs. Moreover, data on the MDAPs are complex artifacts and often times are either incomplete or fuzzy. Therefore, defining vulnerability and consequence parameters for MDAPs is a challenging task that we address below.

MDAPs operate on a multiplex network. At one hand, MDAPs share funding with other MDAPs (as a result, they form a shared-funding network); on the other hand, MDAPs share hardware/software components with other MDAPs (as a result, they also belong to a programmatic network). Therefore, performance of MDAPs can be examined in light of the performance of the individual program (program-centric) as well as its resulting performance in two different networks (network-centric): (1) a **programmatic network** and (2) a **funding network.** In our analysis, we consider both the program-centric and network-centric contributions.

Below, we first discuss how to discover diverse (programmatic and funding) network relationships among the MDAPs and form a multiplex network. Then we define the various parameters for the extended PRA model. Finally, to validate the approach for extended PRA of the MDAP network, we present a case study of critical node analysis for an MDAP enterprise.

Multiplex Network Formation

The interdependency of the MDAPs that influence their performance can be best understood via the programmatic network (Brown, 2014). In a programmatic network, individual MDAPs support other MDAPs by providing software or hardware components. Therefore, our network of interest is based on the programmatic relationships that exist among the MDAPs. We have gathered data on programmatic interdependencies from the DAES reports for the respective MDAPs. Typically, the last page of the DAES report records the inbound and outbound connections.

Apart from their programmatic dependency, the MDAPs are also related via common PE accounts. In our network model, we capture this funding network relationship as well.

Both the programmatic and funding relationships on the same set of MDAPs are superimposed to define a multiplex MDAP network. For example, Figure 3 shows both the funding and programmatic interdependencies among the MDAPs in an MDAP multiplex network in 2009.

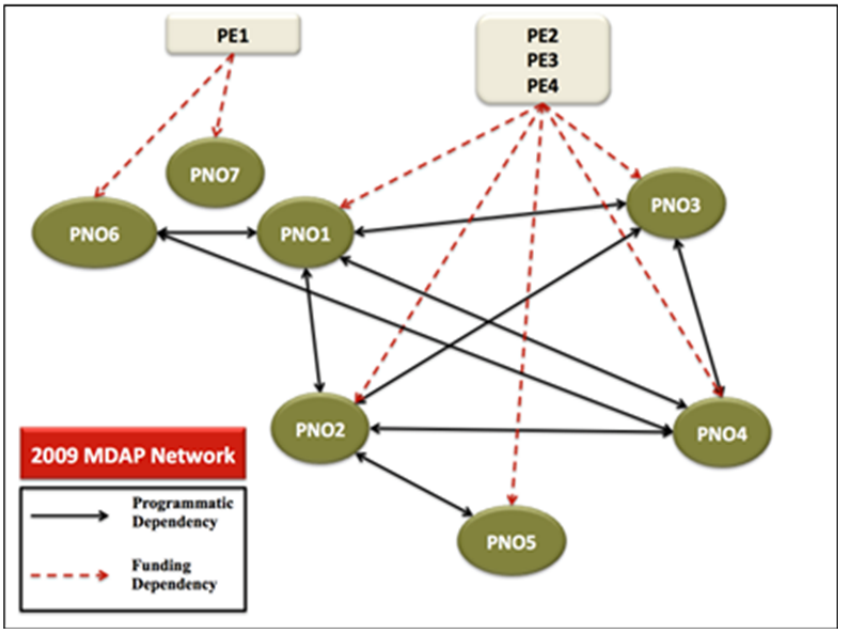


Figure 3. An MDAP Multiplex Network Model

Parameters of Extended PRA Model: Degree (g), Vulnerability (V), and Consequence (C)

We define the extended PRA model parameters as follows:

- **Degree (g):** It is defined by the number of outgoing edges from a node in the programmatic network. Therefore, degree measures the extent of influence of one program (node) on other programs. In an m-node network,

$$g = \sum_{i=1}^{m} n_i \tag{2}$$

- **Vulnerability (V):** It is a measure of weakness of a node in a network. It is defined as the probability of failure of a node if a successful attack is launched on it.

In the MDAP network, we notice that a program may become prone to failure; we call such a program a critical program. Our hypothesis is that breach incidences and other factors mentioned below are indicators of criticality of a program. Program failure is characterized by increased APB breaches and PAUC increase. Moreover, we hypothesize a program's criticality could potentially influence its neighbor's performance (increased breached condition and PAUC increase).

- First, from a program-centric point of view, a program may fail due to its intrinsic poor performance. For example, a weapons procurement cut could lead to intrinsic poor performance (Raja et al., 2012). This program-centric view is captured in the "Program Status" page of the DAES report of the programs.
- Second, APB beaches and the percentage of increase in PAUC is also a measure of a program's performance. These values are recorded in SAR files.

- Third, from a program-centric perspective, the number of its funding and programmatic neighbors influences the performance of a MDAP. For instance, having a large number of funding neighbors (per PE account) makes a program susceptible to potential reduction in promised funding as funds could be siphoned to its neighbors.
- Fourth, having a large number of upstream programmatic neighbors (from which the edges fall on the program) increases its dependency of software/hardware components for successful completion of its tasks.
- Fifth, funding lag also affects the performance of a program and may make it prone to failure.

We propose that the above-mentioned five parameters provide a reasonable estimation for the probability of failure of a program and use these to define vulnerability (Lewis, 2009). Therefore, vulnerability should be considered as the cumulative effect of these parameters. We define the normalized vulnerability based on these parameters using a simple linear function and study its effectiveness:

$$V=\frac{p+b+fNbor+pNbor+diffF}{1+1+1+1+1} \quad (3)$$

Each parameter in the numerator has a maximum value of 1. In the following, the individual parameters are formally defined.

p: It refers to a program's intrinsic performance (captured in DAES reports) and is a linear combination of the factors contributing to Program Status. We use the December DAES report for the last reported month of a year for this computation. The last reported month's data is used as it provides that year's intrinsic performance level of the program. We use the data provided in the "Program Status" page of the DAES reports to compute this metric as described in Table 2.

b: It refers to the number of breaches that occurred in the current year (retrieved from SAR files).

fNbor: It is the normalized number of funding neighbors (retrieved from R docs).

pNbor: It is the normalized number of upstream programmatic neighbors (retrieved from DAES reports).

diffF: It is the normalized differential between received and promised funding amounts (retrieved from SAR and R docs).

Table 2 reports the formulas that we defined to compute these five parameters.

Table 2. Formulas for the Five Parameters Used in the Computation of Vulnerability

Parameters	Formula
p	$\frac{Cost + Schedule + Performance + Funding + Life\ Cycle\ Sustainment}{10+10+10+10+10}$ For the five "Program Status" variables, Cost, Schedule, Performance, Funding and Life Cycle Sustainment, we map the following quantitative values for the colored bubbles: Green: 0; Yellow: 5; Red 10 The value of p is normalized by the maximum numeric values (i.e., 10) for each status variable.
b	$\frac{APB\ SchedBreach + APB\ PerfBreach + APB\ Cost\ (RDT\&E) Breach + PAUC}{1+1+1+25}$ where APBSchedBreach = 1 if APB Schedule Breach occurred, 0 otherwise; APB Cost (RDT&E) Breach = 1 if APB Cost (RDT& E) Breach occurred, 0 otherwise; APB PerfBreach if APB Performance Breach Breach occurred, 0 otherwise **PAUC**: It is captured from the "Unit Cost" section of the SAR. We use the Current year value. Since a "critical" Nunn-McCurdy breach occurs when the program acquisition or the procurement unit cost increases 25% or more over the current baseline estimate, we use 25 as the maximum value for PAUC.
fNbor	$\frac{\sum_{i=1}^{n} fNbor_i}{fNbor\_Max}$ Here, the subscript *i* refers to each PNO account, and fNbor_Max is a predefined large value that is used for normalizing fNbor. We define fNbor_Max as follows, fNbor_Max = Total PE accounts in the network * Total number of MDAPs
pNbor	$\frac{Number\ of\ upstream\ programmatic\ neighbors}{pNbor\_Max}$ pNbor_Max is a predefined large value that is used for normalizing pNbor. We define pNbor_Max as follows: pNbor_Max = Total number of MDAPs in the network
diffF	$\frac{Promised\ Funding - Received\ Funding}{Promised\ Funding}$

- **C (Consequence):** Consequence measures the damage or loss (in dollars) of an asset when failure occurs. Therefore, it should be proportional to the RDT&E funding (from R Docs) and is determined by the breach condition. For example, if a program experiences 100% breach, then its Consequence would be tantamount to its entire RDT&E funding. We define it as follows:

$$C = b^{*} \text{ Funding (RDT\&E)} \tag{4}$$

The breach parameter b from the vulnerability computation is used to compute Consequence.

Case Study: An MDAP Network

We use the extended PRA to identify the most critical nodes for an MDAP enterprise that consists of six MDAPs: PNO1, PNO2, PNO3, PNO4, PNO5, and PNO6. These six MDAPS are funded by four program elements (funding sources): PE1, PE2, PE3, and PE4, as shown in Figures 4–6.

Data for the years 2009 to 2011 are used for this case study. Figures 4–6 show the MDAP enterprise multiplex network for these three years.

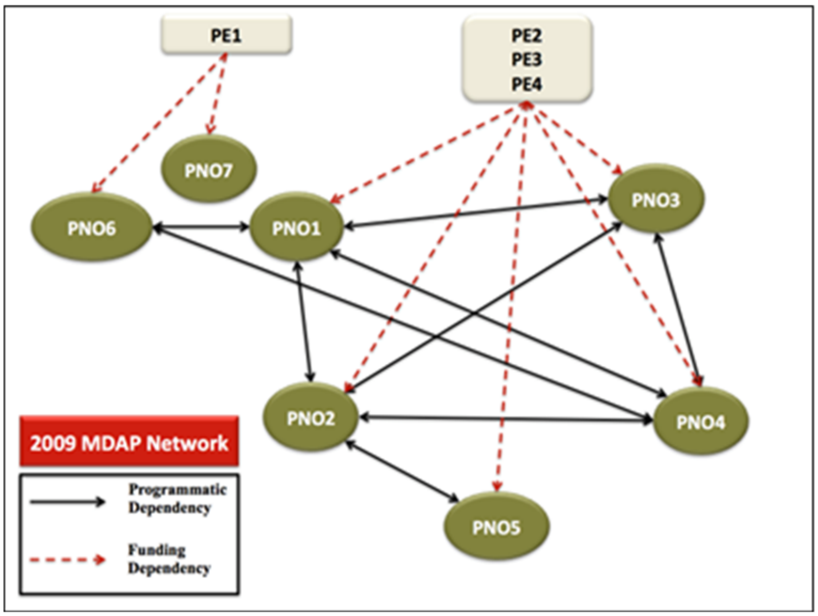


Figure 4. The MDAP Enterprise Multiplex Network, 2009

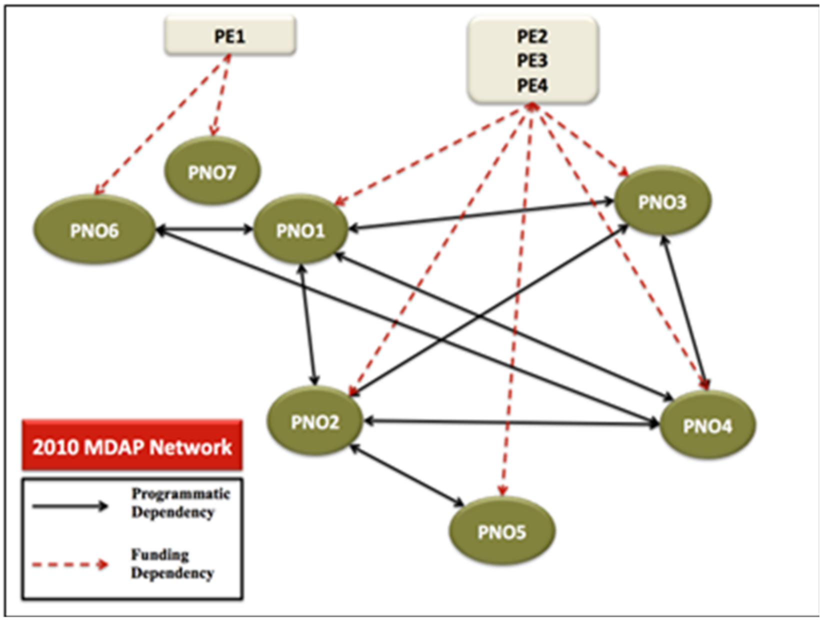


Figure 5. The MDAP Enterprise Multiplex Network, 2010

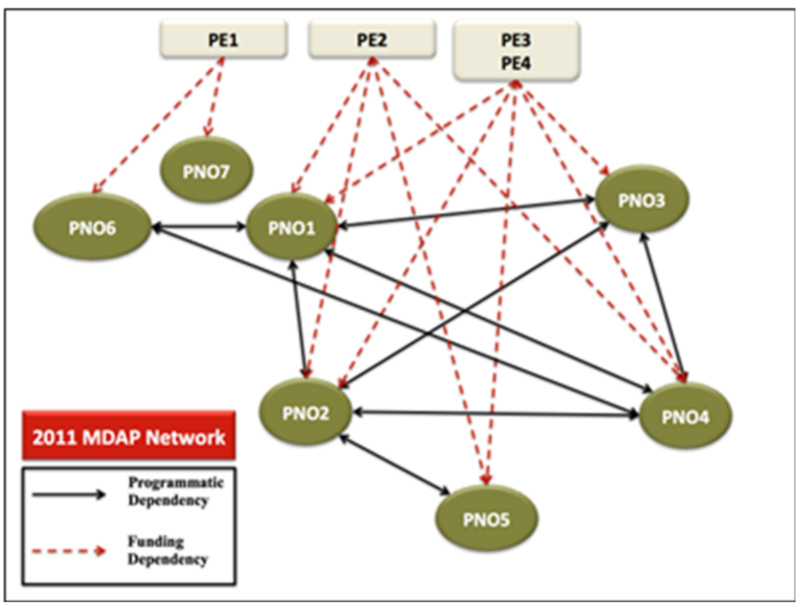


Figure 6. The MDAP Enterprise Multiplex Network, 2011

Detailed calculation of the risk values for each MDAP for three years was performed. Table 3 shows the summary of the results.

Table 3. Critical Node Analysis for MDAP Enterprise Network

	Risk (r)		
	2009	**2010**	**2011**
PNO1	18.11	21.02	97.05
PNO2	0	6.97	3.75
PNO3	6.16	18.89	100.52
PNO4	4.88	11.39	0
PNO5	0	0.17	0.48
PNO6	24.22	24.9	19.7

As an illustration of the calculations in Table 3, we show the detailed calculation of the risk value for PNO1 in 2009 in Table 4.

Table 4. Detailed Calculation of the Risk Value for PNO1 in 2009

Parameters	Calculation
p	$\frac{Cost + Schedule + Performance + Funding + Life\ Cycle\ Sustainment}{10+10+10+10+10}$ $= \frac{0+0+0+0+0}{10+10+10+10+10} = 0$
b	$\frac{APB\ SchedBreach + APB\ PerfBreach + APB\ Cost\ (RDT\&E) Breach + PAUC}{1+1+1+25}$ $= \frac{0+0+0+2.45}{1+1+1+25}$ $= 0.0875$
fNbor	$\frac{\sum_{i=1}^{n} fNbor_i}{fNbor\_Max}$ $= 12/28 = 0.429$
pNbor	$\frac{Number\ of\ upstream\ programmatic\ neighbors}{pNbor\_Max}$ $= 4/6 = 0.6667$
diffF	$\frac{Promised\ Funding - Received\ Funding}{Promised\ Funding}$ $= \frac{215.934(0604280N) - 212.6(SAR)}{215.934(0604280N)}$ $= 0.015439903$
Vulnerability (V)	$V = \frac{p+b+fNbor+pNbor+diffF}{1+1+1+1+1} = 0.2396356$
Consequence(C)	C = 18.894225
Risk(R)	gVC = 18.11091575

From Figure 7, we observe that over the years, PNO1 and PNO3 became the most critical programs in the network. PE6 retained its criticality level, and we do not see significant improvement. A careful analysis of the data for PNO1 and PNO3 in year 2011 reveals that both programs have high breach incidence (that includes increased PAUC). As a result, their consequence values increased as well. Also, these two programs were characterized by higher degrees. All these factors contributed to their high level of criticality. For PNO6, although its degree is relatively small, it has been experiencing schedule and cost breaches as well as increases in PAUC for three consecutive years. The funding budget for PNO1 and PNO6 (over $300 million) is also a contributing factor.

According to 2011 SAR files, PNO1, PNO3, and PNO6 experienced significant PAUC increase and APB breaches, indicating their poor performance level. This observation confirms that our risk computation measure is a step in the right direction.

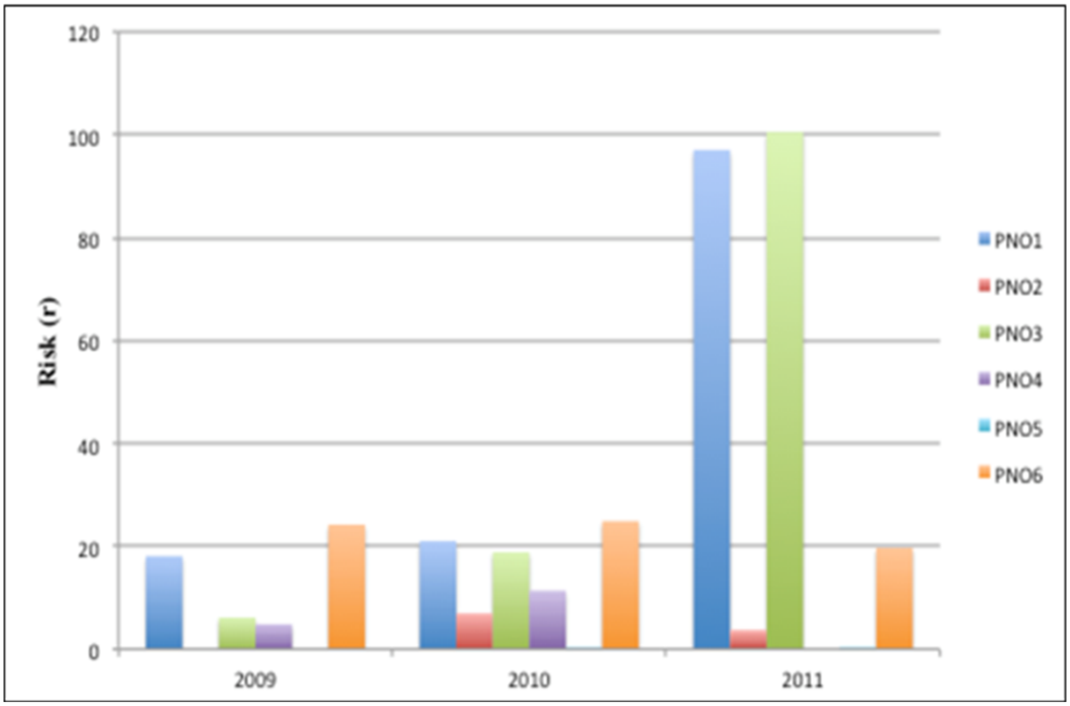


Figure 7. Critical Node Analysis for the MDAP Enterprise Network

Discussion of Extended PRA Model for Risk Computation

The objective for defining the risk parameter (R) in this paper is to capture the effect of multiplex network relations on an MDAP program's performance. As mentioned earlier, breach conditions (from DAES and SAR) are indicators of a program's intrinsic performance, but do not account for the exogenous effects on a program. We have developed the PRA risk model with the potential to capture the network effect on a program's performance. In this model, the intrinsic parameters (p and b) tell us whether a program is "Vulnerable," while the MDAP program's network status accounts for "Criticality." The premise shown by the case study is that this criticality measure helps identify programs that are susceptible to future breaches more effectively than by simply using their intrinsic performance parameters (p or b values).

Manual analysis of MDAP data (done in previous phases of the project) facilitated the process of modeling a small MDAP network for extended PRA analysis. For this modeling, we considered the multiplex nature of the MDAP network and used various performance reports. The results indicate that the extended PRA technique has the potential to successfully identify risky programs and infer the performance of programs.

Also, the PRA analysis uses a network-based composite metric, instead of just individual program PAUC increase and APB breaches, to compute the risk level of a program. For example, both PNO1 and PNO3 have relatively high degree and share the same funding accounts, which makes them susceptible to poor performance. By looking at their increasing PRA risk values in 2009 and 2010, it can be inferred that these two programs are in critical condition. Also by looking at the nearly stable high-risk values of PNO6 in 2009 and 2010, this program should be considered critical as well.

Hence, with the aid of an automated information retrieval mechanism from the performance reports, it is possible to develop an algorithmic tool to identify risky programs. Recognizing the potential of these risky/critical programs to affect the performance of their neighbors could contribute towards predicting cascading effects. As future work, we plan to verify this empirically.

Also, we plan to use this model on another MDAP network to determine if it is able to identify the critical programs; we will modify the parameters of our PRA based model (if necessary) and use this knowledge to define a general model for the entire MDAP network as whole or more realistically, specialized PRA models for classes of similar MDAPs.

Studying the Feasibility of Mathematically Modeling the Phenomenology of MDAP Networks

We also conducted a feasibility study for modeling interdependent networks as a coupled dynamical system and potentially adapting the algorithms for feed-forward networks (Mintchev & Young, 2009; Lanford & Mintchev, 2015) to risk propagation interdependent networks like the MDAP network.

To do this, we would have to determine the network model which includes determining network architecture properties, including various centrality measures, strength of network connections including a precise form of the coupling formalism (strength can be seen as a precise rule that determines dynamical evolution), state features and action options which were already determined in Raja et al. (2012), and a reward optimization model that provides some dynamics to this network. The model would allow us to investigate whether the system has any attractive equilibria, as well as determining the strengths and weaknesses of the basins of attraction. For example, if the steady state of the MDAP network is characterized by only one funded program, with all others having discontinued funding, this is probably not good. We hypothesize that if good equilibria were discovered, an outcome of this analysis could be to recommend a funding strategy that maintains equilibrium or guarantees a rapid convergence toward it.

The specific working hypothesis in the context of the MDAP network is as follows: The programmatic interdependencies between MDAPs have a profound influence on large-scale network performance over an extended period of time.

To determine a network model that is descriptive, predictive, and mathematically sound, we would need a collection of numerical quantities either measured or somehow computed from other measurements recorded in a time series over a sufficiently long period of time. This would involve

- (R1) determination of observable quantities measured numerically (i.e., real numbers on a well-defined scale). The KEY characteristic is to have some a priori evidence that the observables chosen evolve dynamically (i.e., change over time); also, it is absolutely NECESSARY for these to be numerical, or to correspond to some sort of real number scale.
- (R2) finding time series of data on the observables chosen in (R1). Usually a lot of data over a sufficiently long time scale is required to build this historical account of how the observables have changed over time. If the model is to be predictive in the short term, then the variables/observables must have been sampled at a sufficiently high rate.

Evaluating DAES Data

We began by studying the DAES data of several MDAPs collected over a decade with the hope that the sequential monthly data would provide indicators of performance degradation. We have extensive experience with DAES data from our previous work, where we used DAES data to study local and non-local issues that affect the performance of the

MDAP (Raja et al., 2012) and also developed sophisticated text and image extraction tools (Raja et al., 2013, 2014) to automatically extract the DAES data en masse.

Since changes in total cost could be considered as a useful observable, we constructed a few test time series based on the information captured on Top Cost Drivers in the DAES report. Figure 8 captures one such example. It became clear the cost driver time series was not sufficiently volatile enough to facilitate predictability.

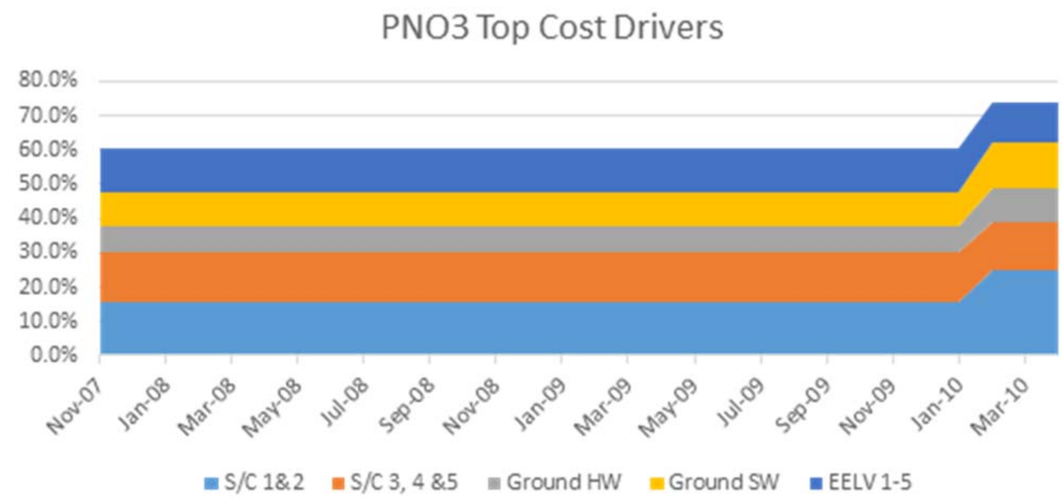


Figure 8. Stacked Area Time Series Data of Five Top Cost Drivers of PNO3

While there is some volatility in January 2010, the volatility is not frequent enough to capture the change in performance risk of the MDAP program over time.

Moreover, we ran into several challenges with the preciseness of the data as far as our goal of building a mathematical model is concerned. Some of our observations are captured below:

In the DAES Program Status page,

- We could not ascertain the quantitative mechanism for color transitions of the red, yellow, or green bubbles that capture the changes in value of Cost, Schedule, Performance, etc. in going from one month to the next.
- The risk in the Risk Summary page describes the risk computation in somewhat of a quantitative way. However, it was still unclear how the risk quantity evaluated; it seems to be coded by a 2-dimensional vector, a (consequence, likelihood) pair; how (if at all) is each of those coordinates computed?

In the DAES page with Top Cost Drivers, Technology Readiness Assessment, Performance (kpps), and Acquisition Program Baseline (APB),

- All of the KPP diagrams seem to be set at T (or threshold); it was not possible to ascertain how these quantities were computed and whether they change over time.
- While the technology readiness assessment box is another potentially interesting measure with regards to building a state space model, the values

did not change for long periods of time and so were at a course level of granularity.

In the Finley charts of the DAES reports,

- The knowledge gained from the Finley charts is that the dependencies are of some programmatic importance—they can affect the course of the subject program—otherwise they wouldn't be mentioned. So of all the potential types of interdependencies that could exist among programs, the Finley charts show those that present a potential risk to the program in question (subject to the limitations of the format and the awareness of the program manager). The dependencies described by the Finley charts generally relate to some component or subsystem in the subject program (or system) that must be provided by, or is somehow dependent upon, the external program (or system). In many cases (actually, most cases) the external entity is a non-ACAT 1D program. There is no requirement for those programs to report their data to OSD via the SAR and DAES. In fact, the data for those programs will be held by the program office or their Program Executive Offices within the military department. This makes getting detailed data about the external program difficult.
- Also, the challenge with the Finley charts is that the nature of the dependency is usually not defined: It could be funding, schedule, or some technical issue. Given the shortcomings of the Finley charts as a way to represent programmatic interdependencies, other more objective representations of system interdependencies have been explored, particularly artifacts that describe the interconnections between the system in question and external systems. These data are in the Information Support Plan (ISP) that each major program generates as part of its milestone approval documentation. The difficulty with the ISP, however, is that the reports are more difficult to obtain, and recent changes in policy have made the data less analytically useful.

The data acquisition challenges could be summarized as follows: Although there are some allusions to the idea that various quantities presented in the reports are quantitatively obtainable through formulas or calculations, there is not much explanation as to how this is actually done or what the numerical values/ranges would be and whether these definitions are consistent across all programs. This information is crucial to building a state space model for the MDAP network. Also, the strategy for determining interdependencies seems to be a difficult. Also, given the time lag (DAES reports are generated monthly) and the level of data captured, often there was not variation in the data from one month to the next.

Analyzing Contract Data

We then deliberated on whether contract data would probably be a better data set for the type of time series based risk analysis we were considering. Instead of focusing on metrics related to contract value (looking for indicators of cost growth), we would instead look at the frequency of contract transactions.

The idea is that when a program is running smoothly, there's probably a baseline rate of contract modifications in the normal course of business (i.e., as funding is added, tasks are completed, deliverables are received, etc.). However, when something traumatic happens, like a test failure or other technical difficulties, we could probably expect significant contractual “churn” as previously-planned efforts are realigned to address the mission-critical issue.

The following is a possible scenario where the “churn” metric might be a more reliable indicator of program distress than cost: Consider a program that is composed of multiple components, each being developed under separate contracts (e.g., a satellite and its ground control segment). If, for example, the satellite has a problem in development (i.e., a test failure), the satellite contract will probably experience cost growth, but the ground control segment might actually experience a decrease in expenditures, as it has to slow down to accommodate delays in the satellite. So whereas costs might increase on one contract, they might be somewhat offset by temporary decreases in the other, which would muddy the “signal” seen at the overall program level. However, each contract would probably have to be re-scoped in order to increase the level of effort for the satellite and reduce the level of effort for the ground segment. Thus, both will incur additional contract “churn” as a result, which should be observable by plotting the frequency of contract modifications over time.

Figures 9, 10, and 11 are the time series of the contract “churn” for the three MDAPs. Each contract transaction reported in the Federal Procurement Data System–Next Generation (FPDS-NG) has an “issue date” indicating when the contract modification was signed. We plotted the frequency of contract actions over time.

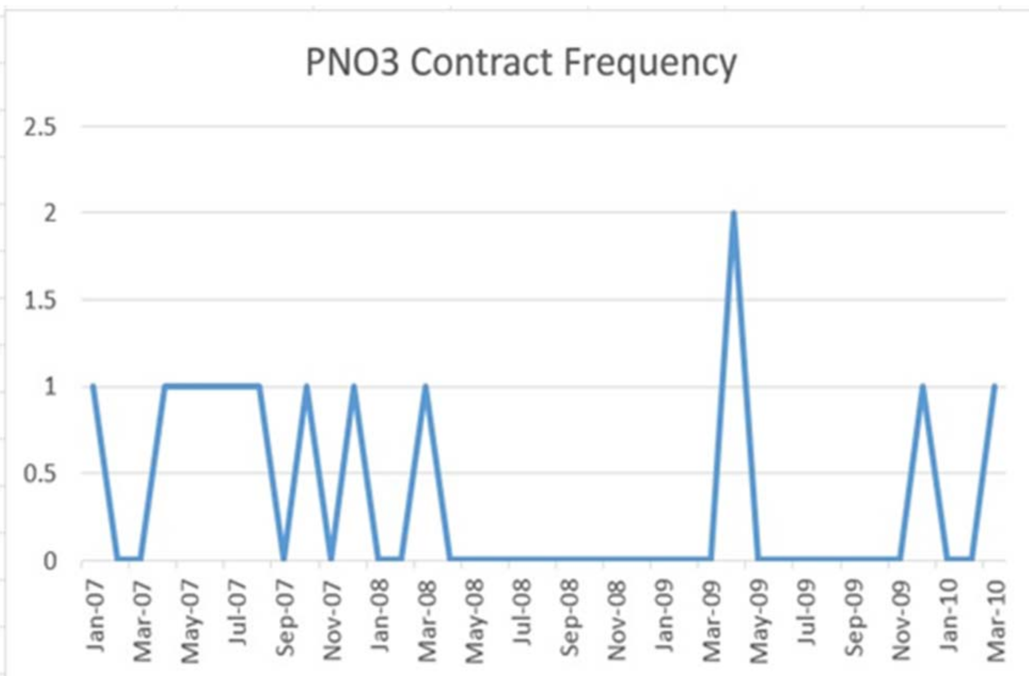


Figure 9. Time Series Data of PNO3-Related Issue Dates

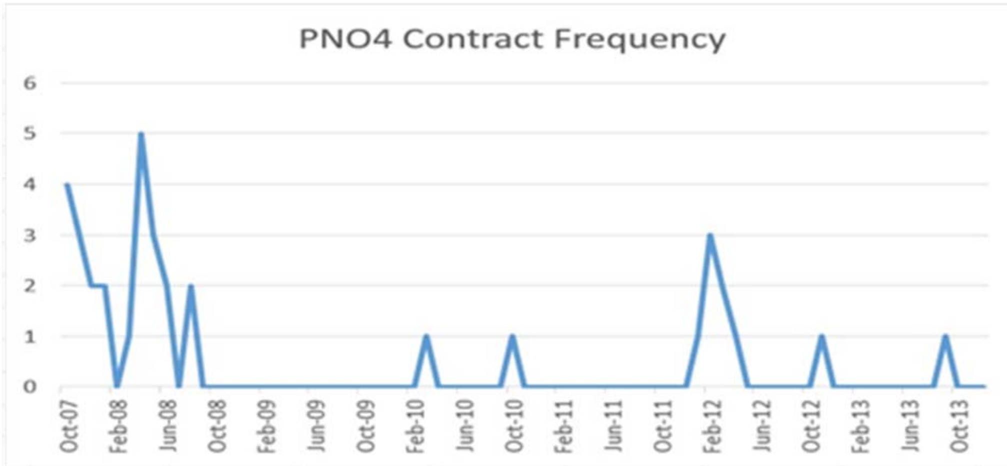


Figure 10. Time Series Data of PNO5-Related Issue Dates

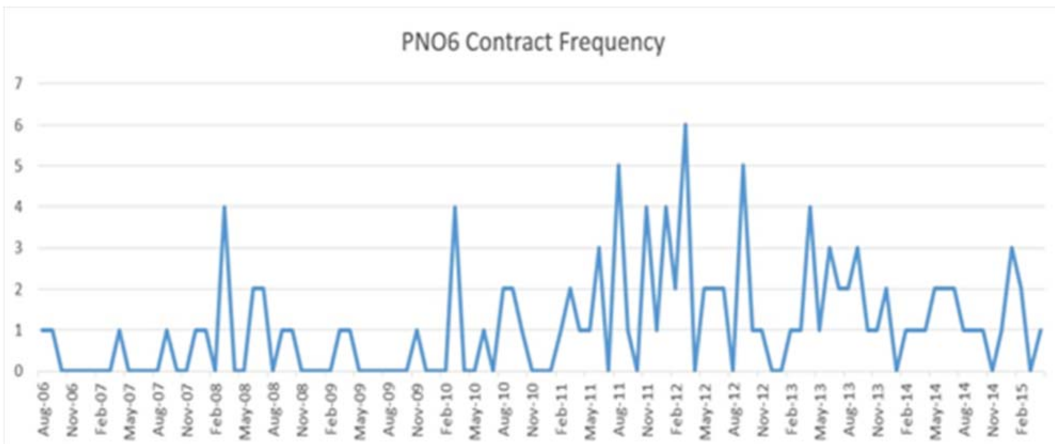


Figure 11. Time Series Data of PNO6-Related Issue Dates

In an effort to determine whether there is any type of correlation between the onset of significant contract churn in Figure 11 and program performance, we examined the breaches reported in the annual SARS data for PNO6. The December 2004, 2005, 2006, and 2007 SAR files show no APB or Nunn-McCurdy breaches, although the notes in the 2005 Threshold breach section state that there was a cost deviation from the key decision point-b approved APB even though there was no change in the total program cost as a result of the action. The 2009, 2010, and 2011 SARS show Schedule and Cost RDT&E APB breaches with varying levels of explanations. The December 2012 SARS indicates no such breach. We are continuing to study the executive summaries as well as SARS of future years in more detail.

Our observation from this examination of churn in contract data is that it does indeed have the volatility that could support the network modeling process. In addition to studying the PNO6 SARS data in greater detail as mentioned above, we are also trying to find contract data over a sufficiently long time scale to support our modeling analysis.

Conclusions and Future Work

In this paper, we have discussed our progress in our ongoing efforts to (1) study the impact of network topological characteristics on risk propagation and our methodology to quantify it, (2) evaluate the critical importance of quantifiable state features in order to assess network dynamics, and (3) describe our investigation into time-series data that could facilitate our analysis.

Our initial results on PRA analysis for a case study and the contract data time series are encouraging, and we plan to further investigate the scale-up of the PRA analysis as well as using the contract data towards building the network model.

References

Albert, R., Jeong, H. & Barabasi, A. (2000). Error and attack tolerance of complex networks. *Nature, 406*, 378–382.

Al-Mannai, W., & Lewis, T. (2007). Minimizing network risk with application to critical infrastructure protection. *Journal of Information Warfare* 6(2), 52–68.

Brown, M. M. (2014). Acquisition risks in a world of joint capabilities: A study of interdependency complexity. In *Proceedings of the 11th Annual Acquisition Research Symposium* (pp. 109–128). Monterey, CA.

Lanford, O., & Mintchev, S. (2015). Stability of a family of traveling wave solutions in a feedforward chain of phase oscillators. *Nonlinearity, 28*, 237–261.

Lewis, T. G. (2009). *Network science: Theory and practice*. Hoboken, NJ: John Wiley & Sons.

Mintchev, S., & Young, L. (2009). Self-organization in predominantly feedforward oscillator chains. *Chaos, 19*(4), 043131.

Raja, A., Hasan, M. R., & Brown, M. M. (2012). Facilitating decision choices with cascading consequences in interdependent program networks. In *Proceedings of the Ninth Annual Acquisition Research Symposium* (pp. 197–220). Monterey, CA.

Raja, A., Hasan, M. R., Rajanna, S., & Salleb-Aoussi, A. (2013). Leveraging structural characteristics of interdependent networks to model non-linear cascading risks. In *Proceedings of the 10th Annual Acquisition Research Symposium* (pp. 293–318). Monterey, CA.

Raja, A., Hasan, M. R., Rajanna, S., & Salleb-Aoussi, A. (2014). A scalable approach to modeling risk in the MDAP network. In *Proceedings of the 10th Annual Acquisition Research Symposium* (pp. 293–318). Monterey, CA.

Szell, M., Lambiotte, R., & Thurner, S. (2010). Multi-relational organization of large-scale social networks in an online world. In *Proceedings of the National Academy of Sciences of the United States of America, 107* (pp. 13636–13641). San Mateo, CA: Morgan Kaufmann.

Acknowledgments

We thank Professor Stan Minchev for his contribution and insights on building a formal model of MDAPs to determine the dynamics of the networks and its effect on performance. We also thank Eui Seong Han for his assistance with the data organization and analysis.

An Optimization-Based Approach to Determine System Requirements Under Multiple Domain-Specific Uncertainties

Parithi Govindaraju—is a PhD student of Dr. William Crossley in the School of Aeronautics and Astronautics at Purdue University specializing in the field of aerospace systems. His research interests center on design optimization and systems engineering. [pgovinda@purdue.edu]

Navindran Davendralingam—is a Research Scientist in the School of Aeronautics and Astronautics at Purdue University. He received his PhD from Purdue University in aerospace engineering in 2011. He is currently conducting research in the Center for Integrated Systems in Aerospace (CISA) led by Dr. Daniel DeLaurentis. [davendra@purdue.edu]

William A. Crossley—is a Professor in the School of Aeronautics and Astronautics, Purdue University, where he has been on the faculty since 1995. His research and teaching interests focus on design and optimization of aerospace vehicles and systems of systems. [crossley@purdue.edu]

Abstract

The task of determining the optimal design requirements of a new system, which will operate along with other existing systems to provide a set of overarching capabilities, is challenging due to the tightly coupled effects that setting requirements on a system's design can have on how the operator uses the system. In this paper, the new system is a strategic military cargo aircraft and the other systems are a fleet of different, existing cargo aircraft; a subset of actual fleet operations from the U.S. Air Force Air Mobility Command defines the example problems in this work. This research builds upon prior efforts to develop a quantitative approach that identifies optimum design requirements of new, yet-to-be-designed systems that, when serving alongside other systems, will optimize fleet-level objectives. The new efforts here address the effect of various uncertainties. The approach incorporates techniques from multidisciplinary design optimization, statistical theory, and robust/reliability-based methods to develop computationally tractable approaches for this kind of problem. The paper also demonstrates the ability to generate tradeoffs between a cost-related metric of fleet-level fuel usage and a performance related metric of fleet-wide productivity. A possible extension for application in commercial air travel also appears in the paper.

Introduction

Nomenclature

AR_X	=	aspect ratio of aircraft type X
B_p	=	maximum average daily utilization of each aircraft (hours)
BPR_X	=	engine bypass ratio of aircraft type X
$BH_{p,k,i,j}$	=	number of block hours for k^{th} trip of aircraft p from base i to base j
$\widetilde{BH}_{p,k,i,j}$	=	distribution of block hours for k^{th} trip of aircraft p from base i to base j
$BH^U_{p,k,i,j}$	=	upper bound of block hours for k^{th} trip of aircraft p from base i to base j
$Cap_{p,k,i,j}$	=	pallet carrying capacity for k^{th} trip of aircraft p from base i to base j
C_{D_0}	=	parasite drag coefficient
$dem_{i,j}$	=	demand from base i to base j in number of pallets
$dem^U_{i,j}$	=	upper bound of demand from base i to base j in number of pallets
E[ ]	=	expectation (arithmetic mean) of a distribution [ ]
$FC_{p,k,i,j}$	=	fuel consumption coefficient for k^{th} trip of aircraft p from base i to base j
$\widetilde{FC}_{p,k,i,j}$	=	distribution of the fuel consumption coefficient for k^{th} trip of aircraft p from base i to base j
$FC^U_{p,k,i,j}$	=	upper bound of fuel consumption coefficient for k^{th} trip of aircraft p from base i to base j
L	=	fleet-level productivity limit (knots-lbs)
$O_{p,i}$	=	indicates if airport i is the initial location (e.g., home base) of an aircraft p
$Pallet_X$	=	number of pallets carried by aircraft type X
P[ ]	=	probability of satisfying expression [ ]
$Prod_{p,k,i,j}$	=	productivity coefficient for k^{th} trip of aircraft p from base i to base j
$Range_X$	=	design range of aircraft type X (nmi)
SFC	=	specific fuel consumption (1/hr)
$Speed_X$	=	cruise speed of aircraft type X (knots)
$Sweep_X$	=	leading edge wing sweep of aircraft type X (deg)
TR_X	=	taper ratio of aircraft type X
$(T/W)_X$	=	thrust-to-weight ratio of aircraft type X
$(W/S)_X$	=	wing loading of aircraft type X (lb/ft^2)
$x_{p,k,i,j}$	=	binary variable for k^{th} trip flown by aircraft p from base i to base j

Research Issue

The *Better Buying Power 3.0* (Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics [OUSD(AT&L)], 2014) document states, “Defining requirements well is a challenging but essential prerequisite in achieving desired service acquisition outcomes.” Typical acquisition processes focus on development at the *system-level* (e.g., aircraft performance), with little explicit consideration for the impact that the new system will have on the holistic performance of a combined set of existing and new systems. Current acquisition processes (how a decision-maker evaluates and acquires systems) are disjointed from considering operations (the way an end user operates these new systems alongside existing ones), resulting in inefficiencies at the higher aggregate level (Taylor & de Weck, 2007; Mane, Crossley, & Nusawardhana, 2007). As an example, consider the acquisition decision-making process within the Department of Defense (DoD) that traditionally involves identification of alternatives, establishment of requirements, estimation of effectiveness, and cost-benefit analyses (Greer, 2010). These action processes do not involve an exhaustive search of the “requirements space” of the new system, where changes in requirements can affect operations due to how the new, yet-to-be-introduced system will be used in conjunction with other existing systems in the fleet.

This research effort seeks to reduce such “handoffs” between the acquisition phase and the operations phase through leveraging quantitative innovations that reduce such handoffs. However, this coupling of the requirements of new systems, and the resulting system’s impact on operations, brings an added dimension of complexity to the acquisition problem. The complexity of dealing with many variables related to interdependent systems, the impact of changing characteristics of such systems, and the uncertainties related to allocations of such systems becomes cognitively impossible to manage without a decision-support framework. Hence, determining the optimal set of requirements for a new, yet-to-be-designed system presents a need for analytical tools to assist decision-makers with quantitatively supported insights.

This paper presents the methodology and formulations of a quantitative approach that identifies the design requirements for a new aircraft under multiple domain-specific uncertainties through an optimization approach. The paper illustrates the approach via examples derived from reported operations of the U.S. Air Force Air Mobility Command (AMC). The approach treats design requirements of new individual systems as decision variables in an optimization problem formulation under various uncertainties to minimize (or maximize) fleet-level objectives—the solution, based on mathematical techniques, identifies the *new aircraft requirement decision variables* that yield the best *fleet-level objectives*.

Two different types of uncertainty are important to this problem: (1) uncertainty in how a designed system “actually” performs in operations as to opposed the predicted performance in the design phase, and, (2) the variations in how much the operator uses the system, as reflected by, say, changing demand for air transportation of military cargo pallets. The uncertainties in how the designed system performs naturally affects the uncertainties in how much the system is being used. Simultaneously considering the system design problem and resource allocation problem under uncertainty captures most of the coupling and interactions present in these two problems, and capturing these can result in fleet-level improvements. Often, high computational expense accompanies quantifying and addressing uncertainty in multiple dimensions, which can make the design problem intractable. Effectively conducting studies that examine several scenarios using different predictions of demand, cost of operating the fleet, and so forth, requires a computationally efficient approach. The authors’ initial efforts to identify an effective approach explored two different strategies—design of experiments and bounding analysis—to understand the effects of considering both demand and design parameter uncertainties in the coupled aircraft design and fleet assignment problem (Govindaraju, Davendralingam, & Crossley, 2015; Govindaraju & Crossley, 2015).

Furthermore, this paper demonstrates an extension of the approach to consider multiple objectives, thereby enabling the assessment of tradeoffs that choices about design requirements may have on fleet-level metrics of interest (e.g., choice of an aircraft may affect fleet-level *productivity* and *fuel burn*—quantities that are at odds with one another). This can allow decision-makers to view this problem in the context of fleet-level fuel consumption as an independent variable. Two key innovations in this approach are that it

1. Considers the holistic implications that setting design requirements may have on the fleet-level metrics.
2. Relegates the mathematical complexities of considering the design requirements, operations of the fleet and manifestations of uncertainties to sound algorithmic approaches, while retaining exploratory and decision-making elements to the practitioner.

Modeling Military Cargo Air Transportation

The transportation of cargo across the AMC service network requires effective deployment of its fleet of cargo aircraft to meet daily cargo delivery requirements while minimizing fuel consumption and related costs. The choice of which aircraft to operate on individual flight legs to meet the cargo delivery obligations within a scheduled time frame determines the total amount of fuel consumed by the AMC fleet. Fleet-wide fuel consumption is tied to the features of aircraft used and the structure of the routes flown. However, the characteristics of the aircraft (e.g., range of the aircraft) also dictate the kind of network that the fleet can serve, thus making it a closely coupled problem. Because of this, there may be an opportunity to identify design requirements for a new aircraft that can reduce the total fleet fuel consumption and/or improve fleet-level cargo delivery performance. This work extends a deterministic decomposition approach (Mane et al., 2007) to allow for the examination of tradeoffs between objectives of productivity (as a measure of mission effectiveness) and fuel consumption when considering the addition of a new, yet-to-be-acquired aircraft to a fleet of existing aircraft under various domain-specific uncertainties. These two competing objectives of productivity and fuel consumption (maximizing productivity increases fuel consumption and minimizing fuel consumption decreases productivity) play a critical role in determining new system requirements—an analyst can perform acquisition assessments by treating fuel consumption as an independent variable in our approach.

Cargo Demand in the AMC Service Network

The AMC service/demand network differs from commercial airline passenger or cargo networks in that cargo demand fluctuates greatly over time and in that cargo demand is asymmetrical, meaning that the demand for cargo from one base to another is usually very different than the demand in the opposite direction between the same bases. Figure 1a shows the fluctuation in the number of pallets transported daily between a representative base pair in the Global Air Transportation Execution System (GATES) dataset for the year 2006. In this plot, the calendar day appears on the horizontal axis, while the heights of the bars indicate the number of pallets transported each day in one direction. Figure 1b presents a histogram of the number of pallets transported per day for the same representative base pair; this reveals that many days had a demand of 20 or fewer pallets on this route. Twenty pallets might be well below the maximum capacity of a single aircraft used to transport this demand. The AMC fleet must have the flexibility to meet fluctuating demand—the comparatively rare, high-demand scenarios, and the typical, nominal demand scenarios—to address fuel efficiency effectively.

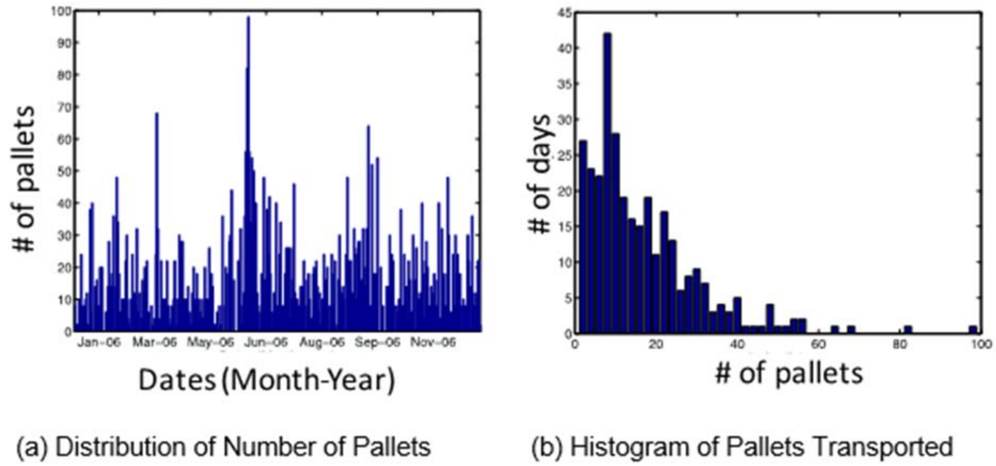


Figure 1. Pallets Transported on a Sample Route From the GATES Dataset

GATES Dataset

The AMC fleet operates on a global network consisting of over 350 bases and in excess of 1750 routes. The GATES dataset provides historical route and cargo demand data, and it contains comprehensive information on palletized cargo and personnel transported by the AMC fleet. From the GATES dataset for 2006, the existing AMC fleet to serve the demand consisted of 92 C-5s, 145 C-17s, and 69 747-Fs. This shows that the AMC transported cargo using C-5 and C-17 aircraft from the strategic fleet and using chartered Boeing 747 Freighter (747-F) aircraft from the Civil Reserve Air Fleet (CRAF) for long range missions. The 2006 GATES data provides a representative cargo flow in the AMC service network and the aircraft used to transport the cargo. For future aircraft design, the demand should be a prediction of future demand; in this work, this historical data takes the place of this future demand prediction.

Each data entry in the GATES dataset represents cargo on a pallet or a pallet-train that the AMC actually transported. Each pallet data entry has detailed information about the pallet transported, such as pallet gross weight, departure date and time, arrival date and time, mission distribution system (MDS), aircraft tail number, aerial port of embarkation (APOE), aerial port of disembarkation (APOD), pallet volume, pallet configuration, and so forth. These data enable the reconstruction of the route network, pallet demand characteristics, and existing fleet size for the fleet assignment problem.

Based on the available dataset, this problem investigation uses the following assumptions:

1. The refined route network from the GATES dataset is representative of all AMC cargo operations
    a. Only routes served by C-5, C-17 and 747-F aircraft are considered. These aircraft types account for a substantial portion (≈75%) of the total pallets transported in the year 2006.
    b. All pallets have fixed dimensions representing the 463L pallet type. Sizing the payload bay and, therefore, the fuselage, of the yet-to-be-acquired aircraft uses these pallet dimensions. In this effort, the problem formulation does not consider any “outsized” cargo capacity requirements.

2. The demand reported in GATES for 2006 is representative of future demand requirements when a new, yet-to-be-designed aircraft would enter into service.

The application problem does not assume any demand growth. The lack of publicly available information coupled with having only one year of operations reported in the GATES dataset prevents the development of a reasonable future pallet demand-forecasting algorithm for the routes operated by the AMC. However, the research methodology is still applicable, and effective, if future demand distributions are available, or if future demand can be estimated using demand forecasting algorithms.

Methodology

We pose the monolithic problem of simultaneously designing an aircraft and its operations as a mathematical programming problem that seeks to minimize (or maximize) a fleet-level objective by searching for the optimal values of a set of decision variables. These decision variables describe the requirements of the new system and the new system design features and determine the assignment of the new and existing systems to meet demand requirements under multi-domain uncertainties. The resulting problem is a stochastic MINLP problem.

- It is stochastic because of the presence of uncertainty in both new system design and pallet demand.
- It is mixed-integer because of the presence of continuous decision variables such as the aircraft design variables of aspect ratio and wing loading, along with integer decision variables such as pallet capacity.
- It is non-linear because of the existence of non-linear objective function and constraints related to the aircraft sizing equations.

Subspace Decomposition Strategy

The monolithic deterministic problem formulation results in an MINLP problem, which is, in general, difficult—if not impossible—to solve; MINLP problems combine the difficulty of nonlinear optimization and the combinatorial nature of mixed integer programs. The decomposition approach, a procedure of solving several domain-specific subproblems linked by a top-level problem, is one procedure that can obtain results for this kind of problem, with some minor modifications. Figure 2 presents the decomposition strategy and shows how information flows between the three smaller subproblems. The subspace problems presented here follow natural boundaries of the domains involved.

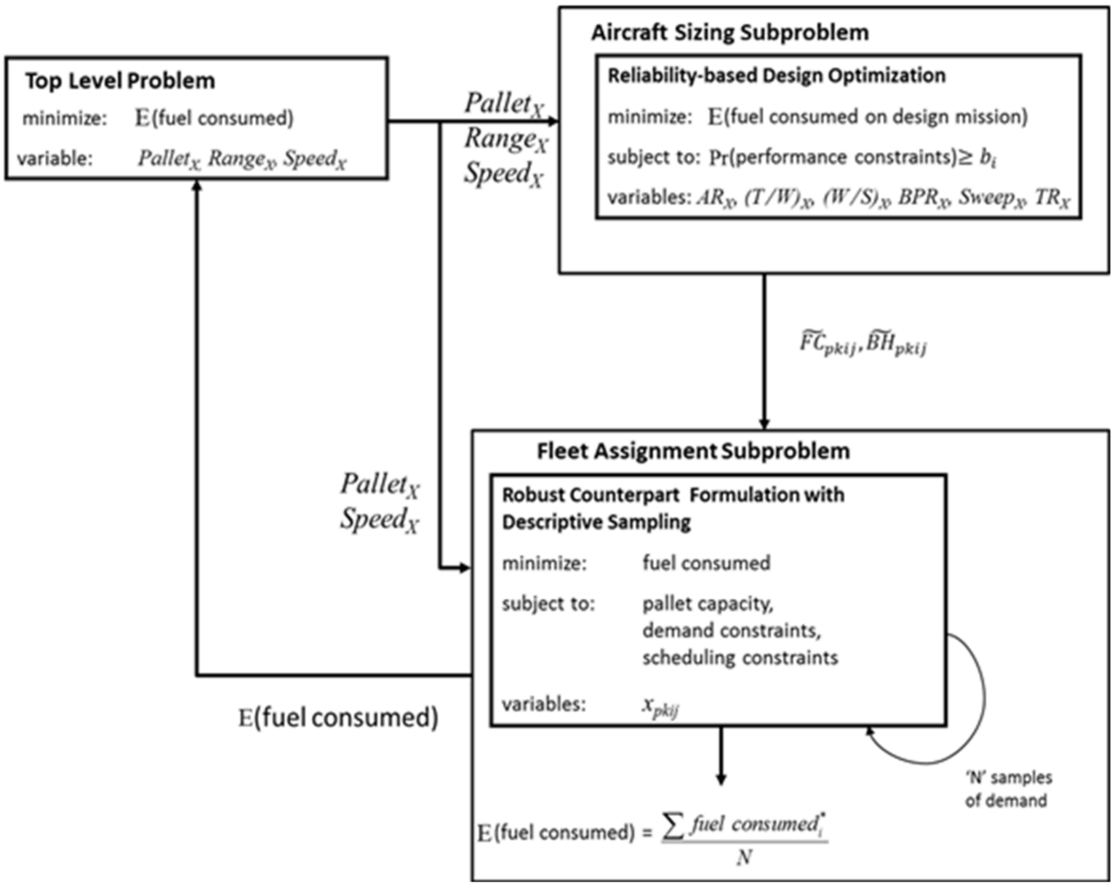


Figure 2. Subspace Decomposition of Monolithic Optimization Problem Addressing Uncertainty in Both Aircraft Sizing and Fleet Assignment

The top-level problem explores the requirements space for the new, yet-to-be-introduced aircraft based on fleet-level metrics. The top-level problem chooses candidate values for the top-level decision variables, which then become parameters for the aircraft sizing subproblem. A reliability-based design optimization formulation is used for the aircraft sizing subproblem. After the aircraft sizing subproblem is solved, the outputs of the aircraft sizing problem and the current values of the top-level optimization problem (namely the productivity coefficients and fuel consumption coefficients, pallet capacity and design range) then become inputs to the fleet assignment problem. A hybrid formulation that combines the descriptive sampling approach and interval robust counterpart formulation solves the fleet assignment subproblem. Here, the assignment problem's objective is to minimize the fleet-level fuel consumption using characteristics of the new, yet-to-be-introduced aircraft (range, pallet capacity, and speed) along with other existing aircraft in the fleet, subject to capacity, demand, fleet-level productivity, and scheduling constraints. The fleet-level values of the performance metrics return to the top top-level problem as the responses of interest.

Top-Level Problem

In this effort, the top-level optimization problem does not include any nonlinear constraints and only has bounds imposed on the top-level decision variables. Equations 1 to 4 describe the deterministic formulation of the top-level problem; the formulation incorporating uncertainty appears later in the paper.

$$
\begin{aligned}
&\textit{Minimize} && \textit{Fleet fuel}\left(Pallet_X, Range_X, Speed_X\right) && (1)\\
&\textit{Subject to} && 14 \le Pallet_X \le 38 && \text{(Design pallet capacity bounds)}\ (2)\\
& && 2400 \le Range_X \le 3800 && \text{(Range at max. payload bounds in nmi)}\ (3)\\
& && 350 \le Speed_X \le 550 && \text{(Cruise speed bounds in knots)}\ (4)\\
& && Pallet_X \in Z^+ \quad Range_X, Speed_X \in R^+ &&
\end{aligned}
$$

Equation 1 describes the objective function that seeks to minimize the fleet-level fuel consumption using pallet capacity, range and cruise speed of the new, yet-to-be-introduced aircraft type X as decision variables. Equations 2 to 4 describe the bounds for the top-level design variables. The values for the bounds were based on strategic airlift requirements and characteristics exhibited by current cargo transport aircraft (Gertler, 2010; Graham et al., 2003). Here, the design requirement decision variable describing payload capacity uses an integer number of pallets, while the design range and design speed decision variables are continuous.

Aircraft Sizing Subproblem

Uncertainty in Design Parameters

The conceptual phase of the aircraft design process relies upon semi-empirical equations and simplified physics models. The limited knowledge available about the system definition at this phase of the design process combined with the usage of low-fidelity modeling tools results in high uncertainty. Aircraft sizing typically determines the size, weight and performance of an aircraft to meet its design mission based on a set of nominal values on operating conditions (e.g., cruise altitude). However, when evaluating the “operating missions” to determine block time and fuel consumed on the flight, there might be a variation in assigned altitude, routing, speed, and so forth, which would alter the block time and fuel consumed. For instance, there is uncertainty in the prediction of the parasite drag coefficient. In this example, a scaling factor k_{C_D} follows a distribution to represent the uncertainty in the parasite drag prediction, so that the “actual” coefficient relates to the “predicted” coefficient in the following manner:

$$C_{D_0\ actual} = k_{C_D} \times (C_{D_0\ predicted})$$

To address the uncertainty related to operations and predictions of the new aircraft performance in the aircraft sizing subspace with reasonable computational expense, the Analysis of Variance (ANOVA) technique, a sensitivity analysis method, determined the subset of the most important parameters that influence the outputs under consideration (Montgomery, 2008). This investigation assumes triangular distributions for the scaling factors of identified parameters listed in Table 1.

Table 1. Triangular Distributions of the ANOVA Identified Uncertain Parameters in the Aircraft Sizing Subspace

Uncertain Parameters (ξ)	Lower limit	Mode	Upper Limit
C_{D_0} multiplier, k_{C_D}	0.90	1.0	1.10
Specific Fuel Consumption, SFC [hr^{-1}]	0.45	0.5	0.55
Oswald efficiency multiplier, k_{e_0}	0.95	1.0	1.05
Cruise altitude [ft]	32000	35000	38000
Pallet mass [lbs]	7200	7500	7800

The aircraft sizing subproblem seeks to minimize the fuel consumption of the new, yet-to-be-introduced aircraft for the values of design range (*Range*$_X$), pallet capacity (*Pallet*$_X$), and cruise speed (*Speed*$_X$) from the top-level problem. With the top-level objective to minimize fleet-level fuel consumption, and the aircraft sizing objective to minimize the fuel consumed by the new aircraft for its prescribed design range, pallet capacity, and cruise speed, a slight disconnect exists between the objectives of these two levels. The difference in the objectives is that, at each aircraft sizing iteration, the minimization of fuel consumption uses a single combination of fixed values for design range, pallet capacity, and cruise speed—this is the typical case in aircraft design where these quantities are set as requirements for some "representative design mission." However, the top-level optimization problem drives the question of "what requirements do we need to set in the first place?" by searching through the decision space of the top-level variables to find aircraft requirements that optimize fleet-level operational aspects of how the aircraft is used.

For example, consider the dimension of design range—as the top-level problem searches across values of range, this naturally changes the set of feasible routes that the new aircraft can fly, thereby changing how the fleet comprising existing and new aircraft serves the overall route network. By doing so, the top-level problem seeks additional fleet-wide fuel savings that these operational aspects reflect as a function of the decision variables. Therefore, the aircraft sizing objective can be viewed as a subset of the top-level problem objective. Because the type of aircraft assigned on individual flight segments drives the total amount of fuel consumed by the fleet, an aircraft designed for minimal fuel consumption will lead to improved fleet utilization that reduces fleet-level fuel consumption, when compared to fleet operations using only the fleet of existing aircraft. The approach in this work poses the aircraft design subproblem in the context of Reliability Based Design Optimization problem to account for uncertainty in the design phase.

The Reliability-Based Design Optimization (RBDO) formulation (shown below) represents the aircraft design under uncertainty problem.

$$\underset{x}{Minimize} \quad \mathrm{E}\left[f\left(x,\xi\right)\right]$$

$$Subject\ to \quad \mathrm{P}\left[g\left(x,\xi\right) \le 0\right] \ge b_i \qquad \forall\ i = 1,\ 2,\ \ldots,\ m$$

x : set of decision variables

ξ : set of uncertain parameters

b_i: desired probability of satisfying the i^{th} constraint

Aggregating the outputs for each realization (sample) of the uncertain parameter allows for the estimation of statistical measures such as expectation and probability, which the objective and constraint function evaluations require. The objective of the aircraft sizing subspace is to minimize the fuel consumption of the new aircraft X using the decision variables listed in Table 2. For each function evaluation of the top-level problem, the current values of $Pallet_X$, $Range_X$, and $Speed_X$ become fixed parameters for the aircraft sizing problem. Table 2 summarizes the decision variables, uncertain parameters and constraints in the aircraft sizing optimization problem.

Table 2. Decision Variables and Constraint Limits in the Aircraft Sizing Optimization Problem

Decision variables, (x)	Lower Bound	Upper Bound
Wing Aspect Ratio, AR_X	6.00	9.50
Thrust-to-weight Ratio, $(T/W)_X$	0.18	0.35
Wing Loading [lb/ft^2], $(W/S)_X$	65.00	161.00
Engine Bypass Ratio, BPR_X	4.50	14.50
Wing Leading Edge Sweep [deg], $Sweep_X$	10.00	35.00
Wing Taper Ratio, TR_X	0.10	0.40
Constraints	**Value**	
Takeoff Distance [ft]	≤ 8500	
Landing Distance [ft]	≤ 5500	
Second segment climb gradient	≥ 0.025	
Top-of-climb rate [ft/min]	≥ 500	

The aircraft sizing subproblem includes performance constraints such as limits on takeoff and landing distances, and also upper and lower bounds for the decision variables. The RBDO formulation optimizes the expected performance metric of interest and ensures that the probability of satisfying the performance constraints is greater than or equal to the user-defined reliability level, b_i, considering the uncertainty present in this subproblem.

Fleet Assignment Subproblem

The fleet assignment subproblem identifies the optimal assignment of the fleet's aircraft to meet demand obligations; this includes allocation of the new aircraft—as described by the solution from the preceding aircraft sizing subproblem—along with existing aircraft in the fleet. The following equations describe the deterministic formulation of the fleet assignment problem; a sampling approach, as described later, will address the uncertainty in the fleet assignment subproblem.

Minimize

$$\sum_{p=1}^{P}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \times FC_{p,k,i,j} \quad \text{(Fleet-level fuel consumption) (5)}$$

Subject to

$$\sum_{p=1}^{P}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \times \left(Speed_{p,k,i,j} \times Cap_{p,k,i,j}\right) \geq L \quad \text{(Fleet-level productivity limit) (6)}$$

$$\sum_{i=1}^{N} x_{p,k,i,j} \geq \sum_{i=1}^{N} x_{p,k+1,i,j} \quad \forall k = 1,2,3...K, \quad \forall p = 1,2,3...P, \quad \forall j = 1,2,3...N \quad \text{(Node balance constraints) (7)}$$

$$\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \times BH_{p,k,i,j} \leq B_P \quad \forall p = 1,2,3...P \quad \text{(Daily utilization limit) (8)}$$

$$\sum_{p=1}^{P}\sum_{k=1}^{K} Cap_{p,k,i,j} \times x_{p,k,i,j} \geq dem_{i,j} \quad \forall i = 1,2,3...N, \forall j = 1,2,3...N \quad \text{(Pallet demand constraints) (9)}$$

$$\sum_{i=1}^{N} x_{p,1,i,j} \leq O_{p,i} \quad \forall p = 1,2,3...P, \forall i = 1,2,3...N \quad \text{(Starting location constraints) (10)}$$

$$\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \leq 1 \quad \forall p = 1,2,3...P, \forall k = 1,2,3...K \quad \text{(Trip limit) (11)}$$

$$x_{p,k,i,j} \in \{0,1\} \quad \text{(Binary variable)}$$

Equation 5 is the objective function that seeks to minimize the fleet-level fuel consumption, where $FC_{p,k,i,j}$ indicates the fuel consumption coefficient of the k^{th} trip for aircraft p from base i to base j. The equation has two parts; the first product inside the square brackets, $x_{p,k,i,j} \times FC_{p,k,i,j}$, represents the fuel consumption of the existing fleet, while the rest of the terms inside the square brackets represents the fuel consumption of assigning the new, yet-to-be-designed aircraft. The fuel consumption characteristics of the new aircraft are a function of aircraft design variables (aspect ratio, thrust-to-weight ratio, etc.) and aircraft design requirements (pallet capacity, design range, and cruise speed). The term $x_{p,k,i,j}$ is a binary decision variable that takes a value of 1 if the k^{th} trip of aircraft p is flown from base i to base j, and it takes a value of 0 otherwise.

Equation 6 accounts for the multi-objective nature of this problem. This forces the fleet-level productivity to be greater than a pre-defined limit, L; the limit is varied and the problem is re-solved for each varied value of the limit to generate a set of Pareto optimal solutions. The term $x_{p,k,i,j} \times FC_{p,k,i,j}$ in Equation 6 refers to the productivity (speed of payload delivered) of utilizing aircraft type p for the k^{th} trip from base i to base j.

Equation 7 is the balance and sequencing constraint that enables the *(k+1)*th trip of an aircraft out of a base, i, to occur only after the k_{th} trip of that aircraft into base i. This constraint ensures that an aircraft needs is present at a base prior to completing a subsequent segment trip out of the same base.

Equation 9 limits flights to only occur within the daily utilization limit, B_p (here, this uses an assumption of 16 hours per day to account for loading, unloading, servicing, maintenance, etc.), of the aircraft, where $BH_{p,k,i,j}$ indicates the block hour of the k^{th} trip for aircraft p from base i to base j.

Equation 9 ensures that the carrying capacity of the combined trips meets or exceeds the pallet demand on each route, where $Cap_{p,k,i,j}$ indicates the pallet carrying capacity of the k^{th} trip for aircraft p from base i to base j.

Equation 10 ensures that the first trip of each aircraft p originates at its initial location (this is considered the aircraft's home or starting base for the day of operations); this initial location is randomly generated. Because the GATES dataset does not clearly indicate the starting location of aircraft each day, the problem formulation here uses a random distribution for each aircraft's starting location. The term $O_{p,i}$ is a binary variable that indicates if base i is the initial location for aircraft p.

Equation 11 ensures that each aircraft p flies at most one trip for its k^{th} segment.

The motivation for the "scheduling-like" formulation is to represent the scheduling and operations decisions made by the Air Mobility Command; it does not explicitly consider pilot scheduling (this 16 hours per day of available aircraft time could represent this, in part), nor does it account for the prioritization of cargo (this is not addressed in this formulation). This formulation, using node balance constraints, allows individual aircraft to make multiple flight segments in one day (as long as these fit within a prescribed time limit), allows for pallets to be carried from their origin to destination on possibly multiple aircraft, and tracks each individual aircraft by "tail number." These features more directly model AMC operations than some of the previous models of the authors and their colleagues when considering passenger airline transportation (Mane et al., 2007; Govindaraju et al., 2015).

Uncertainty in Fleet Operations

The uncertainty associated with the performance of the newly designed aircraft (type *X*) propagates to the fleet assignment subspace through the distributions of the new aircraft's predicted fuel consumption, $\widetilde{FC}_{p,k,i,j}$, and flight block hours, $\widetilde{BH}_{p,k,i,j}$, on given routes in the network; only aircraft "tail numbers" p that are associated with type *X* aircraft have these distributions. Additionally, the AMC service network has inherent pallet demand uncertainty, as described above. Hence, the fleet assignment problem now includes uncertainty in both the performance of the new aircraft and the pallet demand in the service network. In this paper, a hybrid formulation that combines the interval robust counterpart formulation (Lin, Janak, & Floudas, 2004) for user-defined tolerance parameters (δ) and the descriptive sampling technique (Saliby, 1990) solves the fleet assignment problem under uncertainty.

Lin et al. (2004) proposed a robust optimization approach for bounded uncertainty to overcome the large computational expense incurred by scenario/sampling-based frameworks. Their approach produces "robust" solutions that are immune against uncertainties in both the coefficients and right-hand-side parameters of the inequality constraints of the Mixed Integer Linear Programming (MILP) problems. Lin et al. (2004) term a solution to be robust if it satisfies the following conditions:

- The solution is feasible for the nominal values of the uncertain parameters.
- For any value of the uncertain coefficients in the objective function and the uncertain parameters in the right-hand side of the constraints, the solution must satisfy the i^{th} inequality constraint or, at worst, violate the constraint with an error of at most δ x max[1,|p_l|]. In this expression, δ is a user-selected infeasibility tolerance coefficient, and p_l is the right-hand-side limit of the linear inequality constraint.

Applying the interval robust counterpart model to the deterministic formulation of the fleet assignment subproblem described above results in two additional sets of constraints and a modified objective function where an auxiliary variable (*Fleet fuel*) is introduced to enable introduction of the original objective function represented by Equation 5 as a constraint—thereby making it amenable to robust optimization strategies. The reformulation of the original objective function (Equation 5) is now as follows:

$$Minimize\ Fleet\ fuel \tag{12}$$

$$Subject\ to\ \sum_{p=1}^{P}\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \times FC^{U}_{p,k,i,j} \le Fleet\ fuel\left(1+\delta\right) \tag{13}$$

where $FC^U_{p,k,i,j}$ is the upper bound of the fuel consumed by aircraft p on the k^{th} trip from base i to base j. Evaluating the performance of the new aircraft for different samples of the aircraft sizing uncertain parameters (ξ) generates distributions of the performance metrics such as the fuel consumption coefficient. The upper bound, $FC^U_{p,k,i,j}$, is then determined from the distribution of the fuel consumption coefficient $\widetilde{FC}_{p,k,i,j}$ applied to only aircraft p that are of the newly-designed type *X*. δ is the user-defined, infeasibility tolerance parameter that can take values between 0 and 1. For example, setting δ to 0.1 for a particular constraint indicates that 10% violation of the worst-case scenario of that constraint is acceptable. Using Equation 13, if all of the uncertain fuel consumption coefficients for the new aircraft are at their upper bound (i.e., the aircraft burns the most possible fuel from the distribution, ($\widetilde{FC}_{p,k,i,j}$), then the total fuel consumed by the fleet is no more than 10% above the user-defined limit for fleet fuel consumption. The daily utilization limit constraint (Equation 8) is modified as follows:

$$\sum_{k=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N} x_{p,k,i,j} \times BH^{U}_{p,k,i,j} \le B_P\left(1+\delta\right) \quad \forall p = 1,2,3...P \tag{14}$$

where $BH^U_{p,k,i,j}$ is the upper bound of the distribution of block hours of aircraft p (restricted to only aircraft of type *X*) on the k^{th} trip from base i to base j. The deterministic robust counterpart fleet assignment problem now includes Equations 12, 13, and 14 in addition to Equations 5 to 11 from the original deterministic formulation of the fleet assignment problem.

The interval robust counterpart model is also applicable for the demand constraint (Equation 9) in the deterministic formulation, but this leads to a very conservative (protected against the maximum demand scenario) solution because the right hand side constraint limit, $dem_{i,j}$, is set to its upper bound or maximum value, $dem^U_{i,j}$, for each route as shown in Equation 15 below. For this constraint, the GATES dataset provides the values for the upper bound of the pallet demand on each route.

$$\sum_{p=1}^{P}\sum_{k=1}^{K} Cap_{p,k,i,j} \times x_{p,k,i,j} \geq dem_{i,j}^{U}$$
$$\forall i = 1,2,3...N, \forall j = 1,2,3...N \quad (15)$$

Instead, because of the AMC service network's high fluctuations in pallet demand and the on-demand nature of military cargo transport, the approach here employs a descriptive sampling approach (Saliby, 1990) to incorporate the stochastic nature of the demand. The method of descriptive sampling involves a deliberate collection of sample values that closely describes the represented distribution. The descriptive sampling approach samples more values from regions of higher density and fewer values from regions of lower density. The purposeful collection of sample values at specific quantile levels helps to match closely the actual or reported discrete demand distributions using a reduced number of samples, thus reducing the computational expense.

The deterministic robust counterpart formulation is solved multiple times for each demand sample vector generated through the descriptive sampling approach. From these multiple solutions, the expected value of the fleet-level performance metrics (fleet-level fuel consumption and/or fleet-level productivity) now return to the top-level optimization problem as the responses of interest. The robust counterpart formulation accounts for the propagation of uncertainty from the aircraft sizing to the fleet assignment subspace, while the descriptive sampling approach addresses the stochastic nature of pallet demand in the service network.

25-Base Network Problem

This section demonstrates how the subspace decomposition approach can identify the best new aircraft requirements and subsequent aircraft design to address fleet-level metrics under uncertainty. By treating this problem as a multiobjective problem, the approach can also generate tradeoffs between fleet-level metrics of interest; from these best tradeoff solutions, a decision-maker can also observe how the optimum design requirements for the new aircraft change for these different tradeoff opportunities.

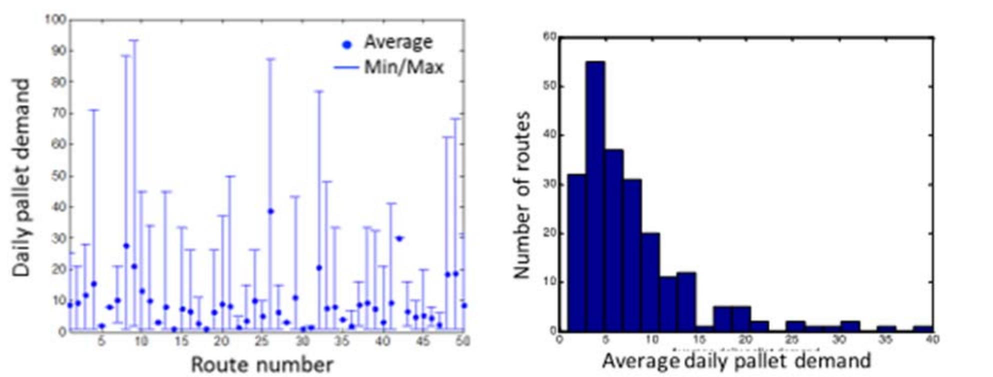
Network Description

This study uses a subset of the AMC route network and fleet, comprising 25 bases and 219 directional routes, to demonstrate the approach. Figure 3 depicts the geographical locations and routes of the 25-base network. For the 25-base network, the existing fleet of AMC comprises 28 C-5s, 44 C-17s, and 21 chartered 747-Fs. The existing fleet serves as a "baseline" to measure the improvements due to the introduction of the new aircraft. This study assumes that five new, yet-to-be-designed-aircraft (all of type X) are introduced into the fleet. This assumption reflects an external decision made by the user or the decision-maker that specifies the number of new aircraft that are added to the fleet.

Figure 3. 25-Base Network

Note. Illustration was generated using http://www.gcmap.com/.

The 25 bases in the network are either the origin or the destination locations that transported the largest number of pallets in the AMC service network for the year 2006. The routes span the continents of North America, Asia, and Europe. Figure 4a shows the average and the minimum/maximum of the directional daily pallet demand for 50 routes in the network. Figure 4b shows the distribution of the number of routes based on the average daily pallet demand. The histogram indicates that the demand distribution is right-skewed and that several of the routes have an average daily demand of less than 20 pallets.



(a) Directional Daily Pallet Demand (b) Histogram of Average Daily Pallet Demand

Figure 4. Pallet Demand Characteristics of the 25-Base Network

Results

In this study, the top-level optimization problem (refer to Figure 2) chooses candidate values for the decision variables of pallet capacity, design range, and cruise speed. These candidate values then become inputs to the aircraft sizing subproblem. The RBDO formulation of the aircraft sizing subproblem uses 50 samples for the uncertain parameters; this is a small number, but it allows for a tractable computational time. The reliability level, b_i, is set to 0.90 for the performance constraints in the aircraft.

After the aircraft sizing problem is solved, the outputs of the top-level subspace such as pallet capacity, and the outputs of the aircraft sizing subspace, the uncertain performance coefficients such as $\widetilde{FC}_{p,k,i,j}$, then become inputs to the fleet assignment subspace. The interval robust counterpart formulation of the fleet assignment problem is solved for 20 samples of demand across the network generated through the descriptive sampling approach. In this study, the infeasibility tolerance parameter, δ_i, is set to 0.10 for all the appropriate constraints. The expected values of the fleet-level performance metrics, calculated from the different solutions of the robust counterpart formulation, now return to the top-level subspace, and this process continues until convergence at the top-level.

Then, to identify tradeoffs between fleet-level fuel consumption and productivity, the entire process repeats with a different limit value on the productivity constraint. Minimizing the fleet-level fuel consumption with several different limits on fleet productivity leads to a number of tradeoff solutions. Figure 6 shows the results from the multi-objective analyses of the 25-base network problem. The plot shows the normalized expected values of the fleet-level metrics. Using normalized fleet-level responses helps to identify the trends and helps to show the relative variations in fleet-level responses for different solutions to the multi-objective optimization problem. The fleet-level responses have been normalized with respect to the lowest expected values from the results of the scenario labeled “Fleet with five new A/C.” Each point in the “Fleet with five new A/C” scenario describes the optimal design of the new aircraft required to meet the specific fleet-level objectives. These results show the collection of optimal aircraft designs that would meet the fleet’s operational needs at each level of permitted fuel consumption or at each level of required fleet-wide productivity.

For three different solutions from the “Fleet with five new A/C” results, Figure 5 contains callout boxes that describe the values of the new aircraft requirement decision variables along with the values of the aircraft design variables. The trends in the fleet-level responses are as expected, with fuel consumption increasing as productivity increases. There appears to be a trend in the “size” of the optimal aircraft along the Pareto frontier for increasing productivity/fuel consumption values. For a normalized expected productivity and normalized expected fuel consumption value of 1.0, the optimal requirement decision variables of the new aircraft X are at the lower bounds for pallet capacity (16) and design range (3800 nmi). Moving from this point on the tradeoff plot towards solutions with increasing fleet-level productivity, the results suggest that larger pallet capacities for the new aircraft X can best meet the fleet-level objectives. There is not substantial evidence to determine whether these trends would generalize to other route networks or other similar design problems; however, the behavior is not unexpected because the aircraft pallet capacity strongly drives the fleet-level productivity metric. Though it is intuitive that a larger aircraft would increase productivity, the optimal design features of the new aircraft X, such as the aspect ratio (AR_X), the wing loading ($(W/S)_X$), the thrust-to-weight ratio ($(T/W)_X$), etc., are reflective of the specific existing fleet and demand characteristics of the service network. For each solution in the plot, the assignments of the fleet of aircraft to routes are different to meet the actual demands better. The introduction of the five new aircraft (of type X) results in fleet-level fuel savings between 2.79% and 6.48% for the same normalized expected fleet productivity values when compared to the case where only the existing fleet operates in the network.

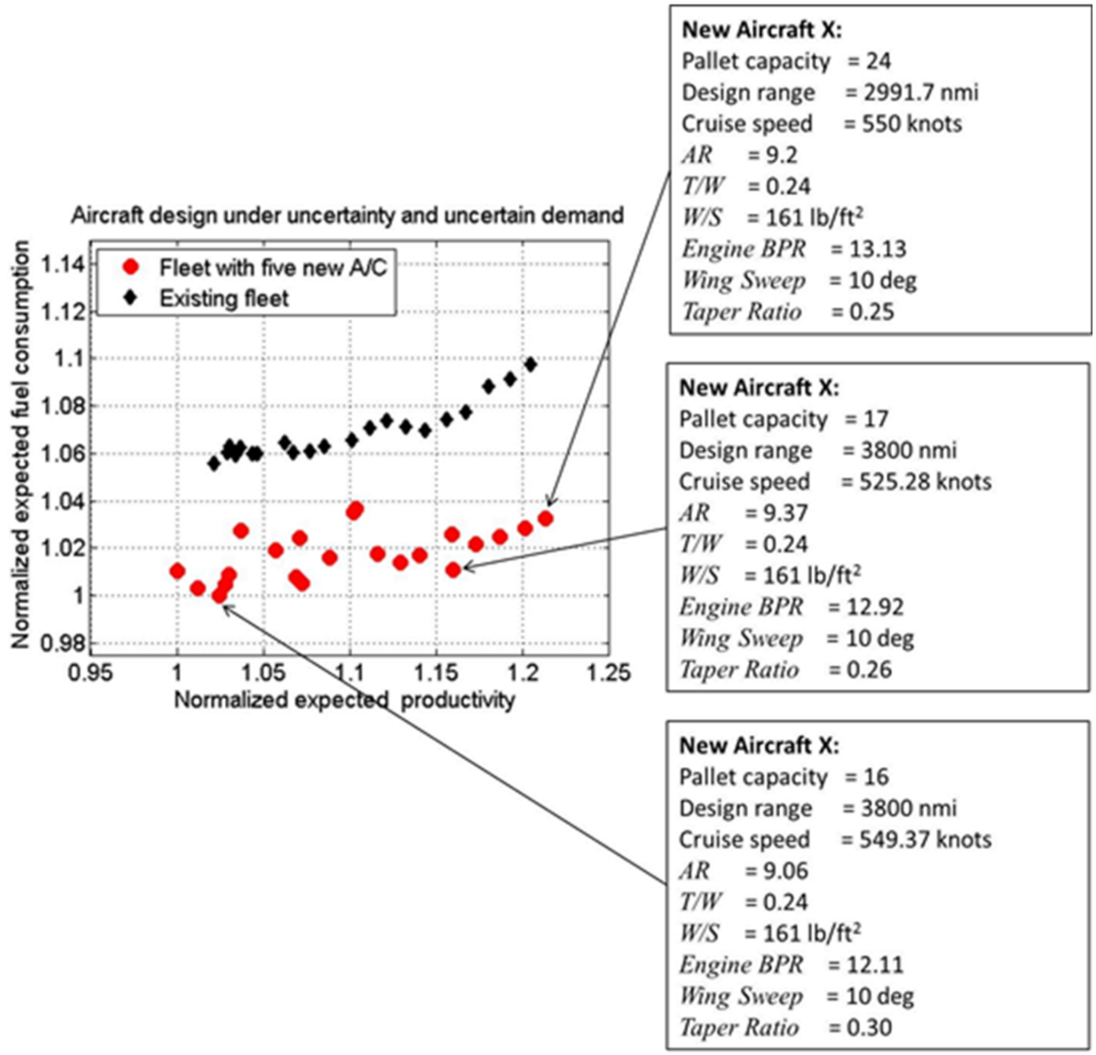


Figure 5. Results From Multi-Objective Analyses of 25-Base Network Problem

The solutions to multi-objective analyses present a way to perform “fuel/cost as an independent variable” type of trade-space analysis; this might be more obvious by switching the axes in the plot from Figure 5. These types of plots can help decision-makers/acquisition planners to analyze the trade-space and select the optimal requirements and design of the new aircraft that would achieve the desired level of fleet fuel consumption and productivity. For instance, a decision-maker can determine the level of fleet productivity available for a specific level of fleet fuel consumption; this fleet-level productivity value can then be translated to a specific (or bounded) level for the mobility airlift requirements that are set by the DoD in terms of tonnage of cargo transported per day. Having established the goals for the fleet-level productivity and fuel consumption, the collection of optimal aircraft designs required to achieve these fleet-level goals can be determined from plots such as those shown in Figure 5.

Decision-makers/acquisition planners can use such results to perform comprehensive exploratory analysis of the design space and identify regions in this design space that present significant viable opportunities to reduce the fleet fuel consumption. For instance, the AMC may need to incur “switching costs” (additional cost for training, maintenance and infrastructure due to the addition of a new aircraft type into the fleet) of integrating a new aircraft type into the fleet for a relatively small decrease in fuel burn; however, the trade-space analysis (Figure 5) can help identify promising designs and “inflection points,” if they exist, where the decision to acquire a new aircraft type could provide significant benefits.

Modeling and Solution Procedure for Commercial Air Travel Applications

In an effort to explore broader applications of the approach for similar acquisition-related issues, the authors conjecture how the approach could help decision-makers consider the best requirements for a new passenger transport aircraft. The nature and structure of uncertainty in commercial passenger air travel differs from the characteristics of data for the AMC service network. Adapting the subspace decomposition framework to commercial applications requires proper attention to the differences in the nature of the uncertainty that manifests in the data.

Similar to the approach used to extract an example problem from the GATES data, data from the Bureau of Transportation Statistics (BTS) T100 Segment database for non-stop monthly passenger demand can provide the basis for a commercial passenger airline problem.

The operations of the commercial air travel industry differ from military airlift operations, such as those managed by the Air Mobility Command (AMC) of the U.S. Air Force. The primary difference lies in the fact that commercial aviation operators such as airlines publish their schedules several weeks in advance of operating the flights, limiting the opportunities for modifications in the face of uncertain passenger demand. However, the AMC has a higher flexibility to modify their flight schedules due to the on-demand nature of palletized cargo transportation. The airline planning process typically involves a chronological sequence of decision-making phases. The planning process starts with schedule planning and development followed by four concurrent routines, namely, crew scheduling, revenue management, airport resource management, and aircraft maintenance routing. The schedule planning and development phase comprises market forecasting, schedule construction, capacity planning, fleet assignment, and schedule evaluation procedures.

For the purposes of strategic fleet planning and acquisition decision-making, the fleet allocation formulation for the commercial air travel case study integrates the schedule creation and fleet assignment procedures into a single mathematical programming problem. Figure 6 shows the modified subspace decomposition framework addressing uncertainty in both the design of the new aircraft and passenger demand for the commercial air travel case study.

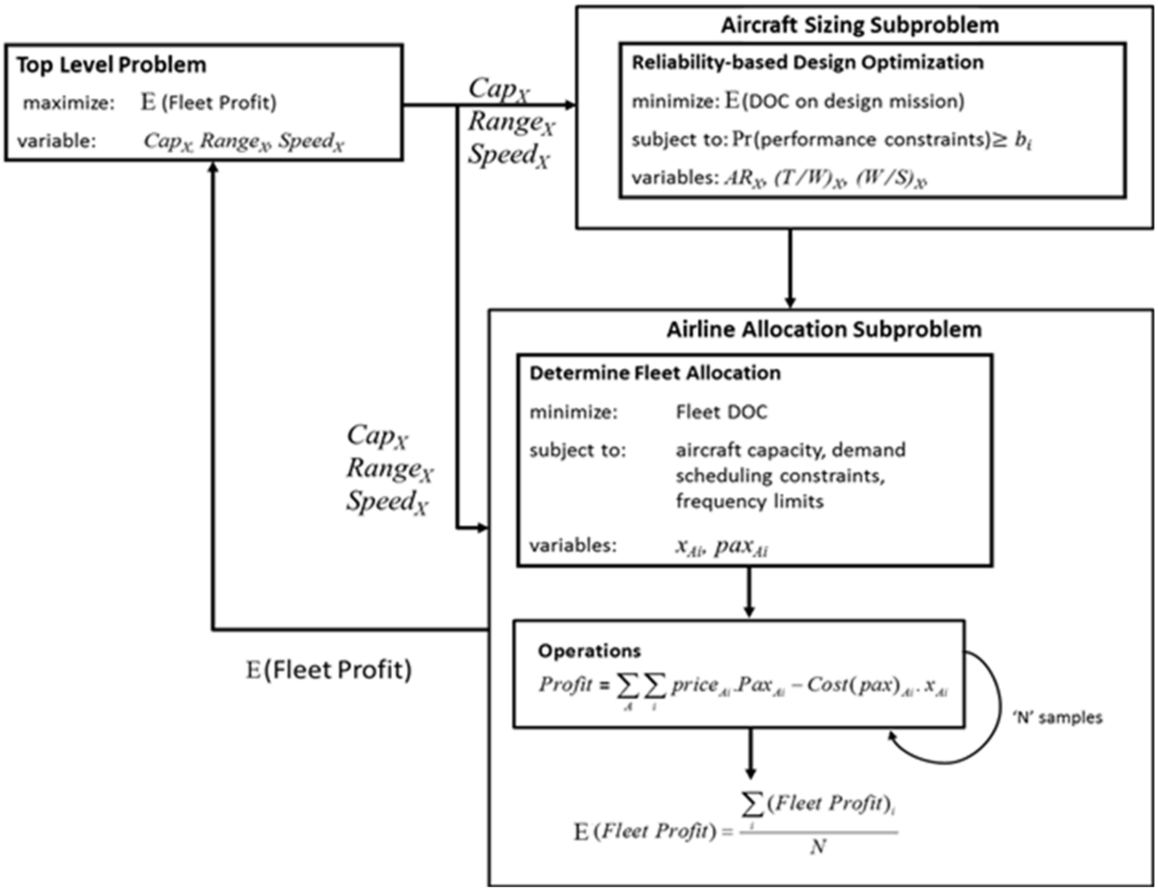


Figure 6. Subspace Decomposition Framework Addressing Uncertainty in Both the Aircraft Sizing and Airline Allocation Subspace for Commercial Air Travel Application

The top-level problem explores the “requirements space” for the new, yet-to-be-designed aircraft using passenger capacity, range and cruise speed as the top-level decision variables. The aircraft sizing subspace accounts for the inherent uncertainty present in the conceptual phase of the design process through an RBDO formulation.

The airline allocation subspace is solved in two steps. The first step involves determining the minimum cost schedule based on the maximum number of passengers transported on each route. To use the framework in practice, the maximum anticipated demand would rely upon internal analysis performed by the airline to predict this; therefore, the demand is a point of input from the analyst and reflects a priori beliefs on the future state of demand. Here, our implementation will use the reported historical demand served on each route as available from the Bureau of Transportation Statistics, and the highest demand served will take the place of what we would expect the airline to predict. Solving the airline allocation problem generates an optimum schedule for the aircraft in the fleet to service the passenger demand in the route network. The second step involves a Monte Carlo simulation to account for the uncertainty in actual/realized passenger demand. The two-step procedure mimics the decision-making process of an airline, where a specific number of seats are allocated first for each route, followed by passengers buying tickets from the airline for traveling on those routes. Using the data from the demand distribution plots, the Monte Carlo simulation calculates the profits for the various simulated instances of passenger demand. From these numerous samples, the average (expected) fleet profit values are then estimated. If required, the airline allocation subspace can be solved for four different quarters of demand data to reflect the seasonal variations in demand and the tactical planning timeframe usually employed by airlines. The expected fleet profit values return to the top-level subspace as the metrics of interest. The process continues until the top-level converges. At convergence, the solution describes the optimal aircraft requirements, the optimal description of the new aircraft, and the optimal allocation of the new and existing fleet of aircraft.

Concluding Statements and Future Work

The approach presented in this paper allows investigation of tradeoffs between fleet-level fuel usage, performance metrics and acquisition alternatives for a conceptual problem based on operations of the U.S. Air Force Air Mobility Command (AMC) under domain-specific uncertainties. The approach, while applied to the AMC case study, appears to be domain agnostic. Results from the AMC case study describe a collection of optimal aircraft design requirements and subsequent aircraft design descriptions that reduce fleet-level fuel consumption while satisfying the operational requirements under uncertainty in the new system design and uncertainty in the service network demand. A reliability-based design optimization formulation addresses uncertainty in the design of the new aircraft. A hybrid fleet assignment formulation that combines the interval robust counterpart model and the descriptive sampling approach addresses both the propagation of uncertainty from the aircraft sizing subspace to the fleet assignment subspace and the demand uncertainty in the service network. The immediately preceding section describes modification of the approach to address a commercial passenger airline application.

The methodology described in this paper can help guide decision-makers and acquisition planners to determine optimal design requirements for new, yet-to-be-introduced aircraft to reduce fleet-level fuel consumption. Solutions from these “design under uncertainty” problems provide insight (expected performance gain and costs incurred) about new systems, and these insights can inform acquisition decisions related to setting the right design requirements for the new system. Addressing uncertainty explicitly in this quantitative approach allows for a more “robust” selection of these new system requirements.

Using the approach to address this as a multi-objective problem enables tradeoffs in the context of “fuel/cost as an independent variable.” Generating the new design requirement and new aircraft design solutions should facilitate discussion and understanding about what features this kind of process should entail under various operational scenarios. The results from the 25-base network problem demonstrate the quantitative framework’s applicability in guiding potential acquisition decisions under uncertainty for the AMC case study, and the computational tractability of the approach to solve large-scale real-world problems. A preliminary framework for adapting the approach to commercial aviation application is presented as well. Future work will focus on extending the decomposition approach to solve the combined aircraft design and fleet assignment problem under commercial aviation specific uncertainties for the commercial travel case study.

References

Gertler, J. (2010). *Air Force C-17 aircraft procurement: Background and issues for Congress.* DIANE.

Govindaraju, P., & Crossley, W. A. (2015). Concurrent aircraft design and trip assignment under both design parameter and demand uncertainty. In *15th AIAA Aviation Technology, Integration, and Operations Conference (ATIO).*

Govindaraju, P., Davendralingam, N., & Crossley, W. (2015). A framework to determine new system requirements under design parameter and demand uncertainties. In *Proceedings of the 11th Annual Acquisition Research Symposium.* Monterey, CA: Naval Postgraduate School.

Graham, D., Bracken, J., Dalfonzo, J., Fedorochko, W., & Hilton, R. (2003). *Sustaining the Civil Reserve Air Fleet (CRAF) program* (No. IDA/HQ-P-3819). Alexandria, VA: Institute for Defense Analyses.

Greer, W. L. (2010). *An application of cost-effectiveness analysis in a major defense acquisition program: The decision by the U.S. Department of Defense to Retain the C-17 transport aircraft* (No. IDA-D-4218). Alexandria, VA: Institute for Defense Analyses.

Lin, X., Janak, S. L., & Floudas, C. A. (2004). A new robust optimization approach for scheduling under uncertainty: I. Bounded uncertainty. *Computers & Chemical Engineering, 28*(6), 1069–1085.

Mane, M., Crossley, W. A., & Nusawardhana, A. (2007). System-of-systems inspired aircraft sizing and airline resource assignment via decomposition. *Journal of Aircraft, 44*(4), 1222–1235.

Montgomery, D. C. (2008). *Design and analysis of experiments*. John Wiley & Sons.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2014). *Better buying power 3.0* (White paper). Washington, DC: Author.

Saliby, E. (1990). Descriptive sampling: A better approach to Monte Carlo simulation. Journal of the *Operational Research Society, 41*(12), 1133–1142.

Taylor, C., & de Weck, O. D. (2007). Coupled vehicle design and network flow optimization for air transportation systems. *Journal of Aircraft, 44*(5), 1479–1486.

Panel 14. The Big Picture of Defense Acquisition

Thursday, May 5, 2016	
11:15 a.m. – 12:45 p.m.	**Chair: Andrew Hunter,** Senior Fellow in the International Security Program, Director of Defense-Industrial Initiatives Group, Center for Strategic International Studies ***Defense Modernization Plans Through the 2020s: Addressing the Bow Wave*** Todd Harrison, Director, Defense Budget Analysis and Senior Fellow, CSIS ***Speed and Agility: How Defense Acquisition Can Enable Innovation*** Peter Modigliani, Division Chief Acquisition Specialist, The MITRE Corporation ***Defense Industrial Base Issue That Can Be Overlooked When Focusing on Major Weapon Systems*** Nancy Moore, Senior Management Scientist, RAND

Andrew Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at CSIS. He focuses on issues affecting the industrial base, including emerging technologies, sequestration, acquisition policy, and industrial policy. From 2011 to November 2014, Hunter served as a senior executive in the Department of Defense (DoD). Appointed as Director of the Joint Rapid Acquisition Cell in 2013, his duties included fielding solutions to urgent operational needs and leading the work of the Warfighter Senior Integration Group to ensure timely action on critical issues of warfighter support. From 2011 to 2012, he served as Chief of Staff to Ashton B. Carter and Frank Kendall while each was serving as Under Secretary of Defense for Acquisition, Technology, and Logistics. Additional duties while at the DoD included providing support to the Deputy's Management Action Group and leading a team examining ways to reshape acquisition statutes.

From 2005 to 2011, Hunter served as a professional staff member of the House Armed Services Committee, leading the committee's policy staff and managing a portfolio focused on acquisition policy, the defense industrial base, technology transfers, and export controls. From 1994 to 2005, he served in a variety of staff positions in the House of Representatives, including as appropriations associate for Representative Norman D. Dicks, as military legislative assistant and legislative director for Representative John M. Spratt, Jr., and as a staff member for the Select Committee on U.S. National Security and Military/Commercial Concerns with the People's Republic of China. Hunter holds an MA degree in applied economics from the Johns Hopkins University and a BA degree in social studies from Harvard University.

Defense Modernization Plans Through the 2020s: Addressing the Bow Wave

Todd Harrison—is the Director of Defense Budget Analysis and a senior fellow in the International Security Program at CSIS. He leads the Center's efforts to provide in-depth, nonpartisan research and analysis of defense funding issues and provides expert analysis on space security issues. Harrison joined CSIS from the Center for Strategic and Budgetary Assessments, where he was a senior fellow for defense budget studies. He has authored publications on trends in the overall defense budget, defense acquisitions, military compensation, military readiness, the cost of nuclear forces, military space systems, and the cost of the wars in Iraq and Afghanistan.

Harrison frequently contributes to print and broadcast media and has appeared on CNBC, CNN, NPR, Al Jazeera English, and Fox News. He has been a guest lecturer for organizations and teaches a class on the defense budget at George Washington University's Elliott School of International Affairs and classes on military space systems and the defense budget at Johns Hopkins University's School of Advanced International Studies. He is a term member of the Council on Foreign Relations and was named one of the Defense News 100 Most Influential People in U.S. Defense.

Harrison previously worked at Booz Allen Hamilton where he consulted for the Air Force on satellite communications systems and supported a variety of other clients evaluating the performance of acquisition programs. Prior to Booz Allen, he worked for AeroAstro Inc. developing advanced space systems and technologies and as a management consultant at Diamond Cluster International. Harrison served as a Captain in the U.S. Air Force Reserves from 1998 to 2003. He is a graduate of the Massachusetts Institute of Technology with both a BS and an MS in Aeronautics and Astronautics.

Abstract

Since the enactment of the Budget Control Act (BCA) of 2011, much attention has been paid to the near-term effects of budgetary constraints on national defense. What has received less attention are the looming budgetary challenges that defense faces beyond the BCA budget caps and the Defense Department's five-year budget planning horizon. Many weapons programs will be at or near their peak years of funding requirements at roughly the same time in the 2020s, creating a modernization bow wave. Just as a large bow wave slows a ship by diverting its energy, carrying a large modernization bow wave is a drag on defense because it leads to program instability and inefficient procurement practices that weaken the buying power of defense dollars.

Current plans for major acquisition programs appear to follow the typical pattern of a modernization bow wave, with funding projected to increase by 23% from FY 2015 to the peak in FY 2022. However, this modernization bow wave is not evenly distributed across the Services and defense-related agencies. Much of the projected increase in modernization funding is driven by Air Force aircraft modernization programs, which are projected to nearly double in costs and account for nearly half of the overall bow wave increase. In contrast, Navy and Marine Corps modernization funding remains relatively flat through the early 2020s and then declines in the later part of the decade, driven mainly by a decline in aircraft procurements. The Army's budget for major acquisition programs is projected to increase 28% in real terms from FY 2015 to the peak in FY 2022, with notable bow waves in funding for ground and communications systems. However, these increases are balanced in part by a sharp reduction in Army aircraft procurements, and the total magnitude of increase in Army funding for major programs dwarfs in comparison to the increase in Air Force major programs.

This CSIS report details the plans for major acquisition programs over the next 15 years and explores the complicating factors that may make the situation more problematic for policymakers. It analyzes a range of options to mitigate the bow wave, including increasing the budget, cutting additional force structure, and making trades among major acquisition

programs. The report finds that while none of the choices available are easy, it provides an opportunity for the new administration taking office in 2017 to better align modernization plans with defense strategy.

Link to full report: https://csis.org/files/publication/160126_Harrison_DefenseModernization_Web.pdf

Speed and Agility: How Defense Acquisition Can Enable Innovation

Peter J. Modigliani—is the Division Chief Acquisition Specialist at the MITRE Corporation. He supports DoD acquisition and CIO executives' strategic initiatives in Agile, cyber, IT, and services acquisition. Previously, as an Assistant Vice President with Alion, he supported the Air Force Acquisition Executive on C4ISR systems. As an Air Force program manager, he developed strategies for billion dollar acquisitions. Pete holds a BS in industrial engineering from the Rochester Institute of Technology and an MBA in IT from Boston College. He is DAWIA Level III in program management. [pmodigliani@mitre.org]

Abstract

The Department of Defense (DoD) leadership demands a more agile, innovative enterprise that can rapidly integrate and deliver leading technologies. In its struggle to keep up with the rapid pace of change in both threats and technologies, the DoD is burdened by complex, bureaucratic processes, policies, and culture that hinder speed and agility. The disjointed budget, requirements, and acquisition domains compound the DoD's difficulties. Many acquisition professionals lack the requisite experience to navigate a disorganized knowledge enterprise to develop strategies and execute processes. Congress and DoD executives have instituted many initiatives to rapidly acquire and deliver capabilities to the warfighters, but these have varying maturity and success.

The DoD can implement key enablers from the enterprise to the tactical levels to replicate the success of government and industry innovations. Schedule should join cost as a top priority for a DoD acquisition enterprise that builds upon and integrates many innovative organizations and initiatives into its activities. This requires bold leadership to reshape the culture and enable top talent to prosper. The DoD should restructure programs and portfolios to enable agile and iterative developments, continue partnerships with established industry, and engage the services of innovative new firms to maintain technological superiority.

Strategic Imperative

Over the last few years, the president, DoD executives, and Congress have sought to ensure the DoD is more agile, flexible, and technologically advanced. Better Buying Power (BBP) 3.0 initiatives include incentivizing innovation and productivity in government and industry, eliminating unproductive bureaucracy, and promoting effective competition. These initiatives are designed to counter the threat that adversaries pose to U.S. technological superiority.

The FY16 National Defense Authorization Act (NDAA) includes a number of provisions to drive speed, agility, and innovation. These include expanding rapid innovation programs and rapid acquisition authorities. One section that holds particular promise directs the creation of a middle-tier of acquisition to promote rapid prototyping and rapid fielding acquisition pathways. These programs rapidly field either prototypes or production units and complete fielding within five years. There are provisions for funding R&D and rapid prototypes. It empowers senior officials to waive laws and policies that impede certain rapid acquisitions. Other types of programs seek to time-box the lengthy requirements process and better align the acquisition and budget systems to support speed and agility.

In March 2016, House Armed Services Committee (HASC) Chairman Mac Thornberry introduced the Acquisition Agility Act to spur the next set of reforms that will ensure that the DoD can respond to rapidly changing threats (Thornberry, 2016). It seeks to enable the DoD to field better technology faster by restructuring major weapon systems, allowing them to rapidly deliver a minimum acceptable capability, then incrementally develop additional components for an open-designed platform.

Key Challenges/Barriers

The Defense Acquisition Enterprise is one of the world's biggest bureaucracies, eclipsed only by the full DoD and federal government as a whole. The enormous burdens imposed by laws, policies, guides, and memoranda from multiple levels of DoD and Service oversight overwhelm programs. All DoD programs follow most of the same processes in the acquisition framework, yet each program spends considerable time and energy identifying the required processes and how to execute them. These processes force program offices to spend far too much time generating paperwork and navigating the bureaucracy rather than thinking creatively about program risks, opportunities, and key elements of their strategies.

Technology Adoption

While the DoD once led technology R&D, global commercial companies now drive innovation (see Figure 1; The White House, 2016). The DoD's R&D budget has declined by over 20% from its peak in 2010, and the defense industry R&D dropped by a third from 1999 to 2012. Many defense firms followed industry trends of stock buy-backs to obtain short-term financial gains, and deferred long-term technology investments. Google, Apple, and Microsoft spend five to six times more on R&D than the five largest defense firms combined (Center for a New American Security, n.d.). As a result, the defense industry moves too slowly to adjust to current technology trends. This has prompted many DoD executives to place high priority on reaching out to new industry partners and breaking down the barriers that prevent organizations from doing business with the DoD.

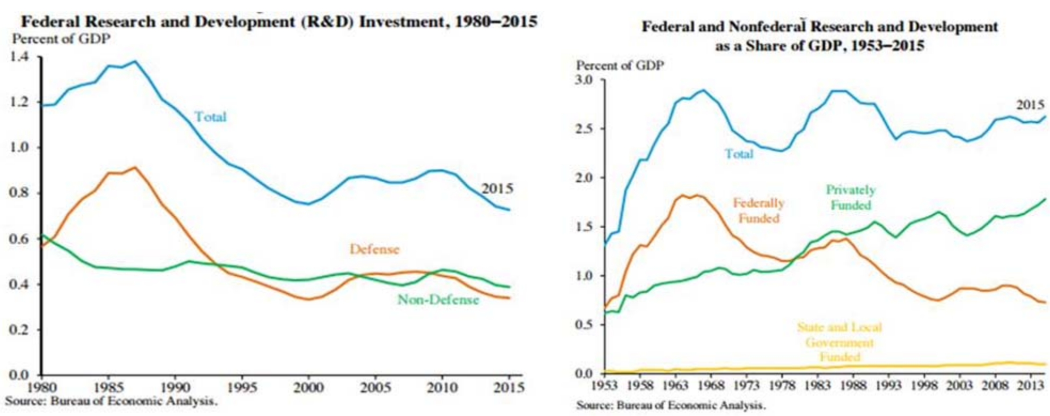


Figure 1. Federal R&D Investments

In *Crossing the Chasm*, Geoffrey Moore (1991) describes the vast gap between early adopters of a high-technology product and the early majority of the market (Figure 2). Drawing on the technology adoption life cycle model and the diffusion of innovations theory, Moore outlines the different expectations of each group and proposes strategies for mainstream product adoption.

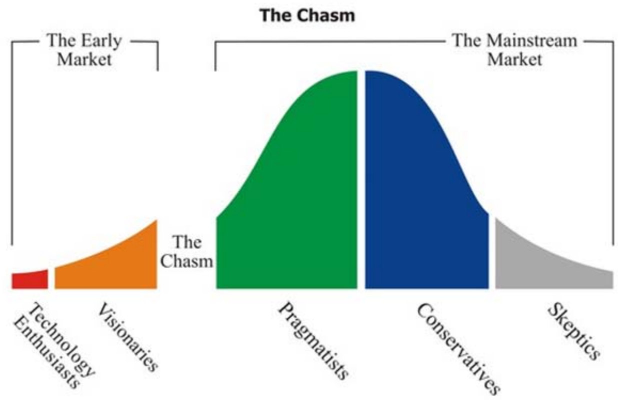


Figure 2. Industry Technology Adoption

The DoD confronts a similar chasm between the emergence of innovative technologies and the integration of those technologies into programs of record (see Figure 3). DoD labs, the Defense Advanced Research Projects Agency (DARPA), Federally Funded Research and Development Centers (FFRDCs), academia, and industry continue to develop exciting new technologies, but the current acquisition system makes it impossible to rapidly and effectively leverage them for the warfighter. One challenge centers on identifying new technology that could remedy operational shortfalls or enable programs to take advantage of opportunities. Furthermore, if a small business demonstrates an operational solution, it must often take part in a lengthy competition with no guarantee of eventual income and, if selected, be subject to rigorous design, testing, and security protocols designed for larger companies.

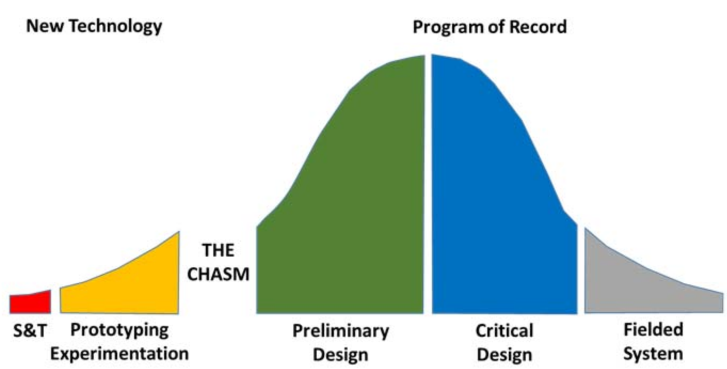


Figure 3. DoD Technology Adoption Chasm

Long development schedules limit the DoD's ability both to provide new capabilities that enable new operational advantages and to retire legacy systems with their increasing costs and risks. Over the last five years, the DoD has paid considerable attention to curbing this growth. DoDI 5000.02 stresses iterative development as a remedy, yet many major systems struggle to implement this approach effectively.

While major systems often take 10–15 years from concept to fielding (see Figure 4), programs only have a *12–18-month window to incorporate new technologies* into the design. During the Technology Maturation and Risk Reduction phase, programs contract with a few companies to develop competitive prototypes. During this phase companies leverage their research and development (R&D) programs and bring in partners to identify the leading technologies to exploit to maximize system performance. The window closes shortly before the Preliminary Design Review (PDR), at which point the key technologies are agreed upon in the design. No further opportunity for technology insertion typically occurs until after the system achieves Initial Operational Capability (IOC), when the program office may seek to upgrade fielded systems or inject improvements via a subsequent increment—which is often managed as another acquisition program.

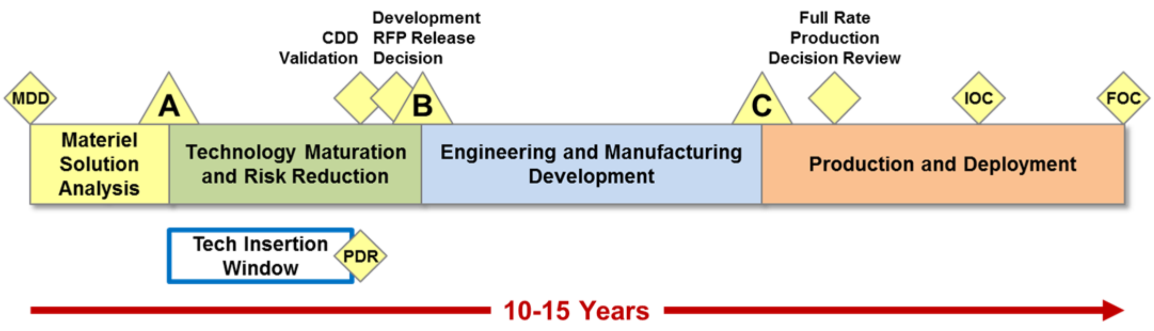


Figure 4. Technology Insertion Window

Key Enablers for Innovation and Technology Insertion

Examining successful government and commercial enterprises, as well as common themes in the DoD's new organizations, initiatives, and legislation enabled identification of some key enablers. Every corner of the Pentagon is investing in organizations focused on rapid capability deliveries. These are led by forward-thinking risk-takers, supported by an innovative culture and subject to only limited bureaucratic constraints. Programs are structured effectively from the start to deliver capabilities fast and avoid common acquisition pitfalls. They emphasize a renewed partnership with industry, particularly with non-traditional high-tech startups. Finally, a focus on delivery, where schedule takes increased priority, drives designs and decisions.

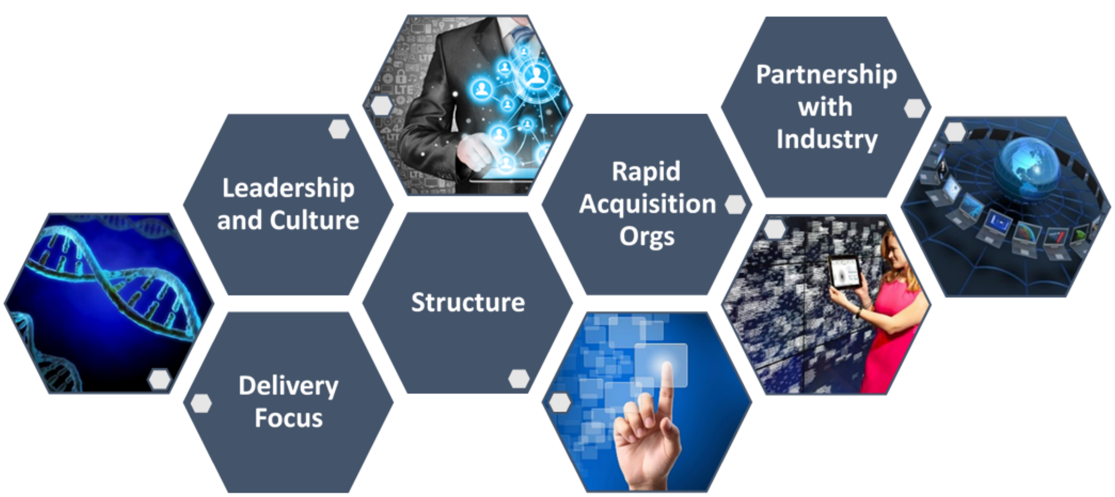


Figure 5. Key Enablers for Innovation and Technology Insertion

Technology Incubators

Incubators and accelerators have proven critical to the development of high-tech startup businesses. These programs mentor and train entrepreneurs in technical and business skills to help them launch a product and scale their business, secure funding, identify partners, hire the right employees, and mature their ideas. Leading incubators include IdeaLab, which assists companies to identify technology solutions to big problems early in the process. Accelerators such as Y Combinator enable speed to market via a defined schedule. Y Combinator alone funded over 1,000 startups with a combined valuation of over $65 billion.

Rapid Acquisition Organizations

Acquisition executives, policy-makers, and process owners can learn from both Silicon Valley and adaptive government organizations how to streamline their processes to enable faster deliveries. The DoD has many organizations and initiatives designed to enable speed, agility, and technological innovation. In 2012 Secretary of Defense Ashton Carter created the Strategic Capabilities Office (SCO):

> to help us to re-imagine existing DOD and intelligence community and commercial systems by giving them new roles and game-changing capabilities to confound potential enemies—the emphasis here was on *rapidity of fielding, not 10- and 15-year programs. Getting stuff in the field quickly*. (Carter, 2016)

The SCO has quickly matured into a major defense organization operating under only limited bureaucratic constraints, with an FY16 budget of $460 million for classified initiatives. Projects address strategic threats and have achieved early product successes in 3D-printed micro drones, self-driving boats, and an electromagnetic railgun.

In other examples, the Pentagon's Rapid Reaction Technology Office (RRTO) executes a series of prototyping and technology demonstration programs to hedge against technology risk, accelerate warfare capabilities, and conduct industry outreach to remove barriers to commercial technology use. The Air Force Rapid Capabilities Office (RCO) focuses on urgent classified projects that must deliver results in accelerated timelines. A flat organization governed by senior Air Force and DoD officials, the RCO operates with active warfighter engagement, small empowered teams, and stable funding. The RCO manages the new Long Range Strike Bomber as one of its premier programs. The Chief of Naval Operations seeks to replicate this model with a Maritime Accelerated Capabilities Office to field mature programs. The Army's Rapid Equipping Force focuses on delivering emerging technologies to deployed operational soldiers. The Joint Improvised Threat Defeat Agency (JIDA) provides COCOMs rapid acquisition and tactical responses to counter improvised threats.

The growth of rapid acquisition organizations gives acquisition executives new avenues to meet their top priority and rapid capability demands. However, these organizations may also have negative effects on traditional acquisition organizations. The DoD's top talent will flock to the rapid acquisition organizations so that they can work on high-priority programs with minimal restrictions and likely achieve greater success. This means that traditional program management offices will have less talent and standing to meet the demands of lengthy bureaucratic processes, compounding program risks, costs, and schedules.

Instead of instituting new programs and organizations to circumvent the acquisition bureaucracy, the DoD could place further emphasis on streamlining and innovation. This would include empowering the acquisition workforce with modern digital tools that effectively leverage their collective intelligence in a knowledge-based enterprise. The DoD should also hold leaders of functional areas accountable for streamlining their policies and processes, and provide the workforce with current guidance, templates, and exemplars on which to model their activities. Arming programs with the right structure and strategies from the start will improve the likelihood of program success and reduce delays in reviews and documentation.

Leadership and Culture

As the DoD seeks to promote rapid, agile, and innovative solutions, it must also adopt the leadership traits and organizational culture of the successful organizations. DARPA, Silicon Valley, and many other government and commercial organizations have recognized how to deliver capabilities quickly to respond to changing conditions in the market or battlefield.

Management thought leader Gary Hamel (2016) has identified key features of high-performing organizations, including

- Small, autonomous teams empowered to make key decisions
- Strong sense of competition and collaboration between operating units
- Significant investment in financial, commercial, and technical skills of employees
- Deeply shared norms and mutual responsibility for unit and enterprise success
- Radically simplified planning and budgeting processes

According to Hamel, a key driver of the bureaucracy is the continual addition of compliance requirements, either by law or policy. Hamel (2016) recommends that organizations assess themselves against three key bureaucracy indicators: the number of management layers, percentage of employee time spent on compliance, and average review timelines.

The DoD must actively engage program champions, stakeholders, and oversight organizations to accelerate decision-making and maintain program momentum. Delegating decision authorities to the lowest possible level and maintaining a short chain-of-command provides rapid, decentralized decision-making. To balance these delegated authorities, portfolio reviews give executives and stakeholders transparency into progress over recent months and plans for the next few months. This governance model would enable capability deliveries months or years earlier than traditional tiered, serial, gate-check program reviews.

Successful leaders set a bold vision, concrete goals, and incentives for successful capability deliveries against an aggressive schedule. They provide simple strategies, free of bureaucratic jargon, incorporating key stakeholder interests. For instance, Boeing's goals for the 727 aircraft design were that the plane must be able to hold 131 passengers, fly nonstop from Miami to New York City, and land on La Guardia runway 4-22 (a short runway). Stakeholders who can clearly articulate the strategy and their role can focus on achieving the desired outcomes. The tight integration of end users and developers enables regular collaboration on operational concepts and development details. This ensures a common understanding of operational requirements and potential technical solutions. While program managers and contracting officers provide official direction to contractors, facilitating user–developer collaboration has proven critical to satisfying users. Acquirers and developers who know specific users by name will have greater impetus to delivering a high-quality system than those who treat users as some abstract, faceless group.

A few DoD organizations embody the traits necessary to overcome the bureaucracy and delight users with groundbreaking innovations. Two organizations with a long track record of success can serve as models for others to replicate.

DARPA has a 50-year history of radical innovation spanning the Internet, satellites, stealth, and unmanned vehicles. DARPA's model includes 100 temporary technical program managers and a mix of high-performing individuals and teams from across government, industry, and academic research centers. Its projects are challenging, focused, and finite, making them attractive to high-caliber talent (Dugan & Gabriel, 2013). DARPA focuses on "use-inspired, basic research" that balances visionary, exploratory, basic research with practical applied research. Project leaders have the authority to reallocate resources, change strategies, and move talent on or off the project as needed, and focus on iterative progress rather than detailed upfront planning. DARPA's flat structure ensures that leaders rapidly become aware of issues and address them.

Special Operations Command (SOCOM) values speed over all other factors when it comes to acquisition (Guerts, 2016). SOCOM collaborates closely with many other organizations to "work at the speed of SOF" [Special Operations Forces] and ensure efficient and effective acquisitions in its dynamic, complex environment. SOCOM opened its own technology incubator, SofWerX, housing it in a 10,000 square-foot open floor building with the look and feel of a tech startup (Erwin, 2016). SOCOM involves innovative firms in frequent engagements, demonstrations, and hackathons.

The culture of the program office also plays a critical role in achieving speed and agility. Many program offices have staff who have applied the same methods for the last 30 years. These are not the innovators the DoD needs. Program managers need acquisition professionals with enough experience to understand the key elements of their function, yet are deeply committed to pursuing new business models. Moreover, a modern workplace environment and suite of collaboration tools would help programs to recruit and retain top talent.

Contracting Officers (COs) must function as strategic partners tightly integrated into the program office, rather than operate as a separate organization that simply processes the contract paperwork. COs cannot treat every contract as an 18–24 month procurement process, but instead must seek to understand the program objectives and design contract solutions. In 2016 the Office of Federal Procurement Policy (OFPP) launched the Digital IT Acquisition Professional (DITAP) Training and Development Program for COs to ensure that the DoD has a cadre of high-performing professionals to acquire leading IT capabilities in the Digital Age, and directed agencies to form Acquisition Innovation Labs to foster a culture of innovation (Rung, 2016).

The different acquisition phases require *different types of leaders*. The early phases call for visionary innovators who can explore the full opportunity space and engage in intuitive decision-making. The development and production phases demand a more pragmatic orchestrator to execute the designs and strategies via collaboration and consensus decisions.[1]

Program Structure

A BBP 3.0 initiative centers on aggressively reducing cycle time by targeting the root causes of schedule delays (Kendall, 2015). The Defense Acquisition Enterprise should promulgate—and act on—the Silicon Valley mantra "Always Be Shipping." Silicon Valley often focuses on getting a Minimum Viable Product (MVP) in the hand of users quickly, then iterate based on active feedback and system performance. The MVP mindset requires some culture and policy changes in Pentagon operations across requirements, budgets, and acquisition domains. A major weapon system that delivers all of its planned capabilities after 10–15 years will not satisfy its customers as much as a system that delivers 60% of its capabilities in 6–8 years, with the program office then involving the customer in iterative deliveries every few years thereafter. This principle holds true for acquisitions from small software programs to the F-22 fighter.

Good program managers know their schedule's critical path and focus sharply on reducing barriers. Sufficient, but not excessive, upfront analysis of requirements, technologies, costs, risks, and alternatives will enable program managers and stakeholders to effectively scope the program. These insights will help structure a program to deliver an MVP as soon as possible, while allowing iterative development over the long term. This balance of speed and rigor will ensure that warfighters obtain useful capabilities faster. Programs should embrace constraints by first adopting fixed schedules, mature technologies, and open architectures that drive design, and then relying on iteration to keep pace with technology advances. Follow-on increments then enable programs to integrate technologies that have since matured to address emerging requirements based on current operations and feedback on previous deliveries. Continual investment in research and analysis enables many government and industry partners to iteratively mature new technologies for a mission area.

The *HASC Acquisition Agility Act* seeks to drive Major Defense Acquisition Programs (MDAPs) to this structure and requires use of open design principles. It proposes that warfighters assign targets for costs and fielding dates and that Milestone Decision Authorities manage programs to these targets. The act differentiates between platforms and their components: Platforms are the major systems and involve slower development, whereas components are structured to be easily and quickly upgraded as technology develops to deliver improvements without waiting for a new system to be approved.

At present, the DoD continues to manage acquisition programs as large, stand-alone systems driven by independent budgets, requirements, and program offices, yet DoD executives and operational commanders seek integrated suites of capabilities. The DoD can enable speed and agility by restructuring, integrating, and managing related programs as acquisition portfolios. The major schedule drivers include securing a budget, defining requirements, program documentation, and awarding a contract. A portfolio structure manages these elements at a capstone level, enabling smaller programs to navigate the acquisition life cycle faster. A portfolio strategy, architecture, and roadmap can shape the

[1] These leadership traits were derived from Geoffrey Moore's concepts in *Escape Velocity*.

continual development and integration of a suite of smaller programs. Dynamic allocation of funds and talent to priority projects optimizes the portfolio performance and achieves a balanced force mix of large, medium, and small systems.

Agile is the leading software development methodology in industry, with growing adoption across the DoD and the federal government. It empowers small, high-performing teams to focus on demonstrating and delivering software rather than on coordinating dozens of documents that must be sent up the chain for approval each step of the way. A tailored version of Agile's guiding principles for the DoD includes the following:

- Small, frequent releases iteratively and incrementally developed
- Reviews of working software instead of extensive documentation
- Rapid response to changes in operations, technologies, and budgets
- Active user involvement to ensure high operational value

While Agile practices are best suited for IT programs, many of them apply to all programs, especially as software plays an increasing role in system performance. Programs have tailored their structure and processes to enable Agile adoption and experienced some early success. The resulting software is often of higher quality and more responsive to users' priority needs. Successful implementation of Agile requires a different culture and set of rigorous processes than the traditional acquisition environment.

Partnerships With Industry

In previewing the FY17 defense budget Secretary of Defense Carter (2016) said, "One of my core goals in this job has been to build and to rebuild bridges between [DoD] and the innovative, strong American technology and industry community." LinkedIn has provided him with ideas on how to overhaul the outmoded DoD personnel system and innovate for the force of the future. Secretary Carter has named Eric Schmidt, Chairman of Google parent company Alphabet, to head the Defense Innovation Advisory Board, which will "address future organizational and cultural challenges, including the use of technology alternatives, streamlined project management processes and approaches—all with the goal of identifying quick solutions to DoD problems" (Defense Media Activity, 2016). The Air Force seeks to harness IBM's Watson's computing power to tackle the "morass of the federal procurement process."

The DoD has established the Defense Innovation Unit Experimental (DIUx) as an initiative to foster increased communication, knowledge, and access to high-tech startup companies in Silicon Valley and Boston. DIUx seeks to build and strengthen relationships and play matchmaker between emerging technologies and operational challenges. By developing outposts where the entrepreneurs operate, the DoD can reach new companies outside the Capital Beltway.

As an example of an innovation partnership, Airbus partnered with micro-manufacturing innovator Local Motors to co-create commercial drones using the Airbus Quadcruiser's hybrid concept as the starting design. Local Motors recently released its Strati roadster, the world's first 3D-printed car. Airbus sought to "speed-up development and manufacturing in aerospace through an open competition based on co-creation and micro-manufacturing." The concept integrates the design of fixed-wing aircraft and quad-copters by combining the business models of a leading commercial firm with a new, distributed network of innovators.

While the DoD continues to engage startups to identify innovative technology solutions, integrating them into major weapon systems programs still requires active participation by traditional defense prime contractors. The large defense contractors argue they have robust networks to identify innovative small businesses with promising solutions and involve them as subcontractors; alternatively, they acquire the technology or company. DoD executives want to use their buying power and operating environment to expand the identification of new technologies and the ability to link those technologies to military applications.

Deliveries Are the Ultimate Measure

The DoD's annual reports to Congress should highlight the military capabilities—aircraft, ships, ground vehicles, space, and cyberspace assets—delivered to warfighters over the past year. At proper levels of classification, the reports should also include a summary of the operational impact of these new systems (with proper classifications) to give operational commands, Congress, and taxpayers a clear understanding of the value these systems provide to their end users relative to the $300 billion per year cost.

The DoD and Government Accountability Office (GAO) already publish the total and unit cost of each major weapon system. Tremendous visibility and incentives would result if the scheduled IOC and Full Operational Capability (FOC) dates for each system were also published and featured prominently on a DoD website. Seeing how many months and years elapse between the program's start—often with the Materiel Development Decision (MDD)—and IOC and FOC would shock many. Supporting tables for program sponsors and acquisition executives could compare the schedule length against their original estimates to identify and monitor schedule drivers and delays.

The MC-12W Liberty Aircraft represents a recent rapid acquisition success story. To address an urgent demand for information, reconnaissance, and surveillance (ISR), the Air Force's Big Safari program rapidly integrated existing sensors and communication datalinks on a commercial aircraft. It delivered Liberty to the theater in less than eight months from funding approval, at a low unit cost of $17 million. Liberty provided a balanced force mix to complement high-end systems such as Joint STARS and Global Hawk. The aircraft flew over 300,000 combat flight hours in Afghanistan and is credited with 73% of all Air Force ISR sorties and the kill/capture of hundreds of high-value individuals in Afghanistan during 2012.

Summary

The DoD's massive size constitutes both its competitive advantage and greatest risk. For the DoD to ensure that U.S. forces remain the premier military in the 21st century, programs must constantly innovate by rapidly incorporating leading technologies. The DoD must remain strategically agile, responding to new threats and opportunities across new domains. Secretary Carter, Under Secretary Frank Kendall, and Congress have pioneered many initiatives on innovation and rapid technology insertion. Each Service and Agency has embarked on related initiatives and set up relevant organizations. Some will require a few years to take root and grow into regular operations.

To succeed, these efforts require committed leadership, but innovation rarely occurs as the result of a top-down, central planning initiative. Instead, achieving the desired results requires a robust ecosystem of technologists, acquirers, and users with environments to model, demonstrate, and test prototypes and solutions. This network of experts across government, FFRDCs, academia, and industry should regularly collaborate online and in person.

DoD leaders need to empower junior officers and civilians to explore new ideas about both technology and business practices. They should assign a team to relentlessly examine every aspect of the acquisition enterprise and expose bureaucratic policies, processes, and barriers that hinder speed and agility. The DoD must also review the current acquisition workforce and identify the outstanding performers to recruit and retain. Then, DoD should partner young, motivated technology enthusiasts with experienced acquisition professionals to mentor each other and tackle challenges. Programs should regularly recognize and promote staff who take risks, embrace new partnerships, and deliver new capabilities to warfighters sooner.

The DoD should structure its programs to apply proven processes for managing schedule-consuming requirements, contracting, and budgets so that they can navigate the acquisition life cycle faster. Systems leveraging open architectures and incremental designs can focus on delivering initial capability quickly, and then iterate improvements over time. The DoD can tailor acquisition processes for each major type of system to streamline each program's path through focused guidance. Partnerships with industry—both traditional defense contractors and startups—will allow the DoD to benefit from their research, technological innovations, and business practices.

DoD executives have laid the groundwork by creating many new organizations and initiatives. As a result, acquisition professionals now have many avenues for pursuing innovative solutions and the leadership support to do so. It will be up to the leaders who join DoD under the next administration to build on their efforts to enable a nimbler acquisition process that meets the needs of a dominant military.

References

Carter, A. (2016). Remarks previewing the FY 2017 Defense budget [Speech]. Washington, DC: DoD. Retrieved from http://www.defense.gov/News/Speeches/Speech-View/Article/648466/remarks-previewing-the-fy-2017-defense-budget

Center for a New American Security (CNAS). (n.d.). Retrieved from http://www.cnas.org/sites/default/files/publications-pdf/CNAS_FutureDefenseIndustry_FitzGeraldSayler.pdf

Defense Media Activity. (2016). Pentagon to establish Defense Innovation Advisory Board. Washington, DC. Retrieved from http://www.defense.gov/News-Article-View/Article/684366/pentagon-to-establish-defense-innovation-advisory-board

Dugan, R. E., & Gabriel, K. J. (2013). Special Forces innovation: How DARPA attacks problems. *Harvard Business Review*. Retrieved from https://hbr.org/2013/10/special-forces-innovation-how-darpa-attacks-problems/ar/1

Erwin, S. I. (2016). Secretive SOCOM opens up to private sector. *National Defense Magazine*. Retrieved from http://www.nationaldefensemagazine.org/blog/Lists/Posts/Post.aspx?ID=2068

Guerts, J. F. (2016). Defense-industrial policy series: Acquisition for Special Operations forces. *Atlantic Council*. Retrieved from http://www.atlanticcouncil.org/events/past-events/defense-industrial-policy-series-acquisition-for-special-operations-forces

Hamel, G. P. (2016). *The $3 trillion prize for busting bureaucracy*. The Management Lab. Retrieved from http://www.garyhamel.com/sites/default/files/uploads/three-trillion-dollars.pdf

Kendall, F. (2015). *Better buying power 3.0.* Washington, DC. Retrieved from http://bbp.dau.mil/bbp4focus.html

Moore, G. A. (1991). *Crossing the chasm: Marketing and selling high-tech products to mainstream customers*. New York.

Rung, A. E. (2016). *Acquisition Innovation Labs*. Washington, DC. Retrieved from https://www.whitehouse.gov/sites/default/files/omb/procurement/memo/acquisition-innovation-labs-and-pilot-for-digital-acquisition-innovation-lab-memorandum.pdf

Thornberry, M. A. (2016). Acquisition Agility Act. Washington, DC: HASC. Retrieved from https://armedservices.house.gov/issues/defense-reform

The White House. (2016, February). *Economic report of the president*. Retrieved from https://www.whitehouse.gov/sites/default/files/docs/ERP_2016_Book_Complete%20JA.pdf

Acknowledgments

Panel 15. Connecting Contracting Strategy to Acquisition Outcomes

Thursday, May 5, 2016	
11:15 a.m. – 12:45 p.m.	**Chair: Major General Kirk F. Vollmecke, U.S. Army,** Program Executive Officer, PEO IEW&S ***Contract Design, Supply Chain Complexity, and Accountability in Federal Contracts*** Adam Eckerd, Assistant Professor, University of Tennessee Amanda Girth, Assistant Professor, The Ohio State University ***An Approach for Modeling Supplier Resilience*** Kash Barker, Associate Professor, University of Oklahoma Jose E. Ramirez-Marquez, Associate Professor, Stevens Institute of Technology Seyedmohsen Hosseini, PhD Candidate, School of Industrial and Systems Engineering, University of Oklahoma ***Antecedents and Consequences of Supplier Performance Evaluation Efficacy*** Timothy Hawkins, Lt Col, USAF (Ret.), Assistant Professor, Western Kentucky University Michael Gravier, Associate Professor, Bryant University

Major General Kirk F. Vollmecke, U.S. Army—became the Program Executive Officer for Intelligence, Electronic Warfare, and Sensors at Aberdeen Proving Ground, MD, in April 2016. MG Vollmecke was commissioned as a Second Lieutenant in May 1984 through ROTC as a distinguished military graduate of Centre College of Kentucky, where he earned a Bachelor of Arts degree in economics and management. He also graduated from the Naval Postgraduate School in 1992, where he earned a Master of Science degree in management with a concentration in acquisition and procurement management. He is a 1999 graduate of the U.S. Army Command and General Staff College and a graduate of the U.S. Army War College in 2004. MG Vollmecke is Acquisition Level III certified in Program and Contract Management. Prior to his current position, MG Vollmecke was the Deputy Commanding General for the Combined Security Transition Command–Afghanistan (CSTC–A) overseeing the security assistance program for the Afghan National Defense Security Forces in support of Operations Enduring Freedom and Freedom's Sentinel.

His acquisition assignments include the Deputy for Acquisition and Systems Management, Office of the Assistant Secretary of the Army (Acquisition, Logistics, and Technology), Washington, DC, in which he provided program management oversight of Army acquisition programs. Prior to that assignment, MG Vollmecke was the Commanding General of the Mission and Installation Contracting Command (MICC), Fort Sam Houston, TX, that provided Army commands, installations, and activities with contracting solutions and oversight across CONUS. Before that, he served as the Deputy to the Deputy Assistant Secretary of the Army for Procurement to ASA(ALT). He also served on the Joint Staff as the J-8 Chief, Capabilities and Acquisition Division; and before his tour on the Joint Staff in 2007, he was the Commander, Defense Contract Management Agency Iraq/Afghanistan supporting Operation Iraqi Freedom.

Other acquisition assignments include Headquarters, Department of the Army Systems Coordinator for the Future Combat Systems (Brigade Combat Team) program; Executive Officer to the Assistant Secretary of the Army(AL&T); Commander, DCMA Boeing Philadelphia; Program Analyst for the Deputy Chief of Staff of the Army for Programs, Program Analysis and Evaluation (PA&E) Directorate; Assistant Product Manager M2/M3 for the Bradley Fighting Vehicle Systems project office; Contingency Contracting Officer assigned to the U.S. Army Forces Central Command–Saudi Arabia under Operation Desert Falcon; and as a Weapon System Contracting Officer assigned to the Army Materiel Command's Communications–Electronics Command (CECOM), which included a deployment to Honduras, Joint Task Force Bravo. Prior to joining the Army's Acquisition Corps in 1991, he served in a variety of mechanized and light infantry battalion staff and company command positions.

Contract Design, Supply Chain Complexity, and Accountability in Federal Contracts

Adam Eckerd—is Assistant Professor in the Department of Political Science at the University of Tennessee. Eckerd conducts research on organizational decision making and the complex relationship between these and other policy decisions and social outcomes. His primary interests lie in understanding how organizations, policy makers, and citizens use information to make decisions and influence public and organizational policy. He has recently co-authored a book*, Rethinking Environmental Justice in Sustainable Cities*, and his work has appeared in journals such as *Public Administration Review*, *Policy Sciences*, *Administration & Society*, *Nonprofit and Voluntary Sector Quarterly*, and *Social Science Quarterly*. [aeckerd@vt.edu]

Amanda M. Girth—is Assistant Professor at the John Glenn College of Public Affairs at The Ohio State University. Her research focuses on government contracting with a specific interest in performance and accountability, which has been published in the *Journal of Public Administration Research and Theory*, *Public Administration Review*, *International Public Management Journal*, and other outlets. Previously, Girth was a manager for a global consulting firm where she supported information technology initiatives at the U.S. Department of State and U.S. Agency for International Development. She also served in Michigan state government. [girth.1@osu.edu]

Abstract

In this manuscript, we argue that supply chain management choices are affected by both the extent to which there is a risk of disruption within the supply chain and external to the supply chain as well. We suggest that the formal governance mechanisms that are favored under different conditions of endogenous and exogenous supply chain risk reflect the risk management preferences of the supply chain partners. In this preliminary study of public sector supply chains, we found evidence suggesting that, as expected, when endogenous risk is low, suppliers tend to bear most of the disruption risk by agreeing to fixed price contracts. Conversely, when endogenous risk is high but exogenous risk is low, buyers (governments) are willing to bear most of the risk by agreeing to cost reimbursement or time and materials contracts. When both endogenous and exogenous risk is high, we found partial support of the proposition that buyers and suppliers are more likely to share risk by agreeing to incentive contracts.

Introduction

Supply chains are complex in at least two fundamental aspects—the complexity or complicatedness of the product, and the uncertainty of information exchange across different organizations (Vachon & Klassen, 2002). Gailbraith (1973, 1977) is credited as describing this complexity and elaborating on ways that organizations can manage uncertainty through better information processing. Flynn and Flynn (1999) apply these concepts to a manufacturing supply chain, noting six drivers of manufacturing environment complexity related to the diversity of management tasks in manufacturing, goals, processes, customers, suppliers, and labor, while Vachon and Klassen (2002) also note complexity in managing supply chains that cross international boundaries. In short, as products become more complicated and as more actors with diverse goals become involved in the production of some good or service, coordination becomes more difficult.

For the most part, this research has focused on complications that reside within the supply chain. These mostly relate to aspects of market uncertainty, either in terms of the resources required in the manufacturing of a product or the stability of demand for the final good. However, there is also uncertainty that is exogenous to the supply chain and while "endogenous uncertainty can be decreased by actions of the firm" (Folta, 1998, p. 1010), "exogenous uncertainty is largely unaffected by firm actions" (Folta, 1998, p. 1011).

Exogenous uncertainty can come from a variety of different sources, such as geopolitical factors or natural disasters, but it ultimately relates to some sort of information that is missing which makes committing resources to production risky, particularly if those investments are highly specific to the product in question.

Firms deal with uncertainty through risk management. In general, when endogenous uncertainty is located in resource acquisition, downstream partners are willing to share risk with upstream partners, and conversely, when risk is in the demand market, upstream partners are willing to share risk with those closest to the market (Flynn & Flynn, 1999). Exogenous uncertainty is more difficult to prepare for and manage, and it may affect any portion of the supply chain (Trkman & McCormack, 2009). Since exogenous risks may not be as manageable, exchange partners will seek a common understanding of the terms of the exchange in order to reduce endogenous costs, namely the transaction costs of the exchange, so that the total costs of the exchange are low enough to deal with the exogenous uncertainty should any unforeseen shocks occur (Weber & Mayer, 2014). The complication is that these risks must be managed ex ante.

Our interest is in a domain where these risks, particularly exogenous risks, may be prevalent: public management. Public sector supply chains are subject to a variety of different exogenous factors associated with the political system, legal institutions, bureaucratic processes, and geopolitics, in addition to endogenous factors such as highly specific products and measurement complexity (Brown, Potoski & Van Slyke, 2006; Dixit, 2002). While managers take care of daily supply chain tasks, a public sector supply chain is ultimately controlled by political authorities who decide to either allocate resources for the purchase of the goods in the supply chain or not. There is no functional market for the goods procured, and political actors' responses to events are often unpredictable. Some public sector supply chains, for things like accounting services or janitorial work, may be relatively resilient to external events, while others, like aerospace or military procurement, may reside in highly turbulent environments (Peck, 2005).

It is well accepted that managerial strategies are different when a supply chain is susceptible to endogenous risk (Manuj, Esper, & Stank, 2014). In this research, we argue that supply chains are also managed differently depending on the susceptibility of the supply chain to being affected by exogenous disruptions. In short, when the risk of disruption due to exogenous factors is high, we argue that risk will be shared between the exchange partners as a means of buffering the supply chain from external events. When exogenous risk is low, the risk management strategy will be determined by the nature of endogenous risk.

Supply Chain Risk

Endogenous

Endogenous risks are generally of two kinds: supply side, or resource-related risks, and demand side, or market-related risks. There are various points at which disruptions can occur in supply chains. Fluctuations in market demand can be expected, but there can be more volatility for certain final goods, and this volatility will affect the entire supply chain. There can be uncertainty in access to resources required for the production of the final good (both undersupply and oversupply), including fluctuations in natural resources, labor, technology, and access to capital for exchange partners. Coordination in the supply chain can be challenging due to the well-known bullwhip effect (Lee, Padmanabhan, & Whang, 1997) which occurs when each exchange partner makes ordering decisions in isolation, causing inefficient repercussions within the supply chain. Further, each of these effects becomes exacerbated as the complexity of the product increases and as more exchange partners are in the supply chain (Flynn & Flynn, 1999).

Exogenous

Exogenous risks exist beyond the focal supply chain, and thus the actors in the supply chain have little ability to control these events. Trkman and McCormack (2009) identify two types of exogenous uncertainty, discrete events and continuous risks, and suggest that different approaches may need to be taken to manage both types of uncertainty. Continuous risks are those that occur at a relatively stable rate and are generally more predictable than discrete risks, and while the exact timing of an occurrence is not known, the risk is more or less stable over time. Continuous risks may be events such as economic downturns, cross-cultural complications, or changing political environments due to election cycles or policymaker interest. Discrete risks are less predictable shocks to the external environment. These risks may include natural disasters, geopolitical events such as terrorism or international conflict, political or broad-based labor strife, or technological changes.

Managing Risk

There are many different specific strategies available for supply chain risk management (SCRM; Tang, 2006). Private sector firms have been shown to follow a few general approaches depending upon the nature of the supply chain risk. For endogenous supply risk, firms can hedge, that is attempt to secure resources from a variety of different suppliers, or they can assume and internalize the risk, by incurring production internally, for example. For endogenous demand risks, firms can postpone or delay production, or they can speculate and maintain inventory until demand stabilizes (Manuj, Esper, & Stank, 2014). Tang (2006) also notes that firms can adapt processes or products to deal with new circumstances or risks, or they can open information exchange to ensure that each partner in the supply chain has access to the same information. In short, the approach to managing endogenous risk is to mitigate the chances of disruption, to the extent possible, and implement strategies to minimize disruptions when they occur in order to ensure the efficient operation of the supply chain.

For exogenous risk, options for information sharing, hedging, postponing, or holding inventory may be much less effective. Hedging will be ineffective because an exogenous shock will be external to any particular supply chain, and it is likely that all or at least most suppliers will be affected by whatever the exogenous shock was. Regardless of the diversity of suppliers, the focal firm’s supply chain will be impacted. Demand focused strategies will be risky because a shock may cause an unpredicted and fundamental shift in the nature of the demand for a product. In this scenario, postponing may lead to shortages because demand may recover, leaving the supply chain unable to meet demand. Conversely, holding inventory will be costly if demand never recovers. Finally, sharing information is unlikely to have an effect if the source of the disruption is external to the supply chain because the shock will be out of the control of any of the exchange partners and no firm is likely to have access to any more information than another.

The main strategy available for managing exogenous uncertainty is therefore to assume and internalize the risk within the supply chain. This can be done by ensuring that the parties to the exchange have a common understanding of the nature of both the exchange and the exogenous risks and coordinate activities to the extent that costs internal to the supply chain, namely the costs of the transaction, are as low as possible, leaving adequate resources to manage any exogenous shocks that occur and enabling the partners to internalize the potential costs of exogenous changes (Weber & Mayer, 2014). The key aspect in managing exogenous risk is thus not to take steps to avoid risky scenarios because exogenous risks cannot be avoided, but rather to take steps to ensure that the supply chain is resilient to the disruptions that do occur. There are a number of approaches that can be taken to improve resilience, such as building flexibility into the contract in case one organization in the chain bears the brunt of the effects of the shock, discussing potential disruptions throughout the supply chain to ensure that there is a common understanding of the things that could potentially occur, and deriving contingency plans and processes in advance should disruptions occur (Brown, Potoski, & Van Slyke, 2015).

Public Sector Supply Chains

Public sector supply chains offer a good context to understand SCRM due to their propensity to be affected by exogenous factors. A key reason why this is the case is that there is only one "consumer" for the end products, and it is a fickle consumer—the polity. In the U.S., the polity is represented in these exchanges by Congress and the President. With one buyer (the government broadly), there is no real market for the end goods, and owing to the non-excludable and non-rivalrous nature of public goods, pricing and, thus, demand are difficult or impossible to ascertain for many of the things that the government purchases. On the supply side, many products are idiosyncratic and involve highly technical or scarce materials, and policies often limit where those materials can come from and what sorts of vendors should be selected.

Internally, this means that there is often considerable endogenous risk owing to the high level of asset specificity that might be required of vendors to produce goods or services for the government, and the technical complicatedness of these goods and services may require numerous subcontractors and sources of supplies for the final product, with each relationship adding complexity to the management of the supply chain (Eriksson, 2015). Externally, public sector supply chains are subject to significant supply and demand uncertainty. Supplies can be disrupted by natural or manmade disasters that reduce the availability of needed resources or by having few potential vendors with the expertise to produce the good or service. Demand can be affected in numerous unexpected ways by disasters or other focusing events, but also due to changes in the political context via elections, interest group activity, or erratic shifts of interest from particular political authorities from either Congress or the Executive branch agencies.

It is typically left to public managers to deal with these risks and be prepared to respond to them. Public managers can take cues from private sector SCRM recommendations in cases where efficient supply chain operation is an overriding goal, but owing to the complex environment of public sector decision-making and the often esoteric nature of the public goods and services that are produced, although always a goal, efficiency is not always going to be the overriding goal. In the private sector, an efficiently operating supply chain that is producing a product that meets a market demand is likely in a good position to be resilient (Tang, 2006), but in the public sector this may or may not be the case. New political authorities may gain control and opt to pursue other priorities, or a shock in some other policy area may convince policymakers to shift resources away. Moreover, wholly inefficient supply chains may thrive in the public sector if the good or service being produced is valued by powerful actors or if the production process benefits powerful actors (Kim & Brown, 2012; Eckerd & Snider, 2016).

This is not to say that the efficiency of the supply chain and the quality of the end product are unimportant, but the overriding goal may be ensuring that the supply chain is buffered from external volatility in order to protect the organizations that are involved in the exchange. For some public products and services, efficiency and product/service quality are likely good buffering techniques in that they are producing a good or service that has wide political support and is in less danger of exogenous risk, but in other contexts, garnering the support of key political actors may be more important, while in others the goal might be to stay "off the radar" as much as possible. Along these lines, Frumkin and Galaskiewicz (2004) find that public managers tend to be more externally focused in their management decisions than private managers do. Governmental organizations "[lack] a single stakeholder group to monitor the organization," such as a board of directors or investors (Frumkin & Galaskiewicz, 2004, p. 289). As such, public managers, as compared with business organizations "are more likely to embrace external referents of accountability to legitimate their operations" and "should be more susceptible to institutional pressures and more likely to be swayed by exposure to environmental pressures that promise an organization greater legitimacy" (Frumkin & Galaskiewicz, 2004, p. 289). Along similar lines, managing exogenous risks may occupy more managerial capacity than dealing with internal issues.

SCRM in the public sector is thus likely more focused on the external environment and particularly focused on buffering the supply chain from those exogenous risks against which there may be some means to do so, at least under some circumstances. As we lay out in detail below, we expect public sector SCRM to have a more procedural than an outcome-oriented focus when exogenous exposure is high. However, we expect that when the key risks are low or endogenous, public sector supply chains will behave similarly to private sector supply chains. This is because when endogenous risk is low, the product or service is likely not particularly complex, so the assets needed for production will not be specific. In these situations, there are likely multiple potential suppliers, and public organizations (like their private sector counterparts) can therefore hedge and be prepared to solicit offers from a variety of different vendors. When products are more complex but exogenous risk relatively low, internalization of risk is more likely. Governments can opt to produce these goods or services internally or when they are procured through a supply chain to internalize the risk to ensure the stability of the supply chain as much as possible (Brown & Potoski, 2003). That is, with low exogenous risk, public managers may be more willing to bear the burden of risk, assured that the external environment is stable for them, and allow their supply chain partners to be buffered from risk as much as possible, or the production process can be managed strictly internally.

However, these strategies are likely to be altered by the extent to which the supply chain is susceptible to exogenous risk. In short, we argue that the more that the supply chain is subject to external risk, the more that the supply process requires buffering and sharing of risk across the supply chain, potentially in ways that are contrary to what we might expect given the endogenous risks of disruption. Before developing our argument of exogenous risk, we first explain risk management techniques as they relate to the public sector.

One of the key ways to manage risk is though the governance of the supply chain relationship (Folta, 1998) and specifically through the structuring of the contract between supply chain partners. There are two general types of contracts, the formal or written contract and the informal or relational contract (Poppo & Zenger, 2002). The relational contract is the ongoing establishment and reinforcement of norms between contracting partners that can be used to facilitate the resilience of the supply chain (Ring & van de Ven, 1994), but our interest here is with respect to the design of the formal contract which better represents an ex ante risk management strategy. In public sector contracting, there are three general approaches to a formal contract, each representing different approaches to dealing with risk. Although there are many variations of these three different types of contracts, a contract falls under an umbrella of being a fixed cost, incentive, or cost reimbursement structure. A *fixed price* contract is just that—the government will propose a price that it is willing to pay for some good or service and solicit bids from vendors to provide the good or service. Fixed price contracts are often short term in nature and place relatively few constraints on vendors. Under *cost reimbursement* contracts, the government will propose a good, service, or task and solicit bids, reimbursing the vendor for the costs incurred in delivering the product. *Incentive* contracts are in between; there will be some minimum commitment of funds provided by the government and services rendered by the vendor, and if the vendor exceeds that minimum, then the vendor receives additional money (or conversely, if the vendor fails to meet expectations, there may be sanctions). Time and materials contracts are another commonly used type of contract, but are considered a form of cost reimbursement contracts (Kim & Brown, 2012). Time and materials are preferable when the buyer needs flexibility, but the supplier has little incentive to optimize efficiency as they are compensated for inputs (Roels, Karmarkar, & Carr, 2010).

The contract that is chosen offers a view of the SCRM strategy. A fixed price contract can be viewed as a hedging strategy and is most likely useful with low complexity products. Simple products or services are likely to have more actual or potential suppliers, enabling the government to dictate the market and put the onus of risk on the supplier. This provides the supplier with flexibility on how the task gets done, but if there are any disruptions, the risk falls on the vendor who agreed to provide a certain amount of some product for the specified price. If disruptions occur and the contract is deemed unsatisfactory, the government can simply find a different supplier. We expect that when a supply chain is not especially prone to endogenous risk, fixed price contracts will be favored, particularly when the supply chain is also not especially prone to exogenous risk. This is because the products in question are likely not complex, the assets are likely not specific, and with a low chance of risk and a buyer's market, the supplier will bear risk.

In situations where endogenous risk is higher, for example when a product or service is complex and buyers are limited, options for hedging are likely to be limited due to a smaller set of potential suppliers, so risk will need to be internalized by the buyer. This can be done by using a cost reimbursement contract, which places the risk on the buyer which is the government in this case. By internalizing risk this way, governments can enable suppliers' willingness to shoulder the costs of highly specific assets and ensure a relatively stable supply of a needed product. There is clear potential for mutual benefit; suppliers face less risk knowing that investment costs will be recouped, while the government saves resources by not having to manage the highly specific assets required. Thus, when endogenous risk is high, we expect cost reimbursement contracts to be favored.

However, these considerations may be changed when a supply chain is susceptible to high levels of exogenous risk, such as projects that are very salient with political authorities. We can conceive of exogenous risks in a variety of different ways, but the public management context offers a relatively clear way to assess one aspect of exogenous risk: the interest that the political system has on the product or service in question (Epstein & Segal, 2000). If a particular project has little salience with the public, then it likely has little salience with Congress, and is therefore less exposed to potential political disruption than a program that is well known. While this may not be the only type of exogenous risk that a public sector supply chain would be exposed to, it is one of the more continuous types of risk for which a strategy may be devised to manage it (in contrast to dynamic risks like natural disasters that can be planned for but are less predictable).

The choice of which contract type is selected can provide an indication of the risk management strategy and it can also send a signal to the external political environment. We expect that one of the main intentions of public managers is using the contract selection as a means through which to buffer the supply chain from unexpected political interest. For salient products that the polity and policymakers focus on, or those situations where exogenous risks are higher, cost reimbursement contracts might be favored if endogenous risks are high, but these contract types may look like “backroom deals” and might garner unwanted attention from the external environment. In other words, the seemingly appropriate contract for the internal dynamics might not be preferred because of the external environment. Therefore, in cases where both endogenous and exogenous risks are high, we suggest that risks can be internalized through an incentive contract, in which the supplier takes on risk by agreeing to some minimum threshold at a fixed cost, while the government takes on risk by agreeing to pay additional costs for any production beyond this threshold. Firms “will charge a premium” for taking on risk, and risk sharing between partners becomes more efficient as uncertainty increases (Jensen & Stonecash, 2005, p. 777), but they are willing to do so because an incentive contract should better buffer the supply chain from exogenous shocks from the political system (Lawther & Martin, 2005). For products that are subject to high exogenous risk but low endogenous risk, we can see two potential strategies. First, if a product is not complex, it is likely that the costs can be estimated well, suggesting that a fixed price contract that keeps government costs low may be the best approach to buffer the supply chain. On the other hand, if the product in question is mission critical, then the government bears some risk in a fixed price contract if the contractor fails to deliver. Although there will be legal recourse to recoup costs, this might matter less than the timely delivery of some important product, and thus incentive or cost reimbursement contracts may be favored. Our expectations are laid out in Table 1.

- *Hypothesis 1: Fixed price contracts will be favored over other contract types when both endogenous risk and exogenous risk are low.*
- *Hypothesis 2: Cost reimbursement contracts will be favored over other contract types when endogenous risk is high and exogenous risk is low.*
- *Hypothesis 3: Incentive contracts will be favored over other contract types when both endogenous risk and exogenous risk are high.*

Table 1. Supply Chain Risk and Contract Design Choice

	Low exogenous risk	**High exogenous risk**
Low endogenous risk	Fixed price	Idiosyncratic to product
High endogenous risk	Cost reimbursement	Incentive

Data and Method

Data

We test our hypotheses of endogenous and exogenous supply chain risk using public sector contracts data from federal agencies in the U.S. Contracts data is derived from the Federal Procurement Data System-Next Generation (FPDS-NG), the only comprehensive source of unclassified federal contracts. The Federal Acquisition Regulation (FAR) requires that contract officers record in FPDS-NG contract actions exceeding $3,000 in value. As a result, this data set offers an exclusive opportunity to study supply chain risk.

The unit of analysis in this study is the federal contract. Because the FPDS-NG captures all contract actions, including when the contract is initiated and subsequent modifications, we execute a process of aggregating the data for each individual contract action associated with a specific contract. (For example, Contract A is initiated and is subsequently modified three times. One modification might increase the initial value of the contract from $30,000 to $50,000. One modification might be a time extension of the contract from 12 months to 13. The final modification might deobligate funds associated with the contract by $5,000. As a result, the value of Contract A is ultimately $45,000, and the time duration is 13 months.)

In this preliminary analysis, our data is comprised of 274,440 contracts from 22 product areas. Federal agencies purchase goods and services using two industry categorizations: the North American Industry Classification System (NAICS) and the Product Services Code (PSC). Table 2 lists the contracts included in this analysis and their corresponding NAICS and PSC classification. The sample encompasses all completed unclassified contracts for these PSC/NAICS from FY2000–2014.

Measures

Dependent variable. Our outcome of interest is a key aspect of how supply chain managers deal with risk: the type of formal contract that governs the relationship. The dependent variable *contract pricing type* is a nominal variable and is coded 1 if the contract is a fixed price contract (specified in the FPDS-NG as fixed price redetermination, fixed price level of effort, firm fixed price, fixed price with economic price adjustment); 2 if the contract is an incentive contract (specified in the FPDS-NG as fixed price incentive fee, fixed price award fee, cost plus incentive fee, cost plus award fee); 3 if the contract is a cost reimbursement contract (specified in the FPDS-NG as cost no fee, cost sharing, cost plus fixed fee); and 4 if the contract is a time and materials contract (specified in the FPDS-NG as time and materials, labor hours). Table 2 reports the distribution of the dependent variable for both products.

Table 2. Dependent Variable: Contract Pricing Type

Contract pricing type			
1 Fixed price	**2 Incentive**	**3 Cost reimbursement**	**4 Time and material**
218,120 79.48%	1,642 0.60%	27,968 10.19%	26,710 9.73%

Put simply, fixed price contracts shift risk primarily to the supplier because the supplier receives a fixed amount regardless of any extenuating circumstances. Cost reimbursement and time and materials contracts are unique contract types, but are similar in that the risk is borne primarily by the government because the government reimburses the supplier for relevant costs regardless of the actual amount of a service received. Both of these contract types are used when requirements are unable to be properly specified. Incentive contracts, which allow suppliers to earn an additional fee based on meeting specified performance objectives, are characterized as a risk-sharing position.

Explanatory variables. We are interested in testing the effects of both endogenous and exogenous risks to the supply chain. We examine endogenous risk—related to complexity in the supply chain—by analyzing *product complexity*. We operationalize endogenous risk using product complexity measures developed by Kim, Roberts, and Brown (forthcoming). The authors surveyed federal acquisition professionals to determine their assessment of ease of measurement and specialized investment ratings. They then combine these factors into a product complexity rating as reported in Table 3. We use their findings to study contracts from 22 product areas, allowing us to assess both endogenous and exogenous risk in our analysis.

Table 3. Product Categories and Complexity Scores

Product Category	PSC	NAICS	Number of Contracts	Product complexity
Solid waste collection	S205	562111	17,907	3.08
Landscaping	S208	561730	37,774	3.16
Laundry and dry cleaning	S209	812320	8,051	3.26
Janitorial service	S201	561720	47,058	3.30
Court reporting	R606	561492	19,940	3.51
Warehousing and storage	S215	493110	2,104	3.61
Security guard and patrol	S206	561612	25,522	3.77
Advertising	R701	541810	5,445	4.43
Auditing	R704	541211	2,682	4.77
Legal service	R418	541110	9,732	4.97
Professional and management training	U008	611430	12,816	5.02
Equipment maintenance and repair	J099	811310	5,971	5.22
Program management and support	R408	541611	12,401	5.62
Logistics support	R706	541614	3,175	5.63
Program review and development	R409	541611	615	5.87
Engineering	R425	541330	55,822	6.76
Computer system development	D302	541512	4,196	7.58
Weapons-basic research	AC51	541710	578	7.60
Defense aircraft - basic research	AC11	541710	880	7.94
Defense aircraft - engineering development	AC14	541330	122	8.46
Weapons - applied R&D	AC52	541710	659	8.60
Defense aircraft - applied R&D	AC12	541710	990	8.66

Exogenous risks—those factors that affect the supply chain but are external to the organizations involved in the exchange, and that the actors have little ability to control—are operationalized using three variables all intended to capture aspects of the potential interest and saliency of the supply chain with political representatives: *contract value*, *market competition*, and *competitive limitations*. Taken together, we expect that heightened exogenous risks will decrease the use of fixed price contracts and drive public purchasers to other contract pricing types in order to share or redistribute risk.

- *Contract value* is the logged value of total dollars obligated to the contract. This is measured in real dollars with 2014 as the base year. Contract value is a clear proxy for political interest, particularly during times of tight budgets (Eckerd & Snider, 2016). While the size of the procurement budget is not necessarily indicative of politically contentious programs, with all else equal, it seems reasonable to assume that larger acquisition projects will draw more political attention. That is, we expect greater political interest and attention to the contract as contract value increases, which may result in selecting different contract pricing types. Specifically, we see two different responses to high value contracts depending on the nature of the endogenous supply chain risk. When endogenous risk is low, we expect that high value contracts will tend to be fixed price. When endogenous risk is higher, we expect more risk sharing in the form of incentive contracts.
- *Market competition* is the logged value of the total number of offers received for the contract. This variable is a proxy for the level of market competition for

the particular product. While this is not a direct measure of political salience, when markets are highly competitive, exogenous attention may be more likely to identify perceived waste or exorbitant costs. An older example of this logic is the infamous Packard Commission's 1986 report identifying the Department of Defense's $435 hammer and $600 toilet seat (Blue Ribbon Commission on Defense Management, 1986). More recently, consider the scrutiny received over the ultimate cost of a fuel station in Iraq (Davenport, 2015). Thus in circumstances where there is much market competition but endogenous risk is low, a situation of high exogenous risk, we expect to see greater use of fixed price contracts that place more risk on the supplier. This is, in part, due to the lower costs of switching suppliers if contract performance is unsatisfactory. However, as with contract value, we expect a different response if there is high endogenous risk. We expect that when market competition is low, but endogenous risk high, cost reimbursement contracts may be preferred to ensure program continuity (thus buffering exogenous interest), but when both competition and endogenous risk are high, we expect incentive contracts to be favored as a mechanism to share risk across the supply chain.

- *Competitive limitations* is coded 1 if the contract is a set aside (e.g., small business, economically disadvantaged business owned by women or veterans, disabled veteran owned, and HUB zone) or 0 if the contract was not designated as a set aside contract. Set aside contracts are contract design tools aimed at leveling the competitive environment for otherwise disadvantaged firms. These political tools can introduce exogenous risk into the supply chain by restricting the supplier market and also potentially drawing interest from political actors who designed the set aside not to meet supply chain efficiency goals, but rather to meet broader social/public policy goals of inclusion and equity (Eckerd & Eckerd, 2016). When endogenous risk is low, we expect that contracts with competitive limitations will be fixed price. When endogenous risk is higher, we expect that incentive contracts will be favored when there are competitive limitations in order to share risk.

Control variables. We control for several factors that can influence the relationship between supply chain risk (whether endogenous or exogenous) and contract pricing type. We control for *contract length*, which is measured as the total length of the contract in years. We include a control for *unrestricted competition* to measure whether the contract was bid with unrestricted competition (e.g., full and open) or restricted competition (e.g., sole source), coded 1 if unrestricted and 0 if restricted. This can affect the number of bidders on a contract. The *Department of Defense (DoD)* accounts for approximately two-thirds of federal contracts. Because the DoD has more experience with contracting in general, we expect that they are more likely to use more diverse contract pricing types. The DoD is coded 1 if the contract agency is the DoD, and 0 otherwise. We also include *year* dummy variables to correspond to the fiscal year the earliest contract action associated with each contract. This is typically the year the first agreement was signed, which is also when the contract pricing type is established. This allows us to control for unobserved policy and/or political changes that might affect the exogenous risk to the supply chain.

Table 4 provides descriptive statistics for each of the independent variables.

Table 4. Independent Variables: Descriptive Statistics

Independent Variable	Mean	Std dev	Min	Max
Product complexity	4.509	1.502	3.08	8.66
Contract value	8.292	4.788	0	21.376
Market competition	1.171	0.791	0	6.908
Competitive limitation	0.327	0.496	0	1
Unrestricted competition	0.697	0.459	0	1
Contract length	1.816	1.627	1	16
DOD	0.503	0.499	0	1

Method

Given the non-ordinal categorical nature of our dependent variable, contract type, we use a multinomial logistical regression model to test our hypotheses. We estimate preferences for incentive contracts compared to fixed price contracts, cost reimbursement contracts compared to fixed price contracts, and time and materials contracts compared to fixed price contracts. Relative-risk ratios are reported for ease of interpretation. Relative-risk ratios and standard errors are generated using a clustering technique to obtain robust standard errors. Standard errors are clustered at the broader independent delivery vehicle (IDV) level (or the unique contract level if the contract is not part of an IDV). Numerous delivery or task orders, each a unique contract, often fall under one broader IDV. Examples of IDVs include blanket purchase agreements, federal supply schedules, and task and delivery order contracts that are government-wide or multi-agency agency contracts. Although these contracts are unique purchases, they are acquired using the same guidelines and pricing associated with the IDV. Because these contracts are similarly structured, we need to account for the similarities amongst these unique but associated contracts. As a result, standard errors are clustered at the IDV level. Approximately 10% of contracts were missing set aside data. As a result, the model is restricted to 246,362 contracts.

Findings

The results of multinomial regression are reported in Table 5. We expect a high level of exogenous risk to be indicated by a combination of a high contract dollar value, low levels of market competition when endogenous risk is high and high levels of market competition when endogenous risk is low, and the presence of competitive limitations. High levels of product complexity are evidence of endogenous risk. When these risk factors are low, we expect fixed price contracts to be favored. We therefore compare each of the other contract types to fixed price contracts which are the most prevalent type of contract selected, by far, with nearly 80% of the contracts in our data being fixed price.

Incentive contracts compared to fixed price contracts. As expected, when endogenous risk is high, incentive contracts are favored over fixed price contracts. Only one exogenous factor, contract value, is statistically significant, suggesting that as contract value increases, so too does the preference for incentive contracts over fixed price contracts. Both findings are consistent with hypothesis 3. We find *competitive limitations* (set asides) are less likely with incentive contracts compared to fixed price contracts. We find a statistically positive relationship between *contract length* and incentive contracts compared to fixed price contracts. We also find that compared to civilian agencies, the DoD is more likely to use incentive contracts than fixed price contracts when endogenous risk is low.

Cost reimbursement contracts compared to fixed price contracts. Again, as expected, when endogenous risk measured by product complexity is high, contracts are more likely to be cost reimbursement than fixed price contracts. All three exogenous factors, *competitive limitations*, *market competition*, and *contract value*, reach levels of statistical significance. Cost reimbursement contracts are favored in situations with fewer *competitive limitations*, less *market competition*, and higher *contract value* than fixed price contracts, which provide mixed evidence for hypothesis 2. The value of the contract and market competition is consistent with our expectations, but competitive limitations is not. We find that cost reimbursement contracts are likely to be longer than fixed price contracts, less likely to be competed with restrictions, and more likely to be DoD contracts compared to fixed price contracts.

Tim*e and materials contracts compared to fixed price contracts*. When endogenous risk, *product complexity*, is high, time and materials contracts are preferred over fixed price contracts. We also see that all three exogenous factors, *competitive limitations*, *market competition*, and *contract value* reach levels of statistical significance. Time and materials contracts are favored in situations with fewer *competitive limitations*, greater *market competition*, and greater *contract value* than fixed price contracts, which again provides mixed evidence for hypothesis 1. We find a statistically positive relationship between contract length and time and materials contracts compared to fixed price contracts. Time and materials contracts are more likely to be competed with restrictions compared to fixed price contracts. We also find that compared to civilian agencies, the DoD is less likely to use time and materials contracts than fixed price contracts.

Table 5. Multinomial Logit Analysis

	Fixed price contract versus					
Independent Variable	Incentive contract		Cost reimbursement contract		Time and materials contract	
	RRR	*Rob std err*[a]	*RRR*	*Rob std err*[a]	*RRR*	*Rob std err*[a]
Product complexity	2.348***	.185	4.19***	0.227	1.992***	0.050
Contract value	1.196***	0.023	1.070***	0.007	1.055***	0.006
Market competition	.0975	0.083	0.651***	0.051	1.3077**	0.039
Competitive limitation	0.743	0.246	0.668***	0.109	0.432***	0.043
Unrestricted competition	2.196***	0.629	2.570***	0.339	1.034	0.087
Contract length	1.149***	0.045	1.071***	0.019	1.060***	0.019
DOD	1.479*	0.357	6.996***	1.212	0.461***	0.042
Constant	1.51e-06***	1.26e-06	5.26e-06***	3.17e-06	0.001***	0.001

N = 246,354
Chi^2 = 5050.254***
Pseudo R^2 = 0.35
*p<.10 **p<.05 ***p<.01; two-tailed tests
[a]112,249 clusters at the IDVPIID level
Year dummies not shown

In order to fully test hypothesis 3, we also compare incentive contracts to cost reimbursement and time and materials contracts. To do this, we compute relative-risk ratios with varying base categories for the two explanatory variables that reach statistical significance across all contract types: *contract value* and *product characteristics*. Results are reported in Table 6. We find evidence that incentive contracts are favored over other contract types when *contract value* and *product characteristics* are high, substantiating H3.

Table 6. Relative-Risk Ratios

Risk factors	Contract pricing types		
	Incentive rather than fixed price contract	Incentive rather than cost reimbursement contract	Incentive rather than time and materials contract
Endogenous: Product complexity	2.348	0.5604	1.1788
Exogenous: Contract value	1.196	1.118	1.133

Discussion

In this research, we argued that the characteristics of supply chain risk would, in part, predict the preferences regarding the nature of the formal contract that governed buyer–supplier relationships in the public sector. We find some support for these hypotheses, but some results are also mixed. First, we expected that, in general, fixed price contracts would be favored when endogenous risk was low, as compared to situations when endogenous risk was high. We expected that fixed price contracts would be especially favored when exogenous risk was also low, and we see some support for this; however, the results are mixed. This may indicate that there are different considerations about managing supply chain risk depending on the nature of the exogenous risk. If exogenous risk is thought of as the potential for political attention (as operationalized by the dollar value of the contract), then our results fit with our expectations in hypothesis 1. If exogenous risk is conceived in other terms, mangers may be following more idiosyncratic approaches to risk management. When endogenous and exogenous risk is high (as measured by market competition and contract value) cost reimbursement contracts are favored over fixed price contracts, generally supporting hypothesis 2.

We find partial support for hypothesis 3. We purported that when both exogenous and endogenous risk is high, incentive contracts would be favored over other contract types. The results show that for the most costly of government contracts, risk sharing between supplier and government purchaser is preferable and incentive contracts are favored.

In no case did we find *competitive limitation*, or use of a set aside, affecting contract choices as we expected. For both cost reimbursement and time and materials contracts, the measure was statistically significant and negative, noting that managers are less inclined to restrict competition through set asides for contracts, wherein the buyer assumes greater risk than the supplier. This might mean that buyers are willing to amplify risk when that risk is shifted to the supplier in fixed price contracts, but are less likely to increase risk through set aside provisions when the risk is borne or partially borne by the government.

We also see evidence that these considerations might be affected by the nature of the relationship between the contracting partners. In all cases, when contract lengths are longer, all contract types are favored over fixed price contracts, suggesting that the longer the contract, the more willing the government is to shoulder risk. Although length of time is an insufficient proxy, this may indicate that like other types of supply chains, as contracting partners work together longer, the relational contract is able to strengthen, mitigating the need to specify risk in the formal contract (Eckerd & Eckerd, 2016).

Limitations

We acknowledge some limitations of this preliminary study. We have only examined 22 product categories and their respective complexity ratings in this analysis, which does not cover the vast array of products and services purchased by the federal government. We also acknowledge that our measures of exogenous risk are incomplete. While each of our measures represent aspects of exogenous risk as they relate to the political process, we recognize the need to develop a more complete picture of exogenous risk that takes into account characteristics that more specifically gauge political salience in addition to susceptibility to natural and or manmade disruptions.

Nevertheless, we believe that this preliminary analysis offers a proof of concept suggesting that there are meaningful relationships between the levels of endogenous and exogenous risk and the choices that are made regarding the formal governance arrangements in public sector supply chains. In short, we argue that the selection of contract type presents public sector contract managers with an opportunity to manage risks.

Conclusion

In this manuscript, we argued that supply chain management choices are affected by both the extent to which there is a risk of disruption within the supply chain and external to the supply chain as well. We argued that public sector supply chains are subject to considerable exogenous risk and that studying supply chain management decisions in the public sector offers a unique opportunity to understand how supply chain managers deal with both endogenous and exogenous risk situations. In this preliminary study, we found evidence suggesting that, as expected, when endogenous risk is low, suppliers tend to bear most of the disruption risk by agreeing to fixed price contracts. Conversely, when endogenous risk is high, we find partial support that government purchasers are willing to bear most of the risk by agreeing to cost reimbursement or time and materials contracts. We also find partial support of our proposition that when both endogenous and exogenous risk is high, governments are more likely share the risk with suppliers by favoring incentive contracts.

References

Blue Ribbon Commission on Defense Management. (1986). *A quest for excellence*. Retrieved March 29, 2016, from http://www.documentcloud.org/documents/2695411-Packard-Commission.html

Brown, T. L., & Potoski, M. (2003). Transaction costs and institutional explanations for government service production decisions. *Journal of Public Administration Research and Theory, 13*(4), 441–468.

Brown, T. L., Potoski, M., & Van Slyke, D. (2015). Managing complex contracts: A theoretical approach. *Journal of Public Administration Research and Theory.* Forthcoming.

Brown, T. L., Potoski, M., & Van Slyke, D. (2006). Managing public service contracts: Aligning values, institutions, and markets. *Public Administration Review, 66*(3), 53–67.

Davenport, C. (2015, November 15). How the Pentagon spent $43 million on a single gas station. *Washington Post*. Retrieved March 29, 2016, from https://www.washingtonpost.com/news/checkpoint/wp/2015/11/02/how-the-pentagon-spent-43-million-on-a-single-gas-station/

Dixit, A. (2002). Incentives and organizations in the public sector. *Journal of Human Resources, 37*(4), 696–727.

Eckerd, A., & Eckerd, S. (2016). Institutional constraints, managerial choices, and conflicts in public sector supply chains. *International Public Management Journal*. Forthcoming.

Eckerd, A., & Snider, K. (2016). Does the program manager matter? New public management and defense acquisition. *American Review of Public Administration*. Forthcoming.

Epstein, L., & Segal, J. A. (2000). Measuring issue salience. *American Journal of Political Science, 44*(1), 66–83.

Eriksson, P. E. (2015). Partnering in engineering projects: Four dimensions of supply chain integration. *Journal of Purchasing & Supply Management, 21*, 38–50.

Flynn, B. B., & Flynn, E. J. (1999). Information-processing alternatives for coping with manufacturing environment complexity. *Decision Sciences, 30*(4), 1021.

Folta, T. B. (1998). Governance and uncertainty: The trade-off between administrative control and commitment. *Strategy Management Journal, 19*(11), 1007–1029.

Jensen, P. H., & Stonecash, R. E. (2005). Incentives and the efficiency of public sector-outsourcing contracts. *Journal of Economic Surveys, 19*(5): 767–787.

Kim, Y. W., & Brown, T. L. (2012). The importance of contract design. *Public Administration Review, 72*(5), 687–696.

Kim, Y. W., Roberts, A., & Brown, T. L. (n.d.). The impact of product characteristics and market conditions on contract type: The use of fixed-price versus cost reimbursement contracts in the U.S. Department of Defense. *Public Performance and Management Review*. Forthcoming.

Lawther, W. C., and Martin, L. L. (2005). Innovative practices in public procurement partnerships: The case of the United States. *Journal of Purchasing & Supply Management, 11,* 212–220.

Lee, H. L., Padmanabhan, V., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science, 43*(4), 546–558.

Leiblein, M. J. (2003). The choice of organizational governance form and performance: Predictions from transaction cost, resource-based, and real options theories. *Journal of Management, 29*(6), 937–961.

Manuj, I., Esper, T. L., & Stank, T. P. (2014). Supply chain risk management approaches under different conditions of risk. *Journal of Business Logistics, 35*(3), 241–258.

Peck, H. (2005). Drivers of supply chain vulnerability: an integrated framework. International *Journal of Physical Distribution & Logistics Management, 35*(4), 210–232.

Roels, G., Karmarkar, U. S., & Carr, S. (2010). Contracting for collaborative services. *Management Science, 56*(5), 849–863.

Trkman, P., & McCormack, K. (2009). A conceptual model for managing supply chain risk in turbulent environments. *Journal of Production Economics, 119*(2), 247–258.

Vachon, S., & Klassen, R. D. (2002). An exploratory investigation of the effects of supply chain complexity on delivery performance. *IEEE Transactions on Engineering Management, 49*(3), 218–230.

Weber, L., and Mayer K. (2014). Transaction cost economics and the cognitive perspective: Investigating the sources of governance and interpretive uncertainty. *Academy of Management Review, 39*(3), 344–363.

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-15-1-0014 awarded by Naval Supply Systems Command Fleet Logistics Center San Diego (NAVSUP FLCSD). The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. government.

An Approach for Modeling Supplier Resilience

Kash Barker—is an Associate Professor in the School of Industrial and Systems Engineering at the University of Oklahoma. Dr. Barker and his students in the Risk-Based Decision Making Laboratory are primarily interested in (i) modeling the reliability, resilience, and interdependent economic impacts of disruptions to critical infrastructure networks and (ii) enhancing data-driven decision making for large-scale system sustainment. He received his PhD in systems engineering from the University of Virginia, where he worked in the Center for Risk Management of Engineering Systems. [kashbarker@ou.edu]

Jose Emmanuel Ramirez-Marquez—Ramirez-Marquez is an Associate Professor in the School of Systems and Enterprises at the Stevens Institute of Technology and the Director of the School's Engineering Management program. Dr. Ramirez-Marquez's work advances systems management and assessment for optimal development of a system through its life cycle. His other interests include reliability analysis, network resilience, and optimization. He received his PhD in industrial engineering from Rutgers University. [jmarquez@stevens.edu]

Seyedmohsen Hosseini—is a PhD candidate in the School of Industrial and Systems Engineering at the University of Oklahoma. His primary research interests include resilience modeling of supply chains, reliability engineering, and meta-heuristic optimization. [m.hosseini@ou.edu]

Abstract

Supplier selection plays a key role in the context of supply chain management. As recent emphasis has been placed on supply chain resilience, so too should such emphasis be placed on resilient suppliers. In particular, this work evaluates how different suppliers enable the supply chain to withstand the impacts of a disruption and return performance to a desired level in a timely manner. The primary measure of supply chain performance is taken to be availability, or the extent to which the products produced by the supply chain are available for use (measured as a ratio of uptime to total time of the use of the product). Available systems are important in many industries, particularly in the Department of Defense, where weapons systems are required in short notice but undergo regular maintenance activities. In addition to availability, suppliers are also measured according to their recovery rate, quality, and delivery rate. Suppliers are evaluated against these four criteria using a multi-criteria decision analysis technique.

Introduction

Supply chain management is becoming increasingly significant to achieve competitiveness in the business environment, as recently the paradigm for corporate management has shifted from competition between individual firms to the competition between supply chains (Cho et al., 2008). In supply chain management, relationships with suppliers have an impact on the success of the strategic goals of a buyer. Hence, it is necessary for a buyer to keep track of these relationships, evaluate supplier performance, and optimize its supply base.

Manufacturing companies need to collaborate with various suppliers to continue their business activities. In manufacturing industries, raw materials and component parts can amount to 70% of the cost of a finished product (Stueland, 2004). In such a circumstance, the acquisition department can have a significant influence on cost reduction, suggesting that supplier selection is among the more critical functions of acquisition.

Supplier evaluation and selection is the process of finding a capable supplier that is able to supply high quality products on time at the right price. Supplier selection is a multi-criteria decision making problem that involves two major tasks: (i) determine the criteria to be considered and (ii) compare the eligibility of suppliers. Generally speaking, the traditional criteria associated with supplier selection can be divided into qualitative and quantitative categories. Quantitative supplier criteria have included transportation costs, purchasing and order costs, delivery time, and product defect rate, while qualitative criteria have included product quality, warranties and claim policies, performance history, technical capability, geographical location, and labor relations (Luo, Rosenber, & Barnes, 2009; Liao & Kao, 2011; Arikan, 2013; Lienland, Baumgartner, & Kunbben, 2013; Yu & Wong, 2015).

Although research efforts have been dedicated to supplier evaluation and selection, accounting for resilience-based criteria for supplier selection has not been well explored (Hosseini & Barker, 2016). The notion of resilience, or the ability of a company or its supply chain to withstand and subsequently recover from a disruption, has become very important in the scope of supply chain management. Supplier disruptions can impose significant losses to the entire supply chain by discontinuing supply flows. For example, a devastating earthquake in central Taiwan in September 1999 had severe consequences for many manufacturing industries and organizations, as total industrial production losses were approximated at $1.2 billion (Papadakis, 2006). Many large scale semiconductor fabrication facilities, estimated to account for roughly 10% of the world's production of computer memory chips, were damaged (Bhamra, Dani, & Burnard, 2011). The impact of the earthquake disaster on the PC supply chain was dramatic, as the supply of computer components was constrained for several months, affecting technology companies such as Dell, Gateway, IBM, Apple, and HP.

In 2011, the Japanese earthquake and tsunami had similar adverse impacts to the global supply chain networks of automobile manufacturers (Manual, 2013). For example, automobile manufacturers attempted to find other sources for a special pigment used in automobile paint after the Japanese earthquake and tsunami disabled the main facility in 2011. The availability of new U.S. automobiles was reduced for several months after the disruption of key suppliers, including the paint supplier. Availability is a key metric not only in industry but also in the DoD. Weapons system availability is critical to the DoD (2005), requiring that such systems be operational at a moment's notice. With smaller maintenance, repair, and overhaul (MRO) inventories and as modern supply chains are increasingly vulnerable to disruptions, it is important to understand how *resilient* suppliers are to such disruptions so that system availability can be maintained.

In this paper, we explore supply chain availability as a measure of resilience and use this measure in a set of supplier selection criteria. The following section offers some background on several components of the research, including the Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS). The section following that discusses the supplier evaluation and selection criteria used in this work, and the section titled Illustrative Example offers an illustrative example of the methodology. Concluding remarks are provided in the final section.

Background

This section provides methodological background to some components of this research, including a paradigm for resilience, recent approaches to comparing suppliers, and a particular approach for the multi-criteria comparison of discrete alternatives.

Resilience Modeling

In the last few years, the concept of *resilience* has been increasingly used to describe the behavior of systems under disruption, and several measures of resilience have been offered (Park et al., 2013; Hosseini et al., 2015). In particular, this work adopts a graphical paradigm of system behavior before, during, and after a disruption is provided in Figure 1 (Henry & Ramirez-Marquez, 2012; Barker et al., 2013; Pant et al., 2014). It is assumed that system performance, measured with function $\varphi(t)$, reduces after a disruptive event e^k and improves to an acceptable level over time (e.g., flow along a network, availability of a system or supply chain). Figure 1 highlights three dimensions of resilience: reliability, vulnerability, and recoverability. The normal behavior of the system in the time interval $t_\text{e} - t_0$, or in its *Stable Original State*, S_0, is described by the system's reliability. The vulnerability dimension of resilience describes the extent to which $\varphi(t)$ degrades to a *Disrupted State*, S_d, during the time interval $t_\text{d} - t_\text{e}$ The recovery of the system to its *Stable Recovered State*, S_f, occurs during the time interval $t_\text{f} - t_\text{d}$.

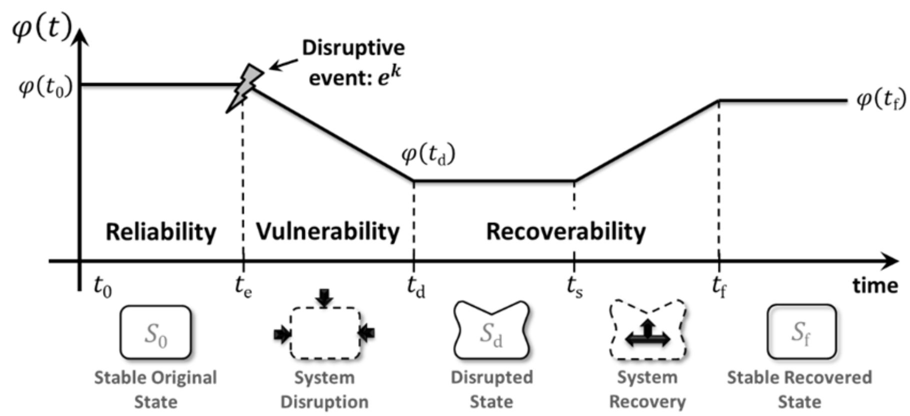


Figure 1. Graphical Depiction of Decreasing System Performance, $\varphi(t)$, Across Several State Transitions Over Time

Supplier Selection Approaches

Various methods have been implemented to deal with supplier selection problems, including multi-criteria decision analysis techniques, mathematical programming, and artificial intelligence, among others. Liao and Kao (2011) combined a fuzzy extension of TOPSIS and multi-choice goal programming to solve the supplier selection problem, allowing decision makers to consider multiple aspiration levels. Kilincci and Onal (2011) employed a fuzzy extension of the analytic hierarchy process (AHP) for supplier selection. Karsak and Dursun (2014) introduced an approach based on integrating quality function deployment and data envelopment analysis for selecting the best among supplier alternatives, studying the interdependence among supplier evaluation criteria with the construction of a house of quality. Deng and Chen (2011) proposed a methodology based on fuzzy set theory and Dempster-Shafer theory to deal with the supplier selection problem. Igoulalene, Benyoucef, and Kumar Tiwari (2015) proposed a fuzzy hybrid multi-criteria decision analysis approach based on combining fuzzy consensus-based possibility measure and fuzzy TOPSIS. Kar (2014) integrated fuzzy AHP and fuzzy goal programming for the supplier selection problem. Lee, Cho, and Kim (2014) combined TOPSIS and AHP based on fuzzy theory to determine the prior weights of criteria and select the best-fit suppliers by taking subjective vague preferences of decision making into account. You, You, Liu, and

Zhen (2015) developed a new multi-criteria decision model based on using interval 2-tuple linguistic variables and an extended VIKOR approach to select the best supplier under uncertainty and incomplete information. Dalalah, Hayajneh, and Batieha (2011) adjusted DEMATEL to deal with fuzzy rating and assessments by converting the relationship between causes and effect of the criteria into an intelligible structural model. Deng, Hu, Deng, and Mahadevan (2014) presented a new form of representation for uncertain information involved with supplier selection, called D numbers, which the authors then integrated with AHP. Fazlollahtabar et al. (2011) proposed a multiobjective mixed integer programming for supplier selection with an objective to minimize total supplier costs including cost, total defect rate, total penalized earliness and tardiness, and total value of purchase.

TOPSIS

TOPSIS, which will be used in this paper for combining supplier performance along several criteria, was developed by Hwang and Yoon (1981) for finding the best among several discrete alternatives given multiple decision criteria. The basic principle of TOPSIS is that the chosen alternative should be the closest to the best (or positive ideal) solution and farthest from the worst (or negative ideal) solution.

Suppose that there are n criteria $(C_1, \dots, C_n)$ which are considered to discern among m discrete alternatives $(A_1, \dots, A_m)$. Let x_{ij} be the performance of the ith alternative for the jth criterion. The weight of importance of the jth criterion is j, such that $\sum_{j=1}^{n} w_j = 1$. TOPSIS is applied to rank the m alternatives with six steps, as follows:

Step 1. Calculate the normalized value n_{ij} for $i = 1, \dots, m$ and $j = 1, \dots, n$ Equation 1 represents one such approach to normalizing the value of the criteria (which could be of different magnitudes) for each alternative.

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \tag{1}$$

Step 2. Calculate the weighted normalized value v_{ij} with Equation 2.

$$v_{ij} = w_j n_{ij} \tag{2}$$

Step 3. Determine the positive ideal solution A^+ and the negative ideal solution A^- with Equations 3 and 4, where S_B and S_C denote the set of benefit criteria and set of cost criteria, respectively. The positive ideal solution has all the best attainable criteria values, while the negative ideal solution has all worst possible criteria values.

Equation 7 suggests that the positive ideal solution consists of those weighted performance ratings that maximize benefit criteria and minimize cost criteria. Likewise, the negative ideal solution, or the weighted performance ratings that represent the smallest from set C^+ and largest from set C^-, is provided in Equation 8.

$$A^+ = \{v_1^+, \dots, v_n^+\} = \{(\max_i v_{ij} \,|j \in S_B),(\min_i v_{ij} \,|j \in S_C)\} \tag{3}$$

$$A^- = \{v_1^-, \dots, v_n^-\} = \{(\min_i v_{ij} \,|j \in S_B),(\max_i v_{ij} \,|j \in S_C)\} \tag{4}$$

Step 4. Calculate Euclidean distance between each alternative and the positive and negative ideal solutions with Equations 5 and 6, respectively, for all i.

$$S_i^+ = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^+\right)^2} \quad (5)$$

$$S_i^- = \sqrt{\sum_{j=1}^{n}\left(v_{ij} - v_j^-\right)^2} \quad (6)$$

Step 5. Calculate the relative closeness to the ideal solution for all i.

$$RC_i = \frac{S_i^-}{S_i^+ + S_i^-} \quad (7)$$

Step 6. Rank the alternatives according to RC_i in Equation 7. The larger the value of RC_i, the closer alternative i is to the positive ideal solution. As such, alternatives are ranked according to descending values of RC_i.

Supplier Selection Criteria

Dickson (1966) introduced 23 supplier selection criteria still found in literature today, including quality, delivery, performance history, and price. Recently, Hosseini and Barker (2016) characterized supplier selection criteria into *primary* (i.e., traditionally used criteria with a history in the literature), *green* (i.e., environmentally-focused criteria recently appearing in the literature), and *resilience* (i.e., dealing with a supplier's ability to withstand and recover from a disruption) categories.

Availability Criterion

The performance function for a supply chain, $\varphi(t) = A_0(t)$, is assumed to be its *availability*, measured as a proportional level of service (ratio of uptime to total time) that can be attained by the products produced by a supply chain. This work makes use of a formulation by Sherbrooke (2004; and extended computationally by Nowicki, Randall, and Ramirez-Marquez, 2012) to redistribute supplies coming from a number of suppliers in meeting demand in a multi-echelon supply chain.

An example is provided in Figure 2, where the supply chain has a central depot, two intermediate locations (e.g., end-item integrators), and six field locations (e.g., sub-assembly suppliers). Each location within an echelon has an input vector that defines the cost, reliability, and maintainability of a spare item at that location. The item's reliability is defined in terms of average number of demands per year, and the item's maintainability is defined as mean time to repair in days. Availability measure A_0, as well as the associated spare strategy for each supplier, was obtained from the algorithm described in Nowicki et al. (2012). The objective of the algorithm is to determine the vendor mix and quantity of spares that either maximizes the operational availability subject to a budget constraint (or otherwise minimizes cost subject to an operational availability target).

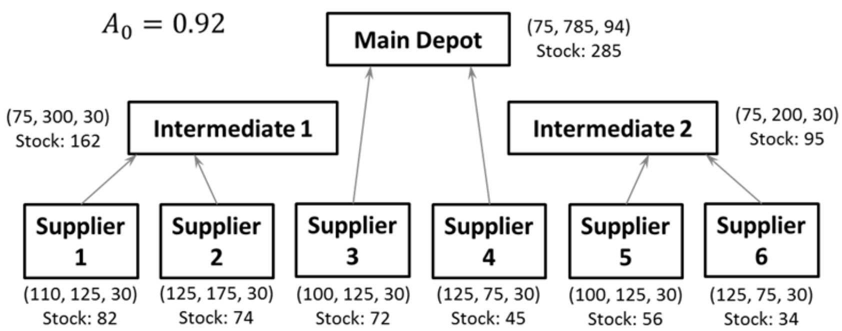


Figure 2. Supply Chain Topology and Characteristics Resulting in an Availability of 0.92

Let E represent the set of echelons in a multi-echelon supply chain, with $e = 0,1,\ldots,|E|$. Let L^e be the set of locations within e, with index $l = 1,2,\ldots,|L^e|$, and let I^{le} be the set of items at location l within echelon e. As the index of an item or product is i, the demand quantity of item i at location l within echelon e in any fixed interval of length t is $N_i^{le}(t)$. And s_i^{le} represents the stock level of item i at location l within echelon e.

To calculate the availability of the multi-echelon supply chain, the expected number of backorders must be identified as the expected amount of unfilled demand that exists at a point in time. Note that unfilled demand is a function of a particular delay scenario, and as such, depends on the number of existing spares at each location; within each echelon they can be used as a surrogate measure for operational availability. Therefore, the amount of backorders for item i can be calculated with Equation 8 (Nowicki et al., 2012).

$$BO\left(N_i^{le}(t)\middle|s_i^{le}\right) = \begin{cases} N_i^{le}(t) & \text{if } N_i^{le}(t) > s_i^{le} \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Note that a backorder of size $N_i^{le}(t) - s_i^{le}$ occurs whenever the number of demands exceeds the inventory on-hand, or $N_i^{le}(t) > s_i^{le}$. As such, the expected number of backorders can be calculated with Equation 9, where x is the random variable.

$$E\left[BO\left(N_i^{le}(t)\middle|s_i^{le}\right)\right] = \sum_{x=s_i^{le}+1}^{\infty} \left(x - s_i^{le}\right) P\left[N_i^{le}(t) = x\right] \qquad (9)$$

Finally, Sherbrooke (2004) demonstrated that the availability of a multi-echelon supply chain denoted by A_0 system can be calculated with Equation 10.

$$A_0 = 100 \prod_{l=1}^{L^E} \prod_{i=1}^{I^{lE}} \left(1 - E\left[BO\left(N_i^{le}(t)\middle|s_i^{le}\right)\right]/n\right)^n \qquad (10)$$

In this study, we would like to identify a backup supplier who can improve the availability of the supply chain when a primary supplier is disrupted. As such, a more resilient supply chain would be able to rebound to an availability value similar to (or improved relative to) baseline availability performance in a timely fashion.

Recovery Time, Quality, and Delivery Rate Criteria

In addition to the availability measure, other criteria are used to compare suppliers. Pairing with availability is *recovery time,* or the amount of time taken to engage an alternative supplier to improve availability. Hence, a supplier with a shorter recovery time (measured in days) is more desirable because it contributes to a more resilient supply chain when combined with availability.

The ability to meet specifications consistently is referred to as *quality*, a commonly used criterion in supplier evaluation. The quality of the product, process, or system is defined here as the percentage of products that meet the expectations of manufacturers.

Dickson (1966) defines *delivery rate* as the percentage of successful deliveries to meet specified delivery schedules. Its meaning is extended into criteria such as freight terms, lead time, delivery capacity, shipment quality, cycle time, and JIT delivery capability.

Availability, recovery time, quality, and delivery rate criteria are integrated together using TOPSIS for the comparison of suppliers that can be engaged when a primary is disrupted. This idea is illustrated with an example in the next section.

Illustrative Example

An example of a three-echelon supply chain of spares illustrates the availability and other criteria to evaluate and compare suppliers. Figure 2 illustrates the baseline supply chain configuration with the stock of spares assigned in each of the echelons.

Recall that each location within an echelon has an input vector that defines the cost, reliability (average demand per year), and maintainability (mean time to repair in days) of spare items at that location. In Figure 2, suppliers 1 and 2 and suppliers 5 and 6 supply to intermediate depot locations, while suppliers 3 and 4 supply to the main depot location. Note that the availability of the spares supply chain is calculated using Equation 10. More information about how the availability of multi-echelons can be calculated can be found in Sherbrooke (2004).

It is assumed that supplier 1 is disrupted and becomes inoperable, as illustrated in Figure 3. The availability reduces from 0.92 to 0.80.

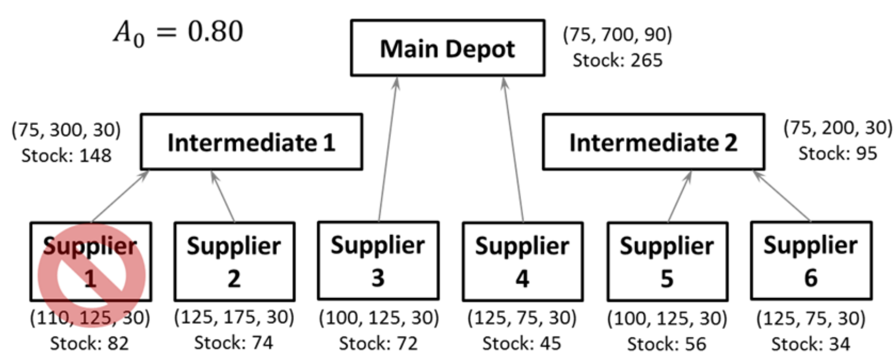


Figure 3. Availability Reduction When Supplier 1 Becomes Inoperable

Assume that three suppliers (A, B, and C) are evaluated as replacements for supplier 1. When their cost, reliability, and maintainability information are individually inserted in the availability algorithm, the supply chain availability resulting from alternative suppliers A, B, and C are 0.95, 0.92, and 0.90, respectively. These availability values, as well as the values of the quality, delivery, and recovery time criteria, are found in Table 1. Figure 4 provides an illustration of the resilience, or the combination of availability improvement and recovery time, of the three suppliers.

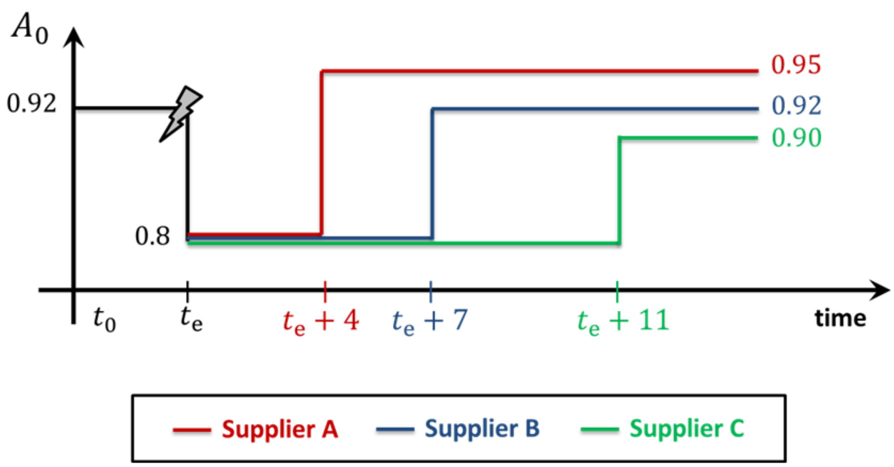


Figure 4. Depiction of the Contributions of the Three Alternative Suppliers to Supply Chain Resilience

Table 1. Criteria Values for the Three Alternative Suppliers to Replace Supplier 1

	Availability improvement	Recovery time	Quality	Delivery rate
Supplier A	0.15	4	0.97	0.82
Supplier B	0.12	7	0.83	0.98
Supplier C	0.1	11	0.89	0.91

Criteria weights of **w**=[0.3,0.3,0.2,0.2] are assumed for availability improvement, recovery time, quality, and delivery rate, respectively. The integration of the four criteria and their weights using TOPSIS results in the ranking provided in Table 2. As such, supplier A would be the best fit to replace supplier 1 in the event that supplier 1 becomes inoperable, according to the four criteria and how those criteria are weighted.

Table 1. Closeness Coefficient and Rank for Each of the Alternative Suppliers

Alternative supplier	RC_i	Rank
Supplier A	0.8934	1
Supplier B	0.5693	2
Supplier C	0.1074	3

Conclusions

The study provides a means to evaluate and select suppliers based on their ability to enhance supply chain resilience when a primary supplier is disrupted. As the availability of particular systems is important, availability is chosen as the primary measure of supply chain performance. Resilience is addressed with the combination of (i) improvement in supply chain availability and (ii) the time required for an alternative supplier to become available to the supply chain. Other criteria, including common supply chain characteristics of supplier quality and delivery rate, were also included. Ultimately, a multi-criteria decision analysis technique, TOPSIS, was used to rank the alternatives across the multiple criteria and their importance.

A small (initial) illustrative example helps illustrate how an algorithm for multi-echelon supply chain availability can be used in a supplier evaluation and selection problem that emphasizes supply chain resilience. Future work will expand this initial illustration to a larger supply chain while performing a sensitivity analysis of criteria weights.

References

Abdolshah, M. (2013). A review of quality criteria supporting supplier selection. *Journal of Quality and Reliability Engineering*, Article 621073.

Arikan, F. (2013). A fuzzy solution approach for multi objective supplier selection. Expert Systems With Applications, 40(3), 947–952.

Bhamra, R., Dani, S., & Burnard, K. (2011). Resilience: The concept, a literature review and future directions. International *Journal of Production Research, 49*(18), 5375–5393.

Chen, C. H., Tan, H. K., Liao, L. R., Chen, H. H., Chan, C. C., Cheng, J. J., Chen, C. Y., Wang, T. N., & Lu, M. L. (2007). Long-term psychological outcome of 1999 Taiwan earthquake survivors: A survey of a high-risk sample with property damage. *Comprehensive Psychiatry, 48*(3), 269–275.

Cho, G., Kim, J., Ho Park, H., Noh, C.-S., So, S.-H., Park, Y. S., & Jung, K. H. (2008). An optimal decision making model for supplier and buyer's win-win strategy in a two echelon supply chain. In *Proceedings of the 41st Hawaii International Conference on System Science*, Waikoloa, HI.

Dalalah, D., Hayajneh, M., & Batieha, F. (2011). A fuzzy multi-criteria decision making model for supplier selection. *Expert Systems With Applications, 38*(7), 8384–8391.

Deng, X., Hu, Y., Deng, Y., & Mahadevan, S. (2014). Supplier selection using AHP methodology extended by D numbers. *Expert Systems With Applications, 41*(1), 156–167.

Deng, Y., & Chen, F. T. S. (2011). A new fuzzy dempster MCDM method and its application in supplier selection. *Expert Systems With Applications, 38*(8), 9854–9861.

Dickson, G. W. (1966). An analysis of vendor selection systems and decisions. *Journal of Purchasing, 2*(5), 17–32.

Du, B., Guo, S., Huang, X., & Guo, J. (2015). A Pareto supplier selection algorithm for minimum the life cycle cost of complex product system. *Expert Systems With Applications, 42*(9), 4253–4264.

Fazlollahtabar, H., Mahdavi, I., Talebi Ashoori, M., Kaviani, S., & Mahdavi-Amiri, N. (2011). A multiobjective decision making process of supplier selection and order allocation for multi-period scheduling in an electronic market. International *Journal of Advanced Manufacturing Technology, 52*(9), 1039–1052.

Henry, D., & Ramirez-Marquez, J. E. (2012). Generic metrics and quantitative approaches for system resilience as a function of time. *Reliability Engineering and System Safety, 99,* 114–122.

Hwang, C. L., & Yoon, K. (1981). *Multiple attributes decision making methods and applications.* Berlin, Germany: Springer.

Igoulalene, I., Benyoucef, L., & Kumar Tiwari, M. (2015). Novel fuzzy hybrid multi-criteria decision making approaches for the strategic supplier selection problem. *Expert Systems With Applications, 42*(7), 3342–3356.

Kar, A. K. (2014). Revisiting the supplier selection problem: An integrated approach for group decision support. *Expert Systems With Applications, 41*(6), 2762–2771.

Karsak, E. E., & Dursun, M. (2014). An integrated supplier selection methodology incorporating QFD and DEA with imprecise data. *Expert Systems With Applications, 41*(16), 6995–7004.

Kilincci, O., & Onal, S. A. (2011). Fuzzy AHP approach for supplier selection in a washing machine company. *Expert Systems With Applications, 38*(8), 9656–9664.

Lee, J., Cho, H., & Kim, Y. S. (2014). Assessing business impacts of agility criterion and order allocation strategy in multi-criteria supplier selection. *Expert Systems With Applications, 42*(3), 1136–1148.

Liao, C-N., & Kao, H.-P. (2011). An integrated fuzzy TOPSIS and MCGP approach to supplier selection in supply chain management. *Expert Systems With Applications, 38*(9), 10803–10811.

Lienland, B., Baumgartner, A., & Kunbben, E. (2013). The undervaluation of corporate reputation as a supplier selection factor: An analysis of ingredient branding of complex products in the manufacturing industry. *Journal of Purchasing and Supply Management, 19*(2), 84–97.

Luo, X., Rosenber, D., & Barnes, D. (2009). Supplier selection in agile supply chains: An information-processing model and illustration. *Journal of Purchasing and Supply Management, 15*(4), 249–262.

Manual, J. (2013). The long road to recovery: Environmental health impacts on Hurricane Sandy. *Environmental Health Perspectives, 212*, 152–159.

Nowicki, D. R., Randall, W. S., & Ramirez-Marquez, J. E. (2012). Improving the computational efficiency of metric-based spares algorithms*. European Journal of Operational Research, 219*(2), 324–334.

Papadakis, I. S. (2006). Financial performance of supply chains after disruptions: An event study. *Supply Chain Management, 11*(1), 25–33.

Sherbrooke, C. C. (2004). *Optimal inventory modeling of systems: Multi-echelon techniques* (2nd ed.). Kluwer Academic.

Stueland, V. J. (2004, April). Supplier evaluations: Best practices and creating or improving own evaluation, Wells Fargo Services Company. In *Proceedings of 89th Annual International Supply Management Conference*.

You, X-Y., You, J-X., Liu, H-C., & Zhen, L. (2015). Group multi-criteria supplier selection using an extended VIKOR method with interval 2-tuple linguistic information. *Expert Systems With Applications, 42*(4), 1906–1916.

Yu, C., & Wong, T. N. (2015). An agent-based negotiation model for supplier selection of multiple products with synergy effect. *Expert Systems With Applications, 42*(1), 223–237.

Acknowledgments

This publication results from research supported by the Naval Postgraduate School Assistance Grant/Agreement No. N00244-15-1-0042 awarded by the NAVSUP Fleet Logistics Center San Diego. The views expressed in written materials or publications do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Antecedents and Consequences of Supplier Performance Evaluation Efficacy

Timothy G. Hawkins,[1] Lt Col, USAF (Ret.)—is an Assistant Professor in the Department of Marketing at Western Kentucky University in Bowling Green, KY. He researches and teaches in the realm of supply chain management, government contracting, and strategic sourcing. He has 20 years of sourcing experience in industry and government. Dr. Hawkins has published articles on opportunism in buyer–supplier relationships, performance-based logistics, collaborative pricing, and electronic reverse auctions in scholarly publications such as *Journal of Supply Chain Management*, *Journal of Business Logistics*, *Journal of Purchasing and Supply Management*, *Journal of Business Research*, *International Journal of Logistics Management*, *Industrial Marketing Management*, *Journal of Business Ethics*, *Supply Chain Management: An International Journal*, *Journal of Marketing Channels*, *Air Force Journal of Logistics*, *Journal of Contract Management*, *International Journal of Procurement Management*, *Journal of Product and Brand Management*, and *Journal of Public Procurement*. His current research interests include performance-based logistics, electronic reverse auctions, procurement ethics, buyer–supplier relationships, strategic sourcing, services procurement, and supplier performance management. Prior to his academic career, Dr. Hawkins worked in Global Procurement for NCR Corporation and served as a warranted contracting officer for the U.S. Air Force. [timothy.hawkins@wku.edu]

Michael J. Gravier—is an Associate Professor of Marketing and Global Supply Chain Management at Bryant University. He received his PhD in Marketing and Logistics from the University of North Texas, an MS in Logistics Management (with a specialization in transportation management) from the Air Force Institute of Technology, and BA with majors in Spanish and Anthropology from Washington University in St. Louis. Prior to his academic career, he spent 12 years as an active duty logistics readiness officer in the U.S. Air Force and worked as a research assistant in the Washington University School of Medicine. He has research interests in supply chain ethics, procurement, logistics pedagogy, logistics in emerging countries, and the evolution of supply chain networks in response to risk factors like obsolescence and changing information needs. [mgravier@bryant.edu]

Abstract

There are numerous weaknesses associated with industrial buyers' collection and use of supplier performance evaluation (SPE) information (a.k.a., past performance information). These weaknesses call into question the efficacy of SPEs. Neither the factors affecting SPE efficacy (i.e., its antecedents) nor the effects of SPE efficacy (i.e., consequences) on suppliers have been empirically explored. Despite the fallibility of SPE schemes, there are no known studies that explore the accuracy of SPEs, nor are there studies examining whether and how inaccurate SPEs affect suppliers—specifically, their performance. The purpose of this research, therefore, is to identify the factors affecting SPE efficacy, then to examine how SPE efficacy, in turn, affects supplier outcomes. This research will employ a mixed method of qualitative interviews and quantitative analysis of survey data collected from suppliers and from assessors of supplier performance.

Introduction

Industrial buyers labor to avoid the deleterious effects of the laws of agency. In industrial buying, the supplier serves as an agent to the principal (i.e., the buying organization). Substantial transaction costs are dedicated to avoid adverse selection—the

[1] Corresponding author

risk of selecting an incapable supplier that otherwise misrepresents itself as capable. Following contract formation, more transaction costs are incurred to monitor supplier performance to thwart supplier opportunism ex post.

Supplier performance evaluation (SPE) became popular in the 1950s (Wieters & Ostrom, 1979), and now SPE is an essential best practice in business-to-business sourcing (Gordon, 2008; Talluri & Sarkis, 2002). SPE is "the process of evaluating, measuring, and monitoring supplier performance and suppliers' business processes and practices for the purposes of reducing costs, mitigating risk, and driving continuous improvement" (Gordon, 2008, p. 4). SPEs are used to (1) prioritize supplier improvement activities, (2) focus management attention on critical suppliers, (3) support supplier selection decisions, (4) communicate dissatisfaction with supplier performance, (5) communicate performance expectations to suppliers, (6) document historical performance, (7) inform the purchasing department of supply base performance, (8) influence suppliers, and (9) continuously improve (Schmitz & Platt, 2003). Specifically, SPEs inform source selection decisions of the likelihood that a prospective supplier will successfully perform the contract (FASA, 1994).

Similarly, the primary purpose of the U.S. federal government's Contractor Performance Assessment Reporting System (CPARS) "is to ensure that current, complete and accurate information on contractor performance is available for use in procurement source selections" (Naval Sea Logistics Center Portsmouth, 2014, p.1). The idea is that by better informing source selection decisions, better best value selections will occur. Integrally related is the supplier's level of performance. If performance levels are assessed and recorded, and if this information is available to buyers during a future source selection, it is believed that suppliers will work harder to ensure satisfactory (or better) performance (OFPP, 2000).

Despite long-term awareness of weaknesses and despite recent, concerted, high-level efforts to improve past performance reporting, the government's past performance evaluations of its suppliers continue to be deficient (GAO, 2014). Too often, they are not properly, timely, or accurately completed. Reports often lack sufficient information to support ratings (e.g., how the contractor exceeded or failed to meet requirements) necessary to withstand a legal challenge, or do not include a rating for all performance areas (OFPP, 2011). Additionally, throughout the rating process, raters are often inclined to inflate ratings in order to avoid conflict with the contractor (GAO, 2009).

Unreliable or inaccurate past performance assessments can harm contractors' reputations and can bias source selections resulting in adverse selection. If past performance information is not reliable, and if buyers and evaluators do not (or cannot) use the information to discriminate between competitive proposals (Kelman, 2010), the effort of collecting and reporting the past performance information is squandered. Likewise, the efforts of prospective suppliers in documenting and of buyer-side evaluators in evaluating inaccurate past performance information during source selections is wasted. Notably, we don't know how much transaction costs by all parties involved are consumed in completing a past performance evaluation. If the effort is significant, and the resultant information is of little value, policy-makers should revisit the policy and its implementing systems. Notwithstanding, buying organizations often use SPE information to identify and rank superior performing suppliers. Of course, the rankings and status are suspect if the underlying SPE ratings are not accurate.

Problems are not unique to the not-for-profit sector. Hald and Ellegaard (2011) found that supplier evaluations change throughout the evaluation process. Underlying data captured in enterprise resource planning (ERP) databases is often flawed. Masses of performance data are condensed into more general ratings sacrificing fidelity. Buyers also commonly use multiple evaluators to rate supplier performance (Hald & Ellegaard, 2011; Buffa & Ross, 2011), which invites different perspectives of supplier performance. To what extent does evaluators' dissonance affect perceived accuracy of SPEs? Additionally, the degree of internal dissonance of supplier evaluations has not yet been examined. Hald and Ellegaard (2011) also reported that performance ratings are sometimes negotiated with suppliers when the accuracy is challenged. However, no one has explored why buyers decide to change their evaluations.

Despite the fallibility of SPE schemes, there are no known studies that explore the accuracy of SPEs. Therefore, further investigation is needed in order to explore the validity of SPE processes. After all, SPE assessments can affect key outcomes such as contract compliance, supplier performance-based payments, supplier reputation, future business awards, incentive awards, and status achievement (e.g., a "preferred" supplier). As such, the effectiveness of SPEs in assisting source selection decisions is questionable (Berrios, 2006). In other words, we do not know the extent to which SPEs validly build the buyer's confidence in its assessment of the risk of doing business with a particular supplier prior to contract award. Furthermore, the impact of deficient SPEs on the industrial supply base is unknown.

Scope and Objectives

The purpose of this research, therefore, is to explain the efficacy of SPE and to explore the effects of SPE efficacy on suppliers. This research will explore the extent to which the supplier performance information collection and usage processes achieve the intended goals of (1) mitigating the risk of adverse selection, and (2) motivating supplier performance. The following research questions will be explored:

1. What factors decrease the efficacy of SPEs?
2. How do suppliers react to inaccurate SPEs?
3. Do SPEs, in general, motivate suppliers to increase performance?
4. How does the accuracy of SPEs affect relationship quality?
5. Why are SPEs often inaccurate?
6. How many man-hours do suppliers invest in responding to SPEs?
7. What communication tactics do suppliers use to manage the SPE process?
8. To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy?

The remainder of this paper is organized in the following manner. The research explores antecedents and consequences to SPE efficacy, and uses two separate approaches to do so. To explore the antecedents, this research builds off of prior research (Hawkins, 2013) to test previously-suggested propositions of buyer-side factors that affect SPE efficacy. To identify the consequences of SPE efficacy on suppliers, an exploratory, qualitative approach is employed. First, a literature review is presented describing the conceptual framework and hypotheses. Next, the study presents the research designs and methodologies.

Literature Review

Past Performance

U.S. federal government contracting serves as the context for this study due to its expansive scope (dollars, industries, and geographies), rigor, established fairness, and standardized procedures. In U.S. federal government contracting, agencies are required to consider past performance information as an evaluation factor in source selections exceeding the simplified acquisition threshold, $150,000 (FAR Part 15)—unless the contracting officer documents a reason not to do so. Necessarily, then, agencies must collect and report contractor past performance information from government contracts (FAR Part 42) surpassing certain dollar values.

It is important to note that in keeping with the government's core goal of transparency and fairness (FAR 1.102), contractors must be afforded the opportunity to comment on the government's assessment of past performance, and any disagreements must be resolved by a reviewing official one level above the contracting officer. Additionally, contractor past performance assessments are increasingly subject to the Contract Disputes Act of 1978 (Lord, 2005). While the courts will not yet direct a particular rating, they will require agencies to adequately support assessments/ratings with sufficient facts. This written justification consumes significant time from the raters, contractors (i.e., rebuttals), and approving officials. As further incentive to conceal true performance, program officials will go to extraordinary lengths to protect their programs. A poorly performing contractor can signal a troubled program, increasing the threat of cancelation (GAO, 2009). Other reasons that truthful performance is not reported include a desire to maintain relations with the contractor, difficulty attributing performance problems to the contractor or to the government, deficient oversight of contractors, deficient contract administration, and the government's lack of contractor performance management (GAO, 2009).

It is also important to note the U.S. Military Departments' recently-emerged practice of ranking government contractors based on performance across multiple contracts. This annual ranking, deemed the superior supplier incentive program (SSIP), relies on performance data from CPARs (USD[AT&L], 2015). The purpose is to incentivize contractor performance, and to recognize those top achievers. Suppliers deemed a superior supplier are eligible for relaxed or more favorable contract terms and conditions. Hence, the efficacy of the SPE process takes on additional meaning by providing firms bragging rights (i.e., marketing material and enhanced reputation) and eased admininstrative burdens.

Supplier Performance Evaluation

It is not surprising that buying firms closely measure their suppliers' performance when 50%–70% of their revenue is spent on goods and services to support the sales (Monczka et al., 2011b). Measuring supplier quality is critical since the cost of poor quality ranges from 10% to 25% of sales, and the cost of poor supplier quality ranges from 25% to 70% of the cost of poor quality (Gordon, 2008). Commercial SPM systems—often web-based and at least partially automated—encompass means to measure, rate, and rank suppliers. One study reported that 97% of firms use a periodic supplier scorecard or assessment for direct materials (CAPS Research, 2011).

SPM pays off; a study by the Aberdeen Group (2005) found that supplier performance of companies with an SPM system improved significantly more than did the supplier performance of firms with no SPM system. Specifically, firms using an SPM system realized 10% greater price savings, 12% better on-time delivery improvement, four times greater quality improvement, and a 4% greater improvement in service. One large telecommunications firm realized a 290% reduction in the number of suppliers and a 260%

reduction in the value of inventory held due to an SPM system (Cormican & Cunningham, 2007). Another study (Limberakis, 2011) found that "best-in-class" buyers (1) are much more likely to benchmark supplier performance against others in the same industry, (2) achieved substantially higher percent on-time delivery (88% versus 48% for "laggards"), and (3) transacted with suppliers that experienced fewer catastrophic failure (2% versus 5% for other buyers). Of the best-in-class buyers, 63% had a supplier benchmarking and performance monitoring information technology system in place. The use of an SPM system was also found to improve buyer–supplier relationships (Prahinski & Benton, 2004). Prahinski and Fan (2007) found that the frequency and content of feedback increase the suppliers' commitment to the buyer, which, in turn, increases supplier performance. Denali Consulting group found that SPM can yield a 3% to 6% cost reduction in total supply chain costs via continuous improvements (Minahan, 2007). A study by CAPS Research (Monczka et al., 2011a) of eight firms found that supplier performance measurement is one of five critical components of effective supplier relationship management (SRM), and that SRM enables vast positive results such as the following: overhead cost reductions, process improvements, increased visibility into actual costs (versus price), year-over-year cost reductions, millions of dollars in savings, product launches on time and on cost, shorter new product development times, total cost reductions of 12%, and quality improvements. Not surprisingly, SPM is a core competence of chief procurement officers (Kern et al., 2011).

Most SPM processes used by buyers integrate subjective and objective evaluations (Simpson et al., 2002; Hald & Ellegaard, 2011). It is assumed that these assessments are accurate; however, as Gordon (2008) pointed out, even the seemingly most-objective performance parameters, such as percent on-time delivery, can be subjective. The supplier evaluation process has rarely been examined, and social and organizational biases have been ignored (Purdy & Safayeni, 2000). Hald and Ellegaard (2011) found that supplier evaluations are shaped and reshaped throughout the evaluation process. They discovered performance data instability as captured in ERP databases. They also found that evaluations were derived by condensing a larger set of performance information to a smaller, more manageable set of numbers. Buyers also commonly use multiple evaluators to rate supplier performance (Buffa & Ross, 2011; Hald & Ellegaard, 2011). Buffa and Ross (2011) noted the importance of supplier evaluation by functionally heterogeneous evaluation teams. Subjective measures among multiple raters invite dissonance in ratings and opinions—either on the same performance observations or across different instances of performance (Buffa & Ross, 2011). Similarly, Perkins (1993) noted that the different members of the buying organization's procurement team perceive the supplier's value delivery differently. While Buffa and Ross (2011) offered an ex post means to accommodate variance among multiple evaluators, there remains little explanation as to systemic sources of the variance. Hence, are there factors that can be managed to mitigate performance evaluators' dissonance? Additionally, the degree of internal dissonance of supplier evaluations has not yet been examined. Hald and Ellegaard (2011) also reported that performance ratings are sometimes negotiated with suppliers when the accuracy is challenged. However, no research has explored why buyers decide to change their evaluations.

Given the above findings, the focal outcome of interest of this study is SPE Efficacy—defined herein as the extent to which SPEs achieve the two stated goals of motivating supplier performance and, during source selection, mitigating the risk of unsuccessful performance (i.e., avoid adverse selection). The ensuing review of the relevant literature identifies the central factors affecting SPE efficacy, then peels the onion back further to unveil their antecedents.

Agency Theory

This research acknowledges multiple perspectives of agency theory as it applies to industrial exchange. The first perspective views the hired supplier as an agent to the buyer to achieve the buyer's objectives. The second perspective examines the buyer internally acknowledging that the buyer is comprised of multiple agents to itself. For instance, employees working in procurement, logistics, financial management, engineering, end users of suppliers' goods and services, and program management represent distinct interests within the firm. Agency theory wrestles with two problems: (1) conflicting interests between principal and agent and (2) difficulty and cost associated with monitoring agents, and the associated uncertainty for not having perfect information (Eisenhardt, 1989).

Beginning with the second perspective, using multiple raters within an organization to evaluate supplier performance can create conflicts of agency. In the case of past performance evaluations, evaluators of performance serve as agents to multiple principals—their employing organization, their local organization or unit, and external stakeholders (e.g., shareholders or taxpayers in the public sector). Problems of agency arise when agents' self-interests differ from his or her employer's goals (Bergen et al., 1992). Agency theory holds that once the principal delegates tasks to agents, there is an asymmetry in information and knowledge such that agents can shirk duties, distort information, and behave opportunistically. To combat these moral hazards, principals can increase monitoring of agents. A less costly approach to control agent opportunism is to align the goals of the agent to that of the principal, particularly using outcome-based contracts (Eisenhardt, 1989). Ex ante, principals can screen potential agents to mitigate adverse selection.

Problems may also emerge when agents must serve conflicting goals of multiple principals—also known as the "hydra factor" (Shapiro, 2005). In this case, the strategy of aligning agents' interests with organizational goals is confounded by conflicting goals. This agency problem might manifest itself in weapon system acquisition when, for instance, a program plagued by technical difficulty is jeopardized if behind schedule or over budget (threat to taxpayers' interest). Such a program could compromise the ability to deliver a system that meets end user needs (threat to end user). Additionally, jobs that are dependent on this program could be jeopardized (threat to program executive officer's and Congress' interest). In this case, an evaluator could be biased toward a favorable SPE in order to protect the supplier and the program from scrutiny. This is an area ripe for further research (Shapiro, 2005).

In agency theory, large organizations of many people and sub-organizations are assumed to act as one homogeneous entity. This is criticized as "misplaced methodological individualism" (Worsham et al., 1997, p. 423). In addition to multiple principals to serve, there may be multiple evaluators (Shapiro, 2005)—particularly on large, complex contracts and where performance occurs in more than one location. In cases of inter-rater disagreement, how is the principle's rating of a supplier (agent) derived? Given these problems of agency, rating dissonance is among the central constructs of this study. The variance in ratings due to multiple evaluators of supplier performance is referred to herein as rating dissonance.

Organizational Behavior

Contract performance often is a complex phenomenon to assess. It can involve many supplier personnel, many buyer evaluators (Wieters & Ostrom, 1979; Palmatier, 2008), multiple internal stakeholders and organizations, and multiple performance criteria at many physical locations. Often, the stakes are high such as implications to profit and future business.

Findings from organizational behavior literature are germane. Academic literature on multiple-rater performance appraisal systems (e.g., 360-degree evaluations in which superiors, subordinates, and peers evaluate the ratee) has examined the underlying premise that more raters offer more unique, valuable information about the employee's performance that would otherwise be lost if relying upon a single rater (van der Heijden & Nijhof, 2004). Additionally, more raters mitigate evaluation bias (Levy et al., 1998). While relying upon multiple ratings is thought to offer more fairness to ratees, variance in ratings is introduced attributable to individual differences in raters (Mount et al., 1998). Thus, different raters often conclude different ratings (Dowst, 1972; Levy et al., 1998), which may be attributed to different backgrounds, observing different instances of supplier performance, and different interpretations of the meaning of performance critieria and rating definitions. These differences take time and effort to resolve and internally agree upon a single rating or narrative.

Multiple raters may be indicators of complexity (e.g., multiple points of failure and multiple locations). Suppliers may be able to more successfully rebut ratings under high complexity. Suppliers may also be more able to offset relatively minor failures with their successes, garnering an overall rating that is acceptable to the supplier. If a supplier can "escape" unscathed in the rating (i.e., no threat), there is little need to increase performance, and little threat of negative performance information being discovered during a future source selection. Given the potential for unreconciled dissonance, it is posited that

> H1: There will be a negative relationship between rating dissonance and SPE efficacy.
>
> H2: Rating dissonance will be positively related to the number of hours to complete the SPE.
>
> H3: The lower the accuracy, the greater the number of hours to complete the SPE.

In federal government contracting, suppliers are provided the SPE ratings and given an opportunity to respond, rebut, agree and otherwise comment. Resolution takes effort expended to explain original positions internally and to seek the facts substantiating the ratings. Thus, supplier disputes, while allowed, are not necessarily welcomed. This phenomenon is not unique to government contracting; suppliers to for-profit businesses may have executive-level relationships within the buying organization and may use those communication channels to voice disagreement with SPEs. Herein, this phenomenon is defined as fear of a supplier dispute. Attempts among multiple raters to thwart a supplier rebuttal may invite internal conflict. Some evaluators may be inclined to inflate ratings to avoid a dispute, while others may take a legalistic, strict approach. If inflated, accuracy suffers. Given the above logic, it is hypothesized that

> H4: The lower the perceived accuracy, the greater the fear of supplier dispute.
>
> H5: There will be a positive relationship between fear of supplier dispute and rating dissonance.

Performance ratings are also constrained by information flow between a rater and ratee. Informational constraints implies that some self/supervisor discrepancies result from differing cognitions about job requirements. When performing any job, an employee must consider what tasks are to be done, how these tasks are to be performed, and what standards are to be used in judging the final outcome. Ideally, these determinations are arrived at in close consultation with the individual's supervisor, thus ensuring identical cognitions about job requirements. In reality, such complete agreement is rarely achieved. The extensive literature on role ambiguity (e.g., House & Rizzo, 1972; Jackson & Schuler, 1985; Rizzo et al., 1970) provides strong evidence that employees often do not have a clear idea of what their supervisors expect (Campbell & Lee, 1988, p. 304).

These findings are particularly relevant in service contracts in which requirements are often not well defined (van der Valk & Rozemeijer, 2009). Different expectations among different performance evaluators of contractor requirements can affect performance evaluations.

Informational constraints can also stem from a supervisor's misunderstanding of the employee's job (Mitchell, 1983). Managers who are recruited from outside the company may have incomplete or inaccurate beliefs about a subordinate's job. Similarly, in situations in which jobs are highly interconnected and interdependent, a supervisor either may be unable to clearly separate the boundaries and duties of different jobs or may do so incorrectly (Kiggundu, 1981). A supervisor's misunderstanding of a subordinate's job also may reflect lack of observation (e.g., Mitchell, 1983). This has implications for a proper amount and method of monitoring suppliers. Insufficient observation can be attributed to the number of other responsibilities a manager has to the inherent nature of one's job. "Thus, it is not surprising that employees and supervisors may come to different conclusions about the employee's effectiveness. If initial cognitions about job responsibilities and standards differ, lack of agreement in ratings is inevitable" (Campbell & Lee, 1988, p. 305). Given that in contracting for services, requirements are often ill defined and given the high level of turnover in buyer-side contract administration (Hawkins et al., 2015), dissonance in supplier performance ratings should be commonplace. Buffa and Ross (2011) identified evaluator turnover as having a potential impact on supplier evaluations over time. Therefore, is it posited that

> H6: There will be a negative relationship between the sufficiency of the requirement definition and rating dissonance.
>
> H7: There will be a positive relationship between the sufficiency of the requirement definition and perceived accuracy of the SPE.
>
> H8: There will be a negative relationship between evaluator turnover and perceived accuracy of the SPE.

Sometimes the employee or the supervisor knowingly gives an inaccurate appraisal. A supervisor may do so to preserve the effectiveness of an interdependent work group (Campbell & Lee, 1988). Academic literature confirms a halo effect in employee performance appraisals (Thomas & Bretz, 1994). The same concern has specifically been raised regarding SPEs (Kelman, 2010). A halo effect could partially explain inflated (i.e., inaccurate) SPEs. Deliberate dishonesty is more likely to occur in self appraisals when they are used for scarce resource allocation decisions (Shrauger & Osberg, 1981). In a supplier relationship context, supplier evaluations may also be tainted by a supplier seeking to preserve its reputation. Suppliers may refute any negative information being recorded regardless of its accuracy. To do so, they often challenge the rating and/or justification, which causes more effort by the buying organization to resolve disagreements. If buying organizations either can't muster the evidence to justify a particular rating and/or consciously decide not to bother with the trouble to debate the rating, accuracy can suffer. Thus, it is hypothesized that

> H9: There will be a negative relationship between perceived accuracy and rating dissonance.

H10: There will be a positive relationship between perceived accuracy and SPE efficacy.

Channel Communication

In channel communication theory, Mohr and Sohi (1995) introduced "distortion." Formality decreases communication distortion. Examining the government's past performance reporting system (CPARS), the reporting is quite rigid and formal. However, the collaboration between multiple raters occurs outside of the CPAR system (i.e., not formal and highly variable). In examining channel communication, often three aspects of communication are explored—formality, bi-directionality, and frequency. If these three facets of communication among exchange members increases, more information is shared, better understandings are attained, and therefore, the accuracy of SPEs should increase. Therefore, it is posited that

H11: There will be a positive relationship between communication frequency and perceived accuracy.

H12: There will be a positive relationship between communication bi-directionality and perceived accuracy.

H13: There will be a positive relationship between communication formality and perceived accuracy.

Weaknesses in evaluators' communications could be linked to resource constraints. Government acquisition personnel are often overworked and understaffed. Combined, this phenomenon is referred to as role overload. Evaluators may simply not have sufficient time to gather the requisite facts and write thorough, sufficient justifications for SPE assessments and ratings. Likewise, evaluators may not have time to reconcile rating dissonance among multiple evaluators. Therefore, it is posited that

H14: There is a negative relationship between role overload and rating justification.

H15: There is a positive relationship between role overload and rating dissonance.

Critics contend that SPEs are often not accurate, and therefore the SPE system (e.g., CPARS) is not useful. If not factual and detailed, the SPEs cannot motivate suppliers to work harder and cannot provide insights that reduce the risk of adverse selection in the future. Hence, absent accuracy, SPEs become less useful. Further, if the SPE scheme is not useful, evaluators will not put forth the effort required to develop a detailed, factual rating justification that will be accepted by the supplier and, if rebutted, internally by the reviewing official., Thus, it is posited that

H16: There is a positive relationship between perceived usefulness and rating justification.

H17: There is a positive relationship between perceived accuracy and rating justification.

H18: There will be a positive relationship between rating justification and SPE efficacy.

H19: There will be a positive relationship between perceived accuracy and perceived usefulness.

Social Exchange Theory

Social exchange theory (SET) is commonly used as a foundation for relationship marketing and buyer–seller relationships (e.g., Dwyer et al., 1987; Kingshott, 2006; Luo, 2002; Morgan & Hunt, 1994; Wilson, 1995). The foundational premises of SET may be summarized as follows. Exchange may involve both social and economic outcomes. These outcomes are compared to other exchange alternatives. Positive outcomes increase trust and commitment and, over time, norms develop that govern the relationship (Lambe, Wittmann, & Spekman 2001). Thus, SET rejects the assumption of universal opportunism and suggests that there is an alternate form of governance—the relationship. Parties to relational exchange, therefore, tend to rely more on trust, commitment, cooperation, satisfaction, and relational norms than strictly on written contracts (Heide & John, 1992). Contracts are incomplete, and can be costly and inefficient to administer as their details increase. Relational exchange renders the exchange more efficient.

Relational aspects have also been found to play a mediating role between suppliers' operational performance measures and a buyer's business performance. Hence, measuring performance alone does not affect business performance. Rather, measuring supplier performance increases socialization mechanisms, which, in turn, increase business performance (Cousins, Lawson, & Squire, 2008). Socialization mechanisms are structures and processes that facilitate contact between the buyers and suppliers such as cross-functional teams, joint sessions, routine supplier conferences, and matrix reporting structures. These interactions enable each party to acquire knowledge of the others' social values and behavioral norms. Interactions entail communications. Communication increases trust (Morgan & Hunt, 1994), a central construct to effective relational exchange.

Research that developed a taxonomy of buyer–supplier relationship types (Cannon & Perreault, 1999) associated higher supplier performance evaluations to more collaborative types of relationships. Such relationships are characterized by greater operational linkages, information exchanges, cooperative norms, and buyer and supplier adaptations to each other (i.e., unique investment and customizations to processes and products for the other party's benefit). With greater channel cooperation, both intra-firm and extra-firm, it is posited that

> H20: There will be a negative relationship between relationship quality and fear of a supplier dispute.
>
> H21: Communication frequency will be positively related to relationship quality.
>
> H22: Communication bi-directionality will be positively related to relationship quality.
>
> H23: There will be a positive relationship between communication formality and relationship quality.
>
> H24: Turnover will be negatively related to relationship quality.

Returning to agency theory, much is said in the management, marketing, and supply chain literatures about supplier monitoring. Since increasing information via monitoring reduces uncertainty and helps prevent agent opportunism, monitoring plays an important role in exchange relationships. As it pertains to SPEs, surveillance is used to collect facts of supplier performance such as quality levels delivered, on-time performance, and generally meeting contractual requirements. These facts may be used to determine performance ratings. Therefore, it is posited that

H25: There will be a positive relationship between surveillance and perceived accuracy.

One relational norm important to effective exchange is fairness (Kumar et al., 1995). Often the concept is referred to as distributive justice, referring to the extent to which each exchange member's cost-benefit ratios are approximately equal. Government buyers in particular have a duty to treat suppliers fairly. In the for-profit sector, fair treatment of suppliers is paramount to effective relationship quality (Kumar et al., 1995). In a SPE context, fairness pertains to the extent to which the supplier is given the performance ratings it deserves. Fair ratings are those that have been earned, no more and no less. Particularly in cases in which requirements are not well defined, the criteria for evaluating supplier performance are not well defined, and/or the ratings used to assess performance are not well defined (or invite wide latitude in interpretation), a supplier must rely on the buyer to be fair. A deviation from a fair rating would insinuate a rating that is not right—or less than accurate.

H26: There will be a positive relationship between fairness and perceived accuracy.

Power/Dependence

Power is among the most significant phenomena in buyer–supplier relationships. It is defined as the ability to cause someone to do something that he or she would not have done otherwise (Gaski, 1984). Power and dependence are two sides of the same coin (John, 1984). In government contracting, extremely high switching costs create dependence of buyers on suppliers after the award of a contract. Additionally, sole source contracts are commonplace which gives rise to buyer dependence (and supplier power). In such cases, particularly when the buyer is less than diligent in its contract administration duties and oversight, buyers may be tempted to use SPEs as leverage to reap concessions from suppliers. In cases where ratings are subtly bargained for some concession, the accuracy of SPEs could be questioned. Therefore, it is posited that

H27: Leverage attitude will be negatively related to perceived accuracy.

Methodology

This research will employ quantitative and qualitative methodologies to examine the antecedents and consequences of supplier performance evaluation efficacy (Table 1). First, the quantitative methodology and results are detailed, then the qualitative procedures and results are described.

Table 1. Research Questions

No.	Research Question	*Research Object	**Research Method
1	What factors decrease the efficacy of SPEs?	B & S	Qt & Ql
2	How do suppliers react to inaccurate SPEs?	S	Ql
3	Do SPEs, in general, motivate suppliers to increase performance?	S	Ql
4	How does the accuracy of SPEs affect relationship quality?	B & S	Qt & Ql
5	Why are SPEs often inaccurate?	B & S	Qt & Ql
6	How many man-hours do suppliers invest in responding to SPEs?	S	Ql
7	What communication tactics do suppliers use to manage the SPE process?	S	Ql
8	To what extent does inter-rater disagreement (i.e., dissonance) affect SPE efficacy?	B	Qt

*B=buyer; S=supplier
**Qt=Quantitative; Ql=Qualitative

The quantitative method will examine data collected via survey of the personnel with the requisite knowledge of contractor performance, CPARS assessing officials. The hypotheses will be tested using partial least squares (PLS) structural equation modeling (SEM). PLS SEM, versus covariance-based SEM, is the valid modeling approach when the model includes formative scales (Hair et al., 2014). PLS SEM also accommodates complex models with a large number of variables, can model non-normally distributed data, and does not pose problems with convergence often found in covariance-based SEM.

The qualitative method entails collecting data via interviews with suppliers whose performance has been rated. According to Yin (2009), a qualitative methodology is appropriate when three conditions exist: (1) The type of research question is exploratory in nature and takes the form of a “why” question, (2) the researcher has no control of the behavioral events being researched (i.e., cannot manipulate behaviors then measure results as in a controlled experiment), and (3) the focus is on contemporary events (p. 8). The research questions surrounding supplier reactions to performance evaluations met all three criteria.

Discussion

Substantial transaction costs are dedicated to avoid adverse selection—the risk of selecting an incapable supplier that otherwise misrepresents itself as capable. Following contract formation, more transaction costs are incurred to monitor supplier performance to thwart supplier opportunism ex post. The effectiveness of a mechanism to monitor and record supplier performance information, a supplier performance evaluation, was the topic of this study.

There are many concerns that the SPEs/ratings are not properly, timely, or accurately completed. Unreliable or inaccurate past performance assessments can harm suppliers' reputations and can bias source selections, resulting in adverse selection. If past performance information is not reliable, and if evaluators don't use it in discriminating between competitive proposals, the effort of collecting and reporting the past performance information is squandered. Likewise, the effort of evaluating and documenting inaccurate past performance information during source selections would be wasted. Anecdotal evidence suggests that buying organizations often do not use past performance information as a meaningful discriminator between proposals.

The purpose of the research, therefore, was to explore the antecedents to and consequences of the efficacy of SPEs.

References

Aberdeen Group. (2005, September). *The supplier performance measurement benchmark report.* Boston, MA: Author.

Bergen, M., Dutta, S., & Walker, O. C. (1992). Agency relationships in marketing: A review of the implications and applications of agency and related theories. *Journal of Marketing, 56*(3), 1–24.

Berrios, R. (2006). Government contracts and contractor behavior. *Journal of Business Ethics, 63,* 119–130.

Buffa, F. P., & Ross, A. D. (2011). Measuring the consequences of using diverse supplier evaluation teams: A performance frontier perspective. *Journal of Business Logistics, 32*(1), 55–68.

Campbell, D. J., & Lee, C. (1988). Self-appraisal in performance evaluations: Development versus evaluation. *Academy of Management Review, 13*(2), 302–314.

Cannon, J. P., & Perreault, W. D., Jr. (1999). Buyer–seller relationships in business markets. *Journal of Marketing Research, 36*, 439–460.

CAPS Research. (2011). *Supplier quality & delivery performance; 2011 supply management benchmarking report.* Tempe, AZ: Institute for Supply Management and W.P. Carey School of Business at Arizona State University.

Cormican, K., & Cunningham, M. (2007). Supplier performance evaluation: Lessons from a large multinational organization. *Journal of Manufacturing Technology Management, 18*(4), 352–366.

Cousins, P. D., Lawson, B., & Squire, B. (2008). Performance measurement in strategic buyer–supplier relationships: The mediating role of socialization mechanisms. *International Journal of Operations & Production Management, 28*(3), 238–258.

Dowst, S. (1972, June). How designers rate suppliers—and purchasing. *Purchasing, 12,* 87–93.

Dwyer, F. R., Schurr, P. H., & Oh, S. (1987, April). Developing buyer–seller relationships. *Journal of Marketing, 51,* 11–27.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2012).

Federal Acquisition Streamlining Act (FASA). (1994).

Gaski, J. F. (1984). The theory of power and conflict in channels of distribution. *Journal of Marketing, 48*(3), 9–29.

Gordon, S. R. (2008). *Supplier evaluation and performance excellence: A guide to meaningful metrics and successful results*. Ft. Lauderdale, FL: J. Ross.

GAO. (2009). *Federal contractors: Better performance information needed to support agency contract award decisions* (GAO-09-1032). Washington, DC: Author.

GAO. (2014). *Contractor performance: Actions taken to improve reporting of past performance information* (GAO-14-707). Washington, DC: Author.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014*). A primer on partial least squares structural equation modeling (PLS-SEM).* Los Angeles, CA: Sage.

Hald, K. S., & Ellegaard, C. (2011). Supplier evaluation processes: The shaping and reshaping of supplier performance. *International Journal of Operations and Production Management, 31*(8), 888–910.

Hawkins, T., Berkowitz, D., Muir, W., & Gravier, M. (2015). Improving services supply management in the defense sector: How the procurement process affects B2B service quality. *Journal of Purchasing and Supply Management, 21*, 81–94.

Hawkins, T. (2013). *Exploring the efficacy of the government's current use of past performance information* (NPS-CM-13-125). Monterey, CA: Acquisition Research Program, Naval Postgraduate School.

Heide, J. B., & John, G. (1992). Do norms matter in marketing relationships? *Journal of Marketing, 56*(4), 32–44.

House, R., & Rizzo, J. (1972). Role conflict and ambiguity as critical variables in a model of organizational behavior. *Organizational Behavior and Human Performance, 7*(3), 467–505.

Jackson, S. E., & Schuler, R. S. (1985). A meta-analysis and conceptual critique of research on role ambiguity and role conflict in work settings. *Organizational Behavior and Human Performance, 36,* 16–78.

John, G. (1984). An empirical investigation of some antecedents of opportunism in a marketing channel. *Journal of Marketing Research, 21*(3), 278–289.

Kelman, S. (2010). Three ways to revitalize the use of past-performance data. *Federal Computer Week*. Retrieved from http://fcw.com/blogs/lectern/2010/05/steve-kelman-past-performance-source-selection.aspx

Kern, D., Moser, R., Sundaresan, N., & Hartmann, E. (2011). Purchasing competence: A stakeholder-based framework for chief purchasing officers. *Journal of Business Logistics, 32*(2), 122–138.

Kiggundu, M. N. (1981). Task interdependence and the theory of job design. *Academy of Management Review, 6,* 499–508.

Kingshott, R. P. J. (2006). The impact of psychological contracts upon trust and commitment within supplier–buyer relationships: A social exchange view. *Industrial Marketing Management, 35*(6), 724–739.

Kumar, N., Scheer, L. K., & Steenkamp, J. E. M. (1995). The effects of supplier fairness on vulnerable resellers. *Journal of Marketing Research, 32*(1), 54–65.

Lambe, J. C., Wittmann, C. M., & Spekman, R. E. (2001). Social exchange theory and research on business-to-business relational exchange. *Journal of Business-to-Business Marketing, 8*(3), 1–36.

Levy, P. E., Cawley, B. D., & Foti, R. J. (1998). Reactions to appraisal discrepancies: Performance ratings and attributions. *Journal of Business and Psychology, 12*(4), 437–455.

Limberakis, C. G. (2011). *The year of the supplier*. Boston, MA: Aberdeen Group.

Lord, R. D. (2005). *Contractor past performance information (PPI) in source selection: A comparison study of public and private sector* (NPS-CM-04-019). Retrieved from http://www.acquisitionresearch.net

Luo, X. (2002). Trust production and privacy concerns on the Internet: A framework based on relationship marketing and social exchange theory. *Industrial Marketing Management, 31*(2), 111–118.

Mitchell, T. R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal., In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement & theory* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Mohr, J. J., & Sohi, R. S. (1995). Communication flows in distribution channels: Impact on assessments of communication quality and satisfaction. *Journal of Retailing, 71*(4), 393–416.

Monczka, R. M., Choi, T. Y., Kim, Y., & McDowell, C. (2011a). *Supplier relationship management: An implementation framework.* Tempe, AZ: CAPS Research.

Monczka, R. M., Handfield, R. B., Giunipero, L. C., & Patterson, J. L. (2011b). *Purchasing and supply chain management* (5th ed.). Mason, OH: South-Western Cengage Learning.

Morgan, R. M., & Hunt, S. D. (1994, July). The commitment-trust theory of relationship marketing. *Journal of Marketing, 58*, 20–38.

Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait, rater and level effects in 360-degree performance ratings. *Personnel Psychology, 51*, 557–576.

Naval Seal Logistics Center Portsmouth. (2014, July). *Guidance for the contractor performance assessment reporting system (CPARS).* Retrieved October 27, 2015, from http://dai-global-developments.com/assets/cpars_guidance.pdf

Office of Federal Procurement Policy (OFPP). (2011, January 21*). Improving contractor past performance assessments: Summary of the Office of Federal Procurement Policy's review, and strategies for improvement* [Memorandum]. Washington, DC: Office of Management and Budget.

Office of Federal Procurement Policy (OFPP). (2000, May). *Best practices for collecting and using current and past performance information*. Washington, DC: Office of Management and Budget.

Palmatier, R. W. (2008, July). Interfirm relational drivers of customer value*. Journal of Marketing, 72*(4), 76–89.

Perkins, W. S. (1993). Measuring customer satisfaction*. Industrial Marketing Management, 22*(3), 247–254.

Prahinski, C., & Benton, W. C. (2004). Supplier evaluations: Communication strategies to improve supplier performance. *Journal of Operations Management, 22*, 39–62.

Prahinski, C., & Fan, Y. (2007). Supplier evaluations: The role of communication quality. *Journal of Supply Chain Management*, *43*(3), 16–28.

Purdy, L., & Safayeni, F. (2000). Strategies for supplier evaluation: A framework for potential advantages and limitations. *IEEE Transactions on Engineering Management, 47*(4), 435–443.

Rizzo, J., House, R., & Lirtzman, S. (1970). Role conflict and ambiguity in complex organizations. *Administrative Science Quarterly, 15*, 150–163.

Schmitz, J., & Platts, K. W. (2003). Roles of supplier performance measurement: Indication from a study in the automotive industry. *Management Decision, 41*(8), 711–721.

Shapiro, S. P. (2005). Agency theory. *Annual Review of Sociology, 31*, 263–284.

Shrauger, J., & Osberg, T. (1981). The relative accuracy of self-predictions and judgments by others in psychological assessment. *Psychological Bulletin, 90*, 322–351.

Talluri, S., & Sarkis, J. (2002). A model for performance monitoring of suppliers. *International Journal of Production Research, 40*(16), 4257–4269.

Thomas, S. L., & Bretz, R. D. (1994). Research and practice in performance appraisal: Evaluating employee performance in America's largest companies. *SAM Advanced Management Journal, 59*(2), 28–34.

Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). (2015, April 9). *Implementation directive For Better Buying Power 3.0—Achieving dominant capabilities through technial excellence and innovation* [Memorandum]. Washington, DC: Author.

van der Heijden, B. I. J. M., & Nijhof, A. H. J. (2004). The value of subjectivity: Problems and prospects for 360-degree appraisal systems. *International Journal of Human Resource Management, 15*(3), 493–511.

Van der Valk, W., & Rozemeijer, F. (2009). Buying business services: Towards a structured service purchasing process. *Journal of Services Marketing, 23*(1), 3–10.

Wieters, D. C., & Ostrom, L. L. (1979). Supplier evaluation as a new marketing tool. *Industrial Marketing Management, 8*(2), 161–166.

Wilson, D. T. (1995). An integrated model of buyer–seller relationships. *Journal of the Academy of Marketing Science, 23*(4), 335–345.

Worsham, J., Eisner, M. A., & Ringquist, E. J. (1997). Assessing the assumptions: A critical analysis of agency theory. *Administration and Society, 28*(4), 419–440.

Yin, R. K. (2009). *Case study research: Design and methods*. Los Angeles, CA: Sage.

Acknowledgments

The author would like to thank the Acquisition Research Program for providing funding support for this research.

Panel 16. Improving Governance of Complex Systems Acquisition

Thursday, May 5, 2016	
11:15 a.m. – 12:45 p.m.	**Chair: Rear Admiral David Gale, USN,** Program Executive Officer, SHIPS ***Complex System Governance for Acquisition*** Joseph Bradley, President, Leading Change, LLC Polinpapilinho Katina, Postdoctoral Researcher, Old Dominion University Charles Keating, Professor, Old Dominion University ***Acquisition Program Teamwork and Performance Seen Anew: Exposing the Interplay of Architecture and Behaviors in Complex Defense Programs*** Eric Rebentisch, Research Associate, MIT Bryan Moser, Lecturer, MIT John Dickmann, Vice President, Sonalysts Inc. ***A Complex Systems Perspective of Risk Mitigation and Modeling in Development and Acquisition Programs*** Roshanak Rose Nilchiani, Associate Professor, Stevens Institute of Technology Antonio Pugliese, PhD Student, Stevens Institute of Technology

Rear Admiral David Gale, USN—is currently assigned as Program Executive Officer (PEO), Ships, where he is responsible for Navy shipbuilding for surface combatants, amphibious ships, logistics support ships, support craft, and related foreign military sales. PEO Ships currently has more than 17 ships under construction and an additional 24 ships and craft under contract.

RADM Gale, a native of Lebanon, NY, enlisted in the Navy in 1976. After service aboard USS *Dupont* (DD 941), he was selected for the Broadened Opportunity for Officer Selection and Training and was commissioned in May 1983. At sea, Gale's assignments included auxiliaries officer in the pre-commissioning crew of USS *Rentz* (FFG 46), engineer officer in the pre-commissioning crew of USS *Chosin* (CG 65), and executive officer on USS *Thomas S. Gates* (CG 51). He was the first commanding officer of the USS *Mason* (DDG 87).

Ashore, he served on the Commander-in-Chief, U.S. Pacific Fleet Propulsion Examining Board, in the Aegis Program Office as the DDG 51 Fleet Introduction officer and surface combatant readiness officer, and as the executive assistant to the program executive officer for Surface Combatants/Aegis Program. He was the Major Program manager for Fleet Introduction and Surface Combatant Lifetime Support (400F) in the Program Executive Office for ships; the military assistant to the Under Secretary of Defense for Acquisition, Technology, and Logistics; and the acting executive director to NAVSEA 21.

After selection to Flag rank in April 2010, Gale assumed duties as Commander, Navy Regional Maintenance Command, where he oversaw the operations and management of the Navy's Regional Maintenance Centers in their execution of surface ship maintenance and modernization. On October 1, 2013, RADM Gale became the Deputy Commander for Surface Warfare, Naval Sea Systems Command (NAVSEA 21).

Gale earned a Bachelor of Arts degree in economics from the University of New Mexico and a master's degree in national resources strategy from the Industrial College of the Armed Forces. He is a graduate of the United States Marine Corps Command and Staff College. RADM Gale's personal awards and decorations include the Legion of Merit, the Defense Meritorious Service Medal, and the Meritorious Service Medal.

Complex System Governance for Acquisition

Joseph Bradley —received his PhD in engineering management and systems engineering at Old Dominion University (ODU) in Norfolk, VA. He holds the degrees of Professional Engineer and Masters of Science in mechanical engineering from the Naval Postgraduate School and Bachelor of Engineering from The Cooper Union. He is currently president of his own small consulting firm working with clients in government and industry. His areas of research include complex system governance, systems theory, competency models, and performance measurement systems. His research has been published in the *Systems Engineering, Naval Engineers Journal*, and *International Journal of System of Systems Engineering*. [josephbradley@leading-change.org]

Additional Authors:

Polinpapilinho Katina—received his PhD in the Department of Engineering Management and Systems Engineering at Old Dominion University (ODU) in Norfolk, VA. He holds an MEng in systems engineering (2011) and a BSc in engineering technology (2009). He is currently a postdoctoral fellow at the National Centers for System of Systems Engineering (NCSOSE). He recently co-edited a critical textbook on Infranomics. His research has been published in the *International Journal of Critical Infrastructure Protection, International Journal of Critical Infrastructures, International Journal of Decision Sciences, Risk and Management, International Journal of System of Systems Engineering,* and *Journal of Requirements Engineering*. [pkatina@odu.edu]

Charles Keating—is a Professor of Engineering Management and Systems Engineering and Director of the National Centers for System of Systems Engineering (NCSOSE) at Old Dominion University (ODU) in Norfolk, VA. He received a BS in engineering from the United States Military Academy (West Point), an MA in management from Central Michigan University, and his PhD in engineering management from Old Dominion University. His current research focus is on complex system governance, system of systems engineering, and management cybernetics. [ckeating@odu.edu]

Abstract

As acquisition processes have become more complex, they appear to no longer be governable by traditional approaches. Missed budgets, delayed deliveries, and expensive canceled systems appear to becoming more prevalent. Numerous investigations have been conducted attempting to elicit the factors that prevented success. Those systems that succeed in terms of usability, budget, and delivery schedule are the rarity and often become case studies themselves as we try to extract the characteristics that differentiate success from failure. A different viewpoint is to look at the acquisition system from the perspective of Complex Systems Governance (CSG). Recent developments in the field of CSG are poised to offer insights into the domain of complex system acquisition. CSG, an emerging field grounded in Management Cybernetics and System Theory, offers a set of nine essential and interrelated functions that enable effective governance—which includes acquisition.

In this paper, after an introduction of our perception of the problem space, we outline the nine essential meta functions and briefly describe the inter-relationships that form a coherent governance scaffold. An exposition of the corresponding CSG reference model is then profiled. We then examine how the meta functions can be applied to acquisition, using the CSG reference model as the framing for an effective governance system. Finally, we offer suggestions and contributions offered by a research thrust in CSG to examine acquisition in a live case setting with implications for the wider acquisition field.

Introduction

As acquisition processes have become more complex, they appear to no longer be governable by traditional approaches. Missed budgets, delayed deliveries, and expensive canceled systems appear to be becoming more prevalent. Numerous investigations have been conducted attempting to elicit the underlying factors that prevented success (Berteau, Levy, Ben-Ari, & Moore, 2011; Francis, 2008, 2009; Rascona, Barkakati, & Solis, 2008). Those systems that succeed in terms of usability, budget and delivery schedule are the rarity and often become case studies themselves as we try to extract the characteristics that differentiate success from failure (Boudreau, 2007; O'Rourke, 2014). Unfortunately, to date there is not a resolution to the problems that delineate acquisition of major systems. Rather than rehash prior approaches or viewpoints, complex system governance (CSG) is offered as an alternative perspective to look at the acquisition system. The hope is that this alternative perspective might provide new insights to an all too familiar problem domain. CSG is an emerging field grounded in Management Cybernetics and Systems Theory. CSG has posited nine meta functions required for effective governance, which will be briefly examined in the next section.

The problems facing practitioners dealing with modern complex systems appear to be intractable. These problems continue to proliferate into all aspects of human endeavor and the systems designed to orchestrate those endeavors. They are not the privilege, or curse, of any particular field or sector (energy, utilities, healthcare, transportation, commerce, defense, security, acquisition, services), as none are immune to the effects of this problem domain. Problems stemming from this domain do not have a precise cause–effect relationship that would make understanding and resolution easy or reducible to the precision demanded by mathematical applications. Arguably, complex systems and their associated problems have been in existence as long as man has been designing, acquiring, operating, and maintaining systems. However, the landscape for modern systems has changed appreciably into a much more "complex problem space." We have previously offered Figure 1 as a visual representation of this problem space (Keating, Katina, & Bradley, 2015) and noted how it (Figure 1) is marked by difficulties encountered across the holistic range of technical, organizational, managerial, human, social, information, political, and policy issues. The different aspects of this "new normal" complex problem space has been previously established (Jaradat & Keating, 2014; Keating, 2014; Keating & Katina, 2011; Naphade et al., 2011) as being characterized by conditions identified in Figure 1. To practitioners of complex systems, this listing is likely recognizable and represents nothing that is not or has not been faced on a routine basis with varying results.

Figure 1. The Complex System Problem Domain Characteristics

This listing in Figure 1 is not presented as exhaustive, but rather it illustrates two important points. First, the issues emanating from this domain continue without consistent resolution methods, thus leaving the door open for new thinking and approaches to address this domain. Second, the conditions identified are not likely to recede in the future, but are more likely representative of the "new normal" for the practitioners dealing with complex systems. As a summary of this domain, we suggest that it is marked by the following five characteristics:

- Uncertainty—incomplete knowledge casting doubt for decision/action consequences
- Ambiguity—lack of clarity in interpretation
- Emergence—unpredictable events and system behaviors
- Complexity—systems so intricate that complete understanding is not possible
- Interdependence—mutual influence among related elements

These conditions are not going away. To ignore them is shortsighted, leaving practitioners (owners, operators, performers, designers) of systems in a precarious position. These conditions are certainly not isolated for complex systems of any particular system or sector, but are rather endemic to complex systems in general. As an illustrative example, we can examine the defense acquisition sector to demonstrate the pervasive nature of the complex system problem domain. Figure 2 is a compilation of challenges facing the defense acquisition sector compiled from several sources (Fauser, 2006; Gansler & Lucyshyn, 2015; Kadish et al., 2006; Mills & Goldsmith, 2014). As evident from the circumstances marking the defense acquisition sector, we can certainly extrapolate those to the complex system problem domain we have established (Figure 1). In addition, we can also project the majority to a wider array of enterprises, sectors, and systems facing similar circumstances.

Effectiveness in dealing with these problem domains beckons for individuals and organizations capable of engaging in a different level of thinking, decision, action, and interpretation to produce alternative paths forward. As one response, CSG is proposed as an emerging field to enable practitioners to build capabilities to better diagnose and effectively respond to deeper level systemic issues that impede system performance (von Bertalanffy, 1950; Skyttner, 2005; Whitney et al., 2015). Thus, CSG seeks to identify and “design through” fundamental system issues such as those identified earlier (Figure 1). Unfortunately, these issues exist at deep tacit levels and appear only as symptomatic at the surface. Thus, efforts to address the problems at the surface level, although providing temporary “fixes,” continually fail to resolve the deeper fundamental system issues. This deeper fundamental system level resolution is necessary to preclude recurrence of the symptomatic issue in another superficial form. For instance, a deep fundamental system issue may appear in one system acquisition program as a budget overrun. However, in another acquisition program, the same fundamental underlying system flaw may manifest itself as a major schedule problem. In both instances, addressing the issues at the surface may provide “temporary” relief but not make the necessary deep system “fix” necessary to preclude future occurrences, albeit in different forms. Exploration and insight at this deep system level is where CSG is targeted to operate with an emphasis on elements of *integration* (continuous maintenance of system integrity), *coordination* (providing for interactions between a system, its entities, and the environment), *communication* (accounting for flow and interpretation of information), and *control* (proving minimal constraints necessary to maintain system performance while maximizing entity autonomy).

Figure 2. Challenges Facing the Defense Acquisition Sector

The purpose of this paper is to examine the challenges and practice implications for CSG. To fulfill this purpose, CSG is developed against the backdrop of the complex system problem domain established above. The remainder of the paper is organized to

1. Provide a brief outline of the nine meta functions required for CSG, and the corresponding CSG reference model, focusing on the responsiveness of this field to enhance effectiveness in dealing with the problems of complex systems.
2. Examine some recent challenges in the defense acquisition field from the CSG perspective.
3. Explore the potential of the CSG field for improving defense acquisition capabilities to more effectively engage the complex system problem domain.

The Nine Meta Functions and the Reference Model for Complex System Governance

A quick appraisal of the situation for dealing with complex systems and their constituent problems appears as dismal as the science of economics. However, CSG is developing as a conceptually grounded field that can provide insights and a fruitful path forward. In this section, we develop a detailed explanation of CSG as "Design, execution, and evolution of the metasystem functions necessary to provide control, communication, coordination, and integration of a complex system" (Keating, 2014, p. 274). The conceptual foundations of CSG are primarily based in Systems Theory (von Bertalanffy, 1968; Skyttner, 2005; Whitney et al., 2015) and Management Cybernetics (Beer, 1972, 1979, 1985) and the field has been built upon their philosophical, theoretical, and methodological underpinnings. Systems Theory has been described as a set of axioms and propositions that define the function of any system (Whitney et al., 2015), while Management Cybernetics has been identified as the science of effective (system) organization (Beer, 1972). Following from the conceptual underpinnings of Systems Theory and Management Cybernetics, the following elements of the CSG definition are elaborated as an essential foundation:

- ***Design***—purposeful and deliberate arrangement of the governance system to achieve desirable performance and behavior.
- ***Execution***—performance of the system design within the unique system context, subject to emergent conditions stemming from interactions within the system and between the system and environment.
- ***Evolution***—the change of the governance system over time in response to internal and external shifts as well as revised trajectory.
- ***Metasystem***—the set of nine interrelated higher level functions that provide for governance of a complex system.
- ***Control***—invoking the minimal constraints necessary to ensure desirable levels of performance and maintenance of system trajectory, in the midst of internally or externally generated perturbations of the system.
- ***Communication***—the flow and processing of information within and external to the system, that provides for consistency in decisions, actions, and interpretations made with respect to the system.
- ***Coordination***—providing for interactions (relationships) between constituent entities within the system, and between the system and external entities, such that unnecessary instabilities are avoided.

- ***Integration***—continuous maintenance of system integrity. This requires a dynamic balance between autonomy of constituent entities and the interdependence of those entities to form a coherent whole. This interdependence produces the system identity (uniqueness) that exists beyond the identities of the individual constituents.
- ***Complex system***—a set of bounded interdependent entities forming a whole in pursuit of a common purpose to produce value beyond that which individual entities are capable (Keating et al., 2015, p. 4).

Foundational to the formulation of CSG is the unique role of the "metasystem." The metasystem construct relies on five essential elements: (1) the metasystem operates at a logical level beyond the elements that it must integrate, (2) the metasystem construct has been conceptually grounded in the foundations of Systems Theory and Management Cybernetics, (3) a metasystem is a set of interrelated functions—which only specify *what* must be achieved for continuing system viability (existence), not *how* those functions are to be achieved, (4) the metasystem functions must be performed if a system is to remain viable—this does not preclude the possibility that a system may be poorly performing, yet still continue to be viable (exist), and (5) a metasystem can be purposefully designed, executed, and maintained, or left to its own (self-organizing) development (Keating et al., 2015, p. 4).

The CSG paradigm can be stated succinctly as follows:

> From a systems theoretic conceptual foundation, a set of nine interrelated functions is enacted through mechanisms. These mechanisms invoke metasystem governance to produce the communication, control, coordination, and integration essential to ensure continued system viability. (Keating et al., 2015, p. 4)

As part of understanding the metasystem and its relationship to the environment, context, and system of interest (Figure 3), the following descriptions are provided to focus our discussion:

- ***Environment***—The aggregate of all surroundings and conditions within which a system operates. It influences and is influenced by a system.
- ***Context***—The circumstances, factors, patterns, conditions, or trends within which a system is embedded. The context acts to constrain or enable the system, including its development, execution, and evolution.
- ***System(s)***—The set of interrelated elements that are subject to immutable system laws and are governed through the metasystem functions to produce that which is of value and consumed external to the system.
- ***Metasystem***—The set of nine functions, which are invoked through mechanisms, to govern a system such that viability (existence) is maintained (Keating et al., 2015, p. 5).

In Keating et al. (2015), we discuss the details of the relationship of these four elements (environment, context, metasystem, system), most of which we omit here for brevity. However, we must note that the separation of the environment, context, system, and metasystem is for convenience and permits analysis. In reality, these four elements exist as an inseparable whole. The separation of these elements always requires judgments. Judgments of boundaries, relevant aspects of the environment, contextual definition, and articulation of the metasystem are always subject to "abstraction error." Therefore, CSG requires purposeful decisions with respect to abstraction of the context, system(s), and metasystem from the environment (Figure 4).

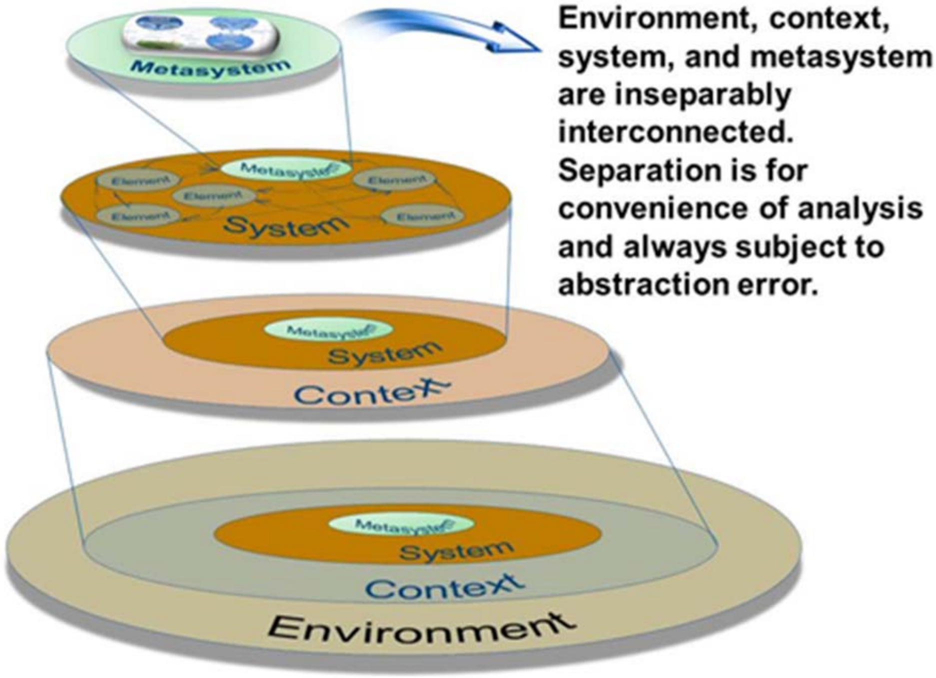


Figure 3. Interconnected Elements of Environment, Context, System, and Metasystem

As noted earlier, the fundamental foundation for CSG is found in Systems Theory and Management Cybernetics, including the philosophical, theoretical, and conceptual underpinnings that serve as a grounding for the field. The metasystem is a construct that defines the set of nine interrelated functions that act to provide governance for a complex system (Figure 4).

The nine metasystem functions included in the metasystem for CSG include the following:

1. ***Policy and Identity*****:** Metasystem Five (M5)—focused on overall steering and trajectory for the system. Maintains identity and defines the balance between current and future focus.
2. ***System Context*****:** Metasystem Five Star (M5*)—focused on the specific context within which the metasystem is embedded. Context is the set of circumstances, factors, conditions, patterns, or trends that enable or constrain execution of the system.
3. ***Strategic System Monitoring*****:** Metasystem Five Prime (M5′)—focused on oversight of the system performance indicators at a strategic level, identifying performance that exceeds or fails to meet established expectations.
4. ***System Development*****:** Metasystem Four (M4)—maintains the models of the current and future system, concentrating on the long range development of the system to ensure future viability.

5. ***Learning and Transformation*****:** Metasystem Four Star (M4*)—focused on facilitation of learning based on correction of design errors in the metasystem functions and planning for transformation of the metasystem.
6. ***Environmental Scanning*****:** Metasystem Four Prime (M4′)—designs, deploys, and monitors sensing of the environment for trends, patterns, or events with implications for both present and future system viability.
7. ***System Operations*****:** Metasystem Three (M3)—focused on the day to day execution of the metasystem to ensure that the overall system maintains established performance levels.
8. ***Operational Performance*****:** Metasystem Three Star (M3*)—monitors system performance to identify and assess aberrant conditions, exceeded thresholds, or anomalies.
9. ***Information and Communications*****:** Metasystem Two (M2)—designs, establishes, and maintains the flow of information and consistent interpretation of exchanges (through communication channels) necessary to execute metasystem functions (Keating & Bradley, 2015).

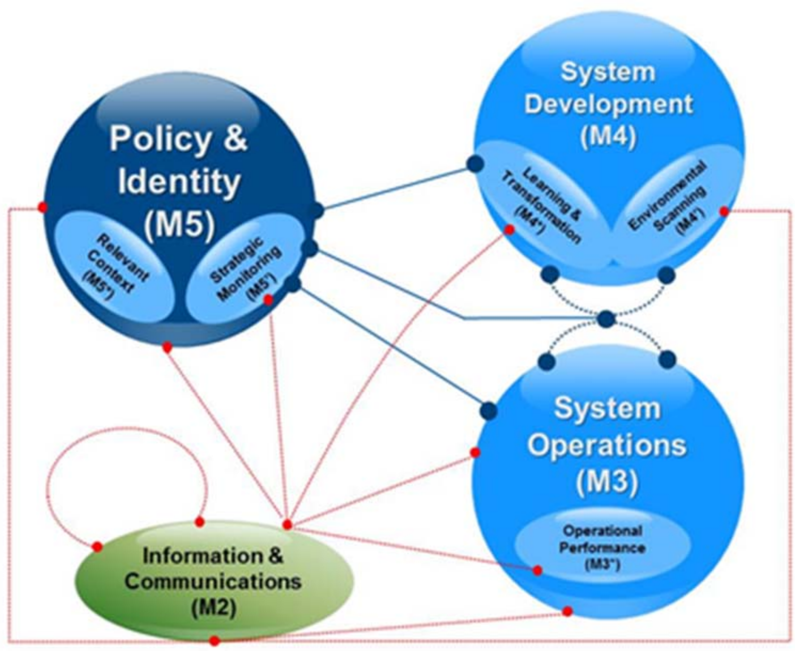


Figure 4. The Nine Interrelated Functions of the Metasystem in CSG

Implementing mechanisms is the final element that forms a CSG triad which also includes *Conceptual Foundations* and *Metasystem Functions*. Conceptual Foundations help to explain and understand *why* systems behave and perform as they do, drawing from the laws and principles of Systems Theory and Management Cybernetics. These laws and principles are immutable and cannot be negotiated away or ignored as if they do not exist. The consequences for violation of the laws are real, carry significant impacts, and are influential in the maintenance of system viability. Ignorance of systems laws will not lessen either their existence or the consequences stemming from their violation. Systems laws and principles operate much as physical principles (e.g., the laws of physics). The laws and principles are (1) omnipresent in explanation of system behavior/performance, (2) cannot be selectively applied or endorsed when convenient, (3) not subject to value judgments regarding applicability, and (4) the principles are value free—meaning that attribution of goodness/badness of the consequences for the performance or nonperformance of a system in accordance with the principles comes from interpretation of the consequence, not the law itself (7, pp. 6–7). Merely naming the nine interrelated metasystem functions with their brief descriptions provides little value, and in fact as a predecessor in the model development, a complete Complex System Reference Model was developed and is highlighted in Table 1. This table is focused on the four primary functions (M2-5) which is inclusive of the subfunctions designated by the prime (′) or star (*) designations (M5′, M5*, M4′, M5*, M3*; Keating & Bradley, 2015).

The Metasystem Functions identify *what* must be achieved to ensure continued system viability. ALL systems must perform these functions at a minimal level to maintain viability. However, viability is not a guarantee of performance excellence, and in fact, we often see performance issues as the system continues to exist. There are degrees of viability, the minimal of which is existence. We turn now to an examination of a selection of high-profile defense acquisitions that might be susceptible to improved outcomes with an advanced understanding of governance.

Table 1. Complex System Governance Reference Model

(Keating & Bradley, 2015)

Metasystem Function	Primary Role	Responsibilities	Implications for Acquisition
Metasystem Five (M5) – Policy and Identity	Primary function is to provide direction, oversight, accountability, and evolution of the System. Focus includes policy, mission, vision, strategic direction, performance, and accountability for the System such that: (1) the System maintains viability, (2) identity is preserved, and (3) the System is effectively projected both internally and externally.	▪ Establishes and maintains system identity in the face of changing environment and context ▪ Defines, clarifies and propagates the system vision, strategic direction, purpose, mission, and interpretation ▪ Active determination and balance for system focus between present and future ▪ Disseminates strategic plan and oversees execution ▪ Provides for capital resources necessary to support system ▪ Sets present and future problem space for focus of product, service, and content development and deployment ▪ Sets strategic dialog forums ▪ Preserves autonomy – integration balance in the system ▪ Marketing of system products, services, content, and value ▪ Public relations planning and execution ▪ External mentorship development (e.g., Board of Directors) ▪ Establishes system policy direction and maintains identity of the system -- executed through strategic direction ▪ Represents the system interests to external constituents ▪ Defines and integrates the expanded network for the system (strategic partnerships) ▪ Evolves scenarios for system transformation and implements strategic transformation direction	▪ Each acquired system has its unique identity that establishes its appropriateness for performing a particular mission, in a particular context, within a particular environment. ▪ Diffusion or ambiguity of system identity can degrade design trade-offs, capabilities, or compromise mission performance ▪ The *acquisition system* also has an identity (tacitly or explicitly known) that guides consistency in thinking, decision, action, and interpretation—ambiguity in this identity is an invitation to inconsistent performance ▪ Acquisition system identity is a reference point that is non-negotiable—if the identity is challenged, the system responds in kind to protect that identity ▪ A weak or muddled acquisition identity fosters inconsistent execution of the system and that which it produces ▪ The articulation, propagation, and maintenance of acquisition identity by active design is a leadership function that cannot be relegated ▪ If the acquisition system is continually disappointing expectations, core identity must be questioned ▪ Acquisition system identity must be compatible with the context, environment and supporting infrastructures (policy, implementing systems, strategies, etc.) ▪ The performance and evolution of the acquisition system must be guided by a strategy for metasystem development that exist beyond traditional system measures in time, depth, and nature

Metasystem Function	Primary Role	Responsibilities	Implications for Acquisition
Metasystem Four (M4) – System Development	Primary function is to provide for the analysis and interpretation of the implications and potential impacts of trends, patterns, and precipitating events in the environment. Develops future scenarios, design alternatives, and future focused planning to position the System for future viability.	▪ Analyzes and interprets environmental scanning results for shifts, their implications, and potential impacts on system evolution ▪ Guides development of the system strategic plan and system development map ▪ Informs the development of the strategic plan ▪ Guides future product, service, and content development ▪ Guides investment priorities ▪ Identifies future relationships critical to system development ▪ Identifies future development opportunities and targets that can be pursued in support of mission and vision of the System	▪ Just as individual systems are developed, so to must the system that provides for acquisition of those systems be developed *(acquisition system)*—this is the acquisition system that in essence is the *System that Acquires Systems* ▪ The governance challenge is acquisition system advancement by purposeful, holistic, and evolutionary development ▪ The acquisition system development should not be left to chance (self-organization), piecemeal (fragmented), limited (non-comprehensive), or sporadic (intermittent) development. ▪ Robust and purposefully designed scanning of the acquisition environment should be engaged across not only technical developments, but also organizational, managerial, human, social, political, policy, and information dimensions. The design, processing, and interpretation of this scanning is an acquisition leadership function. ▪ The acquisition system should evolve by comprehensive system design, based in “deep” system learning generated from detection and correction of system design as well as execution errors.

Metasystem Function	Primary Role	Responsibilities	Implications for Acquisition
Metasystem Three (M3) – System Operations	Primary function is to maintain operational performance control through the implementation of policy, resource allocation, and design for accountability.	▪ Oversight for products, services, value, and content delivery ▪ System planning & control for ongoing day to day operational effectiveness ▪ Develop near term system design response to evolving operational issues and monitor operational performance measures ▪ Operationally interprets and ensures implementation of the system policies and direction ▪ Interpretation and translation of implications of environmental shifts for operations (based on inputs from System Development) ▪ Informs the development of the strategic plan ▪ Determines resources, expectations, and performance measurement for operational performance ▪ Design for accountability and performance reporting for operations	▪ Value provided by the *acquisition system governance* is consumed by acquisition professionals, not the community receiving the ultimate products from that acquisition system. ▪ Care should be taken to measure acquisition system performance beyond the performance of individual systems being acquired—this is a different, higher level (metasystem governance) set of measures (e.g., identity fragmentation). ▪ The audit of governance functions should be both routine and ongoing. Routine should encompass continual evaluation of acquisition health. In contrast, for aberrations (e.g., GAO audit findings), the acquisition system design should be questioned to determine if the issue has "deeper" systemic implications across the governance functions. ▪ The acquisition system should focus on providing maximal autonomy to programs while preserving acquisition system (governance level) integration to meet performance expectations. Over constraint is wasteful of scarce resources and indicative of ineffective/inefficient system design.

Metasystem Function	Primary Role	Responsibilities	Implications for Acquisition
Metasystem Two (M2) – Information and Communications	Enables system stability by designing and implementing the architecture for information flow, coordination, transduction and communications within the metasystem and between the metasystem, the environment and the governed system.	▪ Designs and maintains the architecture of information flows and communications within the metasystem, between the metasystem and environment, and between the metasystem and the governed system ▪ Ensures efficiency by coordinating information accessibility within the system ▪ Identifies standard processes and procedures necessary to facilitate transduction and provide effective integration and coordination of the system ▪ Informs the development of the strategic plan ▪ Identifies and provides forums to identify and resolve emergent conflict and coordination issues within the system	▪ The acquisition system must have a compatible architecture for governance that provides the necessary coupling of entities through information exchange and communications. ▪ The design of information and communication of acquisition system governance operates and must be appreciated beyond the simple information "availability" and "exchange" models. ▪ The acquisition system design must support information flow and communication that is "actionable" for decision and interpretation consistency. ▪ Design, execution, maintenance, and evolution of information and communication for acquisition system governance should be purposeful. It is too important to be left to self-organization or chance development.

Defense Acquisition Challenges

We identified a number of high profile defense acquisitions, primarily through GAO reports and Berteau et al. (2011) that had open sourced analyses of the acquisition program. We used those open source reports to attempt to answer several questions: (1) Does the problem/failure appear to be governance related? (2) Does the language in the report indicate a similar meaning for governance as the CSG meaning? and (3) Is there any concrete indication that the tools and methods of CSG would have helped this program? The results of this analysis are portrayed in Table 2.

Table 2. Analysis of Troubled Programs Through the Lens of CSG

DoD program/ Report Source	Does the problem/failure appear to be governance related?	Does the language in the report indicate a similar meaning for governance as the Complex System Governance?	Is there any concrete indication that CSG would have helped this program?
Zumwalt Class Destroyers (DDG1000) GAO-08-904 [1,2]	Yes	No model/framework of governance – Milestone C suggested – won't help with alignment of perspectives or understanding decisions and actions (communication channel – dialog) among others	Yes – this initiative seems to lack clear vision/strategy. Report suggests that channels of communication are weak (p. 45 for example)
Ford Class Aircraft Carrier (CVN78) GAO-16-847 [26]	Yes	Yes, the report seems to identify many governance issues that can be mapped to metasystem functions within the CSG Reference Model	Yes – contextual assessment to evaluate acquisition culture. The ship is already built, though, so now the asset needs to be protected and maintained.
Total Asset Visibility (Air Force) GAO-08-866 [3, 27]	Yes	Yes, especially the "transformation plans" demonstrating initiative to evolve meta-systemic functioning	Yes – systems thinking likely not present in development, poor coordination of unsuccessful program
Major Automated Information Systems (MAIS) GAO-12-629 [25]	Yes	Yes, GAO seems to have an idea of the metasystem governance expected of a complex system, as well as realistic expectations regarding scope	Yes – some metasystem functions are clearly missing or inadequate, ex. poor coordination and communication (25, pp. 57, 58)
National Security Cutter (Coast Guard/Navy) GAO-16-148 [24]	Yes	Yes, report seems to capture design/execution elements necessary for control/communication/ coordination/integration (but possibly not sufficient?)	Yes – CSG embraces varying perspectives – the CG & Navy did not seem prepared to align perspectives and have poor communications

Results from this preliminary review of the "real world" cases of acquisition suggest that CSG can make a substantial contribution to the acquisition field. Through the lenses of CSG, the deficiencies identified in the programs can be understood at a different level. However, presently the attributions of deficiencies in the CSG of the acquisition programs is little more than an academic exercise in hindsight. We suggest that the true realization of value in the application of CSG to the acquisition field can come from four primary contributions. First, CSG can offer a different set of insights concerning application of the "systems view" to the acquisition field. The inculcation of this systems view can serve to inform a different level of thinking, support a more enlightened decision space, drive different courses of action, and invoke different interpretations. Second, CSG offers a rigorous formulation of the structure and execution of a system (e.g., acquisition program). This structure and its execution ultimately determine the level of performance achieved from a system. Unfortunately, in most instances, the design, execution, and evolution of the nine

CSG metasystem functions are performed on an ad hoc basis. The explicit consideration of the metasystem functions can provide a more “holistic” and rigorous approach to design, analysis, operation, maintenance, and evolution of a governing system (e.g., acquisition system and programs). Third, using the strong systems theoretic basis of CSG can allow a different and deeper level of analysis of acquisition system and program design. This can identify an entirely different view of the surface manifestations of poor performance (e.g., missing cost, schedule, performance expectations). Instead, more fundamental systems based pathologies (i.e., aberrations from healthy system conditions) can be identified and explored from an entirely different (holistic) systems paradigm provided by CSG as presented in this paper. Fourth, CSG can enhance acquisition practitioner capabilities to more effectively perform essential governance functions. All acquisition programs perform governance functions, even if they are not explicitly acknowledged. By accounting for the CSG functions in design and execution of acquisition programs, practitioners can enhance their capabilities to more effectively engage the increasingly complex environments and programs they must direct.

We now shift our attention to the future directions for the inclusion and development of CSG the acquisition field.

Considerations for Future Exploration

Thus far we have presented the case for the potential of CSG contributions for the acquisition field. In this section, we examine specific developmental directions for further inculcation of CSG into the acquisition field. There has been significant literature that has developed the foundations of CSG as an emerging field (Keating, 2014; Keating & Bradley, 2015). However, CSG has not been disseminated or projected to the acquisition field or community of practitioners. CSG has the potential to significantly improve capabilities for practitioners (owners, operators, performers, designers) in the acquisition field. We suggest that the utility of CSG for acquisition can proceed along three interrelated streams of development, including science, technologies, and application (Figure 5).

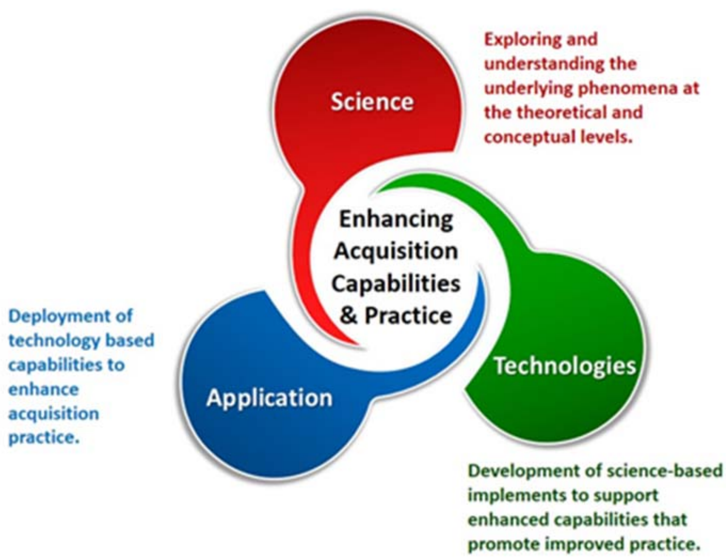


Figure 5. Three Interrelated Streams of Development for CSG in Acquisition

For purposes of this discussion, we take *science* broadly as the search for knowledge to develop testable theory and laws related to a field (e.g., acquisition). The tenets of good science include disciplined inquiry that can withstand the scrutiny of a particular field. The results of science must be theories and laws that can be tested to determine their continued power to provide confirmation or be refuted. For acquisition, this suggests that discovery of new tenets of "acquisition science" may be found at the intersection with the CSG field's foundations in Systems Theory. It would be easy to dismiss development of the science thrust for acquisition as nonessential or a frivolous waste of scarce resources. However, technologies and applications developed without grounding in the underlying science miss an important stable base. While technologies and applications can change rapidly, the underlying theoretical/scientific basis for a field provides long term stability. The importance of this stable science based foundation for the acquisition field cannot be overstated. This is particularly the case given the increasingly turbulent conditions faced by acquisition professionals and programs.

Technology engages science to develop innovations that solve problems and increase the capabilities of practitioners to more effectively function. Thus, technology becomes a bridge between science and application. Finally, *applications* involve putting science-based technologies into action to achieve human purposes (e.g., system acquisition). Ultimately, the applications by practitioners provide utility for science-based technologies. We believe that acquisition research must be engaged and integrated across each of the three levels (science, technologies, applications) if it is to provide sustainable improvement for the acquisition field. The interrelated advancement across these three developmental thrusts for acquisition improvement will (1) accelerate development of each of the other thrusts, (2) provide a grounding to better inform each of the thrust areas such that different directions and insights might be possible, and (3) draw the worlds of science and practice closer to provide a more balanced development of CSG for the acquisition field.

The pursuit of CSG development for acquisition must appreciate the interrelationship and development of science, technology, and application in concert. To look at these three aspects of the development of a field as independent and mutually exclusive of one another is false and somewhat naive. The acquisition field faces a major challenge to pursue parallel integrated paths of development for the science, technology, and application of CSG for acquisition. The easy, and more traditional, research approach is to separate the development of underlying science from corresponding technologies and eventual applications. However, there is much to be gained by permitting the triad to constrain as well as enable one another. The research path that emerges through the integration of science, technology, and application may be very different than had joint development not been considered. It is certainly arguable that the acquisition field currently pursues research that engages a close correlation between science, research, and application domains. However, there is much to gain by pursuit of CSG for acquisition development that explicitly couples science, technology, and applications by design from an integrated systems perspective (Figure 6).

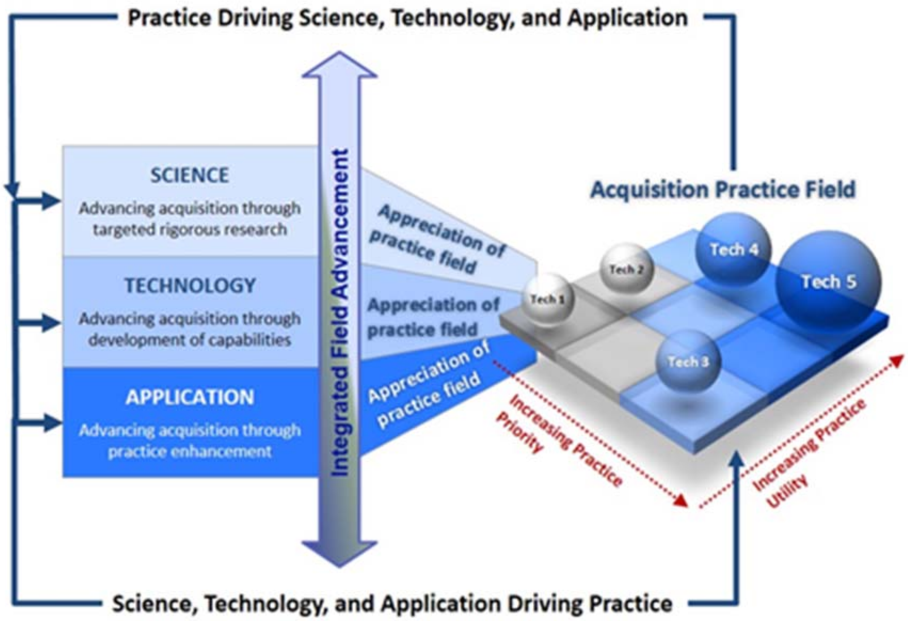


Figure 6. Integrated Development of Science, Technology, and Application for Enhanced Acquisition

The acquisition community is in a position to advance the field by inclusion of CSG in a way that will (1) steer the research agenda for the science and derivative technology developments related to Acquisition System Governance, (2) influence practitioner capabilities through development of science-based technologies to support acquisition governance practices, and (3) provide leadership to pioneer integration of the CSG emerging field to enhance acquisition capabilities and practice.

The major opportunities and impacts of engaging CSG for acquisition are summarized in the following three points:

1. Prod*uce Research Driven Acquisition Governance Technologies to Enhance Practitioner Capabilities and Effectiveness*—Ultimately, Acquisition Governance research can have a substantial impact on the performance of this vital function for government acquisition enterprises. Technologies to leverage scarce resources, provide decision and policy support, and establish effective oversight are hallmarks of effective governance. While emphasis on acquisition reforms targeted to issues of cost, quality, and schedule are necessary, that emphasis alone is not sufficient to provide "holistic" development of acquisition. We argue that it is also a "necessary" condition to emphasize the development of enhanced governance capabilities to truly advance the acquisition field.
2. *Enhance the Capabilities of Acquisition Practitioners*—The acquisition system itself should not be the sole focus of more advanced acquisition governance development. The practitioner and program levels should also be a focus for development. It is shortsighted to develop new governance technologies if the implementing practitioners do not have the compatible "systems thinking" mindset to deploy them consistent with their underlying systems essence. In

effect, governance development should also target enhancing the capacity of practitioners and programs to think more systemically.

3. *Research that Advances Acquisition System Governance*—This emphasis can generate the theory, methods, and deeper understanding of the phenomena associated with acquisition for Government Enterprises. The integration of CSG into the acquisition landscape brings a new perspective, corresponding language, and systems theoretic grounding to acquisition. Unfortunately, the current emphasis too often engages research that directs acquisition to development of systems and technologies that are predominantly outwardly focused—systems, technologies, and products that are acquired through the acquisition system for consumption external to the acquisition system that provided them. This is an essential role for acquisition. However, there is also a corresponding necessity to engage development of systems, capabilities, and technologies that are *inwardly focused* on achieving enhanced effectiveness for the acquisition system, community, and practitioners. We call this emphasis a self-reflexive effort to do "acquisition of acquisition" systems, capabilities, and technologies. In essence, CSG for Acquisition Governance is targeted to realization of this shortcoming in acquisition development. This can be achieved by producing science-based technologies to enhance acquisition practice for consumption by the professionals responsible for the effective design, operation, maintenance, and evolution of the acquisition system.

CSG for acquisition is not offered or pursued as a universal remedy for issues that plague acquisition programs and challenge practitioners. However, we are confident that it will permit practitioners to more effectively deal with the challenges of governance they face on a daily basis in acquisition. CSG integration to acquisition is not intended to replace, relegate, or subjugate the role of the acquisition practitioner. Analysis, interpretation, decision, and action have always been, and will always remain, the purview of acquisition management professionals. What CSG offers to the acquisition field is enhanced capabilities for acquisition professionals responsible for governance of acquisition systems and programs. CSG research seeks to support acquisition practitioners by development and testing of science based technologies and applications that amplify their effectiveness.

References

Beer, S. (1972). *Brain of the firm: The managerial cybernetics of organization*. London, UK: Penguin Press.

Beer, S. (1979). *The heart of enterprise*. New York, NY: John Wiley & Sons.

Beer, S. (1985). *Diagnosing the system for organizations*. Hoboken, NJ: John Wiley & Sons.

Berteau, D., Ben-Ari, G., Hofbauer, J., Sanders, G., Ellman, J., & Morrow, D. (2011). *Cost and time overruns for major defense acquisition programs: An annotated brief*. Paper presented at the Eighth Annual Acquisition Research Symposium, Monterey, CA.

Berteau, D., Levy, R., Ben-Ari, G., & Moore, C. (2011). *Defense industry access to capital markets: Wall Street and the Pentagon. An annotated brief.*

Boudreau, M. (2007). *Acoustic rapid COTS insertion: A case study in modular open systems approach for spiral development*. Paper presented at the IEEE International Conference on System of Systems Engineering (SoSE'07).

Fauser, M. (2006). *Transnational migration–A national security risk? Securitization of migration policies in Germany, Spain and the United Kingdom*. Retrieved from http://pdc.ceu.hu/archive/00004804/01/rap_i_an_0206a.pdf

Francis, P. L. (2008). *Defense acquisitions: Zumwalt-class destroyer program emblematic of challenges facing Navy shipbuilding*. Washington, DC: GPO.

Francis, P. L. (2009). *Defense acquisitions: Cost to deliver Zumwalt-class destroyers likely to exceed budget* (1437909310). Washington, DC: GPO.

Gansler, J. S., & Lucyshyn, W. (2015). Allocating national security resources. In F. Melese, A. Richter, & B. Solomon (Eds.), *Military cost–benefit analysis: Theory and practice* (Vol. 14, p. 18). New York, NY: Routledge.

Jaradat, R. M., & Keating, C. B. (2014). Fragility of oil as a critical infrastructure problem. *International Journal of Critical Infrastructure Protection, 7*(2), 86–99.

Kadish, R. T., Abbott, G., Cappuccio, F., Hawley, R., Kern, P., & Kozlowski, D. (2006). *Defense acquisition performance assessment report*. Washington, DC: Pentagon.

Keating, C. B. (2014). *Governance implications for meeting challenges in the system of systems engineering field.* Paper presented at the IEEE System of Systems Engineering Conference, Adelaide, SA.

Keating, C. B., & Bradley, J. M. (2015). Complex system governance reference model. *International Journal of System of Systems Engineering, 6*(1), 33–52.

Keating, C. B., & Katina, P. F. (2011). Systems of systems engineering: Prospects and challenges for the emerging field. *International Journal of System of Systems Engineering, 2*(2–3), 234–256.

Keating, C. B., Katina, P. F., & Bradley, J. M. (2015*). Challenges for developing complex system governance.* Paper presented at the Proceedings of the 2015 Industrial and Systems Engineering Research Conference.

Mills, S., & Goldsmith, R. (2014, September–October). Cybersecurity challenges for program managers. *Defense AT&L*.

Naphade, M., Banavar, G., Harrison, C., Paraszczak, J., & Morris, R. (2011). *Smarter cities and their innovation challenges. Computer, 44*(6), 32–39.

O'Rourke, R. (2014). *Case studies in DoD Acquisition: Finding what works* (1437928390). Washington, DC: Congressional Record.

Rascona, P. M., Barkakati, N., & Solis, W. M. (2008). *DoD Business Transformation: Air Force's current approach increases risk that asset visibility goals and transformation priorities will not be achieved.* Washington, DC: GPO.

Skyttner, L. (2005). General systems theory: Problems, perspectives, practice (2nd ed.). Singapore: *World Scientific*.

von Bertalanffy, L. (1950). An outline of general systems theory. *The British Journal for the Philosophy of Science, 1*(2), 134–165.

von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications* (Rev. ed.). New York, NY: George Braziller.

Whitney, K., Bradley, J. M., Baugh, D. E., & Chesterman, C. W., Jr. (2015). Systems theory as a foundation for governance of complex systems. *International Journal of System of Systems Engineering, 6*(1), 15–32.

Acquisition Program Teamwork and Performance Seen Anew: Exposing the Interplay of Architecture and Behaviors in Complex Defense Programs

Eric Rebentisch—is a Research Associate at MIT's Sociotechnical Systems Research Center. He leads the Center's Consortium for Engineering Program Excellence and previously led the Lean Advancement Initiative's Enterprise Product Development research. Dr. Rebentisch's portfolio of research projects includes studies of the integration of program management and systems engineering and performance benchmarking of the U.S. shipbuilding industry. [erebenti@mit.edu]

Bryan Moser—is the Founder and CEO of Global Project Design, Inc. (GPD), and Lecturer in the Systems Design and Management program at MIT. His industry engagements in the U.S., Europe, and Japan have provided a laboratory for development of *project design* workshop tools and methods. Prior to GPD, for a decade with United Technologies Corporation, Dr. Moser led research on and implementation of global technology and product systems development. He has published extensively on dispersed teamwork and the interplay of coordination architectures and team behaviors impacting complex project cost, schedule, and performance. [bry@mit.edu]

John Dickmann—is a Vice President at Sonalysts Inc. He served for 22 years as an active duty Navy submarine officer. He is a graduate of MIT's Engineering Systems Division and has researched lean processes and enterprise design for the U.S. Air Force, the Naval Sea Systems Command (PMS-401), and the Office of the Secretary of Defense (Net Assessment). Dr. Dickmann manages systems design projects and conducts operational and strategic technology analysis for defense and industry clients. [jdickmann@sonalysts.com]

Abstract

This research frames complex engineering development programs as sociotechnical systems with program performance driven by interpersonal and inter-organizational dynamics as well as technical system interdependencies. It attempts to address the question of why performance in complex development programs has not improved significantly in the last several decades, despite the development and application of many new and sophisticated tools for managing these programs. A review of the literature on managing complex sociotechnical systems was used to develop a framework and method for instrumenting complex engineering programs and measuring their essential attributes. The proposed framework identifies fundamental elements of engineering programs (relating to, e.g., products, processes, organizations, and people) and the drivers of program performance. The framework is illustrated using a case study of a complex engineering program that spanned multiple technical systems, organizations, and disciplines. The paper discusses the resulting measurement framework and provides examples of the application of the framework to identify management control "levers" for design, engineering, test and evaluation, fielding, and sustainment of complex engineering programs.

Introduction

Large-scale engineering programs are challenging to complete within planned parameters. U.S. Department of Defense (DoD) programs involve outlays of public finds, and are therefore well-documented. They unfortunately often report disappointing outcomes. The U.S. Government Accountability Office (GAO) reported in 2009 that the accumulated cost overrun of the largest 96 DoD engineering development programs reached nearly $300 billion, with an average schedule overrun close to two years (GAO, 2009). This doesn't appear to be an aberration from the early part of the 21st century. The GAO reported previously that combined cost overruns for large development programs (programs totaling more than $1 billion for research, development, testing and evaluation in fiscal year 2005 dollars) initiated in the 1970s exceeded the DoD's initial investment estimate by 30%, or $13 billion (in fiscal year 2005 dollars), with equivalent overruns of 39% in the 1980s and 40% in the 1990s (GAO, 2006). Despite numerous acquisition reform efforts and policy revisions during those years, defense acquisition programs during that three decade period continued to routinely experience cost overruns, schedule slips, and performance shortfalls. Underperformance is not confined just to defense development programs, though. Reports of disappointing performance in large-scale civil engineering programs tell a similar story (Cantarelli et al., 2010).

Poor development program performance seen across a number of different applications and business sectors suggests that there may be underlying, systematic factors that bias programs toward trouble. An example from commercial aerospace illustrates just such a diverse array of challenges. In 2003, Boeing launched the 7E7 program as a refresh and partial replacement for its 767 and 747 families of aircraft. The 7E7 eventually came to be known as the 787 and quickly established a very strong order book from airlines ("Boeing 787 Dreamliner," 2016). By the time it finished development, the program exceeded the estimated costs by three times and took roughly twice as long to develop as estimated. An investigation into the reasons for these outcomes suggested that a likely cause was an excessive growth of development project complexity. The aircraft itself became more complex due to the use of new materials which was in many cases beyond the capacity and experience of Boeing or its suppliers. The organization tasked with developing this new aircraft also became more complex because of significantly increased external development through partners and vendors, outsourced system integration and higher interdependence through a more parallelized development process (Allworth, n.d.). Insufficient ability to handle growth in both types of complexities eventually led to the major project delay and skyrocketing costs, pushing the program into crisis (Denning, n.d.). Although the 787 eventually overcame the crisis, each delivered 787 still generates losses (Gates, 2016). The development of the Boeing 787 aircraft design and production required coordination across the globe of multiple organizations, corporations, and governments, which is representative of many development projects, where estimated costs and development are exceeded.

The same challenges plague complex programs in public sector layered infrastructure initiatives and commercial services deployment. Current program management standards are largely heuristic, experience-based, and generic, with application oriented toward a wide range of sectors, project scope, and activities. They are, however, deficient in addressing the management of interdependencies and cross-boundary interactions in complex programs such as those seen in the 787 development program. These interdependencies and interactions include those that are hierarchical between management layers, lateral between functional groups, and multi-scale based on nested layers of performance domains. While hierarchical and lateral interdependencies are acknowledged in traditional organizational/management literature, they are seldom defined at the level of specificity that is required for program management. Multi-scale relationships are increasingly pervasive in operations, driven by increasing complexity in engineered systems, but they remain largely undefined at a useful level of specificity in either the organizational/management or program management literatures.

The lack of improvement in complex program outcomes may be traced to few new and more effective program management practices and, ultimately, too little innovation in the way that the basic attributes of programs are measured. Prevalent project management standards rest upon heuristic practices. Systems engineering standards address system architecture, methods, and tools, but fall short in defining social and managerial interdependencies with a product system and its design and operating context. Identifying and articulating measurement and instrumentation for program elements at a fundamental level could lead to a richer characterization of the essential attributes of complex programs. The resulting richer datasets might then enable the development of new analytic methods for understanding the underlying drivers of program performance. The framework explained in this paper is part of an effort to move program management from its primary basis in heuristics and collections of best practices toward the design of projects based on a deeper understanding of the interdependencies, behaviors, and performance of sociotechnical networks under systemic complexity.

Current Measurement and Control Systems for Complex Programs

Relatively simple programs (i.e., linear extrapolations from known space to known space, perhaps best characterized by manufacturing), while potentially complicated, may not be particularly complex. That is to say, they may have many elements, but they and the relationships between them are all relatively well-understood and predictable in their behaviors. The associations between tasks, participants, and sequencing may be well-specified, and the imperative is to execute the tasks through clearly-defined relationships. In these types of programs, traditional tools, practices, and methods for managing the program may be entirely adequate to this challenge.

However, when novelty, scarcity, or uncertainty are part of the work space, there are potentially more known unknowns or even unknown unknowns than would be expected in a simple program. This may require learning, innovation, and possibly improvisation from the program team to deal with exceptions to the plan or other unexpected developments. With these emergent behaviors, a linear extrapolation from known practices will be of limited benefit. Learning will require the flow of information from a range of different sources in order to notice, acquire, understand, and synthesize knowledge into needed new forms. This suggests exploration and exploitation of the information (and related resource) networks across the program. This would not be possible without understanding and controlling the dependencies in the system.

Evolution of Systems Projects Control

We can trace the underlying model of work used most commonly in project management back to Taylor and Gantt (Gantt, 1903; Wilson, 2003), with jobs described as sets of discrete tasks, refined over time to reduce variation in repeatable activities of fixed duration and fixed sequence. Both Taylor and Gantt (1903) were managing factories, with the Gantt chart itself originally a table to describe fixed jobs for workers. These same assumptions of standardized durations and sequence were carried forward in the middle of the last century with the advent of critical path (CPM; Kelley & Walker, 1959) and other network techniques.

The complexity of government programs and a shift towards cost control led in that next decade to a more centrally-controlled project management approach. The roots of today’s project control, including earned value, emerged in the 1960s in requirements for defined work breakdown structures (WBS), systems engineering management plans (499), and Cost/Schedule Control System Criteria (C/SCSC). Still, the underling view of task as project “atom” connected through a precedence based network remained.

Amidst a call for simplification compared to prior heavy processes, defense acquisition management was transformed starting in 1991 through the DoD Directive 5000 series (Dillard, 2003). Over the next decade, in parallel with streamlining through adoption of commercial practices in government programs, improved risk management emerged, including attention to technology readiness. The underlying model of project work spread through the introduction of CMMI by SEI and other standards, including PMBOK in North

America. Like the century earlier factory models of work and the mid-century defense approaches, these followed the same underlying model of project work as tasks connected by precedence networks of dependencies.

The emergence of project control beyond large government programs can be partly traced to Fleming and Koppelman, who promoted earned value over two decades as Koppelman built Primavera Systems (now part of Oracle). They commented on the potential of earned value as a concept not only in large defense programs but—in simplified form—for software and other projects (Fleming & Koppelman, 1994, 1996). Their papers and book in the PMI community from the middle to late 1990s introduced earned value to a broader project management audience in industry.

Earned Value

Earned Value management is well-documented by many; this paper assumes the reader has already or can easily gain EVMS basics, so it does not include reiteration of the earned value approach. Instead, we trace the evolution of EVMS from its emergence from the DoD in the early 1990s to recent years.

The A-12 program cancellation in 1991 has been widely accredited to early indicators of cost variation from EVM. However, Christensen's (1994) work showed, too, that after the first third of the program, the cost performance index will stabilize, with limited variability as a project proceeds. Still, as an early indicator of the A-12 Program cost overruns, the method performed as expected (Christensen, 1994).

Evolution of Earned Value: ES and ED

Earned Schedule

In 2003, Walter Lipke, at the time head of the software division in the U.S. Air Force's Oklahoma Air Logistics Center, proposed a change to the schedule measures in EV (SV–Schedule Variances and SPI—Schedule Performance Index) that he called Earned Schedule (ES; Lipke, 2003, 2004). Lipke and others at the time had noticed that the schedule variance (SV and SPI) as used in earned value became less predictive later in a program. Since EV schedule measures are derived from cost (BCWP–Budgeted Cost for Work Performed, and BCWS–Budgeted Cost for Work Scheduled), as an over-schedule project approaches completion, the schedule variance according to SPI approaches 1.0. and SV approaches $0. Instead, in Earned Schedule (ES), as a substitute for budgeted cost as performed versus as scheduled (BCWP and BCWS), Lipke takes the actual duration versus the planned duration for the work performed. Lipke was able to show that in some cases, the earned schedule measures remain meaningful in the latter stages of a program, including showing positive or negative variance at completion if the actual duration differed from the planned duration.

While ES is an improvement over EV and can be implemented using the existing EV metrics, ES still rests upon schedule progress as derived from a portion of original budget spending, with a linear association to schedule progress.

Earned Duration

In a recent and important contribution, Khamooshi and Golafshani (2014) take a step further than Lipke's Earned Schedule approach. They refer to their methods as "Earned Duration" (Khamooshi & Golafshani, 2014). Consistent with Lipke's response to the inaccuracy of EVM schedule-related performance, they propose complete decoupling of the schedule metrics and forecasts from cost related inputs.

The authors point out that as patterns of cost over time, progress over time, and value of scope over time become non-linear, then the assumptions of uniform linear association between spending, progress, and likelihood of ultimate schedule performance become false. Instead, they define earned duration as “Earned Duration of scheduled activity i: EDi, at any point in time, is the value of work performed expressed as proportion of the approved duration assigned to that work for activity (e.g., days)” (Khamooshi & Golafshani, 2014).

In their paper, they also, interestingly, emphasize the dual role of project control techniques: first to ascertain the performance of a project to date as compared to some original baseline plan, and second to determine the accuracy of original estimates, providing a view that will allow comparison and learning across multiple projects.

While the data necessary to calculate earned duration ("ED(t)") goes beyond that required by classic EVM, they argue that the data is available otherwise, as was necessary for original planning and ongoing scheduling by project teams. Importantly, the measures require an estimate of remaining duration, given the current state of the project, as an indicator of progress rather than cost. Whether these estimates are made at the macro or micro level is a project control or architectural decision irrespective of which measures are chosen.

The authors also point out that use of actual performance metrics from an earlier stage of a project as a proxy for expected performance in later stages, given the typically limited information about specific resources, priorities, and other externalities, is “questionable”: “If the stages of the project are different and heterogeneous, which normally is the case, there is no rationale for assuming past performance is a good predictor of the future” (Khamooshi & Golafshani, 2014).

EV Variants and Control Points

Colin and Vanhoucke survey recent literature on earned value and project control techniques, characterizing the set from the original use of Critical Path Method (CPM) from Kelley and Walker (1959) as a bottom-up approach and the more recent earned value and its variants which are described as top-down. They assert that these methods vary in the number and position of control points—positions in a WBS at which are placed monitors for observing and buffers for controlling project flow. EVM as practiced rests upon a topmost WBS element with calculated EV, PV, cost and schedule metrics; they discuss several recent papers which show control points at key points in the project, not necessarily the complete data collected bottom up from each WBS activity. For example, another method by Lipke (2012) places control points along the critical path. They introduce two approaches inspired by Goldratt's Critical Chain method (Goldratt, 1997), in which they explore control along the critical path, along feeding paths into the CP, or instead entire subnetworks which feed the CP.

Their approach recognizes that a program manager (the PM) is burdened by the amount and upkeep of control data and response. By seeking a balance between bottom-up and top-down approaches, they seek to minimize overzealous control that—being unsustainable—causes latent and poor quality control signals. However, their paper does not directly address resource capacities, and only indirectly the capacity of a PM to handle control activities.

Furthermore, as the various methods in Colin and Vanhoucke, Lipke and others rest upon an underlying model of project as network with discrete precedence dependencies, the capacity and quality of interactions across these project interfaces is misrepresented. In selecting the positions of control points, one must ask, “Who calculates, interprets, and reports the information at that interface?" Are their experience, capacity, and accountability aligned with the demanded attention, in timely fashion, to the control point in balance of all the other demands "on their plate"?

Our thinking, when viewing complex systems projects as sociotechnical systems, places an emphasis not only on the PM, but on distributed resources, their condition of awareness and attention, their attempts to work and interact, who make mistakes and correct them, and learn over time. Therefore, not only the PM, but those project participants who would be best positioned to own control points and their source knowledge, should be considered. By analogy, if a control system lacks the capacity to process input signals in real time, it will become saturated and lose its control authority. The capacity of the system can be increased by parallel processing using a system of distributed controllers. However, that approach only works if the distributed controllers are coordinated and the interdependencies between them are managed. We assert that a burden of demands to be aware of and interact for project governance be distributed to align both with capacity and inherent capabilities of resources.

Summary

The legacy of program management and systems engineering tools and methods is rooted in practices suited to factory operations where tasks are assumed to be well-defined, dependencies between tasks are relatively simple, and the flow of work is fairly linear. These assumptions have changed little since a century ago, or at least have not been challenged in a vigorous way. Yet, it is clear that the complexity, both static and dynamic, of development programs and their corresponding sociotechnical systems has increased significantly in both scale and nature over the same time period. As a consequence, it should not be surprising that complex developments suffer poor outcomes and are seemingly uncontrollable. Their program management control systems have not kept pace with their changing nature. The case study introduced in the next section illustrates some of the ways in which management control systems may be challenged in a complex program.

Re-Architecting the Submarine Sonar Sociotechnical System

Our case study examples highlight the challenges of getting people to act with awareness and conviction at critical interfaces. We find that these are especially difficult when change causes patterns of demanded coordination to shift from historical ones. Whether due to habit, old incentives, or gaming—shifts in behavior are necessary to improve both local and systemic performance. This is especially true when program managers are confronted with shifting externalities such as changing technologies, shifting operational performance requirements and fiscal constraints. Successful program leaders understand the need to shift organization and system architectural alignment in these conditions.

The case of Navy submarine sonar system program management in the 1990s is an example of these types of change and the consequent need to recognize and design changed dependencies. To date, many published lessons drawn from this case fall in the category of best practices. How can we move from heuristics to a repeatable, measurable approach that can provide program managers real or near-real time feedback on internal program coordination demands?

Crisis: Loss of Submarine Acoustic Superiority

Military competition in the undersea consists of a constant evolution of operations and technologies to detect and to minimize acoustic emissions. In the early 1990s, U.S. submarines lost their long-standing acoustic advantage against Russian submarines—the ability to detect and track them before detection by them.

Throughout the Cold War, the Navy had invested billions of dollars in acoustic research and advanced sonar systems, concentrating its effort on custom-designed digital signal processing systems with military-unique components and tightly integrated proprietary hardware and software. Specifications were developed by the Navy Laboratory, and detailed design and production of hardware and software were conducted at one of two prime contractors. System development and fielding took a decade or more from conception and cost billions of dollars, with unit costs ranging to hundreds of millions of dollars.

The loss of acoustic superiority coupled with post–Cold War budget cuts created pressure to improve the performance of U.S. submarine sonar systems quickly. Added to this pressure was a wider questioning of the relevance of submarines in the "new world order," which created an organizational crisis, spurring what is arguably the most successful acquisition reform effort in U.S. defense industry history.

Diagnosis and Prescription

The Submarine Force response to this technical and fiscal challenge was to examine the problem from a fact-based perspective. In early 1995, the Submarine Superiority Technology Panel (SSTP), a panel of acoustics experts, was established to examine the technical performance of submarine sonars. It came to two major conclusions: (1) The legacy sonar system technical architecture was ill-suited to leverage Moore's Law for increased signal processing power and (2) the system development and acquisition organizations and the development-acquisition process inhibited experimentation with new algorithms.

The panel observed that there was no viable means to test, evaluate, and integrate advanced algorithms that had been developed by academic researchers through the 1980s. They recommended a commercial-off-the-shelf (COTS) architecture and an "open" development process. By the fall of 1995, the Submarine Superiority Management Council (SSMC) was established to address the findings of the SSTP. The SSMC worked through the spring of 1996, by which time, the core problems of boosting processing power and developing a means to inject new ideas into the sonar system were the main focus of attention. From this time onward, the responsible program managers for submarine sonar systems were working on developing new technical, process, and organization architectures for the development and acquisition of advanced technology for submarine sonar systems.

System Solution

The program managers charged with submarine sonar embarked on an effort to identify viable commercial processing technologies, identify and develop improved signal processing algorithms, and to develop a process to field these improvements quickly. This collective effort was reflected in new technical and organizational architectures—new dependencies and interfaces among the organizations involved in the submarine sonar program and in the functional implementation of the sonar system.

Architecture Changes

Architecture is the overall scheme by which the functional elements of a system (technical, people, organizational) are partitioned to individual subsystems/teams and are arranged with respect to each other. Architecture sets the rules which govern interactions in and among systems, both in *operation and over time*, as they evolve in response to changing technology, operational demands and external constraints (Ulrich & Eppinger, 1995; Moses, 2006; Henderson & Clark, 1990; Garlan & Shaw, 1993; Clark et al., 2004; Clark et al., 2005; Board, 2000).

Change the System Technical Architecture

The major technical change was to separate hardware and software into a layered architecture. This was achieved by the development of middleware, a set of software that served as an interface between commercial processors and proprietary Navy algorithms. This change enabled the program managers to leverage the cost and processing benefits of Moore's Law without the need to change the existing software.

Change the Development and Acquisition Enterprise Architecture

It was recognized that improving schedule performance at the same time as technical architecture changes were implemented required new relationships among management and technical organizations. The cognizant program managers worked together to increase the speed of development by creating an iterative build-test-build process. Their goal was to more closely connect academic and government laboratory research and development with engineering integrators and operational users. This required changing the dependencies and responsibilities for technical tasks and re-allocating decision authorities among participating organizations. Changes to the development process were implemented in parallel with changes in the organizational structure. In an evolutionary (over several years) pattern, PEO-level management evaluated technical information dependencies and information requirements for the evolving system, crafting organizational dependencies and interfaces to address them.

Implementing the System Solution Within the Sociotechnical System

Program Office Architecture

Steps were taken to increase the richness of the dependencies between the Advanced Systems Technology Office (ASTO), responsible for developing advanced sonar technologies, and the submarine sonar systems program office (PMS4252). These changes were documented in the semi-annual PMS4252 Acoustic Program Plan (APP; Naval Sea Systems Command [PMS4252], 1994):

The original dependency:

> "ASTO has traditionally provided one of a kind systems … and then transitioned them to NAVSEA PMS5252 in the form of paper algorithm designs which, in turn, are provided to contractors for implementation."

The new dependency:

> "A common test bed and common, if not identical, deployable hardware will be created so that 6.3 developed capabilities can be easily transitioned to 6.4.
>
> PEO(USW) ASTO and PMS425 are coordinating the development of a concept that would enable expeditious fielding of an advance sonar processing concept into fleet systems."

A bulletized list of specific activities to implement this new coordination was also listed. Included in these architectural decisions was the goal of connecting the advanced development process to operational users (the fleet):

> "The approach of beta testing is fundamentally a means to allow the user (the fleet) to get a feel for a capability (new functions) and quickly provide feedback to the developers as to suitability before significant investment. It also provides a more expedient means to work out the operational concept in a forum well suited to define usability. Additionally, it may prove an effective means to introduce capability in parallel with new systems development, conducted by the commodity manager, PMS425, without long and costly changes to existing MILSPEC systems."

Evolution continued to an Integrated Project Team-Working Group structure which fundamentally changed organizational dependencies within and across industry, academia, and government.

Working Group Dependencies

The new development process was named "Advanced Processing Build" (APB). Early APBs were numbered sequentially (e.g., APB-1); later they were numbered according to the year in which they were developed (e.g., APB-98 for 1998). As indicated above, the ASTO-PMS4252 developed process evolved into an extensive set of interrelated working groups. The program managers actively identified and managed intra- and inter-working group dependencies and dependencies between the WG and participating organizations.

In an early example of this management, one ASTO program manager, in the face of confusion among the working groups, identified specific dependencies and the means by which they should be satisfied. These were laid out in a long directive to the WG, transmitted via email (Zarnich, 1996).

In this note, he specified three working groups, their work tasks, the technical and programmatic uncertainties surrounding them, and his decision rationale. He specified coordination dependencies with other WGs, the Navy Lab, a commercial contractor, and expected dependencies within the WG itself. He also directed the addition of WG members in order to ensure the right level of technical expertise was involved. Included in these directions were WG charters, performance milestones, work products, and deadlines. It is important to note that this directive memo included a traditional project plan in Microsoft Project format.

Resistance to New Dependencies: APB-T(01)

After the first two APB development cycles (APB-1[98] and APB-2[99]), the Program Executive Officer, Submarines (PEOSUBS) directed the extension of the new development process to the Combat Control System (CCS). The new process was named "APB(T)," and the initial development cycle was to be APB(T)-01. This new process was implemented by a different program office using the sonar APB model and involved a similar set of actions at the program manager level. New technical architecture choices were made, new dependencies identified and new organizational architecture (dependencies) implemented.

The new architecture created new dependencies, which caused friction and resistance. Specifically, the architecture changes shifted technical responsibilities and authorities. The Navy lab and its main contractor were particularly affected, and their response was to ignore the new dependencies, which resulted in poor performance of the program in shifting to the new APB model.

Specifically, assigned roles were disregarded, and work products were not delivered, which impacted the work products of other parts of the organization (Navy Program Manager, 2001). As an example, over a two month period, attempts were made to get software artifacts delivered to organizations responsible for integration, but the Navy laboratory was late and, in one case, delivered obsolete software. As an example, one participant noted,

> The nature of the support needed is sufficiently broad, and dynamic, that a cooperative interactive engagement process is more appropriate than simply throwing request lists, and return questions about the items, over the transom to one another. While it's true that the items need to be clearly documented, even in an interactive process, unfortunately it's also true that the process can be ground to a crawl easily—and seemingly very technically and rigorously proper, if the participants don't all want to drive toward timely success. Establishing that motivation for mutual success among the engineering team typically isn't all that difficult—most of them are stimulated simply by the technical issues at hand and their natural desire to solve them. The leadership challenge in those circumstances is usually no more complex than making clear that timely success is the desired outcome and encouraging the cooperative engagement. Hopefully the leaders of the engineering team members are encouraging, rather than discouraging, that positive type of interaction.
>
> … executive-level folks all playing project facilitator, … clearly isn't reasonable (although occasionally it still may feel great). We'll continue to provide encouragement to our troops to strive for the cooperative interaction, hopefully yours will be provided similar encouragement from their bosses. (Navy Contractor, 2001)

Summary

Initial observations on these three simple examples highlight the focus of these Program Managers on dependencies and interactions. Based on interview data and email data records, they were mainly focused on implementing processes that increased outside, or non-traditional, participants in the development process. The expectation was that the new participants and a different process architecture and organization architecture would bring more objective evaluation of technical alternatives and, therefore, result in improved system performance.

However, their focus was on outcomes and on getting the "right" participants connected to each other. It was a very evolutionary process, where membership was increased or decreased based on immediate need, where WGs were established and disestablished as need dictated. The initial examination of this program's history highlights the underlying importance of dependencies and attention to them.

Framework for Design & Control of Coordination

The discussion to this point has highlighted the importance of dependencies and interactions in programs, particularly in a dynamic or complex environment. It further argued that existing measures and control mechanisms do not fundamentally address dependencies and interactions, and therefore there is a significant opportunity to improve upon existing measures and control mechanisms. While this work is preliminary and ongoing, these points will be addressed in the following sections based on work recently completed or underway.

Characterizing and Measuring Dependence

A measurement and control system based on dependence and interaction must start with a characterization of dependence. Starke (2015) identified a set of eight characteristics of activity dependence and 21 corresponding measures derived from a review of the literature and expert discussions (see Table 1). No measure was defined for the characteristic awareness, since awareness of a dependence itself is a precondition to be able to assess the dependence. Starke further attempted to validate the set of characteristics and their respective measures through a survey with 138 participants in a workshop. The work was considered preliminary, but he did establish that all the characteristics and measures were reliable with the exception of *Closeness and Degree of Mutuality*.[1] This characterization of dependence is a first step in the development of a comprehensive system of measurement for programs.

A key objective of a dependence measurement system must be to identify indicators of program behavior that can be linked to superior program outcomes. In the end, if it is to be useful (and used), it must demonstrate its ability to predict future program outcomes more accurately than existing measures. Consequently, it must comply with a number of requirements:

- It must be able to be instrumented so as to be practically and sustainably implemented in a performance measurement system.
- It must have a clear sampling approach, frequency, unit of analysis, etc. in order to produce reliable results.
- It must have a clearly-defined measurement process and ideally be indexed to current measurement and control systems in order to assess its predictive power relative to existing approaches.

This ongoing work is part of a larger effort that aims to revise and further develop these measures and develop an instrumentation method for gathering empirical data on dependence in programs and its impact on program control and performance. This work supports a larger agenda of determining whether it is possible to improve program assessment and control methods and tools beyond those currently in use.

[1] These two characteristics and their corresponding measures in particular have strong pooled dependence traits, and were considered to be not well-suited to the experimental methods used in the validation.

Table 1. Measures of Dependence in Programs
(Starke, 2015)

Dependence Characteristic	Description	Illustrative Measures
Awareness	The extent to which the interdependence is recognized within the process.	*No measures defined*
Closeness	The extent to which the actions of dependent activities have an immediate effect on each other.	Component connection within the product Flexibility of budget Number of activities altering data
Degree of mutuality	The extent to which the dependent activities have equal need for each other.	Difference in amount of needed information Difference in activity priorities Difference in data usage
Feedback mechanism	The way feedback is passed between dependent activities.	Frequency of scheduled information exchanges Frequency of scheduled budget reviews Number of times data is needed
Impact	The extent to which not fulfilling the dependence in the desired manner affects the dependent activities.	Fraction of activity dependent on input Rework caused by faulty input Excess capacity Specification connectedness
Satisfaction criteria	The criteria necessary to fulfill the dependence.	Understanding of what is necessary to fulfill the dependence Consensus on what is necessary to fulfill the dependence
Strength	The amount of required interaction as a direct result of the dependence.	Volume of necessary exchanged information Variability of costs Number of shared components Degree of concurrency
Urgency	The time-criticality for fulfilling the dependence.	Milestone flexibility Variance of activity duration

Coordination in Response to Dependence and Architecture

Having characterized and developed measures of dependence, the next step is to demonstrate how dependence plays a role in program coordination and control processes. Based on Moser, Grossmann, and Starke (2015) and Starke (2015), a framework was developed to demonstrate the mechanisms whereby dependence is satisfied in programs (see Figure 1) Dependence is driven by two sources of need: *Flow* and *Pool* causes. A flow cause of dependence results from the need for results or information from another task. A pool cause of dependence results from the need for a resource shared by another task. They both result in a demand for interaction.

Awareness of the dependence and allocation of attention are the major factors influencing how or if any interaction takes place. The volume, timeliness, cost, and quality of the interaction all have consequences regarding the satisfaction of the dependence. Dependency management, or coordination, may influence the demand itself, the awareness and the allocation of attention, as well as the interaction. Classic dependency management techniques seek to improve the awareness of the dependence (e.g., CPM or DSM) or improve the interaction (e.g., action plans or standardization).

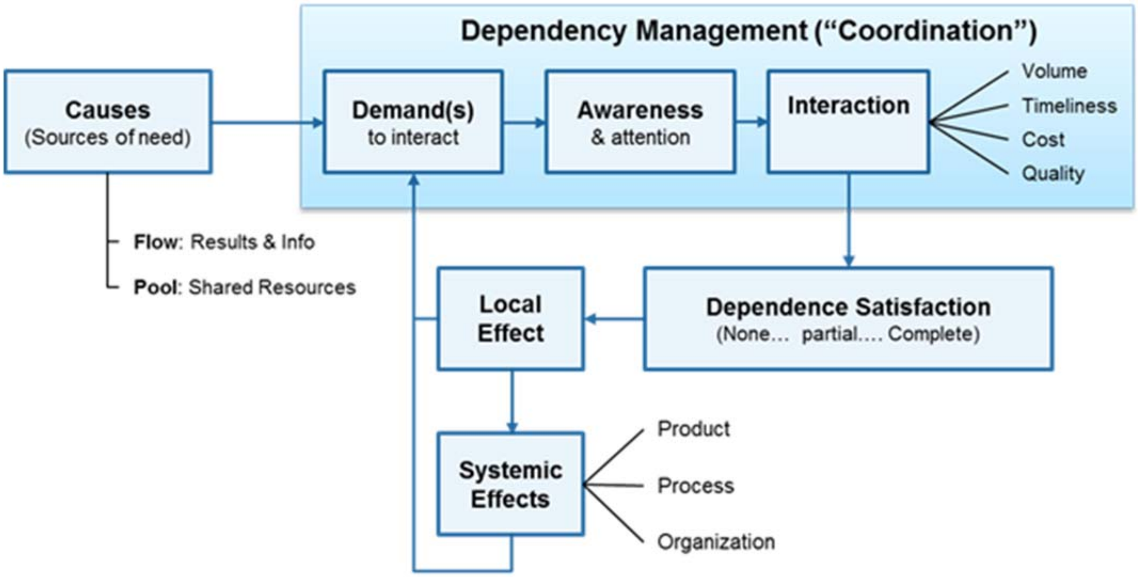


Figure 1. Mechanisms of Dependence From Cause to System Effects

Whether the dependence is satisfied will determine the local effects. This in turn influences the systemic effects. Local effects are the immediate consequences for the tasks (e.g., delay, costs, and rework) and for the individuals (e.g., frustration or establishment of trust). Systemic effects influence the significance of the local effect on product quality, the process as a whole, and the organization. These effects in turn can lead to a change in the remaining demand to interact. If the dependence is fully satisfied, the demand is effectively eliminated, and thus no demand to interact remains. If the dependence is only partly satisfied or not at all satisfied through insufficient interaction, demand to interact may decrease or even increase.

The Health of Interactions as an Early Indicator of Overall Performance

The framework shown in Figure 1 is a simple closed-loop feedback control system. As such, it lends itself to the development of a control system for program management based primarily on dependence. This alone represents a significant departure from existing tools and methods for program measurement, management and control. Current program planning and control practices (if sustainable) are necessary for governance, but may simultaneously act as a straightjacket on learning, depending upon the judgement of program and team leaders to make strategic adjustments. Perhaps an approach that relies on the basic characteristics of dependencies within a program could provide sufficient insights to free up critical program control capacity to enable more effective handling of exceptions, learning, and improvisation within the program.

The emergent and actual performance, in contrast to detailed baseline plans, reflects the gap between detailed control of tasks and strategic management of organization and interactions. Human teams take time and experience, and often fail to learn new habits of interaction. Simply completing one's own work according to finely separated work packages is not sufficient for system performance.

If a dependence-based control system is to make a difference, the treatment (coordination) of the interaction should be driven by the nature of the dependencies amongst tasks, where they fall across the organization, and the pattern of demands they place on teams. Teams would consequently be challenged to adjust their interactions so as to be aware, pay attention, select amongst demands within limited capacity, and to perform. Important questions for teams, implied by the dependence-based framework include the following:

- Have the teams prioritized and paid attention to quality?
- Are defects/issues even noticed?
- If so, how does each team respond and make a decision on how to proceed?

Whether this framework and measurement approach is successful in spurring this kind of activity is the focus of ongoing research, and cannot be conclusively reported at this point in the process. Nevertheless, there is optimism based on not only literature reviews, but also anecdotal empirical evidence. The aim of this study is to collect systematic empirical evidence to assess the validity of the dependence-based approach to program activity measurement and control.

Conclusion, Limits, and Future Research

This paper has demonstrated a set of measures and an emerging measurement process for characterizing the fundamental elements of complex commercial, civil, and defense programs and projects. It focused specifically on the interactions and interdependencies that exist between the product system and social system. It identified implications for the execution of programs and future research relating to program management based on insights gained from this measurement approach.

The validation of this research on engineering projects as sociotechnical systems will require the instrumentation of performance during complex program planning and execution. The intent is to use this paper's representation of dependence to observe projects in progress to test the dependency model's practicality and usefulness. The paper doesn't present an analysis and conclusions because the work is underway.

Future work will require the preparation of a system to measure the demands on and the attention of teams across product, process, and project organization. The responses to dependence will be correlated to local and systemic performance. Additionally, experiments to test the effect of increased awareness of concurrent and mutual dependence on local and systemic performance of the engineering project will be needed. Generating sufficient data to validate this approach from multiple programs will be a lengthy process, but sample identification is already underway. Early experiments using the measures and framework discussed in this paper are promising, but more systematic data will ultimately tell whether this approach addresses the shortfalls in existing methods that have been identified.

References

Allworth, J. (n.d.). The 787's problems run deeper than outsourcing. Retrieved from https://hbr.org/2013/01/the-787s-problems-run-deeper-t

Board, I. S. (2000). *IEEE recommended practice for architectural description of software-intensive systems* (IEEE Std 1471-2000). New York, NY: The Institute of Electrical and Electronics Engineers.

Boeing 787 Dreamliner. (2016, April 4). In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Boeing_787_Dreamliner

Cantarelli, C. C., Flyvbjerg, B., Molin, E. J., & van Wee, B. (2010). Cost overruns in large-scale transportation infrastructure projects: Explanations and their theoretical embeddedness. *European Journal of Transport and Infrastructure Research, 10*(1), 5–18.

Christensen, D. S. (1994). Using performance indices to evaluate the estimate at completion. *The Journal of Cost Analysis, 11*(1), 17–23.

Clark, D., Sollins, J., Wroclawski, J., Dina, K., Kulik, J., Yang, X., … *Chiappa, N. (2004). New arch: Future generation internet architecture.* MIT and USC.

Clark, D., Wroclawski, J., Sollins, K., & Braden, R. (2005). Tussle in cyberspace: Defining tomorrow's Internet. *IEEE/ACM Transactions on Networking, 13*(3), 462–475.

Denning, S. (n.d.). What went wrong at Boeing? Retrieved from http://www.forbes.com/sites/stevedenning/2013/01/21/what-went-wrong-at-boeing

Dillard, J. T. (2003). *Centralized control of defense acquisition programs: A comparative review of the framework from 1987–2003*. Monterey, CA: Naval Postgraduate School.

Fleming, Q. W., & Koppelman, J. M. (1996*). Earned value project management* (1st ed.). Project Management Institute.

Fleming, Q., & Koppelman, J. (1994). The essence and evolution of earned value. Transactions of AACE International, 73.

Gantt, H. L. (1903). A graphical daily balance in manufacture. *Transactions of the ASME, 24*, 1322–1336.

GAO. (2006). *Defense acquisitions: Major weapon systems continue to experience cost and schedule problems under DoD's revised policy* (GAO-06-368). Washington, DC: Author.

GAO. (2009). *Assessments of selected weapon programs.* Report to congressional committees (GAO-09-326SP). Washington, DC: Author.

Garlan, D., & Shaw, M. (1993). An introduction to software architecture. In V. Ambriola, & G. Tortora (Eds.), *Advances in software engineering and knowledge engineering, volume 1.* New Jersey: World Scientific.

Gates, D. (2016, January 5). Will 787 program ever show an overall profit? Analysts grow more skeptical. Retrieved from http://www.seattletimes.com/business/boeing-aerospace/will-787-program-ever-show-an-overall-profit-analysts-grow-more-skeptical/

Goldratt, E. M. (1997). *Critical chain [A business novel].* Great Barrington, MA: North River Press.

Henderson, R. M., & Clark, K. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms [Special issue: Technology, Organizations, and Innovation]. *Administrative Science Quarterly, 35*(1).

Kelley, J. E., Jr., & Walker, M. R. (1959). *Critical-path planning and scheduling*. Paper presented at the Eastern Joint IRE-AIEE-ACM Computer Conference (pp. 160–173). ACM.

Khamooshi, H., & Golafshani, H. (2014). EDM: Earned duration management, a new approach to schedule performance management and measurement. *International Journal of Project Management, 32*(6), 1019–1041.

Lipke, W. (2003). Schedule is different. *PMI CPM Journal*, *The Measurable News*.

Lipke, W. (2004). Connecting earned value to the schedule. *The Measurable News, 1,* 6–16.

Lipke, W. (2012). Speculations on project duration forecasting. *The Measurable News, 3,* 3–7.

Moser, B., Grossmann, W., & Starke, P. (2015). Mechanisms of dependence in engineering projects as sociotechnical systems. In Transdisciplinary Lifecycle Analysis of Systems: *Proceedings of the 22nd ISPE International Conference on Concurrent Engineering* (Vol. 2, p. 142). IOS Press.

Moses, J. (2006). *Overview of organizational structures and system architectures* [ESD.342 class lecture notes]. Cambridge, MA: MIT.

Naval Sea Systems Command (PMS4252). (1994, September 30). *Acoustic program plan*.

Navy Contractor. (2001, March 15). More NUWC support items.

Navy Program Manager. (2001, March 5). NUWC impact to APB(T)-01.

Patanakul, P., Kwak, Y. H., Zwikael, O., & Liu, M. (2016). What impacts the performance of large-scale government projects? *International Journal of Project Management, 34*(3), 452–466.

Starke, P. (2015, June). *A New approach to understanding and measuring task interdependence* (Master's thesis). Munich, Bavaria, Germany: Technische Universität München.

Ulrich, K. T., & Eppinger, S. D. (1995). *Product design and development*. New York, NY: McGraw-Hill.

Wilson, J. M. (2003). Gantt charts: A centenary appreciation. *European Journal of Operational Research, 149*(2), 430–437.

Zarnich, R. W. (1996, September 3). Novel1.doc.

A Complex Systems Perspective of Risk Mitigation and Modeling in Development and Acquisition Programs

Roshanak R. Nilchiani—is an Associate Professor of Systems Engineering at the Stevens Institute of Technology. She received her BSc in mechanical engineering from Sharif University of Technology, MS in engineering mechanics from University of Nebraska–Lincoln, and PhD in aerospace systems from Massachusetts Institute of Technology. Her research focuses on computational modeling of complexity and system ilities for space systems and other engineering systems, including the relationship between system complexity, uncertainty, emergence, and risk. The other track of her current research focuses on quantifying, measuring, and embedding resilience and sustainability in large-scale critical infrastructure systems. [rnilchia@stevens.edu]

Antonio Pugliese—received his BSc and MSc in aerospace engineering from the University of Naples Federico II in Naples, Italy, in 2010 and 2012, respectively. He also received a postgraduate Master in Systems Engineering for Aerospace Applications (SEAMIAero) from the University of Naples Federico II in 2014. He is currently a PhD student in systems engineering at the Stevens Institute of Technology in Hoboken, NJ. He worked on winged atmospheric re-entry systems, focusing on flight quality and thermal protection systems. He also has experience as a systems engineer with attitude control systems and experimental payloads for sounding rockets. [apuglies@stevens.edu]

Abstract

The current methodologies used in risk assessment are heavily subjective and inaccurate in various life cycle phases of complex engineered systems. The increase in complexity has caused a paradigm shift from root cause analysis to the search of a set of concurrent causes for each event and the relevant complexity content of the system. Many of the system's life cycle risks are currently assessed subjectively by imprecise methodologies such as color-coded risk matrix, and subsequently they suffer from unforeseen failures as well as cost and schedule overruns. This research project proposes a novel approach to major improvement of risk assessment by creating a set of appropriate complexity measures (informed by historical case studies) as pre-indicators of emergence of risks at different stages of a systems development process, and also a framework that enables the decision-makers on assessing the actual risk level at each phase of the development based on requirements, design decisions, and alternatives. The goal of this research is to capture the complexity of the system with some innovative metrics, thus allowing for better decision-making in architecture and design selections.

Introduction

Engineered systems have become progressively more complex and interconnected to other various infrastructure systems over the past few decades, and they continue to become more complex. Examples of this can be seen in various fields of engineered systems, spanning from satellites, aircrafts, and missiles to ground transportation systems and sophisticated interconnected power and communication grids. In one perspective, more complexity provides more sophisticated multi-functionality to the engineered system at hand, while in a competing perspective, concurrently can make the system more vulnerable and fragile and prone to failures and emergent behavior. The relationship between excessive complexity in design and operation of complex engineered systems to the risk, emergence, and increased manifestation of failures has been acknowledged by many experts and academics in various engineering design communities. However, there is a lack of comprehensive research that enables the discovery of the relationship between the level of complexity of a design to increased risks and failure of that system. This research is an initial study in understanding, modeling, and suggesting relevant complexity measures in engineering design that can be used and linked quantitatively to the risk assessment of an engineered system.

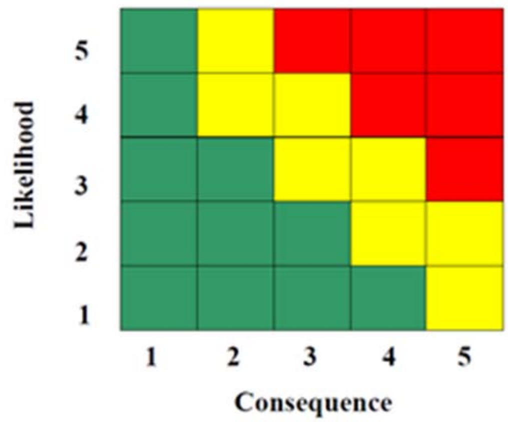


Figure 1. Traditional Risk Reporting Matrix
(DoD, 2006)

Risk can be defined as “a measure of future uncertainties in achieving program performance goals and objectives within defined cost, schedule and performance constraints” (DoD, 2006). In complex engineered systems as well as acquisition programs, often various types of risks exist that manifest themselves at different times throughout the development process. These risks can be technical, programmatic, or strategic in nature and can result in substantial cost overruns, delays, performance issues, reduced adaptability to changing requirements, or even total cancellation of a project. The major challenges of assessing risk using the traditional risk reporting matrices (Figure 1) for complex systems acquisition is that neither the likelihood nor the true consequence of a risk can be objectively established. Substantial uncertainty around the interactions among different components of a system as well as uncertainties across a multiplicity of interfaces. Also, often the symptoms and events after a failure or a problem manifest itself can be seen and are visible (Figure 2); however, the behavior and structure of the engineered system and the architecture and level of complexity of the engineered system that gives rise to such unforeseen events are often unknown. By making the complexity content and the architectural pattern of an engineered system known and explicit, in the next step of research we will be able to find the relationship between the underlying structure and complexity and the manifestation of risks and uncertainties in engineered systems.

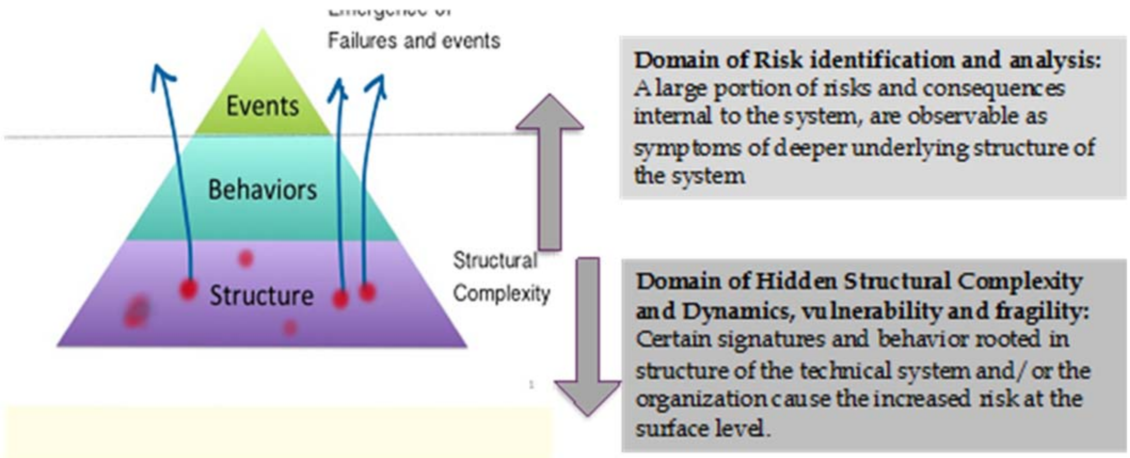


Figure 2. Problem Statement and Assessment of Structural Complexity as an Indicator of Risk and Failure Emergence

The objective of this research project is to create a quantitative and more objective assessment of technical risks and failures in engineered systems. This research aims to explore, formulate, and model the complex risks and failure mechanisms to improve the current inaccurate subjective assessment of risk in different stages of an engineered system development program as well as acquisition programs.

Literature Review

In this section of the paper, an overview of the current literature and state of the art of the complexity and complexity measurement of engineered systems as well as an overview of the literature on risk assessment of the complex engineered systems will be discussed briefly to provide a background of the current ongoing research by the authors. The literature review section begins with an overview of complex systems concepts, followed by various definitions of complexity and emergence, several current existing measures that are often being used in engineering systems designs. The section also presents a brief overview of risk assessment of complex engineered systems.

Risk Management of Complex Engineered Systems

> It is not possible to know exactly how a particular design will perform until it is built. But the product cannot be built until the design is selected. Thus, design is always a matter of decision making under conditions of uncertainty and risk. (Hazelrigg, 1998)

Risk and uncertainty are the hallmarks of all complex engineered systems. The Department of Defense in the *DoD Risk Management Guide* (DoD, 2006) defines risk as follows:

> Risk is a measure of future uncertainties in achieving program performance goals and objectives within defined cost, schedule and performance constraints. Risk can be associated with all aspects of a program (e.g., threat, technology maturity, supplier capability, design maturation, performance against plan). … Risk addresses the potential variation in the planned approach and its expected outcome.

In general, risks have three components, which are the root cause, a probability (or likelihood) assessed at the present time of the root cause occurring, and the consequence (or effect) of occurrence. Often a root cause is the most basic reason for the presence of a risk. Accordingly, risks should be tied to future root causes and their effects (DoD, 2006).

In any complex technical engineering project, risk can be classified as either of technical or programmatic nature, the former concerning performance criteria and the latter focusing on cost and schedule. Both types of risk are often modeled as the product of the probability of an event and its severity (Pennock & Haimes, 2002). In modeling risk, one can also consider the future root cause (yet to happen) of a certain event (Nilchiani et al., 2013), which is where one is supposed to act in order to eliminate a specific risk. Severity and probability are traditionally represented on the widely utilized, color-coded, risk matrix. Figure 1 shows a color-coded risk matrix. Unfortunately, this seemingly quantitative tool hides subjectivity in the estimation of event frequency and severity, and for those reasons is "inapt for today's complex systems" (Hessami, 1999). This not only means that most of the systems that we build today cannot be built with the tools and processes from last century, but also that we have started building in a domain where structural patterns matter, especially for large projects.

Complex Systems

Complexity has been one of the characteristics of many large-scale engineered systems of the past century. Complex engineered systems can provide sophisticated functionality as one side of the coin, and the other side can cause the system to be more prone to unwanted emergent behaviors and more fragility to the engineered system. The field of complexity is rich and spans over the past half century in various fields of knowledge ranging from biological systems to cyber-physical systems. As it has been discussed by several researchers, a strong correlation can be observed between the complexity of the system and various ranges of failures, including catastrophic failures (Cook, 1998; Bar-Yam, 2003; Merry & Kassavin, 1995).

In 1948, Warren Weaver, a pioneer in classifying and defining complexity in systems, described three distinct types of problems: problems of simplicity, problems of disorganized complexity, and problems of organized complexity (Weaver, 1948).

According to Weaver (1948), problems of simplicity are the problems with a low number of variables that have been tackled in the 19th century. An example is the classical Newtonian mechanics, where the motion of a body can be described with differential equations in three dimensions. In these problems, the behavior of the system is predicted by integrating equations that describe the behavior of its components. In the same article, Weaver discusses that problems of disorganized complexity are the ones with a very large number of variables that have been tackled in the twentieth century. The most immediate example is the motion of gas particles, or as an analogy the motion of a million balls rolling on a billiard table. The statistical methods developed are applicable when particles behave in an unorganized way and their interaction is limited to the time they touch each other, which is very short. In these problems it has been possible to describe the behavior of the system without looking at its components or the interaction among them.

Problems of organized complexity are the ones that are to be tackled in the 21st century, and ones that see many variables showing the feature of organization. These problems have variables that are closely interrelated and influence each other dynamically. This high level of interaction that gives rise to organization is the reason that these problems cannot be solved easily. Weaver described them as solvable with the help of powerful calculators, but today's technology is not yet able to solve the most complex of these problems. These are the problems that nowadays we define as "complex."

Predicting the behavior of a system with many interconnected parts changing their behavior according to the state of other components is a problem of organized complexity, and the system itself is a referred to as a complex system.

Cotsaftis (2009) gives a way of determining whether a system is simple, complicated, or complex by looking at its network model (i.e., nodes and edges). The model defines three types of edges: a free flight state vertex V_{ii}, a driven state from outer source vertex V_{ie}, and an interactive state with other system components vertex V_{ij}. The edges are channels along which there is a resource flux p_{ii}, p_{ie}, or p_{ij}. When

$$p_{ii} \gg p_{ij}, \inf p_{ij},, \quad (1)$$

the i^th component is weakly coupled with the others, external and internal. The dynamics of the component can in this case be considered independent from the other components. If the majority of the components satisfy inequality (1) the system is considered to be simple. When

$$p_{ie} \gg p_{ii}, \inf p_{ij}, \quad (2)$$

the i^th component is depending on outside sources. The system can still be partitioned in a set of weakly connected subsystems which dynamics is determined from outside sources. If the majority of the components satisfy inequality (2) the system is considered to be complicated. When

$$\inf p_{ij} \gg p_{ii}, p_{ie} \quad (3)$$

the i^th component is strongly connected to the others, and its dynamics cannot be determined without considering the effects of the other components. Also, the manipulation of the system cannot be performed as in the previous cases, since the internal connections create conditions that reduce the number of degrees of freedom. A system with a reduced number of external control dimensions that satisfies inequality (3) is said to be complex.

This definition is rather qualitative, since not all the nodes in the system have the same importance (in terms of connection number and intensity) and therefore it makes no sense to consider the majority. For this reason Cotsaftis defines the index of complexity as $C_S = n/N$, where n is the number of components that satisfy inequality (3) and N is the total number of components. A complicated system has $C_S = 0$. $C_S = 1$ corresponds to the most complex system possible, but it is also a system where external connections are negligible, and therefore the system is isolated. This is due to the fact that a complex system is describable with a low number of parameters if seen from outside, but has high connectivity in its internal structure.

Considering as an example a sheepdog and a herd of cattle, we realize that the dog has only two degrees of freedom while the herd has $2n$, where n is the number of animals in the herd. By pushing the cattle together, the dog increases their interactions and decreases the number of degrees of freedom of the herd to only two, therefore being able to control it.

The research from these two authors has shown us how complexity and simplicity are interrelated concepts, somehow opposite, but that can also be found in the same system at the same time, depending on the point of view. Madni made a distinction between systemic elegance, which "thrives on simplicity through minimalistic thinking and parsimony" and perceived elegance, which "hides systemic or organizational complexity from the user." If the system is considered to be complex but its complexity can be somehow hidden or resolved, thus making it simpler, then the design can be considered elegant (Madni, 2012). Therefore, in order to achieve a more elegant design, we need to decrease the complexity of the system.

Emergence

Emergence is a major phenomenon related to complex engineered systems. Emergence at the macro-level is not hard-coded at the micro-level (Page, 1999). One example of emergence in natural systems is wetness. Water molecules can be arranged in three different phases (i.e., solid, liquid, and gas), but only one of them expresses a particular type of behavior, which is high adherence to surfaces. This behavior is due to the intermolecular hydrogen bonds that affect the surface tension of water drops. These bonds are also active in the solid and liquid phase, but in those cases they are either too strong or too weak to generate wetness. In this case, the emergence of a property, such as wetness, has been explained at a lower level by looking at the molecules that make up the liquid.

According to Kauffmann (2007), two different types of emergence exists (Kauffman, 2007). The reductionist approach sees emergence as epistemological, meaning that the knowledge about the systems is not yet adequate to describe the emergent phenomenon, but it can improve and explain it in the future. This is the case of wetness, where knowledge about molecules and intermolecular interactions has explained the phenomenon. On the other hand, there is the ontological emergence approach, which says that "not only do we not know if that will happen, [but] we don't even know what can happen," meaning that there is a gap to fill not only about the outcome of an experiment (or process), but also about the possible outcomes.

Longo presents this view with the example of the swimming bladder in fishes (Longo, Montevil, & Kauffman, 2012). An organ that gives neutral buoyancy in the water column as its main function, also enables the evolution of some kinds of worms and bacteria that will live in it. Ontological (or radical) emergence is given by the enormous amount of states the system could evolve into. In these cases we not only are not able to predict which state will happen, but we do not even know what the possible states are.

Gell-Mann also pointed out this difference using the concept of logical depth (Gell-Mann, 1995). When some apparently complex behavior can be expressed with simpler laws that reside at a lower level (e.g., the complicated pattern of energy levels of atomic nuclei that can be described at the subatomic level), the phenomenon is said to have a substantial amount of logical depth.

In our research, the emergence that is going to be tackled is considered to be epistemological emergence, logical depth according to Gell-Mann, where knowledge about the system organizational patterns and internal structure can lead to the explanation of certain phenomena. Unfortunately this concept is not so common in the systems engineering and risk management fields, and therefore this research adopts the industry jargon by talking about complexity and complex systems, but always reminding that we are actually trying to unravel logical depth from a systems engineering perspective.

Definitions and Measures of Complexity

There are various definition of complexity that have roots in various fields spanning from mathematics and biology to engineering design. In a recent paper, Wade (2014) suggests that existing complexity definitions belong to one of three types: behavioral, structural, or constructive. Behavioral definitions view the system as a black box and the measures of complexity are given based on the outputs of the system. Structural definitions look at the internal structure or architecture of the system. Constructive definitions see complexity as the difficulty in determining the system outputs (Wade & Heydari, 2014). In this research we are interested in the modeling behavioral and structural complexity metrics. A summary of behavioral complexity definition as well as structural complexity and some measures are presented in the following sections of the literature review.

Behavioral Complexity Definitions and Metrics

The most famous behavioral complexity metric is with no doubt Shannon's entropy (Shannon, 1948). This metric evaluates the complexity by measuring the entropy of the output message of the system (this metric was initially applied to information systems).

Gell-Mann used Shannon's entropy to define information measure as a metric capable of measuring both the effective complexity, which is the amount of information necessary to describe the identified regularities of an entity, and the total information, which also takes into account the apparently random features (Gell-Mann & Lloyd, 1996). Algorithmic information content and Shannon entropy are used to build this metric. The former is responsible for measuring the effective complexity (knowledge), and the latter the random parts (ignorance). This dual approach is an interesting contribution to the measurement of complexity, since it allows one to group similar entities according to their effective complexity and to measure the diversity of the ensemble as entropy.

Chaisson (2004) proposed a specific energy-based measure of complexity—more precisely, energy rate density, which is "the amount of energy available for work while passing through a system per unit time and per unit mass" (Chaisson, 2015). This metric looks at the system as a black box and measures the net energy amount entering the system. It has been evaluated for multiple entities such as galaxies, stars, planets, plants, animals, societies, and technological systems, and also has been mapped throughout their lifetime showing an increase in complexity (Chaisson, 2014).

Willcox et al. (2011) defined complexity as "the potential of a system to exhibit unexpected behavior in the quantities of interest, regardless of whether or not that behavior is detrimental to achieving system requirements." She proposed an entropy and probability based metric:

$$C(Q) = \exp(h(X)) \tag{4}$$

where *X* is the joint distribution of the quantities of interest, and $h(X)$ is the differential entropy of *X* defined as

$$h(X) = -\int_{\Omega_X} f_x(x) \log f_x(x)\, dx \tag{5}$$

where Ω_X is the support of *X*.

Structural Complexity Definitions and Metrics

There are a few structural complexity measures in current complex engineering systems in recent decades. The metric presented by Cotsaftis (2009) is an example of structural complexity metric, since it looks at the internal structure of the system (i.e., components and interfaces).

Another structural complexity metric was presented by McCabe for software systems (McCabe, 1976). The representation of computer programs using graphs allows one to define the cyclomatic number $v(G)$ as

$$v(G) = e - n + 2p \tag{6}$$

where G is the graph, e is the number of edges, n is the number of nodes, and p is the number of connected components. This same metric has been extended to measure architectural design complexity of a system (McCabe & Butler, 1989).

Sinha presented a structural complexity metric that uses the design structure matrix (DSM) of a system to evaluate its complexity (Sinha & de Weck, 2012). The metric is evaluated using

$$C(n,m,A) = \sum_{i=1}^{n} \alpha_i + \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} A_{ij} \right) \gamma E(A) \quad (7)$$

where n is the number of components in the system, m the number of interfaces, A the DSM, α_i the complexity of each component, $\beta_{ij} = f_{ij}\alpha_i\alpha_j$ the complexity of each interface, $\gamma = 1/n$ a normalization factor, and $E(A)$ the matrix energy of the DSM. Although the proposed metric is very sophisticated, its application sees the evaluation of α_i through expert judgment, and $f_{ij} = 1$ for lack of more information (Sinha & de Weck, 2013). One interesting feature of this metric is the topological complexity $E(A)$, which represents the level of robustness and reliability of the graph network and can be easily evaluated from the DSM through singular value decomposition.

Hybrid Structural-Behavioral Complexity Framework

The goal of this research is to develop a framework for the identification of complexity level of the engineered system and architectural patterns affecting the behavior of the system and various levels of risks. The framework will be applied at the initial design phase, when system requirements are defined, and the system architecture is in its initial development (some hierarchical levels are defined but not all of them).

Our suggested framework is based on two main ideas. The first one is decomposition. According to McCabe, the complexity of a collection of unconnected control graphs is equal to the summation of their complexities (McCabe, 1976). Wade pointed out that in complex systems, reduction by decomposition cannot work since the behavior of each component depends on the behaviors of the others (Wade & Heydari, 2014). This is true for complex engineered systems, but in this research we are tackling logical depth, and therefore we assume that the reductionist approach, as described by Kauffman (2007) can be applied to the problem.

The second idea is that it is possible to measure the complexity of an entity at its boundary. We have seen that various behavioral complexity metrics have been proposed. These metrics consider the system as a black box and only take into account its output. In this research we are going to consider not the output, but the relationship between output and input, as we believe it better describes what the system does.

Framework Application Approach

In order to measure the system complexity, the framework will combine the complexity of components that make up the subsystems at various architectural levels. This combination can be performed applying a structural complexity metric, which considers the system architecture (usually represented as a DSM or adjacency matrix) and the complexity of each component at a certain hierarchical level. The complexity of a subsystem can be evaluated with this approach, assuming that the complexity of its components and its internal structure are known. The process can be repeated upwards in the hierarchy to evaluate the complexity of the system.

At this point, this framework can use all the other structural complexity metrics already available in literature. The existing complexity measures in literature assume that the complexity of each component is already known, or if that's not the case, that it can be evaluated using expert judgment or historical data. In the creation of this framework we have attempted to remove the majority of the sources of subjectivity.

Given that the architecture is not completely defined, there will be some components that are not more than black boxes. The complexity of these components can be measured with behavioral metrics. Of course, historical data about input and output of these components in past projects will be necessary in order to evaluate the metrics, but the subjectivity coming from expert judgment will be removed. Also, there is a difference between using historical data such as input and output, which for engineered systems are physical quantities, and historical data such as rate of failure, or schedule delays due to integration, which depend on the history of the systems they are derived from.

The application of this framework can be divided into five main phases:

1. The architecture needs to be defined. It is important that there is no connection between components (or functions) at different levels, or even between components that are children of different subsystems. The only type of connection allowed for the decomposition principle to be valid is between components within the same subsystem.
2. Once the architecture is defined, it is necessary to characterize the boundary of each component. The interfaces with other components within the same subsystem need to be quantitatively classified, in order to be used in a behavioral evaluation.
3. Once the interfaces are defined and characterized according to their behavior, the complexity of each black-box component can be evaluated using a behavioral complexity metric.
4. The complexity of each subsystem is then evaluated using a structural complexity metric, from the complexity of its components and information about its internal structure.
5. Once the complexity of the lowest level components (i.e., the leaves of the hierarchy tree) is evaluated, it can be combined in a bottom-up approach to evaluate the complexity of the higher level subsystems by repeating the previous steps until the complexity of the overall system is evaluated.

This framework has been built with flexibility in mind, meaning that the interface characterization model, the behavioral metric, and the structural metric are supposed to be plugged in according to the specific characteristics of the enterprise building the system, and the type of system. We have attempted to remove the majority of the subjectivity from the evaluation, since the level of accuracy depends heavily on the level of experience of the experts, but we want to retain the knowledge that any system architect has about the system that its enterprise is comfortable building. Two senior system architects are going to evaluate architectures differently, according to their experience and the experience of the people they worked with, thus naturally picking the best choice for the enterprise they work for. Just as likely, the framework can be adapted to rate as "better designed" the architectures having traits that the enterprise successfully implemented in past projects. Figure 3 shows a summary of the hybrid structural-behavioral framework.

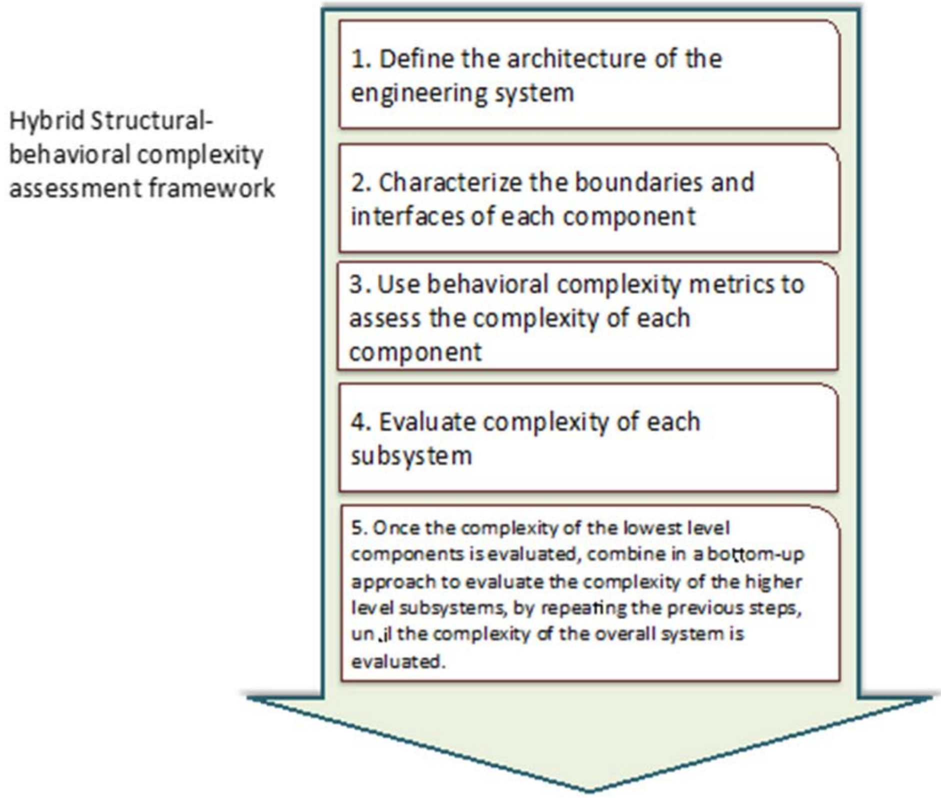


Figure 3. Schematics of the Hybrid Structural-Behavioral Complexity Assessment Framework

Part of this research effort is devoted to generating the modules (interface models, behavioral metrics, and structural metrics) that will then be used in the framework, and also to understanding which set of modules will give the best fit for each specific enterprise.

Interface Characterization Model

The connections between the components of an engineered systems are of various natures and often incommensurable. For example, considering two components having a mechanical and a thermal interface: Is it better to have low mechanical stresses and high thermal fluxes, or vice versa? In order to answer this question, the interfaces need to be classified in a scale that allows comparison between them even when they are of different natures. This will enable the evaluation of many structural and behavioral metrics that include interface complexity.

Currently this model is still under refinement. The assumptions are based on the idea that connections can be ranked in terms of how enabling they are towards a specific goal. As an example, consider the two groups of animals depicted in Figure 4.

Figure 4. Herd of Sheep and Army of Ants
Note. These two groups of animals are examples of constraining and enabling interactions.

Both the herd of sheep and the army of ants are a group of animals that interact with each other. Here the interaction of interest is the purely mechanical one. This type of interaction is constraining in the case of the herd, since it decreases the degrees of freedom of the system. This also happens in the case of the army of ants, but in this case the system has gained in capabilities (i.e., the ability to bridge in mid-air). The emergence of this capability is given by the enabling nature of the mechanical connection. The goal of this part of the research regarding interface modeling is to develop a metric for the evaluation of the level of enablement of any interface towards a specific component, within engineered systems.

Use Case: Satellite Attitude Control System

In order to show how the framework can measure the complexity of a system, we have applied the initial framework to the architecture of an Attitude Control System (ACS) for a satellite. The preliminary architecture is represented in Figure 5.

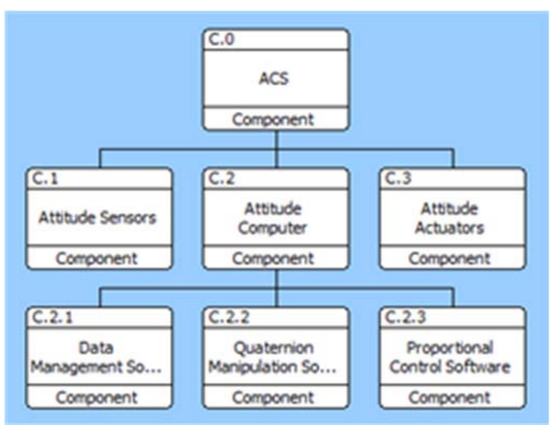


Figure 5. Hierarchical Representation of the Architecture of the ACS

The component C.0, in this case the ACS, is made up of three components—C.1, C.2, and C.3—which are the attitude sensors, attitude computer, and attitude actuators, respectively. For the sake of this example, the architecture of the component C.2 has been laid out only for its software. This architectural level includes components C.2.1, C.2.2, and C.2.3, namely data management software, quaternion manipulation software, and proportional control software. The physical architecture presented in Figure 6 has a one-to-one mapping with the functional architecture, and therefore, for the purposes of this example, they are considered as equivalent.

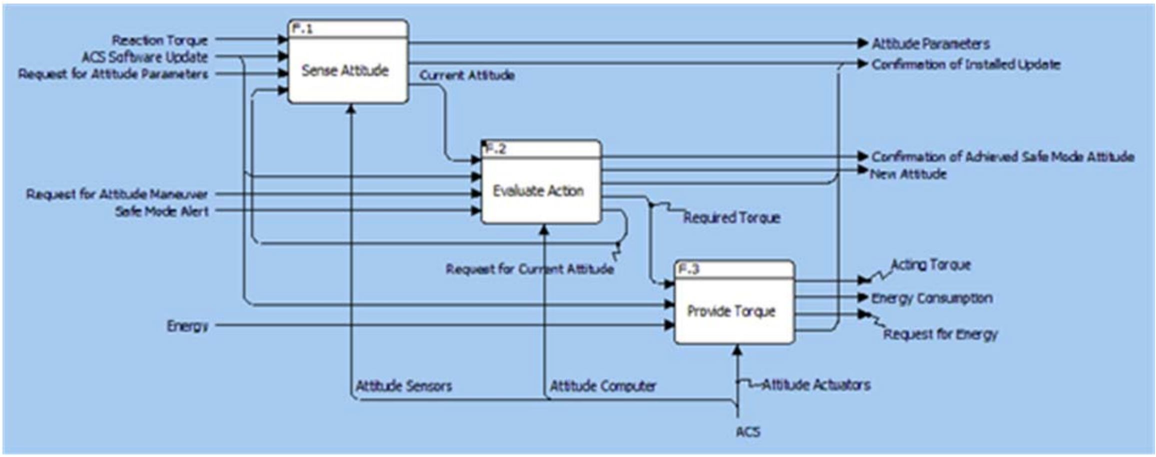


Figure 6. IDEF0 Representation of the F.0 Function Corresponding to the C.0 Component, the ACS

A hierarchical representation of the system architecture is not enough for the application of the framework. The interfaces between the components also need to be defined. Figure 6 shows these interfaces within the F.0 function. The interactions have been defined on the basis of four use cases: attitude maneuver, safe mode attitude maneuver, provide attitude parameters, and ACS software update. The information reported in Figure 6 allows us to build an adjacency matrix for the components of C.0 that can be used in the evaluation of any structural complexity metric.

$$A_{C.0} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{8}$$

In this example, the complexity metric proposed by Sinha & de Weck (2013),

$$C(n,m,A) = \sum_{i=1}^{n} \alpha_i + \left(\sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} A_{ij} \right) \gamma E(A) \tag{9}$$

will be used to evaluate the complexity of the C.0 component $C_{C.0}$. In this case $\alpha_i = C_{C.i}$, $\gamma = 1/3$ can be evaluated using singular value decomposition and taking the sum of the diagonal values $E(A_{C.0}) = 1 + \sqrt{2}$. Equation 9 then becomes

$$C_{C.0} = C_{C.1} + C_{C.2} + C_{C.3} + \frac{1+\sqrt{2}}{3}(\beta_{12} + \beta_{21} + \beta_{23}). \tag{10}$$

Equation 10 still has many unknown variables, which need to be computed. β_{12}, β_{21}, and β_{23} can be evaluated using the interface characterization model. The evaluation of $C_{C.2}$ has the same structural approach of $C_{C.0}$, since its internal architecture has been already defined. The hybrid nature of this framework allows consideration of the most information available, evaluating the complexity of components with already defined internal structure using structural complexity metrics that take the aforementioned structured into account.

$C_{C.1}$ and $C_{C.3}$ can be evaluated using a behavioral complexity metric. This approach is necessary since these components are only defined as black boxes and we only have information about their input and output. Evaluating the complexity of C.1 using an approach based on Chaisson's metric is taken at this stage. The metric considers the energy that the component exchanges. In the case of engineered systems, energy can be exchanged in a variety of ways (e.g., chemical, data, mechanical, thermal). The evaluation of this exchange is also part of the interface characterization model under development in this research.

In order to understand the dependency of the structural complexity on the interfaces, we can modify the architecture of F.0 by adding a connection between F.1 and F.3. In this case, the new component C.O will have an adjacency matrix:

$$A_{C.0'} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{11}$$

This leads to a different value of the matrix energy $E(A_{C.0'}) = 1 + \sqrt{3}$ and thus to a new formulation for the complexity of the component:

$$C_{C.0} = C_{C.1} + C_{C.2} + C_{C.3} + \frac{1+\sqrt{3}}{3}(\beta_{12} + \beta_{13} + \beta_{21} + \beta_{23}) \tag{12}$$

This change in the architecture increases the complexity of the component. Other structural complexity metrics such as the metric proposed by Sinha cannot capture this change properly, since an addition of a single connection between two components leads in this case to two changes in the complexity evaluation. For this reason, in this research we will continue to propose modifications to existing complexity metrics so that the overall framework can lead to more meaningful evaluations.

Summary and Future Work

In this research we propose a framework to perform a quantitative and more objective assessment of complexity level, as a major precursor to assessing objective technical risks and failures in engineered systems. This is part of a larger research vision and objective of a theoretical model of failure mechanisms and risks in engineered systems, which is based on the complexity content of the system. This part of our research focuses on the preliminary design phase complexity assessment and follows and builds upon the previous work by Salado and Nilchiani (2012) on the complexity assessment of requirements and its translation in risks and vulnerability assessment. The new framework suggested, once completed, will be applicable to both development and acquisition programs, as long as the system architecture is partially available.

References

Bar-Yam, Y. (2003). When systems engineering fails: Toward complex systems engineering. *IEEE International Conference on Systems, Man and Cybernetics, 2*, 2021–2028.

Chaisson, E. J. (2004). Complexity: An energetics agenda. *Complexity, 9*(3), 14–21.

Chaisson, E. J. (2014). The natural science underlying big history. *The Scientific World Journal.*

Chaisson, E. J. (2015). Energy flows in low-entropy complex systems. *Entropy, 17*(12), 8007–8018.

Cook, R. I. (1998). How complex systems fail. *Cognitive Technologies Laboratory*. Chicago IL: University of Chigaco.

Cotsaftis, M. (2009). What makes a system complex? An approach to self organization and emergence. In *From System Complexity to Emergent Properties* (pp. 49–99). Springer.

DoD. (2006). *Risk management guide for DoD acquisition*. Washington, DC: Author.

Edmonds, B. (1995). What is complexity? The philosophy of complexity per se with application to some examples in evolution. *The Evolution of Complexity*.

Gell-Mann, M. (1995). What is complexity? Remarks on simplicity and complexity by the Nobel Prize-winning author of *The Quark and the Jaguar. Complexity, 1*(1), 16–19.

Gell-Mann, M. (2002). What is complexity? In *Complexity and Industrial Clusters,* 13–24. Springer.

Gell-Mann, M., & Lloyd, S. (1996). Information measures, effective complexity, and total information. *Complexity, 2*(1), 44–52.

Griffin, M. D. (2010). How do we fix system engineering? *International Astronautical Congress.*

Hazelrigg, G. A. (1998). A framework for decision-based engineering design. *Journal of Mechanical Design, 120*(4), 653–658.

Hessami, A. (1999). Risk management: A systems paradigm. *Systems Engineering, 2*(3), 156–167.

Kauffman, S. (2007). Beyond reductionism: Reinventing the sacred. *Zygon, 42*(4), 903–914.

Longo, G., Montevil, M., & Kauffman, S. (2012). No entailing laws, but enablement in the evolution of the biosphere. In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*, 1379–1392.

Madni, A. M. (2012). Elegant systems design: Creative fusion of simplicity and power. *Systems Engineering, 15*(3), 347–354.

McCabe, T. J. (1976). A complexity measure. *IEEE Transactions on Software Engineering,* (4), 308–320.

McCabe, T. J., & Butler, C. W. (1989). Design complexity measurement and testing. *Communications of the ACM, 32*(12), 1415–1425.

Merry, U., & Kassavin, N. (1995). *Coping with uncertainty: Insights from the new sciences of chaos, self-organization, and complexity*. Praeger/Greenwood.

Nilchiani, R., Mostashari, A., Rifkin, S., Bryzik, W., & Witus, G. (2013). *Quantitative risk—Phase 1.* Systems Engineering Research Center–University Affiliated Research Center.

Page, S. E. (1999). Computational models from A to Z. *Complexity, 5*(1), 35–41.

Pennock, M., & Haimes, Y. (2002). Principles and guidelines for project risk management. *Systems Engineering*, 89–108.

Salado, A., & Nilchiani, R. (2012). A framework to assess the impact of requirements on system complexity.

Salado, A., & Nilchiani, R. (2013a). Elegant space systems: How do we get there? In Proceedings of the *IEEE Aerospace Conference* (pp. 1–12). IEEE.

Salado, A., & Nilchiani, R. (2013b). Using Maslow's hierarchy of needs to define elegance in system architecture. *Procedia Computer Science, 16*, 927–936.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.

Sinha, K., & de Weck, O. L. (2012). Structural complexity metric for engineered complex systems and its application. Gain Competitive Advantage by Managing Complexity: *Proceedings of the 14th International DSM Conference Kyoto, Japan,* (pp. 181–194).

Sinha, K., & de Weck, O. L. (2013). A network-based structural complexity metric for engineered complex systems. *IEEE International Systems Conference (SysCon)* (pp. 426–430).

Wade, J., & Heydari, B. (2014). Complexity: Definition and reduction techniques. *Proceedings of the Poster Workshop at the 2014 Complex Systems Design & Management International Conference* (pp. 213–226).

Weaver, W. (1948). Science and complexity. *American Scientist, 36*(4), 536–544.

Willcox, K., Allaire, D., Deyst, J., He, C., & Sondecker, G. (2011). *Stochastic process decision methods for complex-cyber-physical systems* (Technical report, DTIC document).

Panel 17. Reducing Life-Cycle Costs: Adopting Emerging Manufacturing Technologies

Thursday, May 5, 2016	
1:45 p.m. – 3:15 p.m.	**Chair: Major General L. Neil Thurgood, U.S. Army,** Deputy for Acquisition and Systems Management, OASA (ALT) **Discussant: Michael Schwind,** Vice President of Federal Sector, Siemens PLM ***Benchmarking Naval Shipbuilding With 3D Laser Scanning, Additive Manufacturing, and Collaborative Product Lifecycle Management*** David Ford, Associate Professor, Texas A&M University Tom Housel, Professor, NPS Sandra Hom, Research Associate, NPS Jonathan Mun, Research Professor, NPS ***Materials Testing and Cost Modeling for Composite Parts Through Additive Manufacturing*** Eric Holm, Chief of Command Civil Engineer Inspections, Air Force Materiel Command Vhance Valencia, Assistant Professor, Air Force Institute of Technology Alfred Thal, Jr., Associate Professor, Air Force Institute of Technology Jason Freels, Assistant Professor, Air Force Institute of Technology Adedeji Badiru, Dean, Air Force Institute of Technology

Chair: Major General L. Neil Thurgood, U.S. Army—is the Deputy for Acquisition and Systems Management, Office of the Assistant Secretary of the Army, Acquisition, Logistics, and Technology. Prior to promotion from brigadier general to major general, Thurgood served as the Program Executive Officer for Missiles and Space at Redstone Arsenal, where he was responsible for the development, production, fielding, and life-cycle management of the Army's missile and space-related systems.

Enlisting in the Army in April 1983, Thurgood was commissioned in 1986 and spent much of his career flying Chinooks and serving as an aviator with the 213th Combat Aviation Company, the XVIII Airborne Corps, and the 160th Special Operations Aviation Regiment (Airborne), among other assignments, prior to taking on his current role.

He graduated from the University of Utah with a bachelor's degree in business management and communication. MG Thurgood holds a master's degree in systems management from the Naval Postgraduate School, a master's degree in strategic studies, a doctorate in strategic business and leadership, and several professional certificates including the Legion of Merit (3 OLC), Meritorious Service Medal (4 OLC), and the Air Medal (3).

Discussant: Michael Schwind—is the Vice President of Siemens PLM's Federal Sector. In this capacity, he has responsibility for the execution of business development, strategic partnerships, and consulting for lifecycle management systems to the Department of Defense and other government agencies.

Over the past 30 years, Schwind's professional career has encompassed the engineering and manufacturing lifecycle industries. His tenure with Siemens commenced in 1998 under the former

McDonnell Douglas Corporation. At McDonnell Douglas, Schwind assumed increasing levels of responsibility, progressing from Client Executive to Director of Sales and Marketing, Mid-Atlantic sector. Immediately following the Siemens acquisition in 2007, Schwind assumed the role of VP Federal Sector, Siemens PLM division.

Schwind has been a guest speaker at multiple industry events including at the Naval Postgraduate School, and he is a current member of Surface Navy Association and the American Society of Naval Engineers. Schwind earned a bachelor’s degree in industrial distribution from Clarkson University. He received his certificate in Six Sigma from Villanova University in 2009.

Benchmarking Naval Shipbuilding With 3D Laser Scanning, Additive Manufacturing, and Collaborative Product Lifecycle Management

David N. Ford—received his BS and MS from Tulane University and his PhD from MIT. He is a professor in the Construction Engineering and Management Program, Zachry Department of Civil Engineering, Texas A&M University and a Research Associate Professor of Acquisition with the Graduate School of Business and Public Policy at the U.S. Naval Postgraduate School in Monterey, CA. For over 14 years, he designed and managed the development of constructed facilities in industry and government. [DavidFord@tamu.edu]

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

Sandra Hom—is a Research Associate at the Naval Postgraduate School (Monterey, CA) and specializes in market structures, industry benchmarking research, and knowledge value added analysis. [schom@nps.edu]

Jonathan Mun—is a research professor at the U.S. Naval Postgraduate School (Monterey, CA) and teaches executive seminars in quantitative risk analysis, decision sciences, real options, simulation, portfolio optimization, and other related concepts. He received his PhD in finance and economics from Lehigh University. He has also researched and consulted on many DoD and Department of Navy projects and is considered a leading world expert on risk analysis and real options analysis. He has authored 12 books. [jcmun@realoptionsvaluation.com]

Abstract

Evolving threats and shrinking budgets require that the Navy adopt and implement new technologies effectively and efficiently. The current work estimates the potential cost savings of the adoption and implementation of three advanced technologies: Three Dimensional Scanning (3DLS), Product Lifecycle Management (PLM), and Additive Manufacturing (AM). A review of the capabilities and current uses of the technologies is the basis for modeling their impacts on shipbuilding operations. Knowledge Value Added models were then used to estimate returns on investment without and with the technologies. These results were used to estimate shipbuilding cost savings over the life of the current U.S. Navy shipbuilding plan. Finally, strategic real options were developed and valued to incorporate implementation flexibility into cost savings estimates. Results indicate that the U.S. Navy can save an average of over $2.70 billion per year over 29 years if the potential improvements available through 3DLST, PLM, and AM are fully exploited, regardless of the implementation approach. If implemented fully and immediately, these three new technologies can save the U.S. Navy $3.07 billion, or $3.37 billion if implemented sequentially.

Introduction

The U.S. Navy estimates that it will cost $16.7 billion per year for new-ship construction to become a 306 battle force ship over the next 30 years. It is critical that the Navy capture full benefits of new technologies such as Three Dimensional Scanning (3DLS), Product Lifecycle Management (PLM), and Additive Manufacturing (AM) to reduce costs while meeting mission needs. Research supports the adoption and use of these commercially available technologies, yet does not address their use in naval shipbuilding. Cost savings estimates and strategies for technology adoption and use are important to capturing the full benefits of these technologies.

Our research project examines the use of 3DLS, PLM, and AM by non-shipbuilding industries as a basis for estimating potential naval shipbuilding savings. Secondary research was conducted on the three technologies used by various industries, and three models were developed on the potential cost and efficiency savings that could be derived from the use of those technologies. Recommendations are provided to Navy planners concerning the most effective and efficient strategy for exploiting these technologies.

The U.S. Navy will become a 306 battle force ship over the next 30 years, up from today's battle force of 289. A report of the Navy's 2015 shipbuilding plan covering fiscal years 2015 to 2044, submitted to Congress in July 2014, estimates that the plan will cost the Navy an average of about $16.7 billion per year in constant FY2014 dollars to implement. The Navy plans to buy a total of 264 ships over the 2015–2044 period under the 2015 plan. According to the CBO, given the rate at which the Navy plans to retire ships from the fleet, that construction plan would not achieve a fleet equal to the inventory goal of 306 ships until 2019 under new rules for counting ships that the Navy implemented this year, or until 2022 under the old counting rules. The adoption and full utilization of three advanced technologies (3D Laser Scanning, Additive Manufacturing, and Product Lifecycle Management) can potentially generate significant cost saving in the naval shipbuilding program. Those technologies are described next as the basis for the current evaluation of potential savings.

Product Lifecycle Management

Product Lifecycle Management (PLM) is defined as an

> integrated, information-driven approach comprised of people, processes/practices, and technology, to all aspects of a product's life, from its design through manufacture, deployment and maintenance—culminating in the product's removal from service and final disposal. By trading product information for wasted time, energy, and material across the entire organization and into the supply chain, PLM drives the next generation of lean thinking. (Greives, 2006)

PLM has been used by the automotive, aerospace, and other industries that build very large, very complex products and systems. It was designed to provide stakeholders with current views of every product throughout its lifecycle to facilitate decision-making and corrective actions if necessary.

PLM can be used in shipbuilding to build and maintain the next generation of ships. It spans the entire shipbuilding enterprise and lifecycle to enable shipbuilders to integrate organizational knowledge, automate processes throughout the product lifecycle and improve efficiency, accuracy and execution to reduce time to delivery. PLM can

- Provide shipbuilders and suppliers with access to relevant data.
- Achieve greater performance, lower ownership cost, offer higher fleet availability and reliability, and greater quality and compliance with the latest marine safety and regulatory requirements.
- Make ships easier to build and repair, lowering construction, service, and total ownership costs.
- Link shipbuilders with suppliers linked in the production schedule and all design aspects.

A wide range of industries using PLM are finding that 3DLS is becoming a critical tool to link the gap between physical objects in the real world and in the digital design world. The aerospace, automotive, consumer products, manufacturing, and heavy industries all have benefited from faster time to market, improved quality, and reduced warehousing costs with 3D scanning.

3D Laser Scanning

3D laser scanning technology has been used to achieve significant cost savings, optimize maintenance schedules, increase quality, improve safety and reduce rework. Commercial applications range from maritime and space applications to manufacturing and production. According to industry analysts, the industry's growth is fueled by the growing recognition that 3D aids in the design, fabrication, construction, operations and maintenance processes. Benefits of 3D laser scanning can be applied to shipbuilding.

Laser scanners use infrared laser technology to produce exceedingly detailed three-dimensional images of complex environments and geometries in only a few minutes. Millions of discrete measurements can be captured in every scan using 3D laser scanner technology. The resulting images, a cloud, are millions of 3D measurement points. A complete project may contain hundreds of millions or even billions of points, recreating the complex spatial relationships of the 3D environment.

Often used by offshore oil and gas companies to construct and repair oil rigs, 3DLS is very effective at documenting oil platforms and refineries to assist in engineering, maintenance, and planning processes. The aerospace and automotive industries have used 3DLS for retrofitting floors and measure parts for accurate fit. The DoD has tested 3DLS in several projects, as described next.

Ship Check Data Capture Projects 2005 & 2006

NSRP funded two Ship Check Data Capture projects in 2005 and 2006. Objectives of both Ship Check Data Capture projects were to

1. Develop a process that captures the as-built measurement data in digital/electronic format during a ship check
2. Process the as-built measurement data into 3D CAD models using available commercial-off-the-shelf (COTS) modeling technologies (software and hardware)
3. Provide a building block process for the anticipated development of the capabilities to generate 3D CAD models of the as-built space envelope from the geometric measurement data captured during the ship check.

Ship Check Data Capture 2005

Recognizing the potential of new technologies on the ship check process on the U.S. shipping industry, NSRP funded the Ship Check Data Capture project in 2005. Laser scanning, close-range photogrammetry, and other technologies capturing as-built ship conditions in digital format to create 3D electronic models were evaluated. The project's goals were to determine potential technology synergies producing cost-effective solutions, and prototype a ship check data capture process that could be used by the U.S. shipbuilding industry. It was also anticipated that archived digital data would provide a cost-effective solution to the lifecycle cost management of ships.

Specific benefits from the software and hardware tested include

- Creation of as-built 3D models and validation of as-built models to design models
- Reduction of costly design changes, improved design capability
- Reduced construction rework
- Accurate factory-fabricate in lieu of field-fabricate
- Reduced ship check costs: fewer days, fewer personnel
- Elimination of return visits to the ship for missed measurements
- Obtaining measurements which are difficult or unsafe for human reach (NSRP 2005).

Initial results were so encouraging from this project that a nine-month follow-on project was awarded by the NSRP in 2006.

Ship Check Data Capture Follow-On 2006 Project

The FY06 follow-on ship check project by NSRP evaluated the ship check process developed in the FY05 project further and refined the ship check process to the U.S. shipbuilding and repair industry using available (COTS) technology. In this follow-up project, the team conducted a ship check onboard a surface ship at Bender Shipbuilding & Repair Company and conducted work onboard SSGN 729 to validate the data accuracy/repeatability of the SSGN 729 ship check data collected from the FY05 project.

Performance improvement metrics were developed and tracked to compare the As-Is practice with anticipated project results. This project reported the cost/time savings metrics associated with post processing the ship check data into 3D CAD models compared to creating CAD models using the traditional ship check method with tape measures. Estimated cost savings of 37% and time savings of 39% were realized for ship check data capture/post processing with the available COTS laser scanning technology hardware and software tools results when compared to traditional ship checks using tape measures. The estimated cost savings is 7% above the project goal of 30%, and the estimated time savings is 4% above the project goal of 35%. Further cost savings can be achieved by using laser scanning technology for ship checks from cost avoidance and minimized rework.

The project conclusions were that the technology (hardware/software) was mature enough to support the ship check process. Laser scanners were found to provide a cost effective method to collect as-built data during ship checks as compared to traditional methods. 3DLS provided time and cost saving, and can be applied to the shipbuilding industry.

The ship check process developed in these projects benefits the shipbuilding industry in several ways:

- Reduces or eliminates costly “return visits” to site for measurements normally missed using traditional ship check methods.
- Provides more accurate, complete as-built data for retrofit design projects, resulting in better retrofit designs which ultimately results in cost savings and cost avoidance. With better designs, less construction rework is required (due to interference and fit-up problems and ability to factory-fabricate instead of having to field-fabricate).

3D Scanning in the Navy

NAVSEA deployed 3D laser scanning to improve the efficiency of both shipcheck and shipalt processes in 2005. Shipcheck is the front-end capture and validation of dimensional data, equipment lists, maintenance records, and performance specifications used in shipalt. Traditionally done manually by labor-intensive and costly methods, shipchecks involved using measurement methods such as tape measures, plumb bobs, and often spirit levels. Shipalt is the follow-on alterations, maintenance and modernization of a vessel.

Also in 2005, 3D laser scanning services were used for shipcheck of a three-story hangar bay on the USS *Abraham Lincoln* (CVN 72). Scanning the HVAC, piping, fuel storage tanks, and other structures allowed shipyard engineers to conduct multi-discipline "what-if" scenarios to avoid clashes in the installation of a new deck. Hundreds of hours in labor were saved with scanning versus the traditional methods. 3DIS captures data at up to 2000 points per second and has a range accuracy of 0.2 inches at 55 feet.

3DLS technology was used to assess damage to the USS *San Francisco* (SSN 711) after it collided at high speed with an undersea mountain 350 miles south of Guam. 3D laser scanning was used to evaluate the damaged areas of the submarine's bow. In this case, scanning was invaluable for determining the ship's centerline and collecting empirical data about torpedo tube deformation.

The Naval Undersea Warfare Center (NUWC) began using laser scanning to reverse engineer components with complex geometries in order to enable competitive bidding in 2007. In the past, the Navy did not have sufficient documentation from the original equipment manufacturer (OEM) to competitively procure replacement components which resulted in purchasing very expensive replacements from the OEM. The Navy saved $250,000 by purchasing parts produced with laser scanning through competitive bidding. In addition, the time required to reverse engineer a typical component, including both measurement and modeling time, was reduced from 100 hours to 42 hours with a laser scanner.

3D Laser Scanning in Shipbuilding

Shipbuilding is one of the most complex and demanding of the manufacturing industries, combining aspects of both direct product manufacturing and capital project development. Moreover, shipbuilders often face huge monetary penalties amounting to hundreds of thousands of dollars per day for being off schedule. 3D laser scanning is a cost-effective, accurate, and fast method to help shipbuilders and manufacturers in designing, redesigning, modifying and salvaging ships.

However, only a handful of several progressive shipyards (i.e., Meyer Wert GmbH, Signal International, and Babcock International) use laser scanning technology because it is not currently widely adopted by the shipbuilding industry. Meyer Werft GmbH, a shipbuilder from Papenburg, Northern Germany, uses laser scanners to assist in building cruise liners, tankers and ferries. New ships are constructed from over 60 individual sections called blocks, weighing up to 800 tons each (Leica, 2015). Precise connection interfaces are critical in ship construction and block assembly; mistakes cannot be made, so consistent and accurate measurements are crucial. At every stage of new ship production, a surveying team using laser scanning technology provides services. With more ship parts being prefabricated and then attached to the ship in one piece, 3D surveys such as taking the measurements of a sun shade composed of multiple concave shapes or a 260 m-long waterslide with curves and loops, are critical.

Signal International, a shipbuilder with multiple facilities in the U.S. Gulf Coast, uses a laser scanner on as-built models to check both new production as well as to generate CAD models for refit projects. It uses the technology to assist in the creation of

- Accurate bill of materials
- General arrangements
- Pipe arrangements
- Pipe ISO's by system
- Pipe spool drawings
- Equipment details
- Structural arrangement

Additive Manufacturing

The American National Standards Institute defines additive manufacturing as the "process of joining materials to make objects from 3D model data, usually layer upon layer, as opposed to subtractive manufacturing methodologies" (ASTM, 2013). Additive manufacturing is also commonly referred to as 3D printing. AM differs radically from the currently dominant manufacturing methodologies. Most current methods use subtractive processes (e.g., machining), but AM builds a 3D object by gradually adding successive layers of material that are laid down exactly in their final location. AM does this by fabricating objects directly from 3D computer-aided design (3D CAD) models. The 3D model is disaggregated into multiple horizontal layers, each of which is produced by the machine and added to the preceding layers. Additive manufacturing is often referred to as 3D printing.

In the automotive industry, Ford Motor Co. uses 3D printing in several areas, including the tooling used to create production parts and to build intake manifold prototypes that can be tested for up to 100,000-mile cycles. With traditional manufacturing methods, it would take four months and cost $500,000 to build, while a 3D-printed manifold prototype costs $3,000 to build over four days.

Additive Manufacturing in the Armed Forces

The U.S. Navy has supported research into 3D printing for more than 20 years and has approximately 70 additive manufacturing projects underway at dozens of different locations. One of the active Navy Manufacturing Technology (ManTech) Program projects active in FY14 was the "Non-Destructive Inspection for Electron-Beam Additive Manufacturing of Titanium." In this project, the emerging AM technology of Electron Beam Direct Manufacturing (EBDM) process was evaluated for fabrication of several F-35 Joint Strike Fighter (JSF) components. EBDM is a technology that is considered vital to improving the affordability, reducing lead time and reducing industrial shortfalls inherent in traditional manufacturing technologies. In this Navy Metalworking Center (NMC) ManTech project, an integrated project team (IPT) evaluated the effectiveness of traditional and advanced non-destructive inspection (NDI) techniques, including computed tomography (CT) scanning, traditional radiography, standard hand-held ultrasonic and phased array ultrasonic inspection methods, to establish standardized NDI processes and procedures for production. According to the Office of Naval Research, studies have shown that EBDM technology has the potential to reduce per-part manufacturing costs by 35%–60%when compared to the costs to manufacture complex-shaped parts with traditional manufacturing approaches (Office of Naval Research [ONR], 2015). Product lead time might also be reduced by as much as 80%. The U.S. Army deployed its first mobile 3D printing laboratory in Afghanistan inside a shipping container that is capable of being carried by helicopter in July 2012.

Additive Manufacturing in Naval Ship Building

The Navy Metalworking Center (NMC) is conducting the "Additive Manufacturing for Shipbuilding Applications" project to demonstrate the cost and time benefits of AM to support the construction of Navy platforms. The project is investigating how the use of AM in ship construction can save acquisition costs on several ship classes. More specifically, Ingalls Shipbuilding (Ingalls) and the Integrated Project Team (IPT) will assess and demonstrate the use of AM during ship construction activities, quantify the expected benefits, and provide a recommended path toward implementation. Ingalls has estimated a minimum acquisition cost savings of $800,000 per year by utilizing AM for the construction of DDG, LHA and LPD. Implementation at Ingalls is planned in FY17 for DDG 121, LHA, and all future surface combatants produced there.

Summary

PLM, 3DLST, and AM are technologies that have been applied in other industries to reduce costs and increase efficiencies and have the potential to reduce naval shipbuilding costs. These technologies can save hundreds of millions of dollars in ship maintenance, suggesting that large savings in ship-building are also available.

A Simulation Model of Naval Shipbuilding Operations

Simulating shipbuilding processes requires conceptual and formal models of shipbuilding. These were combined with estimates of technology impacts and the two sets of simulations (without and with the technologies) to model shipbuilding effectiveness. The Knowledge Value Added simulation approach was then used to model the Return on Investment (ROI) of shipbuilding without and with the three technologies. The results were used to estimate shipbuilding costs and potential cost savings.

The U.S. Navy procures new ships through industry contractors. The shipbuilding processes used by those contractors are not uniform. However, the GAO report *Naval Shipbuilding: Opportunities Exist to Improve Practices Affecting Quality* (GAO, 2013) describes the generic stages of shipbuilding that were used as the basis for modeling shipbuilding in the current study. That report's description says, in part, "There are four primary phases in shipbuilding: pre-contracting, contract award, design and planning, and construction, with each phase building upon the work completed in earlier stages." Based on the latter part of this description, the shipbuilding process was modeled as a sequential series of phases. The GAO description continues, "Within each phase, a number of key events have an influence on the overall quality of the ship. In addition, within Navy shipbuilding, additional key activities take place following ship delivery." A review of the report's more specific description of the process reveals that some of the "events" identified occur relatively quickly (e.g., contract award) and are therefore true events, but that many are extended activities that require significant time and resources to accomplish (e.g., detailed engineering design, assembly and outfitting of blocks). These activities describe shipbuilding processes that can benefit from the adoption and use of the three technologies previously described. The GAO description of shipbuilding was condensed[1] into a series of shipbuilding phases as a preliminary step in modeling naval shipbuilding:

- Concept design
- Detailed engineering design
- Pre-construction planning
- Block fabrication
- Assembly and outfitting of blocks
- Keel laying and block erection
- Pre-delivery final outfitting
- System testing and commissioning
- Sea trials
- Post-delivery final outfitting
- Post-delivery tests and trials
- Post shakedown availability

See Housel, Hom, Ford, and Mun (2016) for details. The previously listed phases are the basis of the As-Is model of naval shipbuilding.

Simulating Traditional Shipbuilding Operations

Each of the shipbuilding phases previously described is assumed to have three basic operations: initial completion, quality assurance, and rework. Each operation moves work part way through the phase. The Initial Completion activity moves work from the Initial Completion backlog and Work In Progress (WIP) to the Quality Assurance (QA) backlog and WIP. The QA operation either discovers required rework or approves and releases the work. This moves work from the QA backlog and WIP to either the Rework backlog and WIP (if rework is discovered) or to the stock of Work Completed and Released. The rework operation moves work from the Rework backlog and WIP back to the QA backlog and WIP, where it is inspected again. Figure 1 shows the arrangement of the stocks and flows of each of the shipbuilding phases. In addition to the operations processes previously described, progress through each phase depends on the sizes of the backlogs, the durations required to complete each operation, and the fraction of work that requires rework (Figure 1).

---

[1] Some activities were renamed and descriptions revised to reflect U.S. naval shipbuilding without losing their meaning.

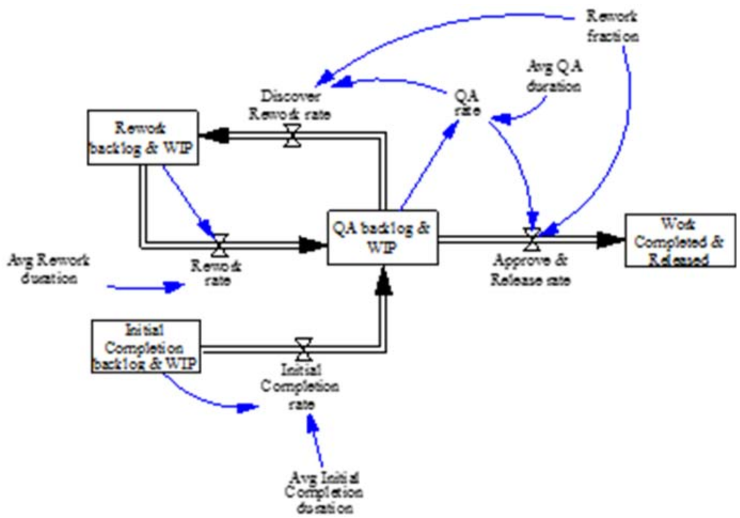


Figure 1. Simulating Shipbuilding: Drivers and Constraints on Shipbuilding Operations in a Single Phase

Each phase operation rate (initial completion, quality assurance, or rework) is driven and constrained by the amount of work waiting to be completed by that operation and the average time required to complete the operation. In the current model these operation durations include process and resource constraints and are assumed to be constant throughout the shipbuilding phase. The rate at which work within a phase is inspected (the quality assurance rate) is disaggregated into the fraction of inspections that discover required rework and the compliment that are approved and released. Progress through each shipbuilding phase in the model is also depends on the completion of work in the preceding (upstream) phase and constrains progress in its downstream phase. Although some overlapping of phases is possible, for simplicity it is assumed that the phases occur sequentially.

Potential Applications of Advanced Technologies to Navy Shipbuilding

Three Dimensional Scanning (3DLS), Product Lifecycle Management (PLM) and Additive Manufacturing (AM) can impact naval shipbuilding in many ways, including

- Integrated Ship Development
- Design and construction document management
- Prototype generation
- Final parts manufacturing
- Manufacturing inspection:
- Radio Frequency Identification (RFID)
- Animated Instructions
- Construction inspection

Several of the technology applications previously listed are already in regular use in industry or fully developed for use in practice. For example, RFID is frequently used to control construction material flows (CoreRFID, 2008). Damen Industries is developing animated electronic construction instructions (Ford et al., 2012), and construction inspection by comparing laser scans of as-built conditions to design documents has been demonstrated (Taylor, 2013). The expected application of the three advanced technologies to specific shipbuilding phases were developed (not shown for brevity).

Shipbuilding operations using the three technologies were simulated for the To-Be conditions. The potential impacts of the use of the three technologies in the shipbuilding phases were quantified in the form of fractional reductions in operation durations and rework fractions. The reduction fractions were combined with the As-Is calibration values for the parameters to generate calibration values for the To-Be simulation. These calibration values were used to simulate shipbuilding operations using the three technologies for the To-Be conditions. Simulation results for the As-Is and To-Be scenarios are shown in Table 1.

Table 1. Simulation Results: Average Completion Rates of Shipbuilding Phases for As-Is and To-Be Scenarios

		AVERAGE COMPLETION RATE (work packages / day)	
No.	PHASE	As-Is Scenario	To-Be Scenario
1	Concept design	0.593	0.8958
2	Detailed design	3.115	4.454
3	Preconstruction planning	1.407	1.741
4	Block fabrication	3.084	9.302
5	Block assemby and outfitting	2.865	11.61
6	Keel laying and block erection	3.439	13.53
7	PreDelivery outfitting	3.439	13.53
8	System testing	2.047	3.508
9	Sea trials	6.34	6.896
10	PostDelivery outfitting	3.273	13.27
11	PostDelivery tests	1.827	1.963
12	PostShakedown maintenance	1.827	1.963

Knowledge Value Added Model of Shipbuilding

The results of the simulations of shipbuilding operations were used as input to the KVA model to estimate the return on investments of the technologies. For both the As-Is and To-Be scenarios the “market” value of the hypothetical ship is assumed to be the estimated total price to the U.S. Navy of the *Arleigh Burke* (DDG51) destroyer, approximately $1.2 billion.[2] This total value was allocated among the 12 shipbuilding phases based on the total learning of each phase. Other values were taken from previous KVA models of naval operations and modeler estimates. The As-Is scenario was modeled using the values previously described.

[2] Estimated prices of *Arleigh Burke* destroyers range were $0.90 billion per ship (1997 dollars based on four ships) and $0.92 billion per ship (1998–1999 dollars based on six ships) with estimates of future ships based on weight up to $1.4 billion per ship.

The use of the three technologies was modeled in the To-Be scenario. Reductions in rework due to improved information quality and availability and the reduced operation durations due to use of richer information by field personnel that provides more specific instructions and designer intent were modeled in the operations simulation model. In addition, the technologies are expected to impact shipbuilding operations in several ways, including

- Increased design scope is required to develop the richer information for field personnel
- Reduce training time of construction personnel due to use of rich construction information
- Reduced unit labor costs as lower skill levels will be required due to providing improved construction and assembly information
- Increased use of automation

The impacts previously listed were incorporated into the KVA model. Note that the value of the ship is unchanged from the As-Is scenario, reflecting the assumption that the same ship is being created with or without the three technologies and the focus of the current work on potential cost savings. Tables 2 shows the returns on investment for the As-Is and To-Be scenarios, the changes in the returns on investment by using the three technologies, and the automation tools applied.

Table 2. Changes in Return on Investment Due to Use of Three Technologies

No.	PROCESS / PHASE	As-is ROI	To-be ROI	Change in ROI	Automation Tools
1	Concept design	-2%	94%	**96%**	AM, PLM
2	Detailed design	561%	1826%	**1265%**	AM, PLM
3	Preconstruction planning	218%	244%	**25%**	PLM
4	Block fabrication	-67%	-31%	**36%**	3DLS, AM, PLM
5	Block assemby and outfitting	-17%	116%	**133%**	3DLS, AM, PLM
6	Keel laying and block erection	-63%	1%	**64%**	3DLS, AM, PLM
7	PreDelivery outfitting	505%	1270%	**764%**	3DLS, AM, PLM
8	System testing	280%	582%	**301%**	3DLS, PLM
9	Sea trials	1018%	961%	**-57%**	PLM
10	PostDelivery outfitting	476%	1243%	**767%**	3DLS, AM, PLM
11	PostDelivery tests	239%	282%	**42%**	PLM
12	PostShakedown maintenance	221%	201%	**-20%**	PLM
	TOTALS	**135%**	**464%**	**329%**	

Table 2 shows that the detailed design and outfitting phases of shipbuilding benefit most from use of the technologies, and that the sea trials and post shakedown maintenance benefit least. Of more significance to the current work, the ROI increases by 329%.

Estimating Shipbuilding Costs and Cost Savings

As used in previous research, costs for the As-Is and To-Be scenarios can be estimated using the definition of Return on Investment (ROI),

ROI = (Benefits – Costs)/Costs,

which can alternatively be written as

Cost = Benefits/(ROI 1).

The previous equation was used with the benefits ($1.2 billion) and Returns on Investment (Table 2) to estimate the costs of each scenario in millions of dollars as follows:

Cost(As-is) = Benefits(As-is) ÷ (ROI(As-is) + 1) = 1,200 ÷ (1.3546 + 1) = $509.64 Million

Cost(To-be) = Benefits(To-be) ÷ (ROI(To-be) + 1) = 1,200 ÷ (4.6409 + 1) = $212.73 Million

Therefore, estimated potential savings for the one hypothetical ship is $296.91 million ($509.64 million–$212.73 million). This represents a savings of 24.74% ($296.91 million ÷ $1,200 million) of the total cost to the Navy. This saving fraction is conservative when compared with the results reported by industry adopters of these technologies described previously in this report (e.g., >30% cost savings for 3D LST alone and up to 80% for AM).

Estimated cost savings in U.S. naval shipbuilding are very contingent on the number and type of ships built. However, a rough estimate can be made based on the 2015 shipbuilding plan described in the first section of this report. According to that plan, the U.S. Navy will purchase 264 ships from 2015–2044 (218 combat ships and 46 combat logistics and support ships) at an average cost of $16.7 billion per year. Based on these numbers the average ship cost will be $1.83 billion ($16.7 billion per year × 29 years ÷ 264 ships). Therefore, savings estimates based on a hypothetical $1.2 billion ship above are considered conservative. Those savings are estimated to be an average of $2.70 billion per year ($296.91 million per ship × 264 ships ÷ 29 years).

Integrated Risk Management and Strategic Real Options Analysis

Integrated Risk Management (IRM) is an eight-step, quantitative software-based modeling approach for the objective quantification of risk (cost, schedule, technical, value), flexibility, strategy, and decision analysis. The method and toolset provide the ability to consider hundreds of thousands of alternatives with budget, schedule, value, strategic, and other program implementation uncertainties, and provide ways to help the decision-maker maximize capability and readiness at the lowest cost and highest returns (both monetized using KVA and nonmonetary strategic value). The variables simulated in the As-Is and To-Be strategies included the uncertain inputs of number of employees, actual learning time in hours, percentage automation achieved, number of times performed per ship, and the average process rates (units per day). These were simultaneously simulated for 1,000,000 trials.

Strategic Real Options

An important step in performing IRM is the application of Monte Carlo risk simulation. By applying Monte Carlo risk simulation to simultaneously change all critical inputs in a correlated manner within a model, researchers can identify, quantify, and analyze the system's risks and uncertainties. Based on the overall problem identification occurring during the initial qualitative management screening process, certain strategic options would become apparent for each particular project. The strategic options could include, among other things, the option to wait, expand, contract, abandon, switch, stage-gate, and choose. Traditional analysis assumes a static investment decision, and assumes that strategic decisions are made initially with no recourse to choose other pathways or options in the future. Real options analysis can be used to frame strategies to mitigate risk, to value and find the optimal strategy pathway to pursue, and to generate options to enhance the value of the project while managing risks.

Figure 2 illustrates the strategic road map for implementation. Strategy A located on the top branch of the strategy tree is a sequential compound option, where the 3DLS, PLM, and AM technologies can be implemented in three phases over a period of 0–9 years, where the second phase will only be implemented if the first phase Proof of Concept (POC) proves to be successful, and the third phase can be implemented only if the second phase proves to be successful. This wait-and-see strategy creates Value of Information, where any kinks in the system's implementation will be worked out over time, focus is placed on one technology implementation in each phase, and costs are stretched out over time providing more flexibility in any budgetary constraints. Sequential compound options are often used in other applications such as

- Stage-gate implementation of high-risk project or technology development
- Prototyping prior to large scale manufacturing
- Low Rate Initial Production (LRIP)
- Technology feasibility tests
- Advanced Concept and Technology Demonstration over multiple stages
- Proof of Concept tests over various stages to determine the most valuable strategies for product rollouts in spiral development
- Government contracts with multiple stages with the option to abandon anytime
- Termination for Convenience (T-for-C) and built-in flexibility to execute different courses of action at specific stages of development
- R&D and phased options to determine most valuable strategy for system of systems technology development

Strategy B in Figure 2 illustrates an alternative course of action where all three 3DLS, PLM, and AM technologies are implemented at once. The cost will be higher (larger up front lump-sum and budgetary approval hurdles), and potential risks will be higher (problems that may arise in implementation of a new set of technologies on a larger scale). Nonetheless, the benefits that will be obtained are faster and more immediate, but these net benefits may or may not supersede the added costs and inherent risks.

Finally, Strategy C is the base case of As-Is model where legacy approaches and technologies are maintained status quo. This strategy can be valued accordingly and the difference in value between Strategies A and C and between Strategies B and C can be readily computed. See Mun (2015) for modeling details.

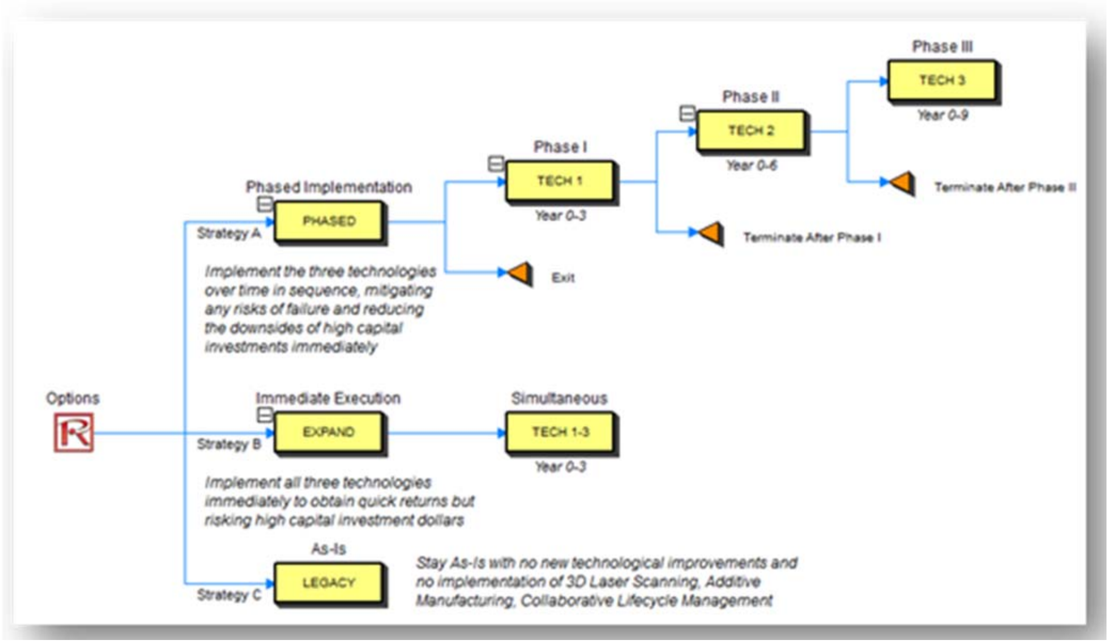


Figure 2. Strategic Real Options

As summarized in Figure 3, the following are some takeaways of the analysis results:

- The As-Is cost is $509.64 million and the To-Be cost after implementing 3DLS, PLM, and AM is $212.73 million, providing the U.S. Navy a cost savings of $296.91 million.
- The $296.91 million when multiplied by 264 ships and allocated over 29 years yields an annual savings of $2.70 billion a year for the U.S. Navy.
- When added flexibility is analyzed, this strategic value increases to $3.07 billion when all three technologies are implemented immediately or $3.37 billion when implemented over multiple stages where risks and uncertainties can be hedged

As-Is Cost	$	509,642,661
To-Be Cost	$	212,733,874
To-Be Cost Savings	$	296,908,787
To-Be Immediate Strategic Value	$	337,632,796
To-Be Sequential Strategic Value	$	370,736,221
Number of Ships		264
Number of Years		29
To-Be Sequential (Strategy A) Strategic Value	$	3,374,978,008
To-Be Immediate (Strategy B) Strategic Value	$	3,073,622,690
To-Be Strategy (Base Case) Strategic Value	$	2,702,893,785

Figure 3. Summary of Strategic Values

Conclusions

Summary of the Study

We reviewed industry applications and tangible benefits resulting from PLM, 3DLS and AM to understand the potential ramifications from these technologies. We then assessed the impacts of using these technologies for naval shipbuilding. A simulation model of shipbuilding operations at the phase level was built and used to forecast the impacts of the technologies on shipbuilding processes. This required both conceptual and formal models of shipbuilding. These were combined with descriptions and estimates of technology impacts on shipbuilding operations and generated two sets of simulations (without and with technology use). The output of the operations simulation model was used to build a Knowledge Value Added model of naval shipbuilding. The KVA model was used to estimate the Return on Investment (ROI) of shipbuilding without and with the three technologies. The outputs of the KVA model were used to estimate shipbuilding costs with and without the technologies. Finally, those costs were used to estimate potential savings over the 29-year naval shipbuilding planning horizon. The uncertain inputs in the model were then subjected to a rigorous Monte Carlo risk simulation and stochastic analysis of millions of simulation trials and these three technologies were divided into various implementation paths. The Analysis of Alternatives using strategic real options were applied and the optimal implementation strategies were recommended.

Results of the Study

The research indicates that Three Dimensional Scanning (3DLS), Product Lifecycle Management (PLM) and Additive Manufacturing (AM) can beneficially impact many phases of naval shipbuilding in multiple operations to reduce costs. Simulation results suggest that the U.S. Navy can save an average of over $2.70 billion per year over 29 years if the potential improvements available through 3DLST, PLM, and AM are fully exploited, regardless of the implementation approach.

However, with the added implementation flexibility of whether the three technologies are to be implemented concurrently, requiring a larger budget and bearing more uncertainties, or the technologies can be introduced over time sequentially where additional value is created. Based on the analysis, 3DLST, PLM, and AM technologies are fully justified, saving the U.S. Navy a base case value of $2.70 billion per year over 29 years. And if implemented fully and immediately, these three new technologies can save the U.S. Navy $3.07 billion, or $3.37 billion if implemented sequentially.

That cost savings estimate and strategic real options assessment will help decision-makers choose how much, when, and how to exploit the benefits and the minimize costs of adopting and implementing the three technologies investigated.

Future Research Opportunities

The research is limited by the relatively narrow focus and assumptions used in the modeling and assessment. For example, the focus on post-technology-adoption does not address the significant challenges and costs of technology adoption, but the same focus does not include the potentially significant benefits of the three technologies during ship operations, maintenance, and repair.

Future research can collect and apply more specific parameter values for improved model calibration. In addition, specific decision-maker flexibility and inherent implementation options can be determined and modeled in more detail to provide a better implementation framework.

References

CoreRFID. (n.d.). Construction & RFID: *The ROI* (White paper). Retrieved from http://www.corerfid.com/Files/White%20Papers/013%20Construction%20and%20RFID%20-%20The%20ROI%20V2.pdf

DoD ManTech Program. (2013, December). *Using manufacturing innovation to improve affordability*.

Ford, D., Housel, T., & Mun, J. (2012, October). *Naval ship maintenance: An analysis of the Dutch shipbuilding industry using the knowledge value added, systems dynamics, and integrated risk management methodologies* (NPS-12-204). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

GAO. (2013). Naval shipbuilding: *Opportunities exist to improve practices affecting quality* (GAO-14-122). Washington, DC: Author.

Greaves, T. (2005, October 11). 3D Laser scanning market red hot: 2005 forecast raised to $178 million: 45% growth projected. *SparView, 3*(31), 1.

Greaves, T. (2006, January 17). U.S. Navy improves readiness with 3D Laser scanning. *SparView, 4*(2), 1.

Kenney, M. (n.d.). *Cost reduction through the use of additive manufacturing (3D printing) and collaborative product lifecycle management technologies to enhance the Navy's maintenance programs* (Master's thesis, Naval Postgraduate School). Retrieved from http://www.acquisitionresearch.net

Leica Geosystems. (2011). *Meyer Werft Papenburg: Modern shipbuilder chooses Leica Geosystems* (Case study).

Mun, J. (2015). *Modeling risk: Applying Monte Carlo risk simulation, strategic real options, Stochastic forecasting, portfolio optimization, data analytics, business intelligence, and decision modeling*. Thomson-Shore.

Naguy, T. (2014, November 19). *Additive manufacturing Air Force perspective*. DoD Maintenance Symposium 2014 Presentation.

National Shipbuilding Research Program ASE Ship Check Data Capture Follow-On Project (NSRP ASE 05-01). (2007, March 7). *Final report FY06*.

National Shipbuilding Research Program ASE Ship Check Project Presentation (NSRP 2005). (2005, December 8). Presentation at the Joint Meeting of Shipyard Production Process Technologies, Business Process Technologies, System Technology and Facilities, Tooling and Risk Management Panels (pp. 139–144). Retrieved from http://www.nsrp.org/panels/bpt/downloads/ShipCheck1205.pdf

NavSource Naval History. (n.d.). USS *Arleigh Burke* (DDG-51). Retrieved from http://webcache.googleusercontent.com/search?q=cache:Pqyq1gqLJmAJ:www.navsource.org/archives/05/01051.htm+&cd=9&hl=en&ct=clnk&gl=us

Office of Naval Research (ONR). (2015). *Navy ManTech project book*. Arlington, VA: ONR.

O'Rourke, R. (2014, August 1). *Navy Force structure and shipbuilding plans: Background and issues for Congress* (CRS Report RL32665). Washington, DC: CRS.

Sparpoint Group. (2007, April 26).Navy scanners saves dollars. Retrieved from http://www.theengineer.co.uk/news/navy-scanners-save-dollars/299398.article

Taylor, C. (2015, August). 3-D modeling on the move. *Construction Business Owner*, 32–33.

USS *Arleigh Burke*. (n.d.). In *Wikipedia*. Retrieved September 15, 2015, from https://en.wikipedia.org/wiki/USS_Arleigh_Burke

Materials Testing and Cost Modeling for Composite Parts Through Additive Manufacturing

Eric S. Holm—recently graduated from the Air Force Institute of Technology (AFIT) with a Master of Science in engineering management. He holds a Bachelor of Science degree in civil engineering from the University of Virginia. Commissioned through the Reserve Officer Training Corps in 2006, he now has 10 years of active service as a Major and as a Civil Engineer officer in the United States Air Force. Major Holm is currently assigned to the Inspector General staff of the Air Force Materiel Command as Chief of Command Civil Engineer Inspections. [eric.holm@us.af.mil]

Vhance V. Valencia—is an Assistant Professor in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT) located at Wright-Patterson AFB, OH. He is the program director for the Graduate Engineering Management program responsible for the program's curriculum and oversight of 20 graduate students annually. Holding two degrees from AFIT, he earned an MS in engineering management and a PhD in systems engineering. He was recently selected for promotion to Lieutenant Colonel. His research interests include additive manufacturing applications for civil engineering, infrastructure asset management, geographic information systems, critical infrastructure protection, and unmanned aerial systems. [vhance.valencia@afit.edu]

Alfred E. Thal, Jr.—is an Associate Professor in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT) located at Wright-Patterson AFB, OH. He is the Curriculum Chair for the Engineering Management thrust area encompassing the Graduate Engineering Management and Graduate Cost Analysis programs. He holds a PhD in environmental engineering from the University of Oklahoma. His research interests include engineering and environmental management, project management, facility and infrastructure management, rapid product development, economic analysis, innovation, sustainability, and project management.

Jason K. Freels—is an Assistant Professor in the Department of Systems Engineering and Management at the Air Force Institute of Technology (AFIT) located at Wright-Patterson AFB, OH. He is a professor of systems engineering specializing in applied statistics, reliability, and lifecycle management. He earned both his MS, studying materials science and engineering, and his PhD, studying systems engineering, from AFIT. His research interests include survival/reliability analysis, systems engineering, optimization, structures and materials, and response surface methods. [jason.freels@afit.edu]

Adedeji B. Badiru—is Dean of the Graduate School of Engineering and Management at the Air Force Institute of Technology (AFIT). He was previously Professor and Head of Systems Engineering and Management at AFIT, Professor and Department Head of Industrial & Information Engineering at the University of Tennessee in Knoxville, and Professor of Industrial Engineering and Dean of University College at the University of Oklahoma, Norman. His areas of interest include mathematical modeling, project modeling and analysis, economic analysis, systems engineering, and efficiency/productivity analysis and improvement. He is the author of over 25 books, 34 book chapters, 70 technical journal articles, and 110 conference proceedings and presentations. [adedeji.badiru@afit.edu]

Abstract

Recent advances in additive manufacturing (3D printing) have introduced new parameters in reducing cost in manufacturing aircraft components. The additive process provides a possible means to reduce an aircraft's lifecycle cost (LCC), but the effects of changed process parameters of additive manufacturing machines on final material characteristics are not well known. This research explores these effects with the intent to motivate greater use and application in aviation. We conduct this study in two parts. First, focusing on fused filament fabrication (FFF) through Mark Forged, Inc.'s Mark One machine, this research creates PA6 dog-bone specimens for (1) a design of experiments (DOE) procedure and (2) a destructive test of a continuous fiber composite specimen from the Mark One machine. Second, this

paper explores cost modeling issues using in the additive manufacturing industry with a specific focus on energy usage. Taken together, this research effort identifies critical factors in additive manufacturing towards revolutionizing the military supply chain.

Introduction

Modern manufacturing processes tend to reflect globalization, a concentration on core activities, shorter product life-cycles, and an increasing focus on customer needs (Baumers et al., 2012). This often results in advanced supply chains which are complex and long (Foran et al., 2005). However, additive manufacturing (AM) can simplify and reduce the supply chain associated with component manufacturing. This can be accomplished by avoiding the tools, dies, and material waste that accompany conventional manufacturing processes (Morrow et al., 2007; Serres et al., 2011). Additionally, and of primary importance to many organizations though, is the fact that AM offers the capability to produce small quantities of customized items at a relatively low average unit cost (Baumers et al., 2011). This is possible because geometric constraints typical of formative and subtractive processes are largely eliminated (Tuck et al., 2008; Baumers et al., 2011), which leads to advanced freeform fabrication (Meteyer et al., 2014) and the capability to create geometrically complex and novel items (Horn & Harrysson, 2012; Mani et al., 2014).

When viewed from a life-cycle perspective, a number of organizations recognize that environmental benefits and performance improvements can be achieved (Horn & Harrysson, 2012; Huang et al., 2013; Huang et al., 2015). For example, the “buy-to-fly” ratio (i.e., mass of raw material needed per unit mass of finished component) ranges from 12:1 to 25:1 for aircraft components made of aluminum and titanium alloys using conventional manufacturing processes (Oak Ridge National Laboratory, 2010; Huang et al., 2015). These high buy-to-fly ratios indicate that 92.3–96.2% of the raw material is wasted, which leads to large energy and environmental emissions footprints (Huang et al., 2015). Thus, AM has the potential to reduce the “cradle-to-grave” environmental impact by reducing waste and minimizing the consumption of natural resources associated with normal manufacturing processes (Morrow et al., 2007; Serres et al., 2011; Huang et al., 2015). Furthermore, the aircraft industry has increased fuel efficiency by incorporating AM components to reduce weight (Oak Ridge National Laboratory, 2010; Huang et al., 2015). Lindemann et al. (2013) cite a cost savings of $3,000 per year for each kilogram reduction in mass, and Huang et al. (2015) estimate that fuel consumption could be reduced by as much as 6.4% if AM was used to its full potential.

Despite these advantages and benefits, a number of limitations have been attributed to additive manufacturing. For example, Ruffo and Hague (2007) list the following limitations associated with AM technology: material selection and characteristics, process productivity, accuracy of product dimensions, surface quality, repeatability, and unit cost at medium and high volumes. However, the low throughput of AM processes is considered to be a primary limitation, which makes it less suitable for high-volume production (Huang et al., 2015). According to Huang et al. (2015), concerns with geometric repeatability, residual stresses, and high surface roughness make AM less appropriate for work requiring high dimensional precision, surface quality, and fatigue resistance. Additionally, during typical AM processes, Schroeder et al. (2015) found that quality concerns, from either operator or machine failures, led to high rejection rates; this means that “industry-standard product quality rates can rarely be achieved.” However, many of these concerns may be addressed in the next 5–20 years (Huang et al., 2015).

Therefore, the primary purpose of this paper is to address a few of the material characteristics. Specifically, the research investigated how variations in two AM process parameters in fused filament fabrication (FFF), also known as fused deposition modeling (FDM), affected the mechanical properties of the two specimens being produced. A secondary purpose of the paper is to broadly review cost modeling issues, primarily from an energy consumption perspective, since research regarding major cost drivers is rather limited (Lindemann et al., 2012) and people tend to focus on purchasing and production costs (Lindemann et al., 2013). Therefore, this paper is a stepping-stone for further research to develop AM composite technology and encourage its use in high-performance applications. The long-term goal is the ability to produce aerospace parts through AM that meet the same service specifications as traditionally manufactured aerospace parts.

Background

Fused deposition modeling is a subset of AM technology using selective deposition processes commonly available in the commercial market under many different brand names. With FDM, a thermoplastic filament is pushed through a computer-controlled extrusion head and deposited on a build plate as a series of layers to form a three-dimensional object. Although FDM is a type of technology, fused deposition modeling and FDM are trademarked by Stratasys, which invented the process (Barnatt, 2013). Other terms used to describe FDM include plastic jet printing (PJP), fused filament modeling (FFM), and fused filament fabrication (FFF). FFF was coined by the RepRap project to avoid legal constraints with using the term FDM. Therefore, when referring to fused deposition modeling, the term FFF is used in this paper.

In FFF, the extrusion nozzle moves in a plane, parallel with the build surface or build plate in the $x - y$ plane (Ahn et al., 2002). A heated extrusion head melts the thermoplastic filament before it passes through the extrusion nozzle, and the AM machine deposits the viscous thermoplastic material onto the build surface as a series of rows. These rows are called rasters or roads. After this deposition, the build plate lowers (or the extrusion head raises) and the machine deposits another layer of thermoplastic material. This process repeats itself until the desired shape is complete (Gibson et al., 2010).

Thermoplastics are the most widely used feedstock in FFF processes, with the most common materials being acrylonitrile butadiene styrene (ABS), polycarbonate (PC), polylactide (PLA), and polyamide (PA). However, thermoplastics are low in strength compared to metals and their mechanical properties. For example, the maximum tensile strength for polymers is about 100 MPa (15,000 psi), and some metal alloys have tensile strengths of 4,100 MPa (600,000 psi) (Callister & Rethwisch, 2012). In comparison, Table 1 shows the tensile yield strength of ABS plastic, nylon-12, carbon fiber reinforced polymer, and carbon fiber made through conventional manufacturing processes (not additive). This decrease in strength limits their use in more high-performance applications such as aerospace, automotive industry, and infrastructure.

Table 1. Tensile Yield Strength for Various Materials
(MatWeb, LLC, 2015; Daniel & Ishai, 2003; Callister & Rethwisch, 2012)

Material	Tensile Strength, Yield (Mpa)	Tensile Strength, Yield (psi)
ABS	42.5 - 44.8	6,160 - 6,500
Nylon-12	9.50 - 170	1,380 - 24,600
Carbon/Epoxy (AS4/3501-6)	2,280	330,000
Carbon Fiber (AS4)	3,700	535,000

To improve the strength of FFF-made thermoplastic parts, carbon fibers can be incorporated to create a composite material called carbon fiber reinforced plastic (Love et al., 2014). A composite is made up of two or more materials exhibiting better material properties than the individual materials comprising the composite (Daniel & Ishai, 2003). Combining carbon fibers with a plastic thus allows for a more durable material. However, the introduction of new materials requires thorough analysis to gain a better understanding of the material's behavior and mechanical properties. This will help engineers predict how the material will perform in various environments under certain life-cycle loads. The ability to know the expected material properties of a part produced through FFF with a high degree of confidence will encourage the use of these materials in more high-performance applications.

Research Method

In 2014, MarkForged Inc. introduced the first commercially available AM machine to create continuous carbon fiber reinforced polymer composites (Black, 2014). The company explains that its goal was to manufacture "end-use parts" but make them "a lot more efficiently" and "use the mechanics of a 3-D printer to automate carbon fiber composite layup" (Black, 2014). For this research, the Mark One 3D printer was selected because it is the only commercially available machine on the market that creates continuous carbon fiber polymer composites using the FFF process. The Mark One has two extrusion nozzles: one for the nylon filament and the second for the continuous carbon fiber towpreg. A carbon fiber towpreg is a bundle of carbon fibers pre-impreganted with a thermoplastic resin to create a filament. When the carbon fiber towpreg passes through the heated extrusion nozzle, the thermoplastic resin melts and the carbon is deposited on a nylon layer. This is different compared to the modified machine used by Namiki et al. (2014) which impregnates the nylon with a carbon fiber towpreg inside the extrusion head.

Since the two most commonly used polyamide (PA) grades are PA6 and PA66, this research used a proprietary blend for a PA6 co-polymer nylon with three types of fiber reinforcement: Kevlar, carbon fiber, and fiberglass. We then used two distinct approaches to characterize the material properties of composite specimens manufactured with the Mark One. First, the material characteristics of the matrix material (PA6 nylon) were investigated through a design of experiments (DOE) method to vary process parameters (input variables) and determine their effect on material mechanical properties (output variables). Second, we performed continuous carbon fiber composite (CCFC) testing of specimens, specifically carbon-reinforced PA6 nylon, produced with the Mark One.

A DOE is a systematic method of conducting controlled tests to evaluate how changes in different factors affect the response of interest. Test specimens were made using the Mark One and tested in accordance with ASTM D638, Standard Test Method for Tensile Properties of Plastics, and ASTM D3039, Standard Test Method for Tensile Properties of Polymer Matrix Composite Materials. Most of the process parameters on the Mark One are fixed, which limited the number of factors to choose for the experiment. Therefore, the factors of interest for the experiment were raster angle orientation and layer height. Table 2 shows the two factors with the various settings for each of six treatments. Since only two factors were selected for testing, it is possible to conduct a full factorial design in which every combination of factors and settings is tested. Raster angle orientation had two settings while layer height had three settings levels. Therefore, this experiment has six different treatments, and three experiments were performed for each treatment for a total of 18 specimens.

Table 2. Raster Angle Orientation and Layer Height for Each Treatment

Treatment	Raster Angle Orientation(°)	Layer height (mm)
1	0/90	0.1
2	0/90	0.15
3	0/90	0.2
4	±45	0.1
5	±45	0.15
6	±45	0.2

In both aspects of this research, the basic geometry and testing procedure for all specimens was the same. Figure 1 shows the dimensions for the nylon tensile specimens according to ASMT D638, Type 1, with a thickness of 4 mm. Tensile testing was conducted on a MTS model 204.52 load cell with a 5.5 kip capacity using a MTS 632.13B-20 clip gage extensometer with a 0.5 inch gage length. Grip pressure was set to 1,000 pounds, and the temperature of the room was 72.3 degrees Fahrenheit with a relative humidity of 45%. Prior to testing, the average width and thickness of each specimen, determined by taking the average of three measurements, was used to calculate the engineering stress and engineering strain during tensile testing. The specimens were tested under stress control until failure; that is, the rate of increasing stress applied to each specimen was the same until failure.

The desired load rate applied to each specimen was based on the cross sectional area of the gage section for each specimen. Equation 1 shows how the desired load rate for each specimen was calculated.

$$L_{rate} = \frac{11{,}000\ psi \times A_{gage}}{300\ sec} \qquad (1)$$

where L_{rate} is the desired load rate for each specimen (lb-f/sec) and A_{gage} is the area of the gage section (in2). Using Equation 1 ensured that a tensile stress of 11,000 psi occurred within 300 seconds (five minutes) of starting the tensile test.

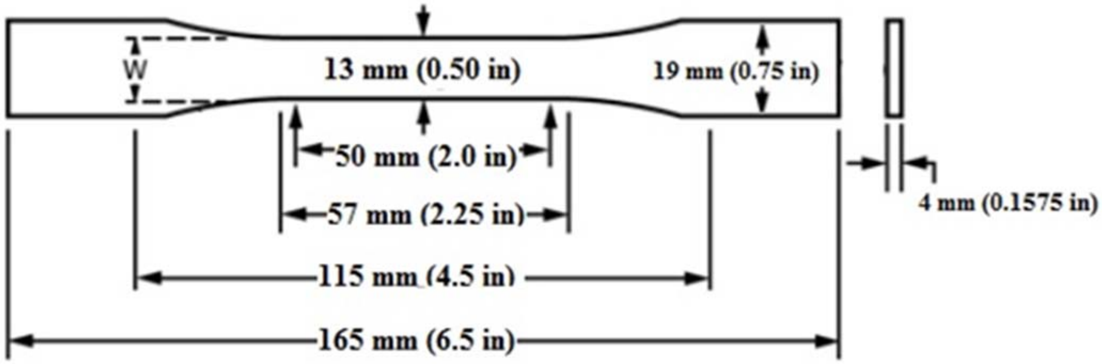


Figure 1. Drawing of Nylon Dog-Bones Tensile Specimen With Dimensions

Material Testing Results and Discussion

The materials testing results are presented in two parts. The first part provides the results from the design of experiments (DOE) procedure. The DOE analyzed the effects of layer height and raster angle on response variables of tensile modulus, yield stress, percent strain at yield, ultimate tensile strength, and percent strain at break. The second part presents the results from the continuous fiber composites from the Mark One machine to describe the composite material's mechanical properties.

Part 1: DOE Analysis

Table 3 shows the mean, along with the upper and lower 95% confidence intervals, of each measured mechanical characteristic for the settings used for each factor. This table provides insight into how each factor and the individual settings within the factors influence the mechanical properties. While descriptive statistics are useful, Table 4 shows the analysis of variance (ANOVA) results for each factor and response along with the interaction. Table 4 includes the R^2 and F-test values from the ANOVA. Higher R^2 values indicate that the factors explain more of the variability in the data; this means that any differences between factors are less likely to be caused by randomness. An overall significance value of $\alpha = 0.05$ was used for the F-test.

The overall F-test values are shown first. If the overall F-test was less than 0.05 for a response, at least one of the two factors explains the variability in the data. The F-test values for each individual factor are shown to the right of the overall F-test values. The F-test on the factors determines if the difference in the mean responses are statistically different. Because there were two factors being tested, the significance value of 0.05 is divided by 2 to get 0.025. Therefore, F-test values less than 0.025 indicate that the process parameter (i.e., factor) is statistically significant in influencing the desired response (i.e., mechanical property). The critical value for the interaction F-test was 0.05 divided by 3 to get 0.0167. Only the interaction of layer height and raster angle orientation on yield stress was found to be statistically significant. Values less than the respective critical value are highlighted in green in Table 4 to indicate statistical significance. Table 5 provides a summary of the statistical significance of each factor and interaction influencing a certain response based on the ANOVA results.

Table 3. Mean and Confidence Intervals of Measured Mechanical Properties by Factor and Level

	Layer Height (mm)			Raster Angle Orientation	
	0.1	0.15	0.20	±45	0/90
Mean Tensile Modulus (GPa)	1.1899	1.1835	1.0695	1.1857	1.1096
Upper 95% Mean	1.2478	1.2519	1.164	1.2383	1.1808
Lower 95% Mean	1.132	1.1151	0.9749	1.1331	1.0384
Mean Yield Stress (MPa)	12.744	12.302	11.917	11.961	12.681
Upper 95% Mean	13.721	12.628	12.602	12.408	13.222
Lower 95% Mean	11.767	11.975	11.231	11.513	12.140
Mean Percent Strain at Yield Stress (MPa)	1.271	1.242	1.323	1.211	1.347
Upper 95% Mean	1.361	1.292	1.459	1.241	1.412
Lower 95% Mean	1.182	1.190	1.188	1.181	1.282
Mean Ultimate Tensile Strength (MPa)	37.735	36.564	35.104	36.519	36.416
Upper 95% Mean	38.618	37.676	35.808	37.577	37.539
Lower 95% Mean	36.852	35.452	34.400	35.462	35.293
Percent Strain at Break	116.762	78.172	35.840	113.231	40.618
Upper 95% Mean	199.489	117.319	83.213	158.973	74.799
Lower 95% Mean	34.034	39.025	-11.533	67.489	6.436

Table 4. ANOVA Results for Each Factor (With and Without Interaction)

	Response	R Squared	F-test	F-test, Layer height	F-test, Raster Angle Orieintation	F-test, interaction
Without Interaction	Tensile Modulus	0.6142	0.0033	0.0059	0.0181	
	Yield Stress	0.4874	0.0217	0.0759	0.0186	
	Percent Strain at yield	0.6788	0.0009	0.0861	0.0002	
	Ultimate Tensile Strength	0.6480	0.0018	0.0007	0.8106	
	Percent Strain at Break	0.6392	0.0021	0.0162	0.0025	
With Interaction	Tensile Modulus	0.7522	0.0024	0.0027	0.0093	0.0702
	Yield Stress	0.7765	0.0013	0.0147	0.0029	0.0069
	Percent Strain at yield	0.7951	0.0008	0.0482	0.0001	0.0673
	Ultimate Tensile Strength	0.7858	0.0010	0.0002	0.7769	0.0508
	Percent Strain at Break	0.6994	0.0069	0.0175	0.0052	0.3344

Table 5. Statistical Significance of Factors for Each Response

	Statistically Significance		
Response	**Layer height**	**Raster angle orientation**	**Interaction of raster angle and layer height**
Tensile Modulus (GPa)	Yes	Yes	No
Yield Stress (MPa)	No	Yes	Yes
Percent Strain at Yield	No	Yes	No
Ultimate Tensile Strength	Yes	No	No
Percent Strain at Break	Yes	Yes	No

A discussion of how each of the two factors influences different responses (i.e., mechanical properties) follows. Both the yield stress (Figure 2(a)) and the tensile modulus (Figure 2(b)) decreased with increasing layer height. Although not shown, the tensile strength also decreased with increasing layer height. This is not unexpected as both tensile modulus and tensile strength are greatly influenced by a material's density. As a material's density increases, so does stiffness (modulus) and strength. This would indicate that smaller layer heights result in larger densities for the items being produced.

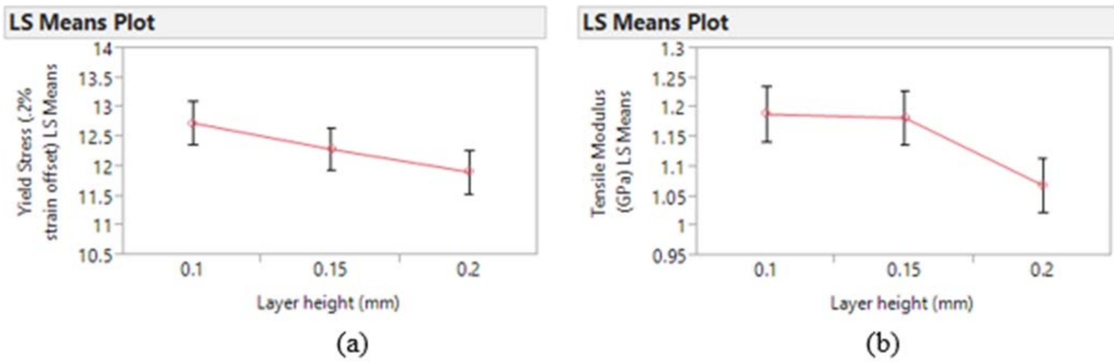


Figure 2. Yield Stress (a) and Tensile Modulus (b) Versus Layer Height

Additionally, the research showed that raster angle orientation was significant in influencing tensile modulus, yield stress, percent strain at yield, and percent strain at break. Figure 3 shows that stiffness was greatest in the ±45 angle orientation versus the 0/90 orientation. Even though the ±45 angle orientation is not directly aligned along the tensile direction, further analysis of the structure found that more layers were resisting in the tensile direction as compared to the 0/90 orientation. The 0/90 orientation only had half of its layers resisting tension since layers with raster angles orthogonal to the tensile force do not contribute greatly to stiffness or strength. This explains why the 0/90 orientation is less stiff then the ±45 angle orientation.

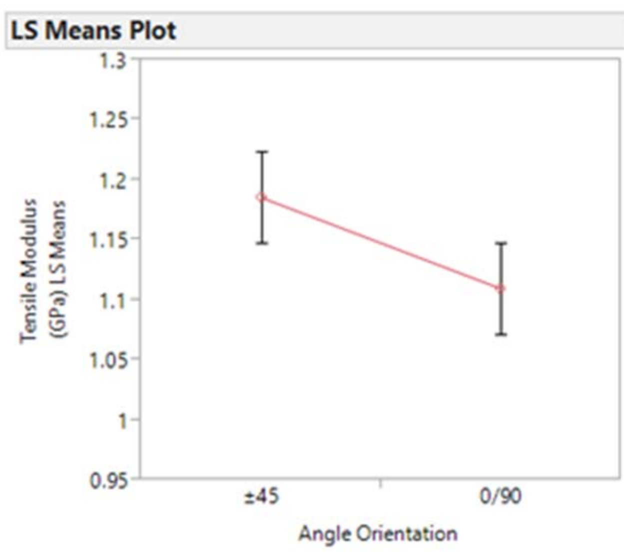


Figure 3. LS Means Plot of Tensile Modulus Versus Angle Orientation

To further investigate the effect of density on the mechanical properties of the PA6 nylon, a one-way ANOVA was performed for a cohort density variable and for each of the mechanical properties measured. The cohort variable for density was defined with two levels: “low density” and “high density.” Low density was defined as being less than 1.095 g/cm3 and high density was defined as being greater than 1.095 g/cm3. Identifying the density level (either “low” or “high”) was determined by visually evaluating the scatter plots of the mechanical properties and density to see if the data formed groups. Figure 4 shows the scatter plot of tensile modulus by density. Visual inspection of the figure reveals two groups of data points, with the points in the upper right quadrant of the figure being the high density group and the points in the lower left quadrant being the low density group. Based on this observation, a one-way ANOVA was performed for the density groups and each of the measured mechanical properties. From the ANOVA results, the differences between the means for low density and high density groups were determined to be statistically significant for tensile modulus, percent strain at yield, ultimate tensile strength, and percent strain at break. The mean yield stress between the low density and high density groups was found not to be statistically different. This suggests that material density alone could be the most influential contributing factor in material strength properties of PA6 nylon.

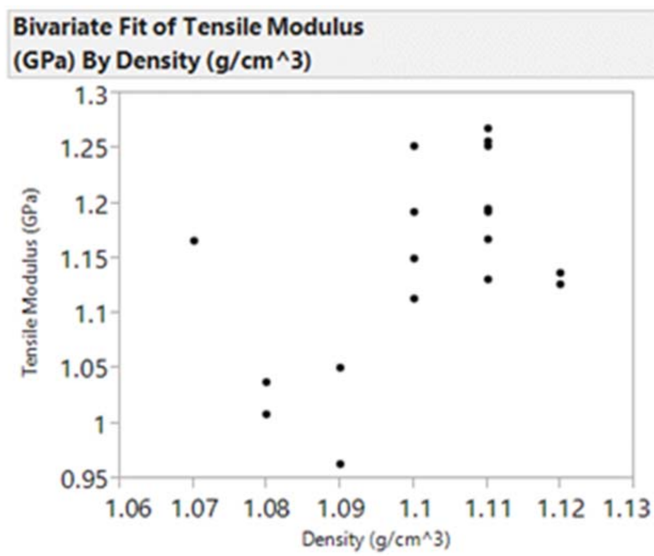


Figure 4. Scatter-Plot of Tensile Modulus by Density for FFF Nylon

Part 2: Continuous Carbon Fiber Composite Testing

In this part of the research, several continuous carbon fiber composites (CCFCs) were manufactured using the same pattern in the nylon-only specimens to determine their mechanical properties. When using fiber, the Mark One defaults to a pre-set layer height of 0.125 mm with no option to change this setting. Figure 5 shows a close-up view of the continuous carbon fiber composite. However, problems arose during the testing of the composite specimens. All but one of the specimens either broke in the grips or slipped in the grips during tensile testing, which voided the results of the test. Therefore, only one test specimen and procedure provided useful data. The specimen was 0.5267 inches wide, 0.1567 inches thick, and 6 inches long. The testing conditions included a room temperature of 71.6°F with a relative humidity of 41%. Figure 6 shows this single specimen after testing, and Figure 7 shows the stress-strain curve from the test. The ultimate tensile strength determined from this single specimen was 121.1 MPa and the tensile modulus was 9.9 GPa.

Figure 5. Close-Up View of Continuous Carbon Fiber Composite

Figure 6. Continuous Carbon Fiber Composite Specimen After Testing

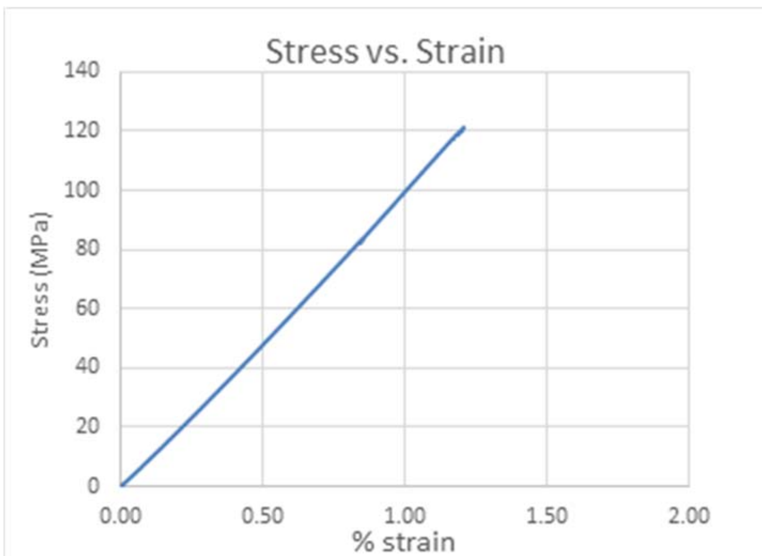


Figure 7. Stress-Strain Curve for Continuous Carbon Fiber Composite

To further analyze the specimen, a scanning electron microscope (SEM) was used to photograph the failure surface from one of the CCFC tensile specimens. Figure 8 shows the fracture surface of a carbon-fiber reinforced nylon composite specimen. The approximate thickness of the fracture is 2.331 mm. Figure 9 shows an alternative view of the fracture surface. In this image, discontinuities are visible between each nylon layer but not between rasters. This indicates that the coalescence of the nylon is not complete between layers but is nearly homogeneous between rasters.

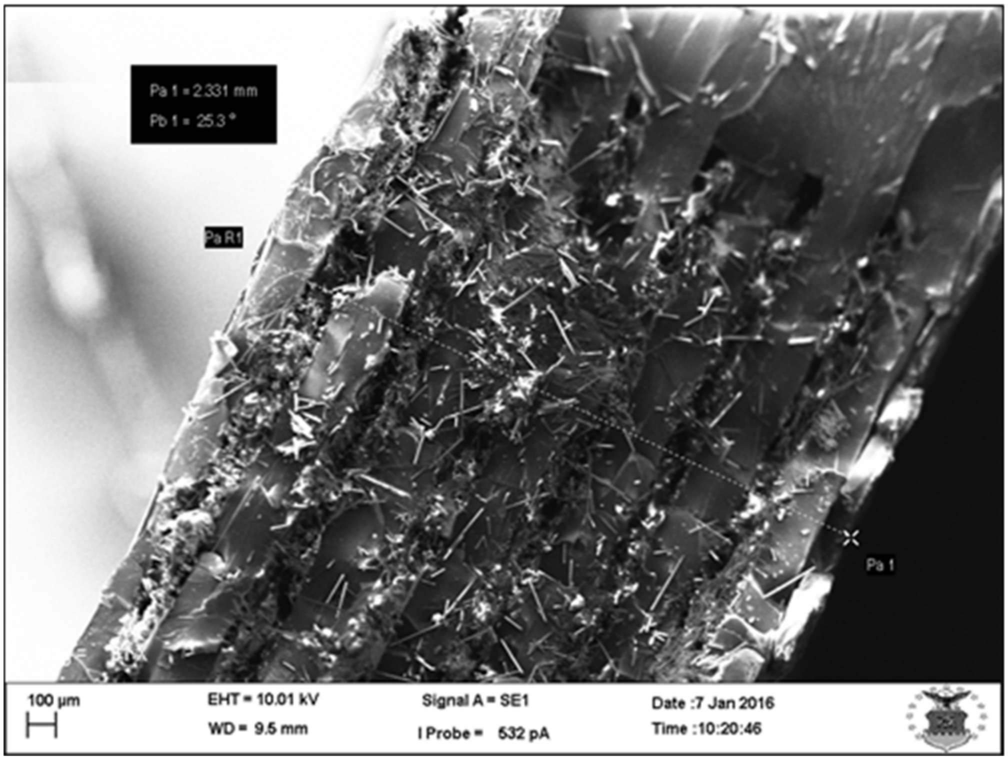


Figure 8. Fracture Surface of Carbon-Fiber Reinforced Nylon Composite

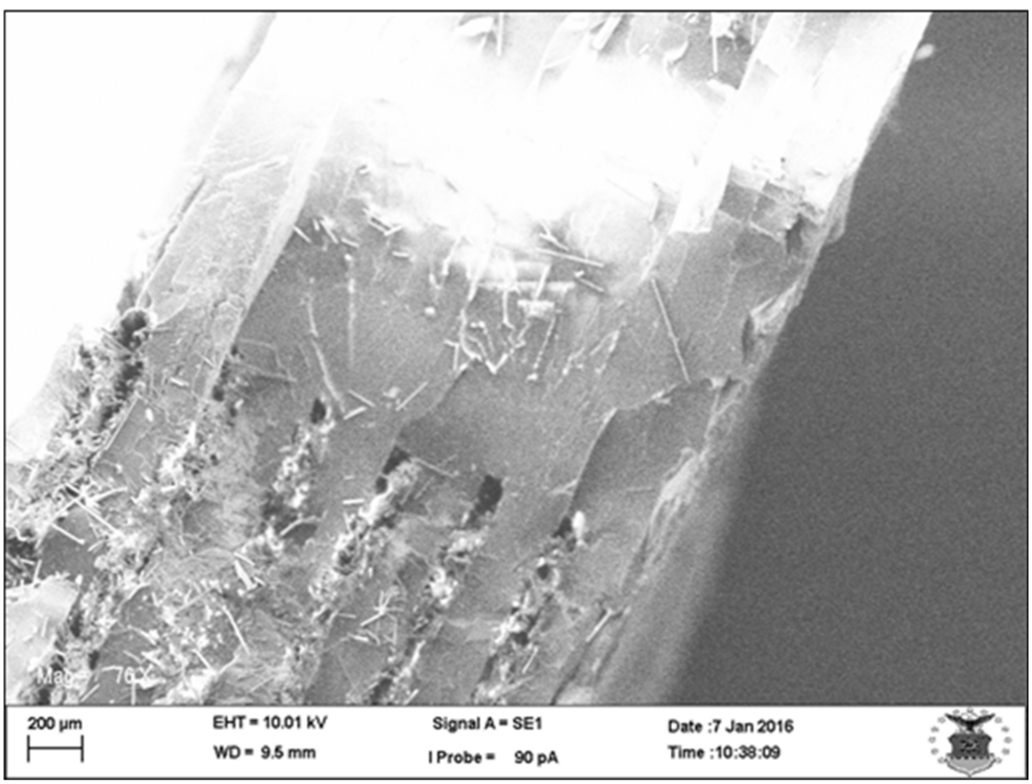


Figure 9. Fracture Surface of Carbon-Fiber Reinforced Nylon Composite

Cost Modeling Issues

When it comes to costs associated with AM processes, there is limited research regarding the major cost drivers (Lindemann et al., 2012) and specifically energy consumption (Meteyer et al., 2014). Focusing on energy consumption, Huang et al. (2015) summarized the existing literature regarding life-cycle energy and found that manufacturing energy consumption varied widely (52.2–4,849 MJ/kg, 26.9–66.02 kWh/kg, and 1.8–3,000 MJ/item). Based on their study, they made several observations:

> First, most studies focus only on the direct energy intensity of AM processes without comparisons to the energy and material requirements of the CM processes that are replaced. Second, most studies have considered polymeric AM technologies, due to their maturity, low cost, and widespread availability. Third, energy intensity estimates for AM processes vary widely across studies, primarily due to different material selections, component geometries, and data collection methods, which preclude direct comparisons of study results. Fourth, none of the studies considered application performance improvements due to changes in component geometries or, by extension, the environmental implications of such performance improvements.

Given the scope of the project, the literature regarding cost issues was limited to energy consumption by AM processes and conventional manufacturing technologies (i.e., bulk-forming and subtractive processes). In most of the research reported in the literature, specific energy consumption (SEC) is expressed in either MJ or kWh per kg (or volume) of material deposited. Only one study was found in the literature comparing these three manufacturing processes. Additionally, only a few studies were found that examined specific AM technology processes.

Yoon et al. (2014) performed a literature review regarding specific energy consumption (SEC) of various processes categorized as bulk-forming, additive, or subtractive manufacturing; they also investigated specific processes as case studies to compare results. The range of values from their literature review are summarized in Table 6; case study values were similar.

Table 6. Specific Energy Consumption

(Data obtained from Yoon et al., 2014)

Bulk-forming processes	0.11–5.82 kWh/kg for injection molding 0.62–7.78 kWh/kg for metal casting
Subtractive processes	2.3–188 J/mm^3 for milling 2.7–36.2 J/mm^3 for turning 9–65 J/mm^3 for drilling 343.4–1982.6 J/mm^3 for grinding
Additive processes	14.5–66.02 kWh/kg for Selective Laser Sintering (SLS) 23.08–346.4 kWh/kg for Fused Deposition Modeling (FDM) 14.7–163.33 kWh/kg for other processes

Yoon et al. (2014) found that the SEC of additive processes was about 100 times greater than bulk-forming processes and that subtractive processes, with SEC values ranging from 1–100s of kWh/kg, and consumed the least amount of energy. However, they noted a clear case of economy of scales. If only one item is being produced, the SEC was lower for additive processes; as the number of items being produced increased, the SEC of bulk-forming and subtractive processes decreased significantly. Cost had an opposite relationship in which the bulk-forming cost was greater than the additive cost when three or fewer items were being produced; when the number of items exceeded three, the additive cost increased sharply. When examining AM processes alone, they also found that there was no significant difference between plastic and metal methods. Therefore, Yoon et al. (2014) concluded that both energy consumption and production cost should be carefully considered, and that both are related to production quantities.

In earlier work, Gutowski et al. (2009) developed an empirically observed relationship between the energy consumption rate (J/kg) and the process rate (kg/h) for manufacturing processes; they subsequently found that processes with process rates less than 0.1 kg/h tend to consume at least 100 MJ per kilogram of material processed. Baumers et al. (2011) found similar results when examining two polymeric laser sintering (LS) processes. However, other LS studies have reported that higher process rates use less energy per kilogram of material deposited (Mognol et al., 2006; Morrow et al., 2006). Baumers et al. (2011) attribute this to better capacity utilization. Related to capacity utilization, Mognol et al. (2006) also demonstrated that AM energy consumption is affected by orientation in terms of the Z height of the item being produced. Telenko and Seepersad (2010) also suggest that Z height and the density of the items being produced affect energy consumption.

Baumers et al. (2010) compared the electricity consumption of selective laser melting and electron beam melting (two metallic AM processes). They showed that efficiencies for parallel processes differ significantly between production maximizing capacity utilization and one-off items. Furthermore, they also found that energy consumption is affected by material selection and layer thickness. Therefore, to substantiate claims that AM is more energy efficient than conventional manufacturing processes, they proposed that summary metrics (e.g., kWh/cm^3 or kWh/g) should be used.

In follow-on work, Baumers et al. (2011) compared the electricity consumption of two polymeric LS processes and demonstrated that energy consumption can be represented by job-dependent, time-dependent, geometry-dependent, and Z-height-dependent categories. Their analysis showed that the majority of LS energy consumption (56–61%) occurred during time-dependent activities. This was consistent with work by Lindemann (2012) showing that machine costs account for 73% of the costs. The calculated energy consumption rate of 36.04kWh/kg for their experiment was consistent with results reported in the literature (Baumers et al., 2011). Their primary conclusion was that productivity is a key factor in determining energy efficiency. Additionally, they suggest that energy efficiency is less for AM processes using a moving head for material deposition than processes using powder bed platforms.

Baumers et al. (2011) provided an overview of electricity consumption with several AM processes and reported energy consumption rates ranging from 61 to 4,849 MJ per kg deposited. Comparing the production of a single item to production maximizing capacity utilization, they concluded that capacity utilization is critical to energy efficient processes. In their experiments, energy savings ranged from 3.17% for FDM to 97.79% for LS processes. For LS and EBM processes specifically, full capacity utilization resulted in much greater energy efficiency compared to producing a single item. On the other hand, full capacity utilization resulted in minimal energy savings for FDM processes (primarily because system warm-up and cool-down are not as critical). Therefore, the use of FDM would be more applicable for serial processes. In summary, their results show that full capacity operation results in less energy consumption per kg of material deposited for all operating scenarios and materials used in their experiments.

After studying a Direct Metal Laser Sintering (DMLS) system, Baumers et al. (2012) found that energy consumption and production cost should not be considered dependent on production quantity. Instead, they suggest that capacity utilization is the primary factor determining process efficiency. After optimizing the build configuration using a volume packing algorithm, they developed a model using parameters for speed, energy consumption, and production cost. The time and energy consumption model they developed, which was validated experimentally, is shown in Equations 2 and 3,

$$T_{Build} = T_{Job} + (\alpha_{Time})(l) + \sum_{z=1}^{z}\sum_{y=1}^{y}\sum_{x=1}^{x} T_{Voxel\,xyz} \quad (2)$$

$$E_{Build} = E_{Job} + (E_{Time})(\mathrm{T}_{Build}) + (\alpha_{Energy})(l) + \sum_{z=1}^{z}\sum_{y=1}^{y}\sum_{x=1}^{x} E_{Voxel\,xyz} \quad (3)$$

where T_{Build} and E_{Build} are the estimated build time and energy investment, respectively, for the complete build operation; T_{Job} and E_{Job} are the time and energy, respectively, associated with machine start-up; E_{Time} is the energy consumption rate (MJ/s); α_{Time} and α_{Energy} are the time and energy, respectively, associated with adding each layer of material; l is the total number of layers; and $T_{Voxel\,xyz}$ and $E_{Voxel\,xyz}$ are the time and energy required to process each voxel (which is a three-dimensional pixel). The total cost estimate can then be expressed as

$$C_{Build} = (C_{Indirect})(T_{Build}) + (w)(P_{Raw\,material}) + (E_{Build})(P_{Energy}) \quad (4)$$

where C_{Build} is the total cost estimate for the build operation, $C_{Indirect}$ is the indirect cost rate, w is the mass of all parts manufactured, P_{Raw} material is the unit price of the material used in the AM process, and P_{Energy} is the price of electricity. Using a full-capacity build experiment consisting of 85 items, Baumers et al. (2012) found that 92.6% of the voxels and 19.8% of the capacity volume were occupied and that 1,059.56 MJ of energy were consumed. This equated to 1.96 MJ/cm³ at a production cost of 5.71 £/cm³. Their results further demonstrated the importance of considering capacity utilization when determining cost and energy consumption metrics to reflect efficient processes.

Baumers et al. (2012) found that the single-step nature of additive processes facilitates the ability to measure energy consumption and production costs. However, since AM processes can simultaneously produce multiple items in a parallel fashion (Ruffo et al., 2006), the degree of capacity utilization affects energy consumption and production cost metrics (Ruffo et al., 2006; Ruffo & Hague, 2007; Baumers et al., 2011; Baumers et al., 2012). Therefore, it is necessary to allocate the total cost and energy consumption to each item being produced in an equitable manner. Baumers et al. (2012) concluded that the quantity and variety of items, in combination with the capability to utilize the available machine capacity, have an impact on process efficiency in terms of both energy and cost. Similarly, Lindemann et al. (2012) showed that AM is more attractive to companies involved in batch production who can maximize capacity utilization.

Conclusions

The experiments conducted during this research showed that both layer height and raster angle orientation impact the mechanical properties of the specimen manufacturing using the FFF process. Specifically, layer height was significant in influencing tensile modulus, ultimate tensile strength, and percent strain at break; and raster angle orientation was significant in influencing tensile modulus, yield stress, percent strain at yield, and percent strain at break. The optimal condition maximizing tensile modulus, ultimate tensile strength, and percent strain at break is a layer height of 0.1 mm and a ±45 raster angle orientation. The optimal condition that maximizes yield stress is a layer height of 0.1 mm and a 0/90 raster angle orientation. The optimal condition that maximizes percent strain at yield is a layer height of 0.2 mm and a 0/90 raster angle orientation. Additionally, the ultimate tensile strength and tensile modulus were lower for the FFF nylon than compression-molded nylon; however, the percent elongation at break was comparable. Finally, the composite specimen that was tested had an ultimate tensile strength of 121.1 MPa and a tensile modulus of 9.9 GPa.

Based on the limited literature review for cost and energy, the following points should be considered regarding energy consumption by AM processes. The most critical factor determining process efficiency, both in terms of energy consumption and production cost, appears to be capacity utilization. This implies that costs and energy consumption must be allocated in an equitable manner, which means that summary metrics (e.g., kWh/cm^3 or kWh/g) must be used. Other factors, which can be related to capacity utilization, include Z height, density, material selection, and layer thickness. However, the time-dependent nature of energy consumption must also be considered; for example, LS and EBM processes benefit greatly from full capacity utilization while FDM processes benefit minimally.

Research Implications

The research shows that the mechanical properties of FFF-manufactured items are impacted by changing the process parameters of layer height and raster angle orientation. In the future, it is likely that engineers will be able to use additive manufacturing to create materials that meet certain performance requirements by specifying a unique treatment of additive manufacturing process parameters. Furthermore, measuring the density of additively manufactured parts could be a non-destructive method of quality assurance. The results from the density investigation revealed that different levels of density showed differences in the mean mechanical properties. The FFF nylon specimens with a “high” level of density showed greater ultimate tensile strength and tensile modulus compared to the FFF nylon specimens with a “low” level of density.

An area for future research is to investigate how different nylon and fiber layup sequences influence mechanical properties. Two possible layup sequences that could be tested are shown in Figure 10. Each sequence has the same number of nylon and carbon fiber layers, with 10 carbon fiber layers and 12 nylon layers. Each layup is also symmetric about the center of the layup to prevent moment forces from influencing testing results. For layup A, each carbon layer is sandwiched between two layers of nylon. For layup B, the layers alternate between two nylon layers and two carbon fiber layers.

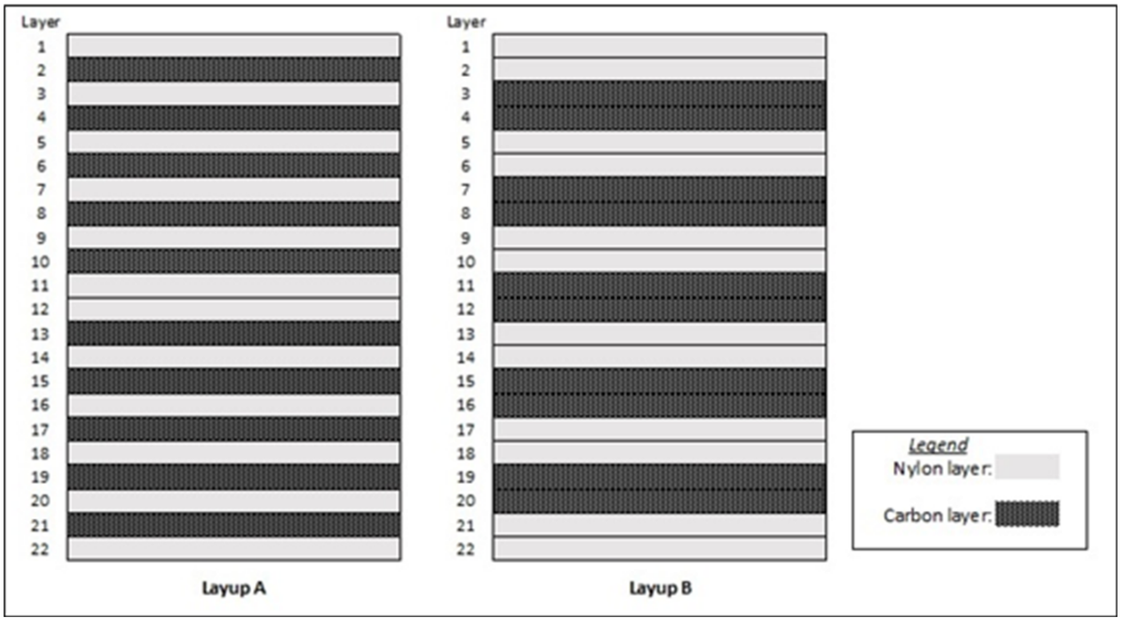


Figure 10. Two Possible Layup Sequences for a Future DOE Experiment

Another area for future research is to develop a better understanding of the relationship between carbon fiber volume fraction and tensile modulus of continuous carbon fiber composites (CCFCs) made through additive manufacturing. The tensile modulus of a single carbon fiber towpreg on a printed nylon layer can be determined through tensile testing. A duplicate CCFC specimen can then be printed to determine the volume fraction of a single carbon towpreg. From this information, a relationship can be made between fiber volume fraction and tensile modulus. This relationship model can be used to predict the tensile modulus of a given carbon fiber fraction. An experiment can then be performed to test the validity of the fiber volume fraction-tensile modulus relationship model.

With these advantages in mind, AM could revolutionize the military supply chain. An AM machine can manufacture needed components or tools in austere areas that are far removed from supply lines. Designs can be made anywhere in the world and sent electronically to a strategically placed machine on the battlefield. Furthermore, in an austere fiscal environment, the military will continue to maintain legacy systems. However, as these systems continue to age, maintaining a supply inventory of spare parts, which become increasingly difficult to obtain, is a challenge (Brown et al., 2014). Instead of going through a lengthy acquisition process to acquire critical replacement parts that have since gone out of production, additive manufacturing can create replacement parts on-demand (Brown et al., 2014).

References

Ahn, S. H., Montero, M., Odell, D., Roundy, S., & Wright, P. K. (2002). Anisotropic material properties of fused deposition modeling ABS. *Rapid Prototyping Journal, 8*(4), 248–257.

Barnatt, C. (2013). *3D printing: The next industrial revolution*. Lexington, KY: ExplainingTheFuture.com.

Baumers, C., Tuck, C., Hague, R., Ashcroft, I., & Wildman, R. (2010). A comparative study of metallic additive manufacturing power consumption. In *Proceedings of the 21st Annual International Solid Freeform Fabrication Symposium.* Austin, TX.

Baumers, M., Tuck, C., Bourell, D. L., Sreenivasan, R., & Hague, R. (2011). Sustainability of additive manufacturing: Measuring the energy consumption of the laser sintering process. Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture, 225*(12), 2228–2239.

Baumers, C., Tuck, C., Wildman, R., Ashcroft, I., & Hague, R. (2011). Energy inputs to additive manufacturing: Does capacity utilization matter? In *Proceedings of the 22nd Annual International Solid Freeform Fabrication Symposium.* Austin, TX.

Baumers, M., Tuck, C., Wildman, R., Ashcroft, I., Rosamond, E., & Hague, R. (2012). Combined build-time, energy consumption, and cost estimation for direct metal laser sintering. In *Proceedings of the 23rd Annual International Solid Freeform Fabrication Symposium.* Austin, TX.

Black, S. (2014, May). 3D printing continuous carbon fiber composites? *High-Performance Composites.*

Brown, R., Davis, J., Dobson, M., & Mallicoat, D. (2014, May–June). 3D printing: How much will it improve the dod supply chain of the future? *Defense Acquisition, Technology and Logistics*, 6–10.

Callister, W. D., & Rethwisch, D. G. (2012). *Fundamentals of materials science and engineering: An integrated approach.* John Wiley & Sons.

Daniel, I. M., & Ishai, O. (2003). *Engineering mechanics of composite materials* (2nd. ed.). New York, NY: Oxford University Press.

Foran, B., Lenzen, M., Dey, C., & Bilek, M. (2005). Integrating sustainable chain management with triple bottom line accounting. *Ecological Economics, 52*(2), 143–157.

Gibson, I., Rosen, D. W., & Stucker, B. (2010). Extrusion-based systems. In *Additive manufacturing technologies: Rapid prototyping to direct digital manufacturing* (pp. 160–186). Springer US.

Gutowski, T., Branham, M., Dahmus, J., Jones, A., Thiriez, A., & Sekulic, D. (2009). Thermodynamic analysis of resources used in manufacturing processes. *Environmental Science & Technology, 43*(5), 1584–1590.

Horn, T., & Harrysson, O. (2012). Overview of current additive manufacturing technologies and selected applications. *Science Progress, 95*(3), 255–282.

Huang, R., Liu, P., Mokasdar, A., & Hou, L. (2013). Additive manufacturing and its societal impact: A literature review. *Journal of Advanced Manufacturing Technology, 67*, 1191–1203.

Huang, R., Riddle, M., Graziano, D., Warren, J., Das, S., Nimbalker, S., Cresko, J., & Masanet, E. (2015). Energy and emissions saving potential of additive manufacturing: The case of lightweight aircraft components. *Journal of Cleaner Production*. http://dx.doi.org/doi:10.1016/j.jclepro.2015.04.109

Lindemann, C., Jahnke, U., Moi, M., & Koch, R. (2012). Analyzing product lifecycle costs for a better understanding of cost drivers in additive manufacturing. In *Proceedings of the 23rd Annual International Solid Freeform Fabrication Symposium*. Austin, TX.

Lindemann, C., Jahnke, U., Moi, M., & Koch, R. (2013). Impact and influence factors of additive manufacturing on product lifecycle costs. In *Proceedings of the 24th Annual International Solid Freeform Fabrication Symposium*. Austin, TX.

Love, L. J., Kunc, V., Elliot, A., & Blue, C. (2014). The importance of carbon fiber to polymer additive manufacturing. *Journal of Materials Research, 29*(17), 1893–1898.

Mani, M., Lyons, K., & Gupta, S. (2014). Sustainability characterization for additive manufacturing. Jou*rnal of Research of the National Institute of Standards and Technology, 119,* 419–428.

MatWeb, LLC. (2015, August 28). Retrieved from MatWeb: Material Property Data website: http://www.matweb.com/

Meteyer, S., Xu, X., Perry, N., & Zhao, Y. (2014). *Energy and material flow analysis of binder-jetting additive manufacturing processes.* 21st CIRP Conference on Life-Cycle Engineering, Trondheim, Norway.

Mognol, P., Lepicart, D., & Perry, N. (2006). Rapid prototyping: Energy and environment in the spotlight. *Rapid Prototyping Journal, 12*(1), 26–34.

Morrow, W., Qi, H., Kim, I., Mazumder, J., & Skerlos, S. (2007). Environmental aspects of laser-based and conventional tool and die manufacturing. *Journal of Cleaner Production, 15,* 932–943.

Namiki, M., Ueda, M., Todoroki, A., Hirano, Y., & Matsuzaki, R. (2014). 3D printing of continuous fiber reinforced plastic. In *Proceedings of SAMPE* Seattle. Seattle, WA.

Oak Ridge National Laboratory. (2010). *Aerospace Workshop planning session summary report*. DOE/ITP/ORNL.

Ruffo, M., & Hague, R. (2007). Cost estimation for rapid manufacturing—Simultaneous production of mixed components using laser sintering. Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture, 221*(11), 1585–1591.

Ruffo, M., Tuck, C., & Hague, R. (2006). Cost estimation for rapid manufacturing—Laser sintering production for low to medium volumes. Proceedings of the Institution of Mechanical Engineers, Part B: *Journal of Engineering Manufacture, 220*(9), 1417–1427.

Schroeder, M., Falk, B., & Schmitt, R. (2015). *Evaluation of cost structures of additive manufacturing processes using a new business model*. 22nd CIRP Conference on Life-Cycle Engineering, Sydney, Australia.

Serres, N., Tidu, D., Sankare, S., & Hlawka, F. (2011). Environmental comparison of MESO-CLAD® process and conventional machining implementing life-cycle assessment. *Journal of Cleaner Production, 19*, 1117–1124.

Tuck, C., Hague, R., Ruffo, M., Ransley, M., & Adams, P. (2008). Rapid manufacturing facilitated customization. *International Journal of Computer Integrated Manufacturing, 21*(3), 245–258.

Yoon, H., Lee, J., Kim, H., Kim, M., Kim, E., Shin, Y., Chu, W., & Ahn, S. (2014). A comparison of energy consumption in bulk forming, subtractive, and additive processes: Review and case study. *International Journal of Precision Engineering and Manufacturing–Green Technology, 1*(3), 261–279.

Acknowledgments

We thank our colleagues, Lt Col Chad Ryther and Dr. Craig Przybyla, from the Air Force Research Laboratory's Materials & Manufacturing Directorate (AFRL/RX) for their expertise and insight greatly assisting this research. Without their support and partnership, this work would not have been possible.

Panel 18. Forecasting and Controlling Costs in Weapons Systems Procurement

Thursday, May 5, 2016	
1:45p.m. – 3:15 p.m.	**Chair: Todd Calhoun,** Director of Program Assessment & Evaluation, Programs & Resources Department, Headquarters Marine Corps ***Costing Future Complex & Novel Projects*** Michael Pryce, Centre for Defence Acquisition, Cranfield University ***Controlling Costs: The 6-3-5 Method—Case Studies at NAVSEA and NATO*** Bruce Nagy, President/CEO, Catalyst Technologies Morgan Ames, Senior Advisor, Catalyst Technologies ***Costing for the Future: Exploring Cost Estimation With Unmanned Autonomous Systems*** Ricardo Valerdi, Professor, University of Arizona CPT Thomas Ryan, Jr., U.S. Army, Professor, U.S. Military Academy

Todd Calhoun—currently serves as the Director of Program Analysis and Evaluation, Programs and Resources Department, Headquarters Marine Corps. In this capacity, he is responsible for providing the Commandant, Assistant Commandant, and Deputy Commandant for Programs and Resources with independent assessments of Marine Corps programs.

Prior to his current assignment, Dr. Calhoun served as a Senior Operations Research Analyst with the Office of the Secretary of Defense, Cost Assessment & Program Evaluation (CAPE) directorate, from 2008 to 2012. While with CAPE, Dr. Calhoun led the Special Operations/Counterterrorism portfolio through four Program Budget Reviews and the 2010 Quadrennial Defense Review. He also oversaw the Civil Affairs and Psychological Operations portfolios while assigned to the Irregular Warfare Division. In his final assignment with CAPE, Dr. Calhoun led the Airborne Intelligence, Surveillance, and Reconnaissance portfolio through the Fiscal Year 2013–2017 Program Budget Review.

Prior to joining OSD, Dr. Calhoun was the Chief Analyst of Marine Corps Systems Command from 2004 to 2008. In this capacity, he was responsible for the development and delivery of a wide range of analyses used in the defense acquisition system. He also led the command-wide Operations Research/Systems Analysis Community of Practice and the Economic & Business Analysis Branch with a mission to provide independent analysis to enable sound investment decision-making. Dr. Calhoun was initially assigned to Marine Corps Systems Command as a defense systems analyst in 1999 while on active duty. He later transitioned to the civil service and was reassigned as an operations research analyst responsible for the development of ground weapons programs. Dr. Calhoun began his Marine Corps career as an infantry officer with 2d Battalion, 3d Marine Regiment in Kaneohe Bay, HI. He graduated from Miami University (Ohio) in 1993 and the Naval Postgraduate School in 1998, and received his doctorate from George Mason University in 2006.

Costing Future Complex & Novel Projects

Michael Pryce—is a Lecturer in defence acquisition at Cranfield University. He teaches across a range of subjects, including the use of costing in acquisition and planning at the master's level. His research focuses on the use of cost as a design tool. He is currently leading a second costing research project with colleagues from UCL, Imperial College, and DSTL and developing the Low Cost by Design (LCxD) network. Prior to joining Cranfield, Dr. Pryce worked at Manchester Business School, where he carried out two NPS-funded projects looking at the role of costing and business models in operations and support. [m.pryce@cranfield.ac.uk]

Abstract

A program of current research funded by the United Kingdom (UK) research councils and supported by the UK Defence Science and Technology Laboratory (DSTL) is reported. The work involves pioneering data collection, analysis, and tool development to support future air combat systems. The role of a community of users and developers of the data and tools is reported, as well as the underpinning philosophy of the work and future prospects for its wider application.

Summary

The United Kingdom National Audit Office *Major Projects Report 2014* states that "Project teams continue to be over-optimistic in their forecasts of both procurement and support costs" and that "Budgets set using over-optimistic forecast costs could result in overall budgets for procurement and support being significantly understated" (Great Britain, National Audit Office, 2015). Correcting the causes of such failings is the aim of this research, with the outputs being intended for use by a wide range of stakeholders.

The research will lead to enhanced methods, tools, and understanding of costing in defence. This will be realised through the creation of a database of historic defence costs, a set of project histories, and other contextual information. An associated viewer tool, and a rating and health check tool for use by practitioners, will be developed.

The project will take work that has previously been at a largely conceptual level through to initial field trials and validation. Later phases leading to full deployment are also planned.

The end users for the work will be in the Aerospace, Defence, and Manufacturing sectors. The successful delivery of this project will allow users to easily organise and access key contextual information that will support more accurate forecasting of project costs and schedules. This will allow more efficient decision making.

During the period of active research, the DSTL provides the initial customer and significant support in informing the development of the work. The project is currently planned to run July 1, 2015, to June 30, 2016.

Project Overview

The research involves the collation of a large, varied data set on historic defence project costs, as well as associated qualitative information in the form of project histories. The work will lead to a wider understanding of the root causes of cost in defence, as well as enhancing knowledge of how to improve the accuracy of early stage project cost estimates.

The research findings will be linked semantically and interpreted through a viewer tool that relates data to projects interactively, based on HTML5, allowing users to contextualise cost data. A second tool to allow project teams to carry out costing "health checks" is also being developed. This will allow users of the viewer to focus their interests on the data and examples that matter. The project's top level structure is shown in Figure 1.

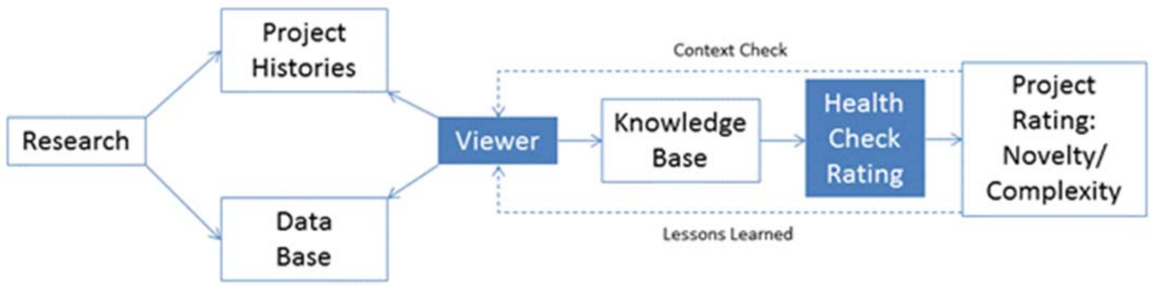


Figure 1. Top Level Structure
Note. Tools are shown in blue.

The project will focus its initial data collection on a number of combat air systems. These have already been partly researched by the team in prior work, and therefore the context is reasonably well understood. This will allow the work to focus on more in-depth data gathering and the development of the key outputs. A key focus of the work will be in tackling variety in the data collected, one of the "3 Vs" of Big Data. The outline programme is shown in Figure 2.

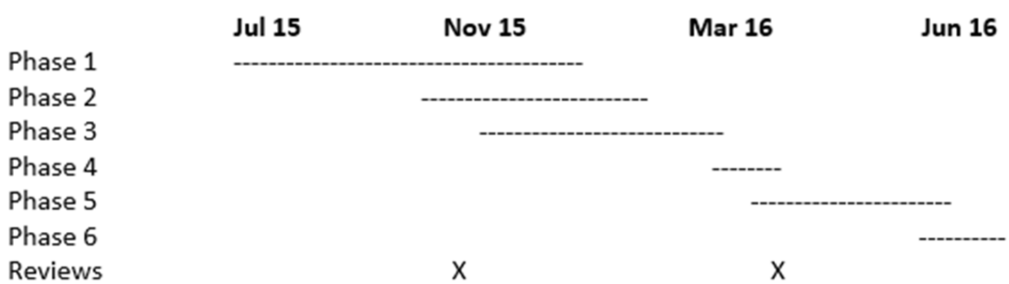


Figure 2. Programme of Work
Note. Project phases: 1 = Data collection, 2 = Data presentation and evaluation, 3 = Initial development and testing of viewer tool, 4 = Development of health check tool, 5 = User trials, 6 = Revision and expansion plan.

The project reviews are carried out by Professor David Kirkpatrick, who acts as an "external examiner" of the thinking behind the project, as well as contributing his own extensive experience in this field to help guide the work and its application.

End-Users

User trials involve a range of potential end users, with informal meetings held ahead of the planned trials. The initial end users for the work are in the defence sector, including the UK Ministry of Defence (MOD), particularly Defence Equipment & Support and Air Command, and UK defence companies such as BAE Systems and Rolls Royce. The work also has the potential to meet already articulated needs from the Australian Department of Defence and the Norwegian MOD.

The successful delivery of this project will allow end users to easily organise and access key contextual information that will support more accurate forecasting of defence project costs and schedules, especially complex and novel ones such as unmanned combat systems.

Technology Readiness

The project will take work that has previously been at a largely conceptual level and develop it to initial field trials. For the tools discussed, this would equate to going from TRL 1 to TRL 6.

Outputs

The project is intended to lead to facilitated workshops, the creation of a network of cost researchers and end users, a number of high-quality academic publications, and the development of a database of defence costs and project histories.

The findings will be used to develop teaching material for the Cost Estimating and Planning module of Cranfield University's Defence Acquisition Management MSc. Students on the MSc. are mainly senior managers from potential end user organisations inside defence. The work will also be relevant to the UK MOD's Financial and Military Capability management environment, which provides the high-level planning framework for UK defence acquisition.

Outcomes

Beyond the completion of the proof of concept work represented by this project, a number of additional outcomes are being pursued:

- Licensing of tools in a wider range of user organisations
- Database to be openly accessible within the costing community
- The development of new decision making approaches in the pre-concept phase
- Better understanding of organisational learning and corporate memory
- Links to existing costing methods that rely on data similarity
- Further development to other domains beyond air, and sectors outside defence

Rationale for the Research

For many years, the UK has struggled to maintain an extensive, coherent and useable set of project costing data for the development of major new platforms. Organizational changes within acquisition, contracting strategies and the high turnover in defence project staff have contributed to both the loss of data and low numbers of experienced staff able to interpret such data (Gray, 2009; Levene, 2011).

These problems have been exacerbated by the focus in the last 15 years on the support and sustainment of existing equipment. This has led to the widespread use of commercial costing tools that, while acceptable for the costing of in-service platforms, are of less use for the development of highly novel platforms, especially when the proprietary nature of the tools' databases means that it is difficult to make allowances and adjustments for novelty. The "bottom up" approach to costing works well for established designs, but is almost impossible for novel ones.

At the same time, the understanding of the need to cost "through life" that has been engendered by the focus on support and sustainment has made the challenges of generating realistic costs at an early stage of a project, when major commitments may be entered into, increasingly important, though no less difficult.

While prior work has shown that the operation and support costs are not necessarily fixed early in the life cycle (Pryce, 2011), the need to capture their likely range, as well as to ensure that the full range of issues outlined by Pugh (1992) for development and production are also addressed, indicated that current data sets and methods in the UK were inadequate. The research was developed on the basis of creating a new method with related data and tools that can be generated dynamically and sustained by a wide community over a long period of time.

This new approach is not intended to deliver a model. Rather, it is aimed at creating a heuristic or "therapeutic" set of data, tools and the necessary understanding to use them that will enable better decisions to be made by project teams in defence acquisition. Philosophically, it differs from many approaches that are currently used, but aims to do so in a complementary way. Such aspects will be explored through later work on a related project, which will seek to add a degree of modelling capability to some of the data being generated, and the idea of Bayesian belief networks will also be explored in this additional work.

The ultimate objective is to help support the "smart customer" inside the UK MOD, and to provide an approach that promotes dialogue between government and industry.

The Project Described

The project is made up of a number of activities and phases as shown in Figures 1 and 2 above. A more detailed description of these is given below in order to understand the intent of each.

Figure 3 shows how the work begins with field research. This consists of the extensive collation of information and data from archives across the United Kingdom. The National Archive at Kew, BAE Systems Heritage archives, the Royal Air Force Museum, Brooklands Museum, and others are the sources of a large amount of historic material that has been lost to the United Kingdom MOD over the years. A key part of the research is not just the collection of this data, but the central recording of the location and source of the material from the archives. Frequently, information from one source helps make sense of information from another. It is these cross connections that provide one of the key strengths of the research.

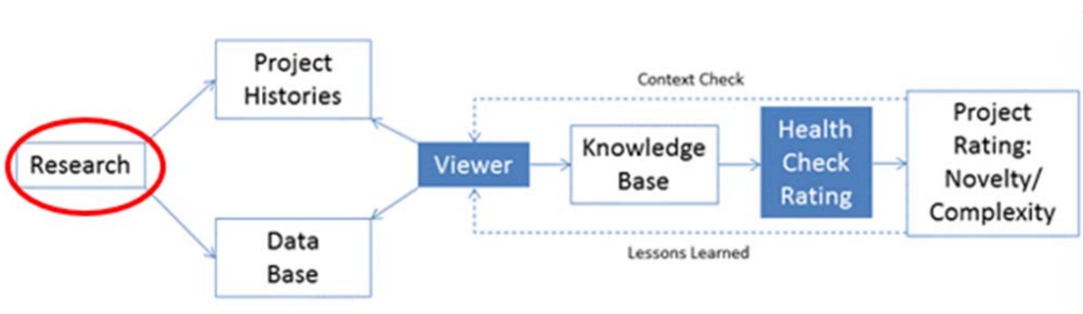


Figure 3. Research

The material obtained from the research is broken down into two basic elements (Figure 4). Qualitative information is used to develop project histories for past platforms. These are not simple case studies, but contextually rich histories of past projects. They can be used as stand-alone histories for one project or platform or related to others to give a programme or portfolio view.

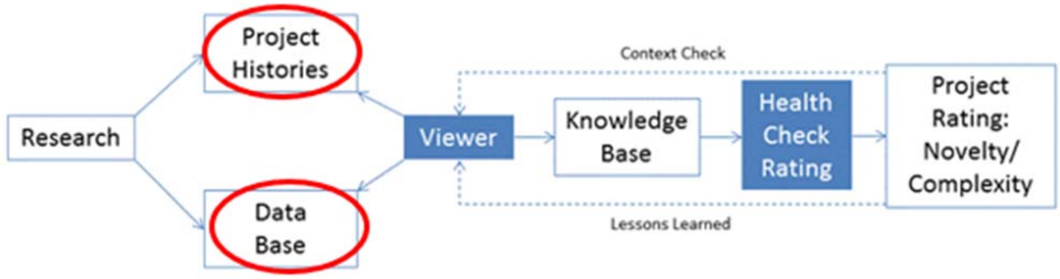


Figure 4. Project Histories and Data Base

The project histories can also be related to the quantitative data sourced from the research. This is data in both raw and semi-processed form. One of the key methods in the generation of the recorded data is its connection to its historic context to ensure factors such as exchange rates, inflation, etc. are properly accounted for, and that suitable recording of the derivation of the presented data ensures it is fully explicable to end users.

The end users explore the histories and data obtained from the research using a viewer tool. This is intended to provide their main method of generating a knowledge base for their own immediate needs (Figure 5). The viewer tool acts as the main, IT based resource that project teams can utilise over time to help them work with the data and histories. It is intended to show the data, histories and associated project timeline information in a single screen. A conceptual version is shown in Figure 6, and a mock-up is shown in Figure 7.

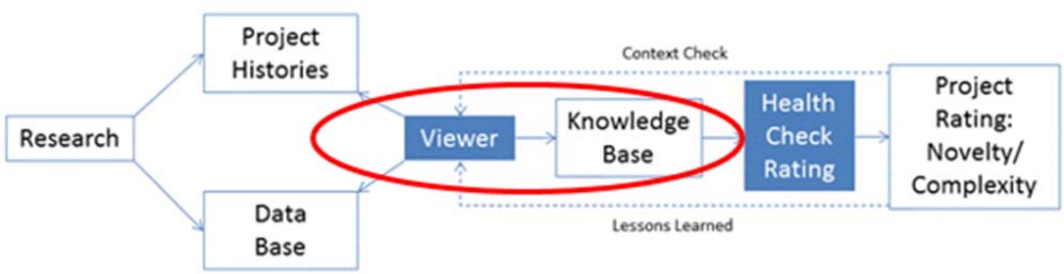


Figure 5. Viewer Tool and Knowledge Base

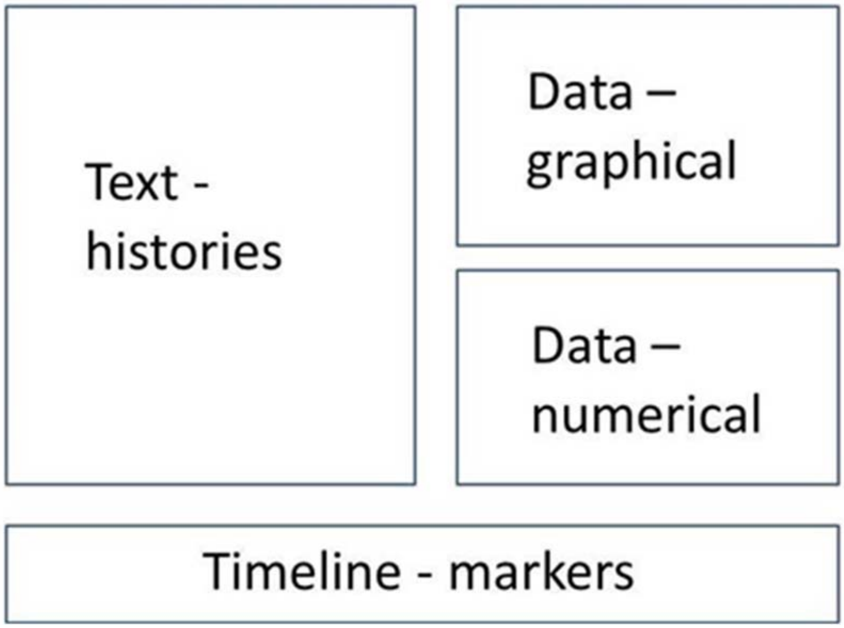


Figure 6. Viewer Tool Concept

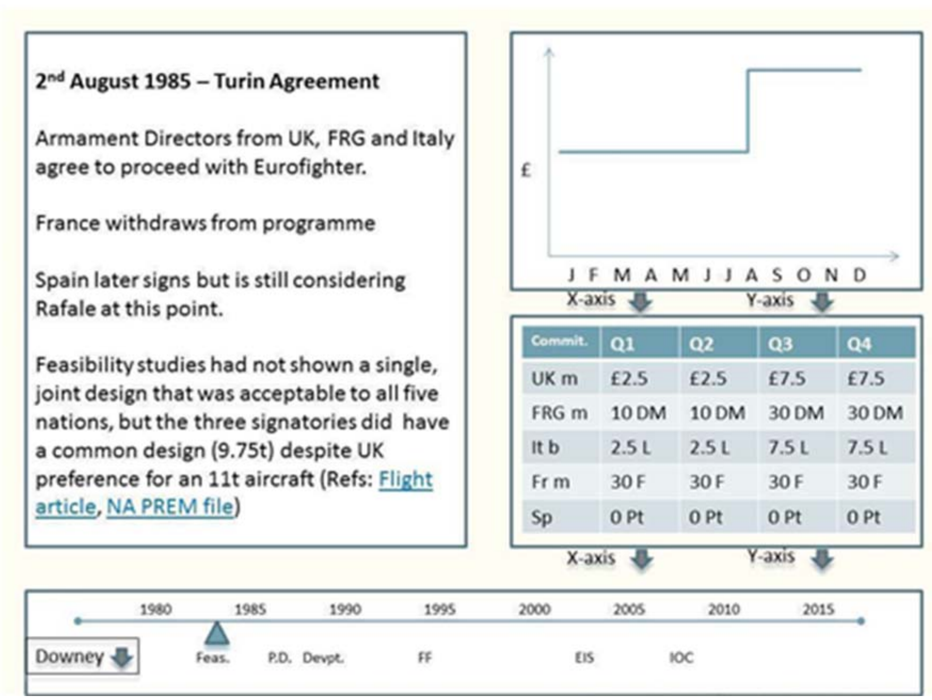


Figure 7. Viewer Tool Mock-Up (Indicative Data)

The viewer tool is being developed using HTML5. This is intended to make it platform agnostic and "future proof." A search function, using meta-tagging and standard web search engine methods, allows intuitive, adaptable use with minimum functional training.

HTML5 will also allow for development of the viewer tool by project teams using in-house or external resources that only require common, standard technologies. Such an approach also means that the viewer tool can run in secure/shared/stand-alone modes defined by the user's own IT environment. Desktop/tablet use is possible, as is networked and online use.

Before using the viewer tool, the data and histories that are most relevant to the user are identified through a health check tool. This is intended to identify the characteristics of the user's current project and to help map this across to the histories and data shown by the viewer tool (Figure 8).

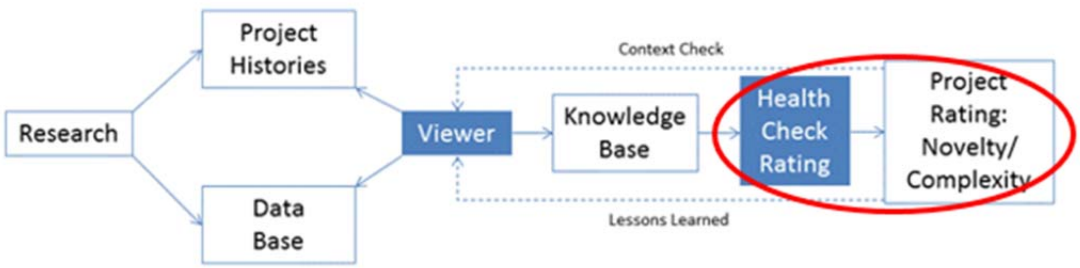


Figure 8. Health Check Tool and Project Rating

The health check tool uses qualitative/subjective inputs and relies on "polling" of the user project team to identify the characteristics of their current project. This is intended to capture diverse views using a rating system conceptually similar to the "Cooper Harper" style hierarchy used for piloting aircraft. Alphanumeric outputs are used to identify "true likeness" projects from which learning can occur and to help identify if the context of the user's project is similar to the one(s) explored using the viewer tool. These activities are usually carried out through an on-site facilitation with the user (Figure 9).

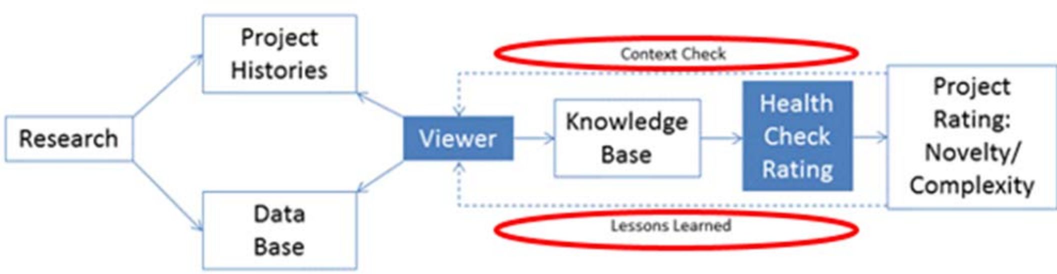


Figure 9. Context Check and Lessons Learned

Discussion

The description of the project illustrates how the project is being developed in an essentially linear fashion. However, this masks the interactive, incremental and innovative ways that the tools, data, histories and basic research can be used.

For users, their knowledge base can be constantly reinforced and updated by repeated use of the tool, forming a loop that links their learning from the tool with the changed perception and performance of their project. This can be allied to adaptation of the viewer tool to their own specific needs. Both aspects are intended to go some way to compensating for the loss of skills and data that has happened in the UK MOD costing community in the last 20 years with regard to cost forecasting for complex and novel projects.

For the research team, involvement in users' activities through data provision, tool set up and facilitation helps to refine the further development of the research outputs, as well as helping to focus the collection of further data. It has already been found that much more data and information exists in archives than was expected, and that the corporate memory on projects such as the Eurofighter Typhoon and F-35B (i.e., projects still in development but decades old) can be usefully refreshed using archival sources.

Although the work is currently focussed on air combat systems, a second project, funded by the DSTL (called Air Systems Programme Data) is now beginning that will start to address how an approach that works in the air domain can be extended, notably to the maritime sector. This will also widen academic participation, with the research team involving staff from University College, London (David Andrews), Imperial College (Dr. Michael Weatherburn), and industry (Andrew Dakin, 649 Ltd.).

The approach taken to this project mirrors many of the techniques becoming increasingly prevalent in the emerging field of digital humanities (Burdick et al., 2012), while aiming to avoid some of the identified pitfalls. As such, it represents a new approach to costing, in which the subjective understanding of contextualised information and data is used to assist decision making, rather than the supposed objective evidence given by "hard" data.

As the work continues, it is becoming increasingly clear that the information and data provided on a screen is only the tip of the iceberg of potential benefits. Re-building, through digital humanities type techniques, the institutional capacity to understand cost as part of the early stage design process is a trans-disciplinary activity. To benefit from it requires a diverse set of engagements with users in a truly open and innovative manner.

References

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnap, J. (2012). *Digital_humanities*. Cambridge, MA: MIT Press.

Gray, B. (2009). *Review of acquisition for the secretary of state for defence: An independent report by Bernard Gray.* Ministry of Defence.

Great Britain, National Audit Office. (2015). *Major projects report 2014 and the equipment plan 2014 to 2024.* Ministry of Defence.

Levene, P. (2011). *Defence reform: An independent report into the structure and management of the Ministry of Defence*. Norwich: TSO.

Pryce, M. (2011). *Costing complex products, operations & support* (MBS-CE-11-196). Monterey, CA: Acquisition Research Program, Naval Postgraduate School.

Pugh, P. G. (1992). Working top-down: Cost estimating before development begins. In *Proceedings of the Institution of Mechanical Engineers* (Vol. 206, pp. 143–151).

Acknowledgments

The work presented in this paper is based on research funded by the United Kingdom's Economic and Physical Sciences Research Council (EPSRC), Grant Ref. EP/K503927/1, with an in-kind contribution from the United Kingdom's Defence Science and Technology Laboratory (DSTL).

Tim Jefferis at the DSTL has been important to the development of the work reported here as the DSTL lead user for the work, providing an initial focus for its outputs as well as carrying out some of the data gathering. Alexandra Bush of Isis Innovation Ltd. has played a key role in the refinement of the IP issues underlying the project. Staff at BAE Systems (Mick Porter), BAE Systems Heritage (Peter Hotham and Tony Wilson), Brooklands Museum (Chris Farara), and the National Archive have also been supportive in providing material and support for the work.

All views expressed, and errors made, remain the author's.

Controlling Costs: The 6-3-5 Method—Case Studies at NAVSEA and NATO

Bruce Nagy—is the CEO/founder of Catalyst Technologies. The company specializes in metadata-based cost control solutions. He is a graduate of The Citadel and the Naval Postgraduate School, having earned four degrees in the fields of electrical engineering, mathematics, and biological science. A commissioned U.S. navy officer, he was awarded research monies as a principal investigator for the Office of Naval Research in Artificial Intelligence. He spent over 30 years in the fields of management metrics, processes, and algorithms, having designed and released 18 commercial products used by the Department of the Navy to overcome operations and acquisition challenges. [bruce@catalysttechnologies.com]

Morgan Ames—is an acquisition management subject matter expert. His Navy career included U.S. Naval Academy (1974), NAVPGSCOL (1981), over 20 years as a surface warfare office–qualified engineering duty officer with proven acquisition management, COTR, and T&E/OPEVAL specialties. Assignments included SPAWARSYSCOM; EOP/ONDCP; and Navy-USCG C4ISR liaison. A 20-year defense industry career included business development for major systems integrators on three Navy ship classes; NMCI and AADC Capability, and onsite contractor support services of OUSD(AT&L)/Homeland Defense; DCNO-N6; Navy CIO; and AFRICOM-J2. He participated on NDIA studies addressing FORCEnet and Air and Missile Defense, and contributed to the SECNAVINST 5000.2e update. [morgan@catalysttechnologies.com]

Abstract

This paper discusses a critical gap in the U.S. Navy acquisition process. This gap is caused by the absence of workforce alignment metrics and metadata algorithms in two areas: (1) As applied to determining estimated cost per work breakdown element as part of the bid analysis process; and (2) after the award is granted, as part of the task management functions to ensure expectations associated with the original estimate can be reliably fulfilled. In this paper, the integration of workforce alignment metrics that are processed by a statistically-based, metadata-driven algorithm is referred to as the 6-3-5 Method. The 6-3-5 Method was specifically engineered using results from numerous case studies from NAVSEA- and NATO-based programs focused on creating adequate cost control measures in support of the acquisition process. Based on these case studies, it is shown that the Department of the Navy's acquisition process has a significant gap in its ability to adequately provide cost control. The case studies demonstrate how the 6-3-5 Method fills this gap, ensuring the future financial health and competitive status of the U.S. Navy to adequately address emerging threats to U.S. national defense.

Introduction

The need for Navy leadership to evolve its current cost control solution has become even more pressing with current discussions about financial uncertainty, talent management and better use of workforce innovation. For example, the cover of the August 2015 issue of National Defense magazine reads, "Pressure Mounts to Fund Ohio Replacement." At the Navy League's Sea-Air-Space (SAS) 2015 Symposium, Admiral James A. Winnefeld, Jr., Vice Chairman of the Joint Chiefs of Staff, during his banquet address, stated that "budget cuts are causing significant uncertainty" (Winnefeld, 2015). *Defense News'* August 2015 interview with Brad Carson, acting Department of Defense (DoD) personnel and readiness chief, focused on the need to take advantage of the unique talents of military and civil service employees (Carson, 2015).

Secretary of the Navy (SECNAV) Ray Mabus' recent speech at the same Navy League event stated a need for "innovation" as an inherent part of Navy culture to combat these troubling economic times more effectively (Mabus, 2015). These insights from DoD and Department of the Navy (DoN) leadership are supported by research findings led by Google, Inc. These finding reveal the need for an organization to create a culture that uses talent more effectively through collaborative processes that promote workforce innovation (Duhigg, 2016).

This paper introduces the 6-3-5 Method that integrates metadata statistics and assessment heuristics to promote better use of "big" data and workforce collaboration as applied to defense acquisition and related management activities. The 6-3-5 Method consists of a metadata-based algorithm, statistical analytics, a heuristic methodology, various measurements, and a visual approach to display results. Three case studies, demonstrating the use of the 6-3-5 Method, add proof to the need to address the challenges described by DoD and DoN leadership, as well as validate Google research results regarding the benefits from collaboration processes that promote workforce innovation.

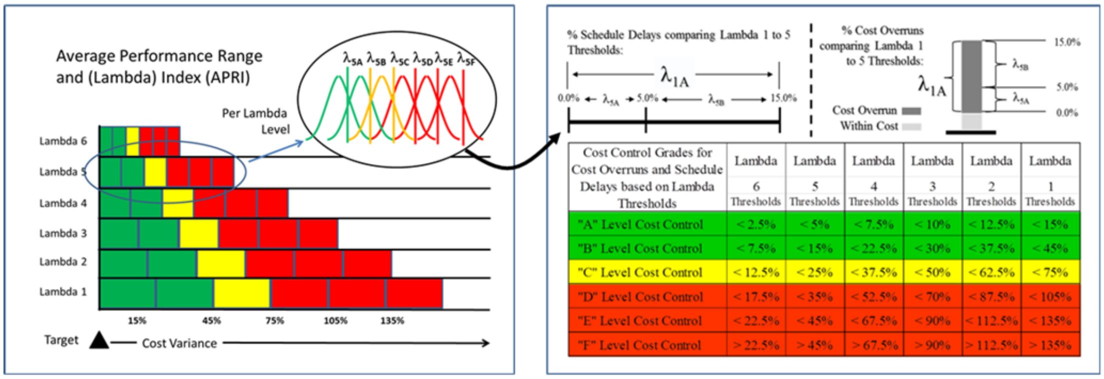
The case studies demonstrate how the 6-3-5 Method uses performance data to align more effectively workforce talent to goals, reduce cost variances and improve cost control. The methodology provides that ability to assess strengths and weaknesses in the architectural framework of an organization's cost control approach. Also, it provides specific recommendation solutions and clear direction to fill identified gaps or weaknesses in the framework to increase the likelihood of successful outcomes (i.e., actuals equaling estimates without compromise), whether in the form of Task Planning Sheet (TPS) deliverables assigned to government civil service employees or Ships Work List Item Number (SWLIN) deliverables for overhaul/new construction performed by prime contractors/shipbuilders.

The paper presents the 6-3-5 Method and related case studies in the following order:

- Overview of the 6-3-5 Method
- How the 6-3-5 Method supports the acquisition process based on a Program Executive Office (PEO) Carriers Case Study
- How the 6-3-5 Method supports workforce management based on a NATO Program Case Study
- How the 6-3-5 Method supports the workforce management based on a Naval Warfare Center Case Study

The first case study involving PEO Carriers provides examples of two significant cost control issues that are addressed using the 6-3-5 Method. This case study exemplifies how the 6-3-5 Method can be applied to any DoD request for proposal (RFP) process involving the review of bids for cost estimation accuracy. In this case study, the 6-3-5 Method emphasizes a better use of the cost control data that is required by programs that are required contractually to provide earned value analysis.

The mathematics is based on Van Trees's work on "Detection, Estimation and Modulation Theory" (Van Trees, 2001). The application is based on viewing data to identify an average performance range, where reliable performance reduces cost variances. The algorithm implementing the 6-3-5 Method processes the data, determines the highest/maximum likelihood that the average performance is true and not a false positive. The algorithm's mathematical basis has been peer reviewed and supported by California State University faculty. This analysis provides an accurate cost analysis of a bid or financial estimate. This paper introduces how detection and estimation mathematics can be applied to metadata (data about data) and algorithm processing based on an Average Performance Range and Index (APRI) table, as described in Figure 1.



Cost Control Grades for Cost Overruns and Schedule Delays based on Lambda Thresholds	Lambda 6 Thresholds	Lambda 5 Thresholds	Lambda 4 Thresholds	Lambda 3 Thresholds	Lambda 2 Thresholds	Lambda 1 Thresholds
"A" Level Cost Control	< 2.5%	< 5%	< 7.5%	< 10%	< 12.5%	< 15%
"B" Level Cost Control	< 7.5%	< 15%	< 22.5%	< 30%	< 37.5%	< 45%
"C" Level Cost Control	< 12.5%	< 25%	< 37.5%	< 50%	< 62.5%	< 75%
"D" Level Cost Control	< 17.5%	< 35%	< 52.5%	< 70%	< 87.5%	< 105%
"E" Level Cost Control	< 22.5%	< 45%	< 67.5%	< 90%	< 112.5%	< 135%
"F" Level Cost Control	> 22.5%	> 45%	> 67.5%	> 90%	> 112.5%	> 135%

Figure 1. Average Performance Range and (Lambda) Index Structure Used by 6-3-5 Method

There are two challenges for an acquisition management team that are discussed in this paper. The first challenge involves bid assessment.

Bid assessment is challenged not only when assessing new technology costs, but also when applying it to upgrades to existing technology already deployed within the fleet. In this paper, past performance analysis of a bid is turned into metadata and algorithmically analyzed using an APRI table. The analysis results in determining statistical confidence intervals, more conventionally discussed as statistical confidence, as a more effective means to determine the accuracy of a bid or estimation cost. Specifically, a PEO Carriers case study is reviewed involving an aircraft carrier overhaul, where SWLINs within the bid are statistically analyzed to determine the likelihood of issues not meeting contractual cost control requirements in compliance with the Nunn-McCurdy Amendment (Nunn-McCurdy Act, 1983). In this case study, the 6-3-5 Method provides analysis to the acquisition management team before bid acceptance based on past performance history using the same “actuals” as compared to estimated data that supported earned value calculations. (Albeit outside the scope of first case study: the 6-3-5 Method also can provide analysis using heuristics that can later be verified with performance data.)

The first case study demonstrates how applying a feature of the 6-3-5 Method provides an accurate gap analysis for each SWLIN cost estimate described within the bid, where each line item is described by a statistical confidence interval, having both an upper bound and lower bound. This interval can be mapped to the contractual cost control requirements needed to be satisfied by the offering prime contractor. The 6-3-5 Method also provides a structured approach to ensure the offeror is responding adequately to issues to ensure the proper due diligence necessary to fill those identified gaps before bid acceptance.

The second challenge for an acquisition management team involves the need to ensure that there is adequate due diligence by the offering prime contractor to resolve issues identified as statistically not meeting cost control contractual requirements. This due diligence is essential to complete before the contract award is granted to ensure minimal cost variance during implementation. The first case study emphasizes this key concept.

Providing a structured due diligence approach that can be statistically analyzed in terms of confidence is a necessary government procedure that can no longer be left, “at best,” to ad hoc processes. All too often, prime contractors increase their profits when Engineering Change Orders (ECOs) are generated. Each ECO causes a decrease in cost control for the government acquisition management team, which can result in significant cost overruns and schedule delays. The size of the cost the government must pay for an ECO is proportional to the size of the profit gained by the contractor and inversely proportional to decrease in cost control for the program. The only exception is with Fixed Price contracts, which have other issues this paper addresses in terms of quality compromises needed by the government versus contractor profit margin targets (FAR, 2016).

Applying the 6-3-5 Method allows the acquisition management team to determine objectively the rigor of due diligence that was applied to minimize potential impact to the government, thereby reducing the use of ECOs or minimizing their effects on cost variance and potential schedule delays. This first case study shows how this type of rigorous statistical analysis fills a critical gap in DoN's cost control solution, and should be considered as the first step to manage acquisition costs adequately.

The second case study demonstrates the required cost control objectives (CCOs) within a program and the procedural management steps needed to achieve adequate due diligence using the 6-3-5 Method. Specifically, this case study involves a NATO-sponsored development and acquisition program. It outlines key aspects of the 6-3-5 Method in terms of workforce management, alignment and innovation. This NATO case study views use of the 6-3-5 Method from the prime contractor's point of view, focusing on risk analysis of the project plan and validation of an effective mitigation strategy to reduce ECOs. It introduces the use of workforce alignment metrics and a metadata-based algorithm, compliant with the 6-3-5 Method, to promote and enable workforce innovation as an effective solution to addressing unknowns during task assignment and implementation. The workforce alignment metrics and metadata-based algorithm provide the acquisition team with a measurable, objective way to ensure adequate due diligence is being performed before an ECO needs to be generated or alternatively, a mitigation strategy is implemented.

Use of workforce alignment metrics and the related metadata-based algorithm becomes the focus of the third and final case study involving deliverables listed in Task Planning Sheets and assigned to a branch within a NAVSEA naval warfare center of excellence. The case study emphasizes the need for workforce innovation at the task execution level, when daily challenges by the workforce are encountered and their ability to succeed relates directly to the quality of service provided. This type of government environment in which a branch of civil service employees must do assigned work can be equated to a firm fixed-price contract between the branch and related PEO to provide the agreed upon service and meet the expectations described within the Task Planning Sheet.

Methodology Overview

The 6-3-5 Method consists of a metadata-based algorithm, statistical analytics, a heuristic methodology, various measurements, and a visual approach to display results.

Fundamentally, there are only two types of metrics, *a priori* and *a posteriori*, that apply to decision-making. Using the 6-3-5 Method, these metrics are used to create six metadata tags. A metadata tag provides intelligence in terms of what a data value means. The 6-3-5 Method requires a total of six metadata tags to support the acquisition effort's task management process, from estimation to product/service delivery. All six metadata tags ensure leadership is making informed decisions that have the highest likelihood of having successful outcomes and maintaining cost control. Each metadata tag is created from either an *a priori* or *a posteriori* metric type. The metadata tagging for *a priori* or *a posteriori* metric types are described in Figure 2.

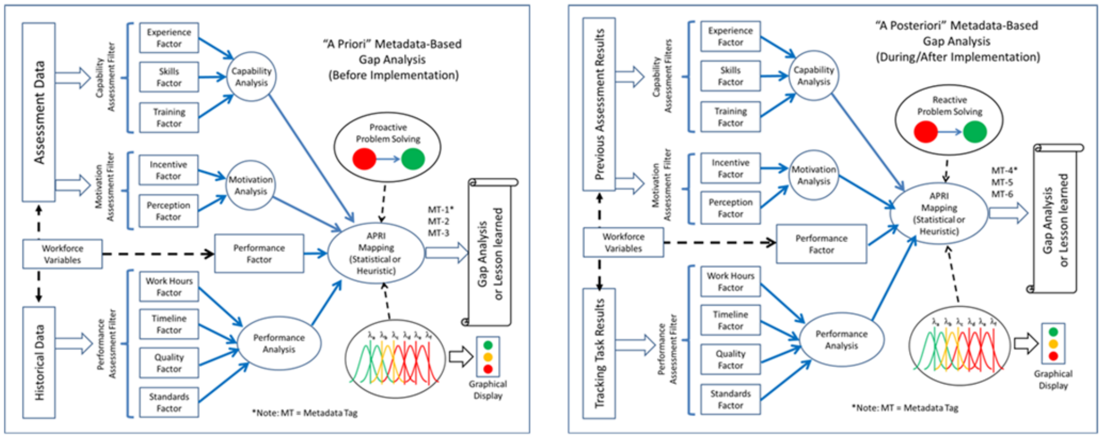


Figure 2. *A Priori* and *A Posteriori* Metadata Tagging Used by the 6-3-5 Method

The first three metadata tags (MT) support *a priori* (proactive) measurements. These three tags focus on providing intelligence to uninterpreted raw data values in order to (1) prevent forecasted issues, (2) eliminate impact of existing issues, or (3) when prevention and elimination cannot occur, minimize cost variance factors:

MT-1. Statistically-Based Lessons Learned Metadata: This is metadata that supports the analytics to do a statistically-based gap analysis. This analysis determines whether there are any lessons learned before proceeding forward in implementing an action. That action could be accepting a prime contractor's bid or allowing tasks to be implemented, during which time and money are consumed. The 6-3-5 Method uses past performance measurements or characteristically similar data (determined by an algorithm's process flow described in Figure 6) mapped to an APRI table via MT-4 that are converted into higher level tagging that support statistically-based lessons learned metadata. These metadata tags are then analyzed mathematically to perform gap analysis on the statistical likelihood of success in meeting cost control expectations. Low statistical confidence identifies potential for cost overruns. If the statistical confidence is in the red zone of the APRI table, significant cost overruns with statistical confidence are forecast. This metadata measurement forecast provides management with actionable data that can be used to proactively prevent/mitigate cost variances using the proactive due diligence (MT-3).

MT-2. Proactive Assessment Metadata: This is metadata that supports the analytics to do a heuristically-based gap analysis. The workforce self-assessment answers are mapped to an APRI table that is converted into metadata. The metadata is analyzed heuristically to provide a gap analysis of potential cost overrun issues, independent of MT-1 results, again based on the likelihood of successfully meeting cost control expectations. When performance data is mapped to the assessment metadata tags via MT-4, the measurements are, once again, tagged respectively with a statistical confidence of the gap identified and each's impact to cost. This metadata assessment measurement provides actionable data for management to use to proactively prevent/mitigate cost variances using the proactive due diligence, addressed by MT-3.

MT-3. Proactive Due Diligence Metadata: This is metadata that validates "best" solution selected for resolving an issue. When *a priori* metadata tags, either from MT-1 (statistics-based) or MT-2 (heuristics-based), indicate a gap in achieving reliable results, this issue initializes the due diligence process when following the 6-3-5 Method. Due diligence metadata tags are used to ensure adequate rigor is achieved in resolving these cost control issues and preventing/minimizing cost variance. The metadata from either MT-1 or MT-2 is inserted into a problem-solving format in accordance with the 6-3-5 Method Problem Solving Collaboration Approach later described (Case Study 2). Using these constructs results in solutions that are mapped into an APRI table and converted into due diligence metadata.

The due diligence metadata determines the degree to which the solution fills the gaps originally identified. If the gaps are not adequately filled (i.e., complete prevention of the cost variance issue), then the due diligence metadata highlights those core areas of concern, where a mitigation strategy is identified to minimize impact. If gaps are filled as validated by the metadata tags, then management also receives validation from the measurement via a recommendation as to the best corrective action to preclude/mitigate cost variance. When past performance data becomes available, this data gets mapped to the due diligence metadata using the APRI table via MT-4. The result provides a higher level tag that determines statistical confidence of the corrective action, specifically forecasting the likelihood of success when implementing the determined solution.

The final three MTs, based on *a posteriori* (reactive) metric types, focus on providing intelligence to untagged (or "raw") data values in order to recover or minimize factors that have been measured as having impact (i.e., increasing cost variance during or after implementation).

MT-4. Performance Tracking Metadata: This is metadata created from performance tracking measurements. Specifically, these are estimates compared to actuals regarding Full Time Equivalent (FTE) hours and schedule start and finish dates. Performance data collected is mapped into an APRI table, where the results are converted into metadata tags that support both heuristic- and statistic-based analysis used by MT-1, MT-2, MT-3, MT-5 and MT-6. In support of the DoN's cost control decision-making, use of intelligently tagged performance tracking measurements that can lead to higher level tagging constructs is recommended. Even earned value management techniques have significant limitations in helping decision makers know core issues and "best" corrective actions to provide highest probability of success. All the other MTs are statistically dependent on this MT-4's APRI tagging, which can be translated to core issues via tags, to provide insights into "best" corrective actions and reveal the statistical confidence, again as metatags, of a potential solution for success. Metadata tables displaying APRI tagging and statistical confidence are shown in the NAVSEA case studies (Case Studies 1 and 3).

MT-5. Reactive Due Diligence Metadata: This is metadata that validates "best" solution selected for resolving an issue. When *a posteriori* metadata tags, either from MT-4 (performance data) or MT-6 (lessons learned data), indicate an issue, the due diligence process is initiated when following the 6-3-5 Method. It is similar to MT-3, with the exception that this is a reactive or after the fact. Due diligence metadata tags are used to ensure adequate rigor is achieved in resolving these cost control issues and recovering/minimizing cost variance. The metadata from MT-4 or MT-6 is inserted into a problem-solving format in accordance with the 6-3-5 Method's Problem Solving Collaboration Approach. Using these constructs results in solutions that are mapped into an APRI table and converted into due diligence metadata.

The due diligence metadata determines the degree to which the solution fills the gaps originally identified. If the gaps are not adequately filled, then complete recovery of the cost variance issue, the due diligence metadata highlights core areas of concern, where a mitigation strategy is identified to minimize impact. If gaps are filled as validated by the metadata tags, then management also receives validation from the measurement as to best corrective action to recover cost variance impact. Using MT-4, a statistical confidence can forecast the likelihood of success for the identified corrective action.

MT-6. On-the-Job Lessons Learned Metadata: This is metadata that supports the analytics to do a heuristically-based gap analysis. This tag can be translated to organizational learning, both heuristically and statistically. The workforce lessons learned assessments are mapped to an APRI table that is converted into metadata. The metadata is analyzed heuristically or statistically, based on MT-4, to provide objective lessons learned. Because of the metadata tagging constructs, the lessons learned can be translated for use throughout the organization and is not limited to the project or team generating the learning. The on-the-job lessons learned metadata focuses on "what worked" and "what didn't work" with regard to the (1) customer, (2) organization, (3) teams, and (4) individuals performing the work. This metadata is also valuable in identifying a need for solutions to proactively prevent future cost control issues throughout various projects and activities within the organization. The main difference between MT-6 and MT-3 is that MT-6 involves the archiving of lessons learned for others to use at some future date, where MT-3 is lessons learned to address an immediate tactical need.

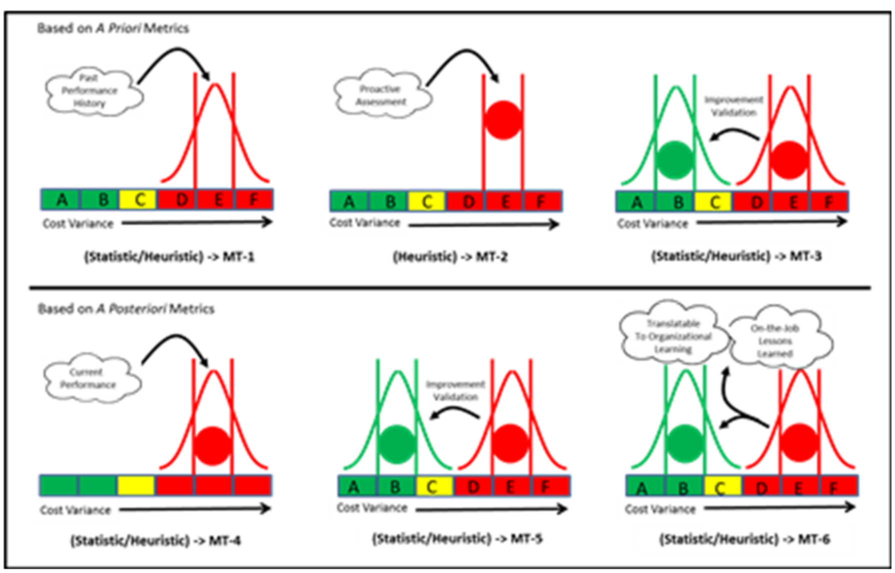


Figure 3. Metadata Tags From APRI Mapping Examples

Figure 3 graphically describes how MTs use APRI mapping. The six MTs require context for use. The three Cost Control Objectives (CCOs) provide a context in which the previous six metadata tags are used. Providing this context ensures management knows how to deploy and maintain the previously described six metadata tags to optimize spending control and reduce cost variance.

> CCO-1—Set and manage customer/business expectations: The first CCO ensures expectations are sufficiently defined based on seven categories. The first five categories deal with productivity needs and the last two focus on efficiency needs for acquisition programs and operational workflow management. The categories are: (1) Requirements, (2) Quality, (3) Process, (4) Technology, (5) Culture, (6) Cost (Including Workforce Allocation Hours), and (7) Schedule/Timeline
>
> CCO-2—Reliably achieve those expectations: The second CCO ensures that those defined expectations are reliably achieved, without compromise. Steps within this CCO include gap analysis and due diligence before, during and after implementation to ensure cost variances are minimized during the life of the project.
>
> CCO-3—Learn to continually do better: The third and final CCO focuses on Continuous Process Improvement to ensure the workforce is continually learning to be better at providing reliable, quality services/products, including how to better collaborate and innovate when overcoming challenges. There are four categories that represent accumulated lessons learned. Those categories are: (1) customer, (2) business/organization, (3) team, and (4) individuals associated with the quality and reliability of the work performed.

With the three CCOs, a context for using the six metadata tags is described. Yet, MT-3 and MT-5 require the use of the 6-3-5 Method's problem-solving constructs. The 6-3-5 Method's problem-solving approach is in the form of five Due Diligence Steps (DDS) to metrically ensure rigor in handling the issues identified as cost control gaps. The sequence of how these steps are described is in Figure 4. These steps are structured to ensure that a team is integrating the appropriate MT into one of the three CCOs previously discussed. Following these steps not only ensures the proper used of *a priori* and *a posteriori* metrics, but also that the three CCOs are continually achieved and due diligence is being rigorously applied when necessary.

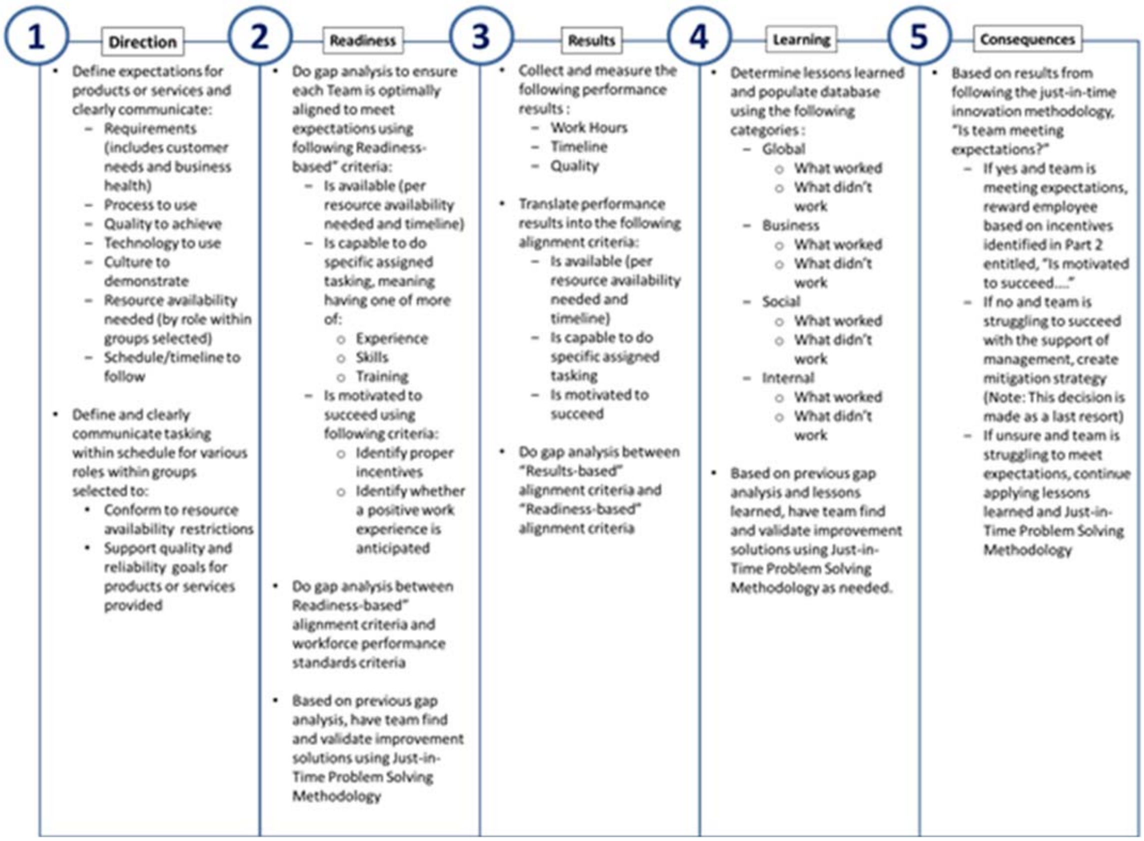


Figure 4. The 6-3-5 Method's Due Diligence Steps

These five steps also establish and maintain an environment that catalyzes the workforce to collaborate and discover innovative solutions to uncovered challenges that would have impacted/compromised the reliability and quality of the products/services being delivered.

> DDS-1: Provide Direction—Step 1 focuses on having management use a checklist to make sure that adequate direction is provided for any follow-on assessments and analysis. It also includes using lessons learned, either statistically or assessed heuristically, to identify weaknesses in the direction. Step 1 is crucial to include if CCO-1 is to be achieved. Metadata tags will use MT-1 for statistic-based gap analysis and MT-3 to validate problem-solving solution.
>
> DDS-2: Readiness of Workforce to Succeed—Step 2 is the due diligence process to ensure, within reason, that the workforce is (1) set up to succeed, (2) handle the unexpected, and (3) able to support each other when faced with severe challenges. Once Step 1 involving direction is complete, where strengths and weaknesses of the direction provided are known based on lessons learned, the implementers need to proceed with due diligence is resolving issues. A detailed due diligence process is described in Case Study 2 focused on workforce innovation and alignment. Even if no lessons learned issues arise from Step 1, to ensure "best chance to succeed," the 6-3-5 Method supports workforce self-assessment of assigned tasks in terms of their experience, skills, and other factors associated with reliable, quality results. (Specific factors and algorithm structure for this self-assessment

process is outside the bounds of this paper.) Step 2 is crucial to include if CCO-2 is to be adequately achieved. Use of metric visibility regarding due diligence rigor will be provided within the case studies. Metadata tags will use MT-2 for heuristic-based gap analysis and MT-3 to validate problem-solving solution.

DDS-3: Measure Results—Once tasks are completed, typical performance results are measured, including Full Time Equivalent (FTE) hours and schedule (calendar duration of tasking) comparing what was estimated to what actually occurred. Step 3 is crucial to include if CCO-2 is to be achieved.

DDS-4: Learn from Results—This step is another form of due diligence focused on learning from results. The 6-3-5 Method requires that due diligence include deliberate learning. The goal to Step 4 is to shift the focus of the learning from cost overruns and schedule delays to internal factors (i.e., ways to better provide direction, more accurate workforce alignment assessments that solve cost variance issues, including performance). Step 4 is crucial to include if CCO-3 is to be achieved. Metadata tags will use MT-4 for data understanding, MT-5 to validate problem-solving solution, and MT-6 to capture learning.

DDS-5: Apply Consequences—In the real world or a classroom, consequences are part of the educational process. Step 5's focus is to use what is learned in Step 4 to determine rewards for success and next step learning for those failed expectations. Step 5 is crucial to include if CCO-3 is to be achieved. Metadata tags may use MT-5 to validate problem-solving solution and MT-6 to capture learning.

Given the statistical processing and heuristics involved within the 6-3-5 Method, it was necessary to develop a metadata-driven "Workforce Alignment to Business Expectations" (WA2BE) Algorithm incorporating all 6 MTs, 3 CCOs, and 5 DDSs. The algorithm has been applied to waterfall project management and Agile software development styles. Independent of the version, the algorithm consists of two parts. The first part (Figure 5) uses *a priori* metrics to identify gaps in a proposed cost. The first part is before task execution but can be run up until the time tasks are assigned for the resource talent. The second part (Figure 6) of the algorithm uses *a posteriori* metrics that are applied after the resource talent is performing the assigned task. The algorithm uses performance data during and after implementation.

Uniquely, the WA2BE Algorithm knows how to translate "similar" historical data and make it relevant to a current project or operations. "Similar" is based on various types of complexity parameters of skill set and workload. The following three case studies demonstrate how various aspects of the 6-3-5 Method incorporating the metadata-based algorithm is applied to the acquisition and management process, and the results achieved from the application.

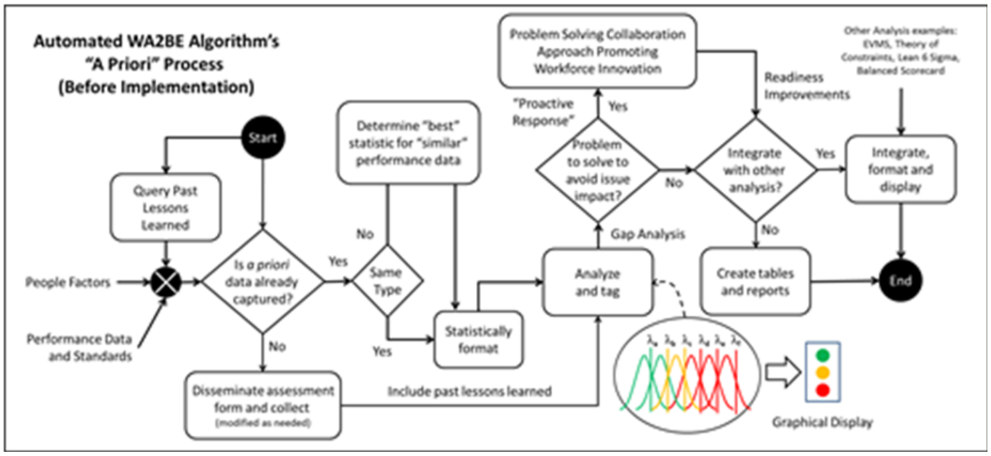


Figure 5. WA2BE Algorithm Process Flow Before Implementation

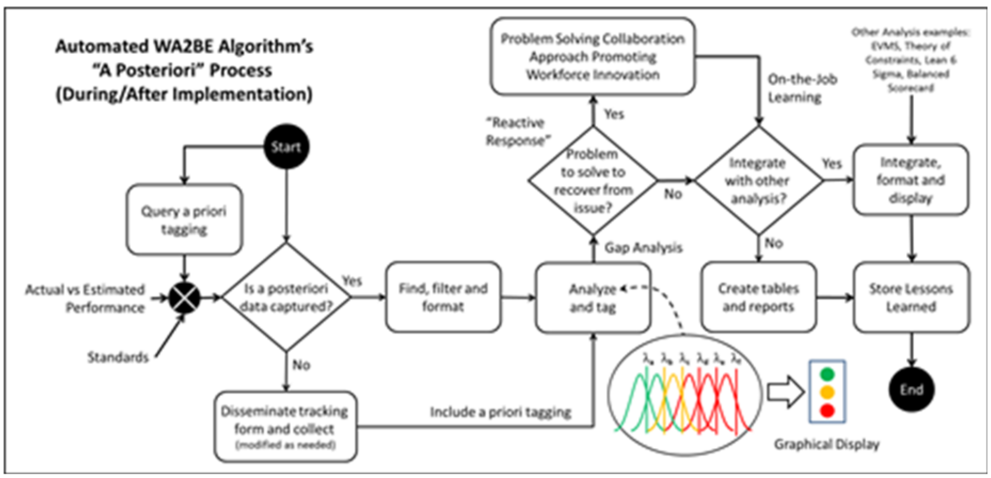


Figure 6. WA2BE Algorithm Process Flow During/After Implementation

PEO Carriers Case Study

In this NAVSEA-aligned PEO case study, statistically-based lessons learned metadata is used to support CCO-3. In this case study, previous RCOH data allowed the WA2BE algorithm to compare identical SWLIN types resulting in gap analysis of contract performance reliability. Note that in cases where the same type of data is absent for past performance analysis, then the WA2BE algorithm can translate and use "similar" past performance data to identify issues prior to contract award with statistical confidence.

The statistical analysis, implemented by the WA2BE algorithm, analyzed various permutations within the historical performance data to determine "best" estimation characteristics. A Lambda Level 6 (Figure 1) was selected for the algorithm's analysis. An important note: Once the statistically-based lessons learned are applied to identify gaps, the five steps involving due diligence discussed previously can then be applied to eliminate or mitigate cost control issues.

In the April 2015 online edition of the Navy League's *Seapower* magazine, The Honorable Frank Kendall, Under Secretary of Defense for Acquisition, Technology, and Logistics, stated, "We will not start programs we cannot afford" (Kendall, 2015). The U.S. Government Accountability Office (GAO) recommends cost control methods described in its

Defense Major Automated Information Systems: Cost and Schedule Commitments Need to Be Established Earlier (2015) report. That study emphasizes the need to have programs establish their initial acquisition program baselines (APBs) involving cost and schedule within two years of officially reporting that work has commenced. Industry's response to comply with this study is to create smaller, more manageable programs or program segments. Unfortunately, as this white paper describes through its case studies, even establishing an APB within two years to comply with these findings is still not an adequate cost control approach. This becomes obvious when coherent metadata tagging is created as a basis for comparison between past overhauls and a planned or in-process overhaul, as is described below.

To understand why a two-year APB approach will not fully satisfy Under Secretary Kendall's stated need, this case study examined a Nimitz-class aircraft carrier class refueling and complex overhaul (RCOH) of USS *Carl Vinson* (CVN-70). The RCOH occurred in 2005 as reported in the July 14, 2009, issue of *Defense Industry Daily* (DID) (CVN 70, 2009). That article stated, "In November 2005, [at that time] Northrop Grumman Newport News [Shipbuilding (NNS)—now Huntington Ingalls Industries] in Newport News, VA was awarded a $1.94 billion cost-plus-incentive-fee [CPIF] contract for accomplishment of the FY 2006 mid-life refueling and complex overhaul (RCOH) of the Nimitz Class aircraft carrier USS *Carl Vinson* [CVN 70]." As that article further states, "NAVSEA's official cost figure for the CVN 70's entire RCOH is $3.1 billion. As of April 2007, they [NAVSEA/PEO Carriers] told DID that the program was on budget." Thus, within less than a two-year period (November 2005 to April 2007), the cost had increased by over $1 billion.

Timeline for the USS *Carl Vinson* RCOH:

- 4th Quarter CY2005—"Workforce Alignment to Business Expectations" (WA2BE) Algorithm was applied and used a grading system of an "A" through "F" for each of the CVN-70 Carl Vinson's RCOH Ships Work List Item Number (SWLIN). Grades of C, D, E, and F indicated that cost overruns would potentially exceed contract RCOH goals. Using a stoplight dashboard, C is represented as "yellow," where D, E, and F are displayed with "red." The grades are color-coded in Table 1. The algorithm had an overall statistical confidence of 99% that the RCOH's costs would exceed its contract goals (coded "red"):
    - Sample Source: The statistical analysis was based on 684 SWLIN forecasted grades using CVN-68 and CVN-69 RCOH historical data that was provided by PEO Carriers. There were seven SWLINs that had no values available for analysis.
    - Metadata Summary (Contributors to cost overruns—Cs, Ds, Es, and Fs are tags to indicate the degree in which actuals will potentially exceed contractual agreement—the lower the grade, the higher the potential): Table 1 is a simple example of metadata analysis using previous overhauls per SWLIN categorization as linked to shops assigned, trades assigned, and related management/operations. These types of analytical summaries based on *a priori* metrics allow for better bid negotiations.

Table 1. CVN-70 RCOH Metadata Summary Analysis of Bid From Newport News Shipbuilding

Grades / Performers	A	B	C	D	E	F
Management and/or Operations	0%	0%	31%	8%	40%	21%
Shops	0%	0%	41%	59%	0%	0%
Trades	0%	0%	57%	41%	2%	0%

- 4nd Quarter CY2005—Newport News Shipbuilding (NNS) Awarded $1.94 Billion CPIF Contract
- 2nd Quarter CY2007—NAVSEA/PEO Carriers' Official [Actual] Cost is $3.18 Billion [Reported by DID], obviously exceeding the original contract goals.

Figure 7 shows three figures, where each figure represents a row in Table 1. The three figures describe probability density functions determined using the APRI analytics and maximum likelihood criteria to statistically reduce false positive results. The RCOH bid contractually needed to be within +/- 10%. 684 SWLINs had a statistical confidence of 80% or greater that they had either a marginal or poor estimate in terms of meeting the contractual 10% threshold. This analysis involved shops, trades, and management. This is why the WA2BE Algorithm determined with a statistical confidence interval of 99% that the costs would exceed the contractual threshold. Use of the WA2BE Algorithm demonstrated that this RCOH bid should not be accepted without all 684 SWLINs reviewed for improvement. Starting with the "F" graded SWLINs, the review needs to use the algorithm's analytics to determine if any changes were effective (creating an "A" or "B" grade) in having the U.S. Navy avoid getting "stuck with the bill"—again!

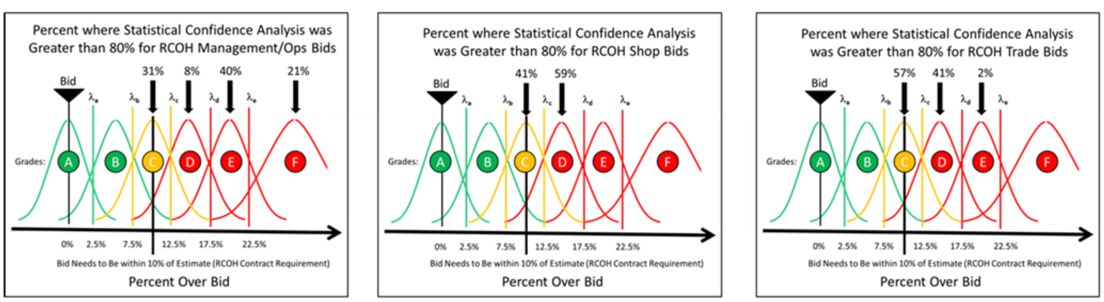


Figure 7. Probability Density Function View of Metadata in Table 1

Figure 8 provides an example of how the results from using five due diligence steps within the methodology are statistically validated. The key is to be able to graphically view the results of the due diligence efforts with stoplight displays before proceeding with bid acceptance or task management. PEO Carriers did not follow the five due diligence steps. Therefore, it is difficult to determine whether they accepted the bid under Figure 7 conditions or those of Figure 8.

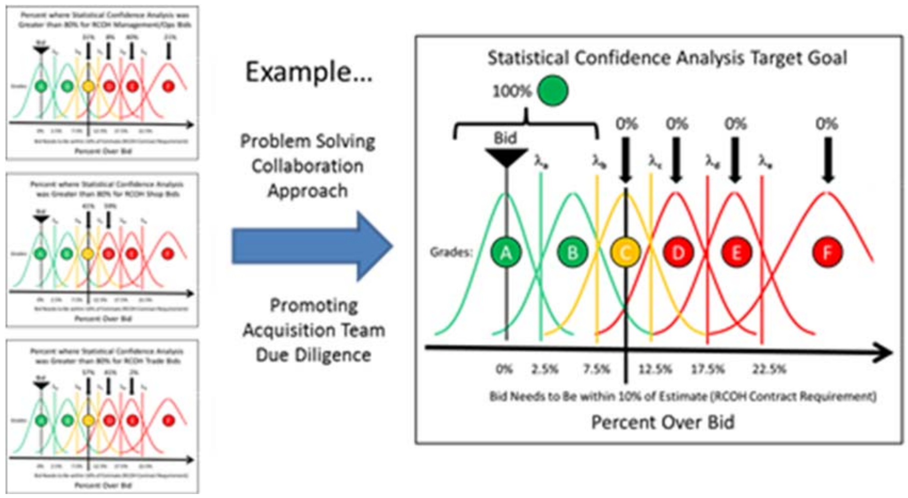


Figure 8. Target Goal of a Bid Statistical Analysis

To emphasize when using the 6-3-5 Method, Figure 8 provides an example of what the statistical analysis of any RCOH bid must be before the contract is accepted. The following case study describes how the 6-3-5 Method supports the achievement of this goal when given issues represented in Figure 7.

NATO Program Case Study

The 6-3-5 Method's WA2BE Algorithm applied heuristic processing to a NATO-sponsored program developing Air Traffic Control Systems (ATCSs) for various nations, including Belgium. The prime contractor developing the systems was a Hughes Aircraft spinoff company. The ATCSs being developed required the use of touchscreens (a new technology at the time) and had the complexity of integrating new hardware development with complex software coding. As described below, the prime contractor encountered many difficulties creating an accurate baseline for the Belgian-variant ATCS acquisition—until the WA2BE algorithm was applied. These problems were addressed using a formal workforce collaboration problem-solving approach supported within the 6-3-5 Method as described in Figure 9.

The 6-3-5 Method's collaboration problem-solving approach was successfully used by the prime's management team to reduce cost overruns dramatically from 35% to 3%. When following the 6-3-5 Method, first the implementation of the metrics allowed for visibility into issues. The key problem discovered was a requirement document that did not provide enough detail for the workforce to assess their alignment. That means that there was no alignment, which caused more than 80% of the task to be in the red, as described in Figure 10.

After the requirements document was written to allow the workforce being assigned tasks to self-assess, the following corrective actions were taken based on the results from the workforce applying the 6-3-5 Method's collaboration problem-solving approach (Figure 9) to improve cost control (i.e., workforce alignment):

- Increase the team's confidence, knowledge, or training about performing the Tasks per management's productivity and efficiency expectations (a) informal training, (b) formal training, (c) completing similar Tasks.
- Modify the resources (e.g., internal resources, external resources, time, labor-hours) or working environment defined for use with the assigned Tasks.

- Modify the way the team is accomplishing their Tasks by using a different approach.
- Modify the support group/other team members to assist each other in doing assigned Tasks more reliably.
- Increase incentives for the team to perform the complexities of the Tasks.

It is important to note that had the Nunn-McCurdy Act applied to this NATO program, it would have been well within the 15% requirement, and even when the Amendment was at 10% (Nunn-McCurdy Act, 1983). The DoN would not have been stuck with another unexpected bill, as described in Case Study 1. Instead, this program used all six MTs as they continually cycled through the five DDSs, achieved all three CCOs, enabling NATO prime contractor management to realize greater cost control by meeting budget, schedule, and quality expectations more reliably. The NATO program was being due diligently supported (see Figure 10).

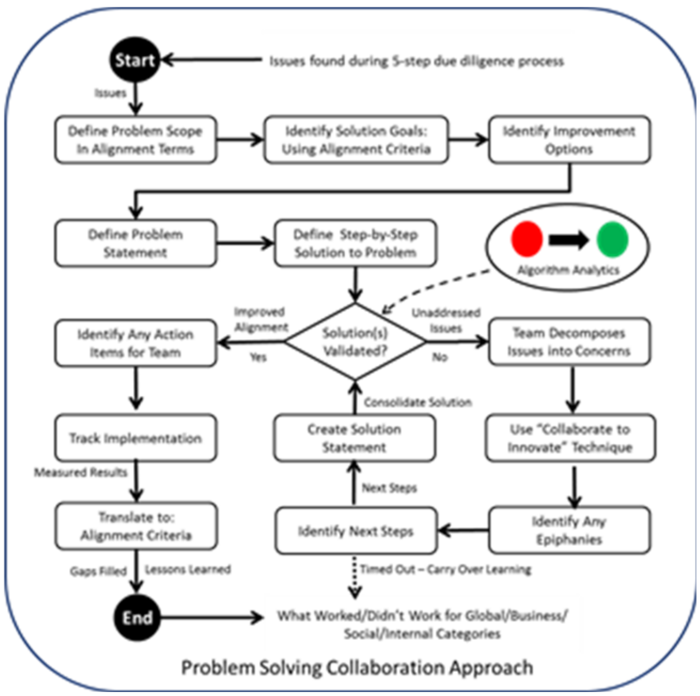


Figure 9. Problem Solving Collaboration Approach

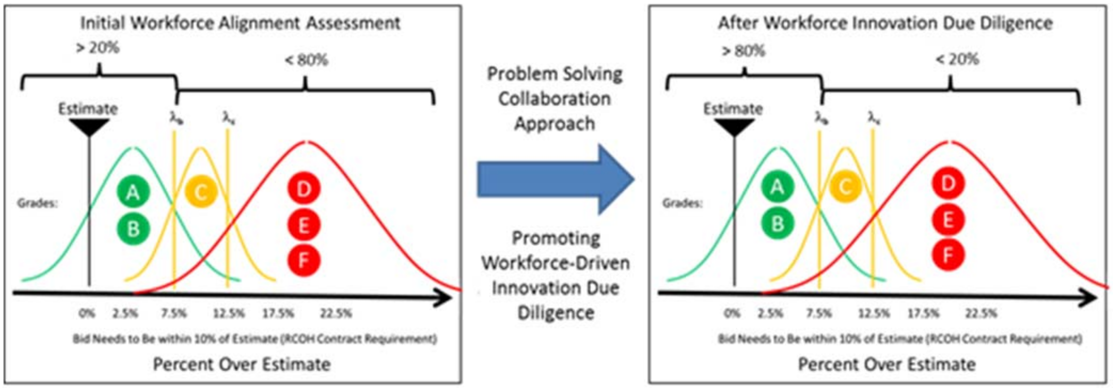


Figure 10. Heuristic Assessment Analysis Statistically Mapped to APRI Table Before and After Problem Solving Collaboration Approach Validating Improved Cost Control Resulting in Cost Overrun Reduction From 35% to 3%

The WA2BE Algorithm's implementation of the five-step method supports a rigorous due diligence approach involving team problem solving through collaboration that promotes team innovation. Figure 9 provides how collaboration can lead a team to innovative solutions and lessons learned. Using this process flow results in one of three results:

1. A solution that does not require innovation
2. A solution that does require innovation
3. A next step to discover a solution

If the problem is not solved in time, the algorithm supports the team with lessons learned documentation of the steps used and the related outcomes to discovering the solution. In the case where the issue happens again, the workforce can access documentation to use as a running start at finding the solution within the needed time frame. If a solution is found, the lessons learned provides immediate assistance to support the workforce in knowing how to recover from any impact that would increase cost variance. Lesson learned documentation is in the form of what worked and what did not work.

The hypothesis is offered: Can any team be due diligent about reducing impact of unknowns to a project without considering innovative options/solutions to the issues? Would this be considered a significant workforce collaboration goal? The algorithm's implementation of the 6-3-5 Method is based on the assumption that workforce collaboration, and when needed, workforce innovation, is essential to success when attempting to achieve rigorous due diligence.

Highlighting a difference in approach as compared to the CVN-70 RCOH case study, the prime's executive leadership for the Belgian ATCS chose to make the time investment to perform the necessary due diligence—despite already being behind schedule—to address the "red" stoplight issues. First, the algorithm was used to achieve CCO-1 (Setting and managing customer/business expectations for the program). Next, the algorithm was used to support CCO-2 (Reliably achieving those expectations). The algorithm allowed the team to assess their reliability in meeting expectations, with objective, metric precision. Once assessed, the algorithm identified gaps in workforce alignment, prioritizing these gaps in terms of criticality (i.e., greatest budget, schedule and quality impact to the program), allowing the team to focus on worst case scenarios. This was all done through its tagging process. The algorithm facilitated constructive discussions by having tags that provided simpler descriptions of core issues that allowed the performing team to find innovative solutions, when needed. This gave the performing team an advantage in meeting expectations more reliably, producing wiser, more efficient and when needed, innovative solutions.

The WA2BE algorithm's connectivity relationships and related metadata tagging made it easier for the team to innovatively solve issues—turning an average group into a high performance team.

For example, the NATO-sponsored ATCS development program required each system being tailored for various countries be designed with touchscreen user interfaces. Unfortunately for the Belgian ATCS program, the prime's software development team didn't have access to touchscreen user interfaces at their workstations, nor was funding available to upgrade their workstations. For the Belgian team, the touchscreens were only available on the Test and Evaluation (T&E) Test Bed, which was time shared by multiple NATO-sponsored programs for formal integration testing. The algorithm identified "lack of funding to buy touchscreen user interfaces" as a symptom of the Belgian team's inability to meet cost and schedule. Additionally, like all symptoms, it did not present itself as a solvable problem for the team to address. In fact, it caused them to feel victimized given other circumstances associated with the prime's executive leadership. This is a very typical workforce response when they feel that they are set up to fail by management.

The complete symptom presented itself as follows: For the Belgian ATCS software team, workflow involved code development at their workstations, followed by software installation onto the T&E Test Bed. Once installed, they ran their code using the timeshared Test Bed's touchscreens to find "bugs." Once found, they needed to end their Test Bed session to return to their workstations to "hopefully" fix the code. Once potentially fixed, they had to reinstall their updated version onto the Test Bed, again timesharing with other ATCS programs, and check their code using the T&E Test Bed's touchscreens. Then the process repeated to resolve any new bugs found. Until all bugs were fixed, formal integration tests could not be performed. Because of the timesharing challenges, this caused the Belgian team to incur uncontrollable cost overruns and schedule strips.

The algorithm forecasted the number of days these symptoms would push the schedule, thus causing a failure in meeting requirements. The algorithm's metadata process converted these symptoms into workforce alignment gaps that identified core issues based on one of the five alignment factors in meeting business expectations, breaking down the problem into addressable and solvable "chunks."

These problem "chunks," customized to the team's specific workforce alignment needs, were then filtered into a solvable format for the team to more easily and efficiently apply their collective wisdom and innovation. This translation into smaller, solvable issues allowed the team to creatively improve workforce alignment. In other words, the algorithm's use of metadata guided the team to innovate a solution based on their alignment needs in meeting schedule, without affecting cost or quality.

Whether applying the WA2BE algorithm to an S-Curve analysis, Earned Value Management (EVM), etc., the new baseline/target was assessed with the quality of data having mostly "As" and "Bs," with a few "Cs" that were being mitigated, as opposed to its previous lower grades (Earned Value Management Systems, 2007). The "garbage in, garbage out" syndrome was eliminated. The new baseline was significantly different in another way; it conformed to the provided funding, once thought insufficient. The data supporting any type of analysis, S-Curve, EVM, etc., now consists mostly of all "green" lights—indicating, "quality in, quality out." As indicated in this case study, the Belgian team faced and innovatively conquered a common issue in today's development environment, "lack of funding." During the life of the program, they continued to solve issues using the Just-in-Time Problem Solving Collaboration Approach instead of feeling victimized.

The transformation of the Belgian team in its ability to meet expectations was so dramatic that the prime contractor began applying the WA2BE Algorithm to its other NATO-sponsored programs.

Naval Warfare Center Case Study

Performing under Navy fixed price support service contracts, the WA2BE algorithm was applied within various branches of two NAVSEA Warfare Centers of Excellence (COEs). For both centers, since historical data was not available to apply the statistical analyses, assessments were made via MT-2 and heuristically analyzed using the WA2BE Algorithm. MT-2 supported DoN civil service employees and related contractors to create the data that was translated into metadata and mapped to an APRI reference, as was done in Case Study 2. Case Study 3 will compare the use of WA2BE Algorithm by branches from two different COEs.

For both warfare centers, the first challenge was in the achievement of CCO-1 (Setting and managing expectations). At one center's branch, a more effective WBS was created, with a nomenclature that supported greater understanding of the relationships within the branch's organization. At the other center's branch, where deliverables were defined in Task Planning Sheets (TPSs), the focus was on documenting workflow and defining FTE hours consumed. For both branches, once metrically determined as having CCO-1 complete with the appropriate metadata analysis, the algorithm supported management to achieve CCO-2 (Reliably achieving expectations).

Without prior history, the algorithm generated forms based on the metadata for both branches to self-assess workforce alignment to deliver their assigned fleet work products with reliability and quality. In CCO-2, the algorithm used the completed forms to identify gaps in workforce alignment and then used the Just-in-Time Problem-Solving Methodology to guide the groups in addressing their respective issues without compromising time, money or quality. One center's branch manager went through detailed preparation to use the Just-in-Time Problem-Solving Methodology based on the alignment metrics. Another center's branch manager chose not to go through the training. The results were profoundly different in a manager's ability to grasp innovative due diligence performed by his team whenever dealing with challenges.

It is important to note that when a branch manager chose not to receive training, the solutions produced by his team to improve quality of deliverables were suppressed. Presumably, the branch manager was concerned that his senior leadership would view this data as an inability to manage his group effectively. This is another example of SECNAV's comments and concerns regarding a long standing "zero-defect" culture.

Significantly, when the team's solutions to deal with "red" stoplight issues were not supported by management, the identified issues manifested within six months to a year. This resulted in formal negative feedback from the branch's Navy customer. Again, this emphasizes two points: (1) even without historical data, the algorithm assessment using the metadata approach was shown to be able to forecast issues accurately six months in advance, and (2) when "red" stoplight issues are identified and solutions provided within the metadata structure, they need to be addressed and implemented before impact.

Timeline for the branch that did not support team's solutions to "Red" stoplight issues:

- CY2007—"Workforce Alignment to Business Expectations" (WA2BE) Algorithm used a grading system of an "A" through "F" for each of the deliverables called out by the TPSs. Grades C, D, E, and F indicated potential quality issues. Using a stoplight dashboard, C is represented as "yellow," where D, E, and F are displayed as "red." The grades are color-coded in Table 2. Using the metadata structure, the algorithm produced assessments that were answered by branch team members and analyzed based on APRI and metadata relationships. (The just-in-time innovation solutions to improve the quality of the work generated by the branch workforce assigned to the deliverables were discarded.)
    - Sample Source: The analysis is based on 181 deliverables and their related assessments. There were four deliverables that had no assessments available. Appropriation source were Operations and Maintenance, Navy (O&MN)—funding 60 deliverables; Other Procurement, Navy (OPN)—funding 40 deliverables; Ship Construction & Conversion, Navy (SCN)—funding 54 deliverables; and Navy RDT&E—funding 31 deliverables being analyzed by the algorithm.
    - Metadata Summary (Contributors to Quality Issues—Cs, Ds, Es, and Fs are tags to indicate the degree in which actual performance will potentially have quality issues—the lower the grade, the higher the potential): Table 2 is a simple example of metadata analysis, where each type of money connects to its related deliverables, related processes, people assigned within the process, and related alignment metrics per tasks/activities along the process. This data is tagged and summarized below. These types of analytical summaries allow for better management of monies within branches, divisions, departments, commands, and various other echelons within the DoN.

Table 2. Metadata Summary Analysis of a NAVSEA CEO Branch's Quality of Deliverables by Appropriation

Grades / Appropriations	A	B	C	D	E	F	TBD
O&MN	28%	30%	13%	7%	20%	0%	2%
OPN	33%	5%	0%	5%	58%	0%	0%
SCN	28%	13%	17%	24%	13%	0%	6%
RDT&E	0%	65%	0%	0%	35%	0%	0%
All Appropriations:	24%	25%	9%	10%	29%	0%	3%
Work Quality:	49% (Good Quality)		9% (Marginal)	39% (Poor Quality)			3% (TBD)

- CY2008—The NAVSEA/"PEO Customer" had significant complaints involving quality of work performed by that same specific branch that used the algorithm approximately six months earlier but discarded the team's solutions to improve the quality of work.
- CY2014—Executive leadership for the entire Center was replaced due to Navy customer complaints regarding the quality of work.

Significantly, when the team's solutions to deal with "red" stoplight issues were not supported by management, the identified issues manifested within six months to a year. This resulted in formal negative feedback from the branch's Navy customer. Again, this emphasizes two points: (1) even without historical data, the algorithm assessment using the metadata approach was shown to be able to forecast issues accurately six months in advance; and (2) when "red" stoplight issues are identified and solutions provided within the metadata structure, they need to be addressed and implemented before impact.

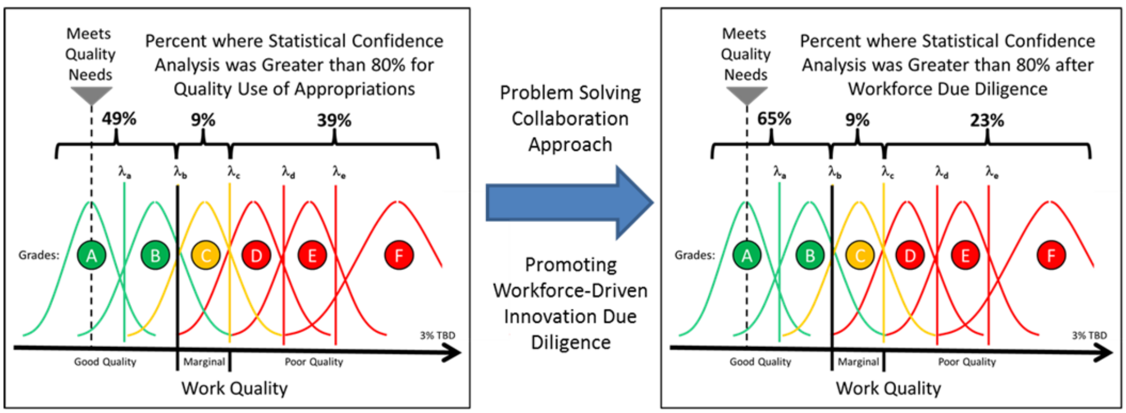


Figure 11. Heuristic Assessment Analysis Statistically Mapped to APRI Table Before and After Problem Solving Collaboration Approach Validating Proposed Solution to Improving Workforce Quality

As Deming (1993) once stated, "A bad system will beat a good person every time." For the branch manager who did not allow "red" stoplight issues to be solved and implemented by their teams, was this a display of a "zero-defect, never failing" mentality, and was it pervasive throughout his chain of command? In support of understanding the culture of the command, there is another instance when the center used the algorithm to support implementing NAVSEA-wide initiatives. In CY2014, when a high-level civil service employee remarked about his group's "red" stoplight issues, "Don't throw me under the bus for my answers." He preferred to avoid stating that "the emperor has no clothes" and any consequences that might ensue.

The algorithm using its metadata structure accurately determined that the workforce alignment was so poor that his implementation team had a very low likelihood of meeting the defined expectations of the command in supporting the NAVSEA-wide initiatives. Once again, in Case Study 3, the concern was what executive leadership might think of his answers. When executive leadership reviewed his assessment, their predominant concern was what impression a visiting NAVSEA admiral might think of their command's organization. This supports the belief that a pervasive "zero-defect, never failing" mentality was being evidenced throughout the chain of command.

Figure 11 demonstrates that while 49% supported by this branch are of good quality, 51% had a statistical confidence of 80% or greater that the work quality would be marginal or poor. When civil service work quality suffers, the impact can seriously hinder the U.S. Navy's readiness to support its warfighting missions.

Failure-focused, "zero-defect" cultures must end if cost control is to improve. Unfortunately, ignoring or at worst, discarding data that accurately forecasts outcomes does not avoid issues; it only delays having to deal with them. Such delays make any corrective action, after the fact, more costly.

The algorithm's metadata approach to identify and overcome issues, instead of being victimized by them, directly contrasts against a non-zero defect mentality. In Case Study 3, one center's branch manager and his supervisory task leads took the training and benefited from the algorithm's analytics. Another center's branch manager skipped the training and rejected the benefits of the just-in-time innovation solutions offered by his team. Can an assumption be made that if the branch manager had chosen to attend the training instead of skip it, would he have allowed his team to follow through with their just-in-time innovation solutions?

This case study emphasizes how the WA2BE Algorithm fills a gap as a needed solution for use by various U.S. Navy shore commands as part of their cost control approach, supporting just-in-time innovation. All commands and programs within the DoN need stabilized budgets, where innovation is used to maintain budget adherence reliably to meet expectations. This algorithm supports this type of cultural change.

Conclusions

The data collected at the SWLIN/TPS deliverable levels, once metadata is tagged and processed through the algorithm, becomes very revealing. The three case studies reveal crucial gaps in the DoN's current ability to achieve an effective cost control solution. Two of these case studies were performed under NAVSEA contracts and the third in support of a NATO program. The case studies demonstrate how the metadata-based algorithm implementing the 6-3-5 Method gave leadership greater proactive control over internal and external factors that affected cost variance. Specifically, the algorithm identifies and collects workforce alignment data, making relevant links between data points, providing heuristics (an approach to achieve Artificial Intelligence) to create tags that summarize the value of those links, and processes analytically the metadata tags to provide statistical confidence and other higher-level metadata related to cost control.

Through application of the algorithm's automated analysis, as shown via tables of cost control data collected from these studies, leadership is able to make more informed decisions and an aligned workforce is encouraged and enabled by management to contribute through innovation and by lending its fully invested and supportive voice to recommend process improvements. These solutions offer management improved decision-making capability to better tailor the program for success.

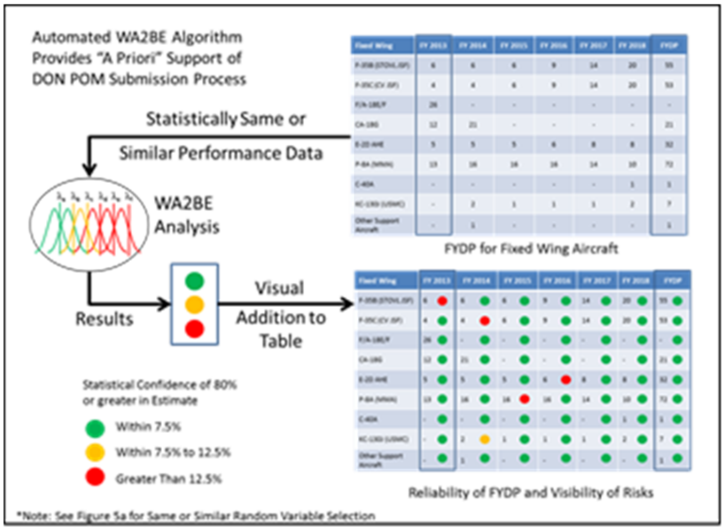
As per H. James Harrington, PhD, president and chairman of the American Society for Quality (ASQ) and president and chairman of the International Academy for Quality (IAQ), "Measurement is the first step that leads to control and eventually to improvement. If you can't measure something, you can't understand it. If you can't understand it, you can't control it. If you can't control it, you can't improve it" (Harrington, 1987, p. 43). Through the 6-3-5 Method's metadata analysis via the WA2BE Algorithm based on either statistic or heuristic assessment measurements for each program and aggregating in the roll-up from lower to higher Echelon command levels, and ultimately Budget Submitting Offices and

OPNAV Resource Sponsors. These 6-3-5 Method-based analyses can feed the related line items in the DoN's POM submission to DoD in support of the federal government's Planning, Programming, Budgeting and Execution (PPBE) process (Office of the Under Secretary of Defense for Acquisition, Technology, & Logistics [OUSD(AT&L)], 2013).

Table 3 provides an example of how a stoplight approach, based on the WA2BE Algorithm's analytics, can support the DoN's POM submission process. The table represents a graphical view of an original estimate submission before and after applying the algorithm's statistical confidence analytics. The WA2BE Algorithm is able to use either same or similar past performance data related to each DoN POM submission line item to determine the statistical accuracy of each item's budget estimate. The statistical accuracy of the analysis is represented by graphical stoplights, where each stoplight represents confidence of 80% or greater with a margin of error, defined by the Lambda value within the APRI table. In an environment where budgets need to accomplish more with less funding, accurate estimation using *a priori* metadata is a vital necessity/capability for all DoD activities. For areas in the "red," the 6-3-5 Method provides a proven problem-solving collaboration approach, as described in the previous case studies and figures. The collaboration approach uses *a priori* metrics and either heuristic assessment or statistical confidence to validate the reliability that the proposed solution will cause "red" DoN POM submission line items to turn "green," assuring accuracy of estimated submission.

Admiral James A. Winnefeld Jr., Vice Chairman of the Joint Chiefs of Staff, during his address to the Navy League SAS 2015 banquet, stated that budget cuts are causing significant uncertainty for our future warfighting capability. More money is always a desired state; yet even with more money, the reliable control of spending needs to be addressed. To summarize many sentiments discussed by speakers at SAS 2015, including Admiral Winnefeld, there is a level of uncertainty with regard to a stable budget that has now become a DoD legacy "posing serious future problems." In Case Study 1, the Nunn-McCurdy Act was discussed (Nunn-McCurdy Act, 1983). The current Act requires that Congress be informed of any major Acquisition Category (ACAT) program running over 15%. Just a decade earlier, the threshold number was 10%. These percentages represent uncertainty in any Program of Record's financial outcome. The goal should be to have less uncertainty, for example, a 5% threshold—not more uncertainty—with regard to the mandate of the Nunn-McCurdy Act.

Table 3. An Example of How 6-3-5 Method Supports DoN POM Submission in Knowing Reliability of Submission/Risks per Line Items



Because of its proven accuracy to use *a priori* metrics to identify and resolve issue well in advance, the WA2BE Algorithm is recommended as an effective means to minimize DoN budget uncertainties. Case Study 2, regarding the reduction of cost overruns from 35% down to 3%, ensures that when an APB is set, it is also stable with a workforce that is aligned with a high statistical confidence of meeting expectations. Case Study 2 also demonstrates the ability for the DoN to establish and maintain a culture of innovation and learning using the WA2BE algorithm throughout its acquisition and in-service engineering/life cycle support programs, as well as its centers of excellence.

The 6-3-5 Method deployed using the metadata-based WA2BE Algorithm ensures that the workforce remains optimally aligned to meet fleet and force capability expectations reliably. As described in Case Study 3, this ability applies to the federal civil service workforce throughout all Navy shore commands and their industry service support partners, as well as major system integrators delivering products. It ensures that just-in-time innovation becomes the new cultural norm. Just-in-time innovation, as established and visible through the algorithm's metadata analytics, will set a standard for adequate due diligence in handling the challenging budgetary issues related to delivering reliably on-time, within-schedule, and at the quality required. As demonstrated in the case studies, the 6-3-5 Method making data more intelligent starting with APRI tagging is a needed solution in today's "keep-up or fall-behind" information-based society in order to sustain our Navy operational capabilities, forward presence and warfighting advantage.

Based on the findings contained in this white paper, it is recommended that DoN leadership evaluate and deploy the WA2BE Algorithm and its metadata constructs, according to the 6-3-5 Method, to estimate and manage expenditures reliably, while improving the quality and health of programs in support of existing Joint Capabilities Integration and Development System (JCIDS); Planning, Programming, Budgeting and Execution (PPBE), including Program Objective Memorandum (POM) submission; and Defense Acquisition System (DAS) processes, including principal Major Defense Acquisition Programs (MDAPS), for example, the Ohio-class SSBN replacement and naval Joint Strike Fighter variants (Secretary of the Navy, 2011).

When combining the results from all three case studies, the SWLIN/TPS deliverable analysis evidence indicates that application of the 6-3-5 Method would improve the health of Navy Programs of Record by supporting the respective Acquisition Program Management Offices and those naval warfare systems' centers of excellence responsible for in-service engineering and life-cycle sustainment support. The case studies demonstrate critical gaps that are affecting the recapitalization and sustainment of the world's finest Navy. This paper demonstrates how the 6-3-5 Method can immediately, effectively, and efficiently satisfy these DoD and DoN needs, addressing these critical gaps through use a metadata-based, statistics and heuristic assessment algorithm resulting in improved cost control capability.

References

Carson, B. (2015, August 30). Acting DoD personnel and readiness chief, interview with Defense News editor Vago Muradian. *Defense News*.

CVN 70 *Carl Vinson's* mid-life RCOH refueling & maintenance. (2009, July 14). *Defense Industry Daily.*

Deming, W. (1993). Deming four day seminar in Phoenix, Arizona (via the notes of Mike Stoecklein). The W. Edwards Deming Institute Blog. Retrieved March 1, 2016 from http://blog.deming.org/2015/02/a-bad-system-will-beat-a-good-person-every-time/

Duhigg, C. (2016, February 28). What Google learned from its quest to build the perfect team. *The New York Times*. Retrieved from http://www.nytimes.com/2016/02/28/magazine/what-google-learned-from-its-quest-to-build-the-perfect-team.html

Earned Value Management Systems (EVMS). (2007, September). Government Electronics Information Technology Association. American National Standards Institute/Electronic Industries Alliance, ANSI/EIA-748-B.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2016).

GAO. (2009). *GAO Cost estimating and assessment guide: Best practices for developing and managing capital program costs* (GAO-09-3SP). Washington, DC: Author.

GAO. (2015). *GAO Defense Major Automated Information Systems: Cost and schedule commitments need to be established earlier* (GAO-15-282). Washington, DC: Author.

Harrington, H. (1987). *The improvement process: How America's leading companies improve quality.* New York, NY: McGraw-Hill.

Kendall, F. (2015). Sea-Air-Space Exposition celebrates 50th anniversary. *Seapower*. Retrieved from http://www.seapowermagazine.org/

Mabus, R. (2015, April 15). Secretary of the Navy luncheon address, Navy League Sea-Air-Space 2015, Gaylord National Harbor Resort and Convention Center, National Harbor, MD.

Nunn-McCurdy Act, 10 U.S.C. § 2433 (1983).

Office of the Under Secretary of Defense for Acquisition, Technology, & Logistics (OUSD[AT&L]). (2013, January 25). *The Planning, Programming, Budgeting, and Execution (PPBE) Process* (DoD Directive 7045.14). Washington, DC: Author.

Secretary of the Navy. (2011, December 1). Department of the Navy implementation and operation of the Defense Acquisition System and the Joint Capabilities Integration and Development System (SECNAVINST 5000.2E). Washington, DC: Author.

Van Trees, H. L. (2001). *Detection, estimation, and modulation theory, part I*. New York, NY: John Wiley & Sons.

Winnefeld, J. A. (2015, April 14). Vice Chairman of the Joint Chiefs of Staff banquet address at Navy League Sea-Air-Space 2015, National Harbor, MD.

Costing for the Future: Exploring Cost Estimation With Unmanned Autonomous Systems

Ricardo Valerdi—is an Associate Professor in the University of Arizona's Department of Systems and Industrial Engineering. His research focuses on cost estimation of complex systems, test & evaluation, human systems integration, enterprise transformation, and performance measurement. Dr. Valerdi is Editor-in-Chief of the *Journal of Cost Analysis and Parametrics*, served on the Board of Directors of the International Council on Systems Engineering, and is a Senior Member of IEEE. [rvalerdi@arizona.edu]

CPT Thomas R. Ryan, Jr., U.S. Army—is an Officer in the U.S. Army with a Master of Science in Systems and Industrial Engineering and will be an instructor in the United States Military Academy's Department of Systems Engineering. His research interests are STEM education, cost estimation of complex systems, and human systems integration. He served in both Iraq and Afghanistan as an Infantry Officer and brings a unique perspective, user and customer, to estimation of unmanned systems.

Abstract

Cost, schedule, and quality may not drive a technology, but they shape the chances of that technology becoming actualized. In recent years, the DoD, one of the leading customers of unmanned systems, has continued to struggle with management of cost and schedule causing programs to deliver products that are "good enough," delayed months to years, or even worse, decommissioned. Cost estimation techniques in use today are vast and based on techniques unrelated to emergent systems. One of the most prevalent requirements in the unmanned systems arena is autonomy. The acquisition community will need to adopt new methods for estimating the total cost of ownership of this new breed of systems. Singularly applying traditional software and hardware cost models do not provide this capability because the systems that were used to create and calibrate these models were not Unmanned Autonomous Systems (UMASs; Valerdi, Merrill, & Maloney, 2013). Autonomy, although not new, will redefine the entire way in which estimates are derived. The goal of this paper is to provide a method that attempts to account for how cost estimating for autonomy is different than current methodologies and to suggest ways it can be addressed through the integration and adaptation of existing cost models.

Introduction

Life Cycle Models

When designing a product, the recommended practice is to consider design decisions and their impact throughout the entire life cycle. This is a holistic approach that allows the engineer to examine all phases, and ensure that the stakeholders' (e.g., operators, testers, and maintainers) needs are met (Blanchard & Fabrycky, 2010). This is the same approach that should be taken when identifying product costs, thinking holistically throughout the life cycle. For purposes of discussing the realm of Unmanned Autonomous Systems (UMASs) we focus on two life cycle standards: DoD 5000 (Hagan, 2011; Mills, 2014) and ISO/IEC 15288 Systems Engineering—System Life Cycle Processes (ISO/IEC, 2002).

Both product life cycle standards are organized into discrete phases. Each phase has a distinct role in the life cycle and helps separate major milestones throughout the life cycle of a product. These life cycle stages help answer the "when" and are useful in identifying development, production, and operational costs.

DoD 5000 Acquisition Life Cycle

Although there are many commercial customers being identified and pursued within the UMAS arena, the largest acquirer of autonomous systems is the DoD. The DoD 5000 is a useful framework to apply to a product, as it forces engineers to produce specific sub-products in each of the five phases (Hagan, 2011):

1. In the first phase, Materiel Solution Analysis, the DoD requires an initial capabilities document and an analysis of alternatives study.
2. During the second phase, Technology Development, the goals are to produce a demonstrable prototype that will allow the customer to make decisions in the risk, technology, and design.
3. The third phase, Engineering and Manufacturing Development, forces the engineer to again demonstrate prototype articles, conduct integrated testing (Developmental, Operational, and Live Fire Test and Evaluation), Prepare for both the Critical Design Review and the proposal for product continuation.
4. During the fourth phase, Production and Deployment, engineers are now preparing low-rate and full scale production.
5. The final phase, Operations and Support, consists of activities such as maintaining capabilities, logistical support, upgrades, customer satisfaction, and prepare for proper disposal.

The five phases and major milestones are shown in Figure 1.

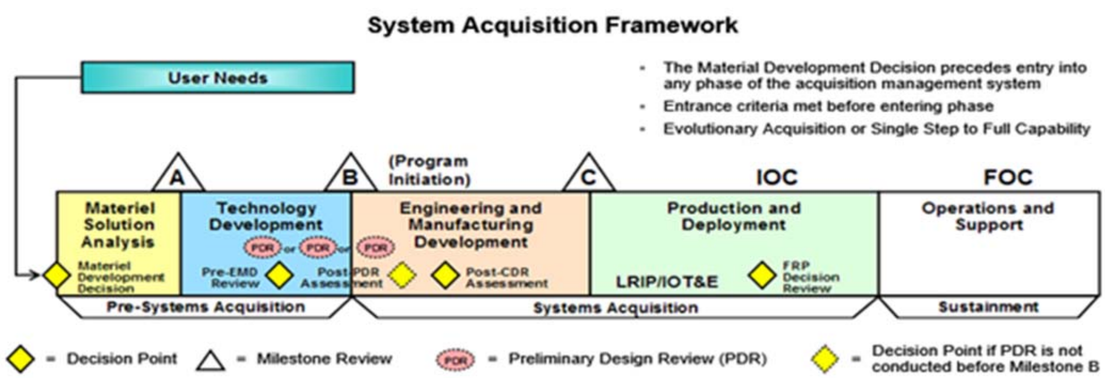


Figure 1. DoD 5000 Acquisition Framework
(Spainhower, 2003)

ISO 15288 Life Cycle

A definition of the system life cycle phases is needed to help define the boundaries between engineering activities. A useful standard is ISO/IEC 15288 Systems Engineering System Life Cycle Processes (ISO/IEC 15288). However, the phases established by ISO/IEC 15288 were slightly modified to reflect the influence ANSI/EIA 632 Processes for Engineering a System has on COSYSMO's System Life Cycle Phases, and are shown in Figure 2.

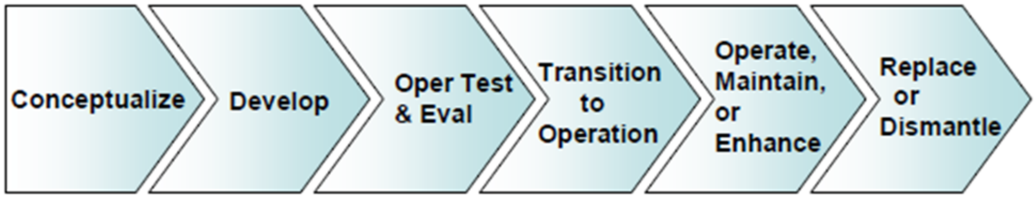


Figure 2. COSYSMO System Life Cycle Phases

Life cycle models vary according to the nature, purpose, use, and prevailing circumstances of the product. Despite an infinite variety in system life cycle models, there is an essential set of characteristic life cycle phases that exists for use in the systems engineering domain.

1. The Conceptualize stage focuses on identifying stakeholder needs, exploring different solution concepts, and proposing candidate solutions.
2. The Development stage involves refining the system requirements, creating a solution description, and building a system.
3. The Operational Test & Evaluation stage involves verifying/validating the system and performing the appropriate inspections before it is delivered to the user.
4. The Operate, Maintain, or Enhance involves the actual operation and maintenance of the system required to sustain system capability.
5. The Replace or Dismantle stage involves the retirement, storage, or disposal of the system.

We revisit these life cycle models later in this section and decompose various types of costs into their respective phases to demonstrate Total Cost of Ownership.

Cost Estimation Methods

The exploration of new cost modeling methods involves the understanding of the cost metrics relevant to the UMAS as well as an understanding of their sensitivity to cost from a production and operational standpoint. In this light, this section provides an overview of different cost estimation approaches used in industry and government. Significant work has been done to understand the costs of aircraft manufacturing (Cook & Grasner, 2001; Markish, 2002; Martin & Evans, 2000) but these studies only deal with manned commercial and military aircraft. Nevertheless, they provide useful insight on how one could approach the estimation of the UMAS life cycle cost.

Case Study and Analogy

Recognizing that companies do not constantly reinvent the wheel every time a new project comes along, there is an approach that capitalizes on the institutional memory of an organization to develop cost estimates. Case studies represent an inductive process whereby estimators and planners try to learn useful general lessons by extrapolation from specific examples. They examine in detail elaborate studies describing the environmental conditions and constraints that were present during the development of previous projects, the technical and managerial decisions that were made, and the final successes or failures that resulted. They then determine the underlying links between cause and effect that can be applied in other contexts. Ideally, they look for cases describing projects similar to the project for which they will be attempting to develop estimates and apply the rule of analogy that assumes previous performance is an indicator of future performance. The sources of case studies may be either internal or external to the estimator's own organization. Home-grown cases are likely to be more relevant for the purposes of estimation because they reflect the specific engineering and business practices likely to be applied to an organization's projects in the future. Well-documented case studies from other organizations doing similar kinds of work can also prove very useful so long as their differences are identified.

Bottom-Up & Activity-Based

Bottom-up estimating begins with the lowest level cost component and rolls it up to the highest level for its estimate. The main advantage is that the lower level estimates are typically provided by the people who will be responsible for doing the work. This work is typically represented in the form of subsystem components, which makes this estimate easily justifiable because of their close relationship to the activities required by each of the system components. This approach also allows for different levels of detail for each component. For example, the costs of an airplane can be broken down into seven main components: center-body, wing, landing gear, propulsion, systems, payloads, and assembly. Each of these components, such as the wing, can be decomposed into subcomponents such as winglet, outer wing, and inner wing. This decomposition is illustrated in more detail in Figure 3. This can translate to a fairly accurate estimate at the lower level components. The disadvantages are that this process is labor intensive and is typically not uniform across products. In addition, every level introduces another layer of conservative management reserve which can result in an overestimate at the end.

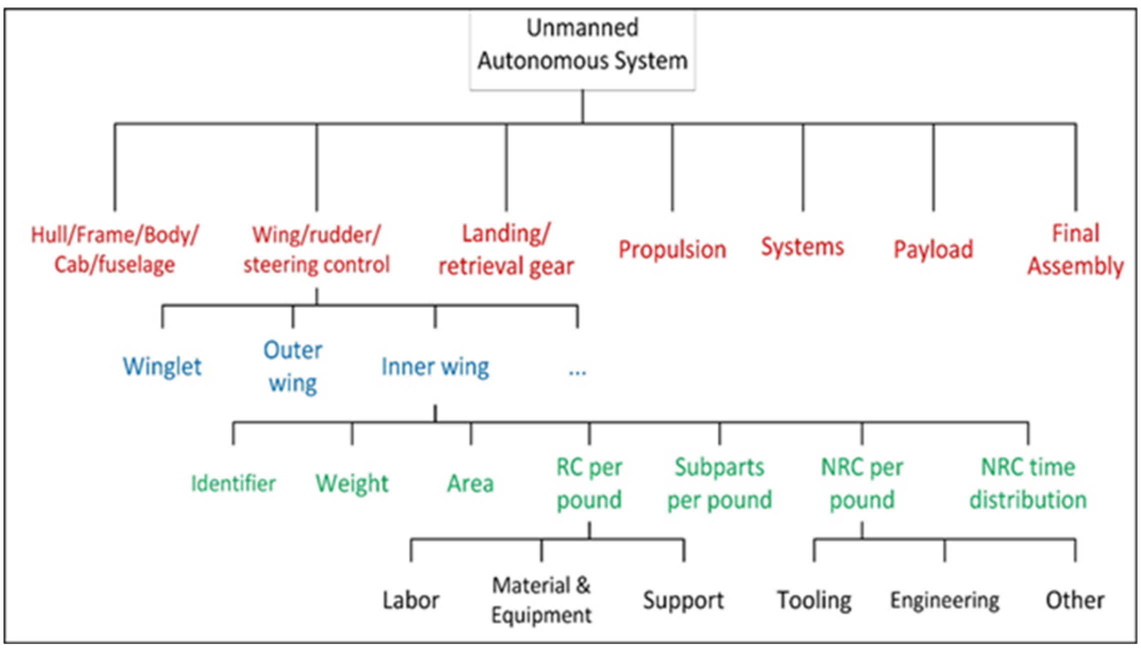


Figure 3. Product Breakdown Structure of a Typical UMAS

Parametric Modeling

This method is the most sophisticated and most time consuming to develop but often provides the most accurate result. Parametric models generate cost estimates based on mathematical relationships between independent variables (i.e., requirements) and dependent variables (i.e., effort or cost). The inputs characterize the nature of the work to be done, plus the environmental conditions under which the work will be performed and delivered. The definition of the mathematical relationships between the independent and dependent variables is the heart of parametric modeling. These relationships are commonly referred to as cost estimating relationships (CERs) and are usually based upon statistical analyses of large amounts of data. Regression models are used to validate the CERs and operationalize them in linear or nonlinear equations. The main advantage of using parametric models is that, once validated, they are fast and easy to use. They do not require a lot of information and can provide fairly accurate estimates. Parametric models can also be tailored to a specific organization's characteristics such as productivity rates, salary structures, and work breakdown structures. The major disadvantage of parametric models is that they are difficult and time consuming to develop and require a lot of clean, complete, and recent data to be properly validated. Despite the wide range of estimation approaches available for commercial and military aircraft, no parametric models have been created specifically for a UMAS. This could be attributed to the fact that UMASs have not been around for very long and, as a result, there are insufficient data available to validate such models. Before proposing a framework for such a model, unique issues pertaining to the UMAS life cycle are discussed.

UMAS Product Breakdown Structure

It is widely recognized that creating a work breakdown structure (WBS) or product breakdown structure (PBS) is the most complete way to describe a project (Larson, 1952). The level of detail required to properly utilize, or manage with, the PBS such as the one shown in Figure 3 is a crucial component to assigning costs to a product's subcomponents. In this section, we discuss some of the commonalities and shared considerations of designing a WBS/PBS within an unmanned system at the system level. Budgeted amounts for various unmanned and autonomous systems are shown in Tables 1–4 at the 2nd or 3rd level of a WBS/PBS.

Table 1. Air System (UAS)

(DoD, 2014a)

Unmanned Aerial Vehicle – Global Hawk	Unit Cost ($M)	Number of Units	Total Cost ($M)	Program allocation*
Aerial Vehicle	69.84	45	3,143.16	66.60%
Ground Control Station	21.82	10	218.21	4.62%
Support Element	n/a	n/a	1,357.84	28.77%
Projected Total Cost	n/a	n/a	4,719.21	100.00%

*Since the program allocation was only available for the Global Hawk, we applied the same ratios to other unmanned programs.

Table 2. Ground System (UGS)

(DoD, 2014b)

Unmanned Ground System COTS/GOTS	Unit Cost (M$)	Number of Units	Total Cost ($M)
Ground Vehicle	*3.39*	4	*13.56*
Ground Control Station	*0.23*	4	*0.94*
Support Element	n/a	n/a	*5.86*
Projected Total Costs	n/a	n/a	20.36

Table 3. Ground System (UGS)
(DoD, 2014b)

Small Unmanned Ground (SUGV)	Unit Cost (M$)	Number of Units	Total Cost ($M)
Ground Vehicle	*0.180*	311	*55.90*
Ground Control Station	*0.012*	311	*03.88*
Support Element	n/a	n/a	*24.15*
Projected Total Costs	n/a	n/a	83.93

Table 4. Marine System (UGS)
(DoD, 2014c)

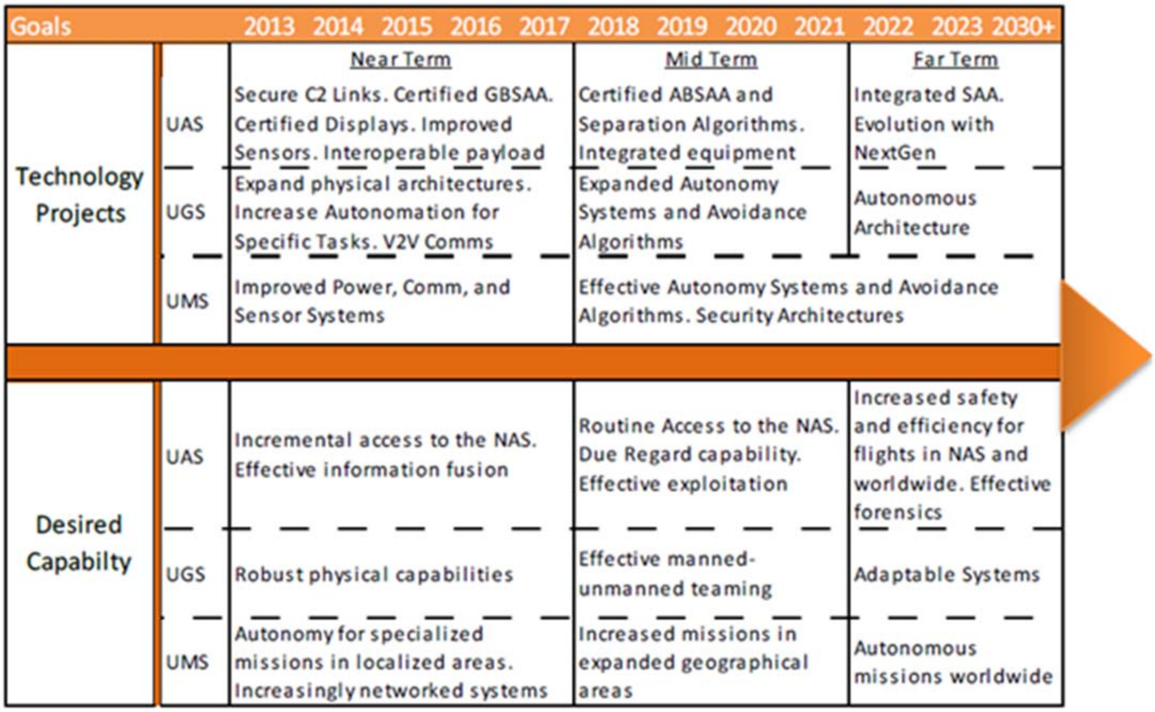
Modular Unmanned Scouting Craft Littoral (MUSCL)	Unit Cost (M$)	Number of Units	Unit Cost ($M)
Maritime Vehicle	0.700	*13*	*9.03*
Surface Control Station	*0.048*	*13*	*0.62*
Support Element	n/a	n/a	*3.90*
Projected Total Costs	n/a	n/a	13.56

That Ground Control Stations are the user controls (i.e., the video game-like interface to maneuver vehicle)
**Italicized numbers* = extrapolation based off of RQ-4 Global hawk program ratio
**unaltered numbers are from the Exhibit P-40 Presidential Budget FY2015 or equivalent cost data

One observation from the UMAS examples provided in Tables 1–4 is the range of unit costs. On the high end, the Flyaway Unit Cost of the Global Hawk Unmanned Aircraft System is $92.87 million (DoD, 2014a, p. 177). On the low end, the Modular Unmanned Scouting Craft Littoral is $700,000 (DoD, 2014c). Another observation from these examples is the wide range of units purchased; as few as four COTS/GOTS packages to convert manned systems to unmanned and as many as 311 Small Unmanned Ground Systems (DoD, 2014b).

Special Considerations

The unique physical and operational characteristics of UMASs require special consideration when exploring cost modeling approaches. In Figure 4, the DoD has laid out its desires for the UMAS over the next 30 years. The DoD has organized its requirements by air, ground, and maritime operational environments, as well as projected the types of exploration initiatives that should allow for success of these autonomous systems. Figure 4 is not meant to be totally exhaustive, but to guide the general direction of the military's UMAS vision.



Goals		2013 2014 2015 2016 2017	2018 2019 2020 2021	2022 2023 2030+
		Near Term	Mid Term	Far Term
Technology Projects	UAS	Secure C2 Links. Certified GBSAA. Certified Displays. Improved Sensors. Interoperable payload	Certified ABSAA and Separation Algorithms. Integrated equipment	Integrated SAA. Evolution with NextGen
	UGS	Expand physical architectures. Increase Autonomation for Specific Tasks. V2V Comms	Expanded Autonomy Systems and Avoidance Algorithms	Autonomous Architecture
	UMS	Improved Power, Comm, and Sensor Systems	Effective Autonomy Systems and Avoidance Algorithms. Security Architectures	
Desired Capabilty	UAS	Incremental access to the NAS. Effective information fusion	Routine Access to the NAS. Due Regard capability. Effective exploitation	Increased safety and efficiency for flights in NAS and worldwide. Effective forensics
	UGS	Robust physical capabilities	Effective manned-unmanned teaming	Adaptable Systems
	UMS	Autonomy for specialized missions in localized areas. Increasingly networked systems	Increased missions in expanded geographical areas	Autonomous missions worldwide

Figure 4. Operating Environment Technology Development Timeline (2013–2030)
(DoD, 2013)

Mission Requirements (DoD, 2013)

The mission requirements are specified tasks with which the UMAS must comply in order to perform. These requirements are shaped by the operational environment (OE), or venue by which the UMAS will perform its intended functions or capabilities that can be physical and situational. The physical environment can consist of air, ground (surface and sub-surface), and marine (surface and submersible.)

System Capabilities

In essence, what will the UMAS do for the customer? These functions must also include current capabilities such as attack, logistical, and reconnaissance. This area also includes any of the "-ilities" that a UMAS might need to adhere to that are not specified in its mission requirements. These may include manufacturability, reliability, interoperability, survivability, and maintainability.

Payloads

A final consideration for the UMAS is its payload. This could also be categorized as special equipment. For example, a logistical UMAS (or cargo transportation system like the SMSS™) needs to have a tow system or recovery package in addition to the ground vehicle; or if it is an attack/reconnaissance system—it needs to support munitions, missiles, or gun platforms.

Although many more areas can be identified for consideration when engineering a system for autonomy, this section was meant to highlight the WBS/PBS in more detail rather than the technical capabilities of the UMAS itself. The cost to build and produce a system is a bottom line decision for the producer (and the engineer), but the DoD needs and expects that a WBS represent all phases of the life cycle. By accurately representing the system in a more complete WBS/PBS, the cost estimates will have more fidelity and a higher confidence, because estimators will be able to link the lowest level of that structure to a group of cost drivers within a cost model.

Cost Drivers and Parametric Cost Models

Cost drivers are characteristics of projects that best capture the effort, typically measured in Person Months, required to complete them (Boehm, 2000). As mentioned in the Parametric Modeling section, developing these characteristics, or drivers, is data and labor intensive. The developer of the model must establish a strong mathematical relationship, usually a form of regression, between an identified characteristic and its impact on the project. The number of cost drivers for each type of estimate will vary according to the type of component (hardware, software, etc.).

Each cost driver has a scale, usually of five levels, which allows the user of the model to best represent characteristics of the product. For example, a cost driver can be described using Very Low, Low, Nominal, High, or Very High—each one of these choices has a value that will either increase or decrease cost (Valerdi, 2008). Each level is clearly defined so the user can estimate the complexity of a system as realistically as possible. The key for success with utilizing parametric modeling and its drivers is to fully understand and be realistic with assignment of scale values.

Cost Drivers for Estimating Development Costs

Our proposed method for system level estimation is to combine five different parametric models that best represent the amount of effort required to successfully build, test, produce, and operate an Unmanned Autonomous System (UMAS). These include (1) Hardware, (2) Software, (3) Systems Engineering and Program Management, (4) Performance-Based Characteristics, and (5) Weight-Based Characteristics.

Each of the five models is subsequently described and should be considered when developing a complete life cycle estimate; however, it is not mandatory to utilize all five since each UMAS will have unique cost and performance considerations.

Hardware

SEER-H is a hybrid model that utilizes analogous estimates, as well as harnessing parametric mathematical cost estimation relationships specific to hardware products. SEER-H aids in the estimation of hardware development, production, and operations costs (SEER-H® Documentation Team: MC, WL, JT, KM, 2014). Unlike the other estimation tools available, SEER-H has an exhaustive suite and could be used to estimate many technical areas. The number of cost drivers in SEER-H is extensive; therefore we focus on only three within the Mechanical/Structural Work Elements category:

- Material Composition—the material that will dominate the system and its difficulty to acquire
- Certification Level—the amount of Test & Evaluation with demonstration required for the materials utilized
- Production Tools and Practices—how ready the materials are for production

Material Composition

This SEER-H driver is categorized by the predominant material used to build the system, sub-system, or the system's components, as shown in Table 5. The estimator should also consider some of the materials that may not dominate, but are identified as critical. The total cost may be a combination of critical and dominant materials.

Table 5. Material Composition Rating Scale
(SEER-H® Documentation Team, 2014)

Material	Key Property
Aluminum/ Malleable Metals	Metal alloy, easily manufactured. Example: Aluminum, magnesium, copper, aluminum-lithium.
Steel	Hard, rigid metal alloy, resistant to rust. Example: Steel, Stainless Steel.
Commercially Available Exotic	Commodity available exotic materials. Example: Titanium, precious metals, boron, higher end composites.
Other Exotic	Requires very complex metallurgical processes, available only through special orders. Example: Metal matrix composites, particulate strengthened composite materials, research materials.
Composite	Commodity available, continuous filament or particulate strengthened composite materials. Example: Graphite or boron epoxy, fiber glass.
Polymer	Nonmetallic compound, easily molded, may be hardened or pliable. Example: Plastics, thermoplastics, elastomers.
Ceramic	Very Strong, brittle. Example: Ceramic, clay, glass, tile, porcelain

Certification Level

Certification level represents the requirements imposed on the manufacturer by the customer, as shown in Table 6. This parameter quantifies the additional cost associated with the customer's certification requirements; therefore, any extra certification, inspections, or intangible property security controls, etc., will increase cost.

Table 6. Certification Level Rating Scale
(SEER-H® Documentation Team, 2014)

Rating	Description
VERY HIGH	Very high level of qualification testing including fatigue, fracture mechanics, burst, temperature extremes and vibration testing. Example: Manned Space Product.
HIGH	High level of qualification testing including fatigue, fracture mechanics, burst, temperature extremes and vibration testing. Example: Space Product.
NOMINAL +	Qualification testing for mission requirements including static and dynamic load testing, wind tunnel testing and all other tests required for military aircraft. Example: Military Airborne/ Aircraft Product.
NOMINAL	Qualification testing in accordance with FAA requirements, as specified for commercial or general aviation aircraft. Example: Airborne/ Aircraft Product.
NOMINAL -	Qualification testing in accordance with U.S. Army Mobility requirements, or U.S. Navy specifications. Testing includes meeting shock, vibration, temperature and humidity requirements. Example: Military Ground-Mobile or Sea Product.
LOW	Nominal qualification testing for mission requirements covering equipment located in controlled environments (temperature, humidity). Example: Military Ground System
VERY LOW	Minimal testing required (functional check-out). Example: Commercial Grade Product.

Production Tools and Practices

This parameter describes the extent to which efficient fabrication methodologies and processes are used, and the automation of labor-intensive operations. The rating should reflect the state of production tools that are in place and already being used by the time hardware production begins (see Table 7).

Table 7. Production Tools and Practices Rating Scale
(SEER-H® Documentation Team, 2014)

Rating	Description
VERY HIGH	Production tooling associated normally with large-scale production (20,000 units or above). Highly sophisticated tools, die casts, molds. High degree of mechanization, robotics manufacture, assembly and testing. High degree of integration between computer-aided manufacturing and design. Example: Die casting, multi-cavity molds, progressive dies and other sophisticated tools.
HIGH	Production tooling normally applied to medium scale, averaging 20,000 unit production. Tools are custom designed with simple dies. Some degree of mechanization, numerically controlled machine tools, some integration with computer aided design. Example: Simple die casting, complex investment casting, custom die sets for sheet metal fabrication.
NOMINAL	Production tooling facilitates production of 1,000–2,000 units. Complex tools, simple dies and castings. Little mechanization, few numerically controlled machining operations. Some automated links with CAD. Example: Complex sand castings, investment castings of some complexity, and simple custom die sets are included in the tooling category.
LOW	Tooling designed for the production of up to 1,000 units. Standard tools, casts, dies, and fixtures are supplemented with some custom tools and jigs. Occasional or experimental use of automated links with CAD. Example: Sand castings, investment castings and simple custom die sets. Many Aerospace/ DoD programs are in this category.
VERY LOW	Minimum tooling required to produce up to about 50–100 units. Many operations of manufacture, assembly and test are by skilled labor. The use of standard tools and fixtures is predominant. No automated links. Example: Simple sand castings are in this category.

Software

The recommended parametric estimation tool for UMAS software aspects is the Constructive Cost Model (COCOMO II). This model has 30 years of refinement, and is an industry and academic standard for parametric modeling (Boehm, 2000). The number of cost drivers in COCOMO II vary from 7 to 17 depending on the life cycle phase of the project in which the estimate is being performed (Boehm, 2000). Since less information is known at the beginning of the project, the COCOMO II model provides fewer parameters to rate. As more information is known about the software project, the number of parameters increases. This section is not meant to replace the COCOMO II User's Manual,[1] but rather provide relevant details about the relevant cost drivers. Three drivers are relevant for UMAS software estimation:

- Size—measured by number of lines of lode
- Team Cohesion—weighted average of four characteristics
- Programmer Capability—how efficient programmers are as a whole

Size

Size is in units of thousands of source lines of code (KSLOC) is derived from estimating the size of software modules that will constitute the application program. It can also be estimated using unadjusted function points (UFP), converted to SLOC, then divided

[1] http://csse.usc.edu/csse/research/COCOMOII/cocomo_main.html

by one thousand. Equation 1 is the basic COCOMO II algorithm which includes Size as the central component to calculating effort in Person Months (PM).

$$PM = A \times (SIZE)^E \times \prod_{i=1}^{n} EM_i \quad (1)$$

Team Cohesion

This parameter accounts for the human component in software design. These elements are not limited to but contain differences in multiple stake-holder objectives, cultural backgrounds, team resiliency, and team familiarity (see Table 8). The focus is how the design team interacts externally within the project.

Table 8. Team Cohesion Rating Scale

Characteristic	Very Low	Low	Nominal	High	Very High	Extra High
Consistency of Stakeholder objectives and cultures	Little	Some	Basic	Considerable	Strong	Full
Ability, willingness of stakeholders to accommodate other stakeholder's objectives	Little	Some	Basic	Considerable	Strong	Full
Experience of stakeholders in operating as a team	None	Little	Little	Basic	Considerable	Extensive
Stakeholder teambuilding to achieve shared vision and commitments	None	Little	Little	Basic	Considerable	Extensive

Programmer Capability

This parameter also deals with a human aspect of software engineering; however, it differs from team cohesion in the direction of the focus. In this parameter the assessment is on the internal workings of the team's capability as it relates to the team's efficiency, thoroughness, internal communication, and cooperation (see Table 9).

Table 9. Programmer Capability Rating Scale

PCAP Descriptors	15th percentile	35th percentile	55th percentile	75th percentile	90th percentile	
Rating Levels	Very Low	Low	Nominal	High	Very High	Extra High
Effort Multipliers	1.34	1.15	1.00	0.88	0.76	n/a

Systems Engineering and Project Management

To estimate the Systems Engineering and Project Management required effort for a UMAS, we use the Constructive Systems Engineering Cost Model (COSYSMO). This parametric model's output accounts for integrating system components and will quantify intangible efforts such as requirements, architecting, design, verification, and validation (Valerdi, 2008). This model also depends on 18 size and cost drivers.[2] By introducing some of the most important drivers we capture the most important cost considerations of a UMAS. The three most relevant systems engineering cost drivers are as follows:

---

[2] http://cosysmo.mit.edu

- Number of System Requirements—number of specified functions a system must perform to meet the user's needs
- Technology Risk—how mature or demonstrated the technologies are
- Process Capability—how well/consistent the team/organization performs in terms of the Capability Maturity Model Integration (CMMI)

Number of Requirements

The Number of Requirements parameter asks the estimator to count the number of requirements for the UMAS at a specific level of design (see Table 10). These requirements may deal with number of system interfaces, system specific algorithms, and operational scenarios. Requirements are not limited to but may be functional, performance, feature, or service-oriented in nature depending on the methodology used for specification. Of note, requirement statements usually contain the words "shall," "will," "should," or "may."

Table 10. Number of Requirements Rating Scale

Easy	Nominal	Difficult
Simple to implement	Familiar	Complex to implement
Traceable to source	Can be traced to source with some effort	Hard to trace to source
Little requirements overlap	Some overlap	High degree of requirement overlap

Technology Risk

The Technology Risk parameter asks you to evaluate a UMAS's sub-system's maturity, readiness, and obsolescence of the technologies being implemented (see Table 11). Immature or obsolescent technologies will require more systems engineering effort.

Table 11. Technology Risk Rating Scale

	Very Low	Low	Nominal	High	Very High
Lack of Maturity	Technology proven and widely used throughout industry	Proven through actual use and ready for widespread adoption	Proven on pilot projects and ready to roll-out for production jobs	Ready for pilot use	Still in the laboratory
Lack of Readiness	Mission proven (TRL 9)	Concept qualified (TRL 8)	Concept has been demonstrated (TRL 7)	Proof of concept validated (TRL 5 & 6)	Concept Defined (TRL 3 & 4)
Obsolescence			Technology is the state-of-the-practice Emerging technology could compete in future	Technology is stale New and better technology is ready for pilot use	Technology is outdated and should be avoided in new systems Spare parts supply is scarce

Process Capability

Like some of the COCOMO II parameters, this COSYSMO example focuses on the consistency and effectiveness of a project team performing the systems engineering processes. The assessment of this driver may be based on ratings from a published process model (e.g., CMMI [2002], EIA-731 [ANSI/EIA, 2002], SE-CMM [Boehm, 2000; Clark, 1997], ISO/IEC 15504 [2003, 2012]). It can alternatively be based on project team behavioral characteristics if no previous external assessments have occurred.

Table 12. Process Capability Rating Scale

	Very Low	Low	Nominal	High	Very High	Extra High
CMMI Assessment Rating	Level 0 (if continuous model)	Level 1	Level 2	Level 3	Level 4	Level 5
Project Team Behavioral Characteristics	Ad Hoc approach to process performance	Performed SE process, activities driven only by immediate contractual or customer requirements, SE focus limited	Managed SE process, activities driven by customer and stakeholder needs in a suitable manner, SE focus is requirements through design, project-centric approach – not driven by organizational processes	Defined SE process, activities driven by benefit to project, SE focus is through operation, process approach driven by organizational processes tailored for the project	Quantitatively Managed SE process, activities driven by SE benefit, SE focus on all phases of the life cycle	Optimizing SE process, continuous improvement, activities driven by system engineering and organizational benefit, SE focus is product life cycle & strategic applications
SEMP Sophistication	Management judgment is used	SEMP is used in an ad-hoc manner only on portions of the project that require it	Project uses a SEMP with some customization	Highly customized SEMP exists and is used throughout the organization	The SEMP is thorough and consistently used; organizational rewards are in place for those that improve it	Organization develop best practices for SEMP; all aspects of the project are included in the SEMP; organizational reward exists for those that improve it

Performance-Based Cost Estimating Relationship

One important consideration of every product is its ability to perform the specified requirements well. The model that best captures the performance characteristics of a product was created by the Army for Unmanned Aerial Vehicle Systems, but can be modified to fit other autonomous systems (Cherwonik & Wehrley, 2003). The methodologies for estimating performance are not restricted to this list, but should fit in similar categories for air, land, sea, or space (see Table 13).

Table 13. Performance-Based Characteristics Rating Scale

(Cherwonik & Wehrley, 2003)

Performance Based Categories	Descriptions
Vehicle or Body of UMAS	Define and measure how well the vehicle or body of the UMAS performs its intended requirements.
Sensors	Define and measure how well the UMAS can interact and react with its intended (or unintended) environment
Control System	Define and measure how efficient the command and control system interacts with UMAS

The cost drivers that are recommended for performance measurement are based on an aerial platform, but are modified in this section to provide ideas on what areas to consider (see Table 14).

Table 14. Performance Cost Drivers

(Cherwonik & Wehrley, 2003)

Performance Drivers	Description/ Use of driver
Operational Environment Constraints	Define and measure the physical boundaries guiding the UMAS.
Endurance	Define and measure the amount of time or distance the UMAS can perform its intended task prior to needing human interaction.
Sensor Resolution	Define and measure the sensitivity, accuracy, resiliency, and efficiency of the UMAS sensors.
Base of Operations	Define and measure how constrained the UMAS by its logistical requirements and the resources required for effective operations.

The Army's performance-based Cost Estimating Relationship is shown in Equation 2:

$$\underline{\text{UAV T1R1}}\ (\text{FY03\$K}) = 118.75 * (\underline{\text{Endurance}} * \underline{\text{Payload-Wt}}.)^{0.587} * e^{-0.010(\underline{\text{FF Year-1900}})} * e^{-0.921(\underline{\text{Prod 1/0}})} \quad (2)$$

Where:

UAV T1R1 = Theoretical first unit cost of UAV air vehicle hardware normalized for learning (95% slope) and rate (95% slope), via unit theory. In FY03 $K.

Endurance = UAV air vehicle endurance in flight hours

Payload-Wt. = Weight of total payload in pounds. Total payload includes all equipment other than the equipment that is necessary to fly and excludes fuel and weapons.

FF-Year = Year of first flight

Prod 1/0 = 1 if air vehicle is a production unit.
= 0 if air vehicle is a development or demonstration unit.

Weight-Based Cost Estimating Relationship

A final consideration for estimating the cost of the UMAS is its weight. Weight may already exist as an important cost driver in other estimation models such as hardware and performance; however, we feel that this particular estimation relationship is strong enough to also be a stand-alone component. When operational implementation is considered for a given autonomous system, weight plays a critical role in the success or failure. Some drivers, modified from the source to apply to the UMAS, are shown in Table 15.

Table 15. Weight-Based Cost Drivers
(Cherwonik & Wehrley, 2003)

Weight-Based Drivers	Description/Use of driver
Weight of total system	Define and Measure the weight of total system as it relates to its intended objectives (this does not include ordnance or other attachable options).
Payload Weight	Define and Measure the amount and type of ordnance or any additional attachable option that is deemed mission critical.
Sling-load or Recovery Operation Capacity* *if applicable	Define and Measure the amount of weight the UMAS can support as a sling load or in a tow capacity, in addition to its nominal capacity.

The Army's weight-based Cost Estimating Relationship is shown in Equation 3:

$$\text{UAV T1R1 (FY03\$K)} = 12.55 * (\text{MGTOW})^{0.749} * e^{-0.371(\text{Prod } 1/0)} \quad (3)$$

Where:
UAV T1R1 = Theoretical first unit cost of UAV air vehicle hardware normalized for learning (95% slope) and rate (95% slope). In FY03 $K.
MGTOW = UAV air vehicle maximum take-off weight in pounds.
Prod 1/0 = 1 if air vehicle is a production system
= 0 if air vehicle is a development or demonstration model.

Proposed Cost Drivers for DoD 5000.02 Phase Operations & Support

Logistics—Transition From Contractor Life Support (CLS) for Life to Organic Capabilities

Managing logistic support is complex and not easy to summarize into a single parameter. However, all systems require maintenance which can be described within the range provided in Table 16. The goal of this parameter is to allow life cycle planners to nest their system engineering plan into DoD requirements and minimize contractor life support.

Table 16. Logistics Cost Driver

Uniformed Servicemen Only	> 2 years transition	2–5 year transition	< 5 year transition	CLS Only
System was designed in a manner that current life support is sufficient for operational use.	Very few contractors (1–5) needed at Colonel (O-6) level command units to ensure proper life support.	Few contractors (6–10) needed at Colonel (O-6) level command units to ensure proper life support.	Contractors are needed at every level of command Captain (O-3) through Colonel (0-6). Minimum 1 at each level.	System is so technologically advanced that operational use will require a permanent contractor presence.

Training

The development costs for a UMAS can be significant, but one area of consideration is how quickly and efficiently users can be trained to employ the system. With the increasing levels of autonomy, this warrants its own cost driver.

Table 17. Training Cost Driver Considerations

Minimal impact	Medium Impact	High Impact	Extreme Impact	Unknown Impact
Training fits current TRADOC[3] through-put. And requires minimal certification (example system is a modified version of a previously integrated systems – autonomous raven)	Training program is similar to a current DoD method; however, needs to be a stand-alone block of instruction or course. Can use existing facilities and infrastructure currently provided.	Training program is not similar to any current DoD method. Needs to be a stand-alone course. Needs facilities and infrastructure not currently provided.	Training program is not similar to any current DoD method. Needs to be a stand-alone course. Needs facilities and infrastructure not currently available.	Training systems are still being developed and will require extensive integration

The planning for and implementation of such training considerations in Table 17 will be challenging. The DoD acknowledges these challenges and offers a perspective of expectation management displayed in Figure 5. The training objectives attempt to lay out how the UMAS and other emergent systems will be inculcated into the existing training system. As engineers build their systems understanding, these strategies will help with system implementation in areas that are not implicitly the system being procured.

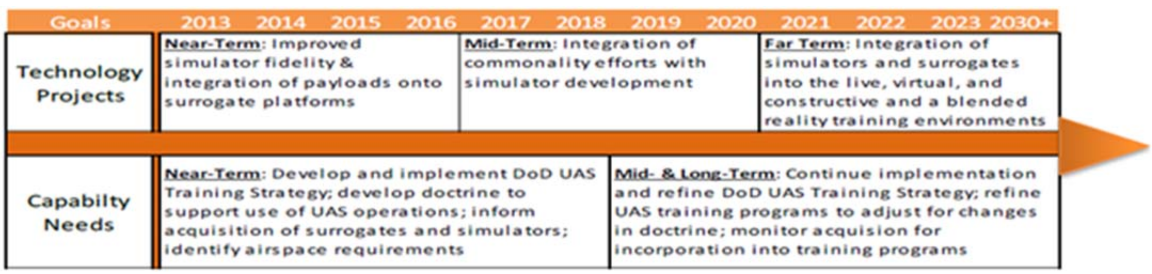


Figure 5. UMAS Training Objectives (2013–2030)
(DoD, 2013)

Operations—Manned Unmanned Systems Teaming (MUM-T)

The goal of the DoD's investment in the UMAS is to enhance the warfighters' capability while reducing risk to human life, maintaining tactical advantage, and performing tasks that can be dull, dirty, or dangerous (DoD, 2013).However, all of the systems will require some level of manned-with-unmanned cooperation. The more these two worlds efficiently work together, the better the operational outcome.

[3] U.S. Army Training and Doctrine Command

Table 18. Manned Unmanned Systems Teaming Cost Driver

Very Low Teaming	Low Teaming	Nominal Teaming	High Teaming	Very High Teaming
Meets no joint interoperability requirements, and generates data that needs to be transferred to a common operating picture	Meets minimum branch specific interoperability requirements, but is not compatible with all systems employed by its home branch	Meets branch specific interoperability requirements, and shares information with manned systems, branch specific.	Meets all branch specific interoperability requirements, as well as one or more joint requirements. Also shares information with manned systems.	Meets all Joint interoperability requirements, and shares a common operating picture other manned and unmanned systems can utilize

Considerations for Estimating Unmanned Ground Vehicle

For a large scale project that requires the integration of multiple engineering disciplines, specifically in the field of the UMAS, no single estimation tool can completely capture total life cycle costs. By applying the proper estimation models, or a combination of these models, the estimator can ensure complete coverage of each program element and their relative cost impact across the UMAS project life cycle.

The example used to illustrate the cost estimating process is the Lockheed Martin Unmanned Autonomous Ground System, Squad Mission Support System (SMSS™). By utilizing the product work breakdown structure (P-WBS) cost experts can then apply an estimation tool at the appropriate level. The sum of each sub-estimate is then integrated into the overall project level estimation. Considerations for which level within the P-WBS requires estimates is unique to each UMAS project. Contractual requirements will be the determining factor on how detailed the estimate needs to be.

In response to the critical need for lightening, the soldier and marine infantryman's load in combat as well as providing the utility and availability of equipment that could not otherwise be transported by dismounted troops, the Squad Mission Support System is being developed by Lockheed Martin. The SMSS™ can address the requirements of Light Infantry, Marine, and Special Operating Forces to maneuver in complex terrain and harsh environments, carrying all types of gear, materiel, and Mission Equipment Packages (MEP).

The SMSS™ is a squad-sized UGV platform shown in Figure 6, about the size of a compact car, capable of carrying up to 1,500 pounds of payload. Designed to serve as a utility and cargo transport for dismounted small unit operations, it possesses excellent mobility in most terrains. The SMSS™/ Transport lightens the load of a 9–13 man team by carrying their extended mission equipment, food, weapons, and ammunition on unimproved roads, in urban environments, and on cross-country terrain. Control modes include tethered, radio control, teleoperation (NLOS and BLOS), supervised autonomy, and voice command. TRL level is 7–9.

Figure 6. Squad Mission Support System (SMSS™)
(Lockheed Martin, n.d.)

As shown in Table 19, the five proposed cost models adequately capture all of the P-WBS elements of the SMSS™. In some cases, the cost of individual elements can be captured by more than one cost model. To ensure that costs are not double counted, the estimator should decide which of the cost models will be used for each WBS element. This decision could be based on the amount of fidelity provided by each cost model or the ability of the cost model to capture the WBS element's characteristics that influence cost.

Table 19. Types of Estimates Needed per Product Breakdown Structure Element

Type of Model Recommended

Ref. #	WBS Element (Mills, 2014)	Hardware (SEER-H)	Software (COCOMO II)	Systems Engineering (COSYSMO)	Weight-based CER	Performance-based CER
1	Squad Mission Support System (SMSS)					
1.1	Common Mobility Platform Vehicle					x
1.1.1	Vehicle Integration, Assembly, Test, and Checkout			x		
1.1.2	Hull/Frame/Body/Cab				x	x
1.1.2.1	Main Chassis Structure	x			x	
1.1.2.1.1	Frame and Hull				x	x
1.1.2.1.2	Hood				x	
1.1.2.1.3	Deck Panels				x	
1.1.2.1.4	Skid Plate				x	
1.1.2.2	Electronics Box Structure				x	
1.1.2.3	Front Brush Guard				x	
1.1.2.4	Rear Brush Guard				x	
1.1.2.5	Front Sensor/Component Mount				x	x
1.1.2.6	Rear Sensor/Component Mount				x	x
1.1.2.7	Equipment Rack				x	
1.1.2.8	Pack Racks/Tail Gate				x	
1.1.3	System Survivability			x		x
1.1.4	Turret Assembly		x	x		x
1.1.5	Suspension/Steering					x
1.1.6	Vehicle Electronics		x	x		
1.1.7	Power Package/Drive Train	x				x
1.1.8	Auxiliary Automotive	x		x	x	x
1.1.9	Fire Control					x
1.1.10	Armament				x	x

1.1.11	Automatic Ammunition Handling				x	x
1.1.12	Navigation and Remote Piloting					x
1.1.12.1	Navigation Unit		x		x	
1.1.12.2	Robotics Subsystem		x			x
1.1.12.3	Autonomy Subsystem		x			x
1.1.13	Special Equipment			x		
1.1.14	Communications		x		x	x
1.1.15	Vehicle Software Release		x			
1.1.16	Other Vehicle Subsystems			x		
1.2	Remote Control System				x	
1.2.1	Remote Control System Integration, Assembly, Test, and Checkout			x		x
1.2.2	Ground Control Center Subsystem			x		
1.2.3	Operator Control Unit (OCU) Subsystem				x	
1.2.4	Remote Control System Software Release		x			

Once the appropriate cost models are determined for each WBS element, the cost can be calculated as the sum of the outputs of the five cost models, as shown in Equation 4.

$$\begin{aligned} &Cost\ (convert\ all\ individual\ outputs\ to\ \$K) \\ &\quad = (Hardware) + (COCOMO\ II) + (COSYSMO) \\ &\quad +(Weight\ Based\ CER) + (Perfomance\ Based\ CER) \end{aligned} \quad (4)$$

The expected unit cost would be in the range of $1 million to $100 million, depending on the capabilities and complexities of the UMAS. This is based on the historical results from the unit cost of the Global Hawk Unmanned Aircraft System ($92.87 million) and Modular Unmanned Scouting Craft Littoral ($700,000). If the estimated cost falls outside of this range, careful analysis should be done to ensure that the capabilities of the UMAS being estimated are truly beyond the scope of the historical data.

Another basis of comparison could be the two cost estimating relationships described in this section which consider flight hours and maximum takeoff weight. While these cost drivers would only be relevant for Unmanned Aerial Vehicles, they can serve as sanity checks when performance and weight are important considerations.

For the purposes of this section of the report, we are unable to provide a comparison of actual costs versus estimated costs to validate our proposed cost modeling approach. One reason is the proprietary nature of the data. Another is the lack of fidelity that is available to compare UMAS costs using the same cost elements, namely vehicle, ground control station, and support elements.

Additional Considerations for UMAS Cost Estimation

Test and Evaluation

Many systems engineering and project management experts advise concurrent planning of test and evaluation (T&E) during the earliest phases of a project (Blanchard & Fabrycky, 2003). In similar fashion, estimating the cost of these activities should also begin earlier rather than later. As budgets are allocated and costs are estimated, some key considerations on how the UMAS may be tested might be analytical testing, prototyping, production sampling, demonstration, and modification (Blanchard & Fabrycky, 2003). The current practice in many organizations is to focus most of the cost of product development and when the project reaches the T&E phases use the remaining funding. This often leads to reduced testing and schedule slippages.

2 Demonstration

Demonstration is one of the unique aspects of T&E because there are many categories or sub-sets of demonstrating a product's capability. The two that are most important are demonstrating systems integration and demonstrating full operational capability. The costs associated with these are very different, and will also vary by type of UMAS. Some questions to consider when estimating the UMAS, but specifically demonstrating the UMAS, are as follows:

Level of Autonomy:

a. At what level of autonomy is the UMAS designed to operate?
b. How will the level of autonomy influence safety, reliability, and integration to other systems?

Systems Integration:

a. Will these demonstrations coincide with the design reviews or be separate events?
b. What key system capabilities will your team want to demonstrate?
c. Will you focus only on risky technology or demonstrate solutions to previously developed concepts?

Full Operational Capability:

a. Who is your audience? Depending on whether it is government or commercial this will play a huge factor in where and how you demonstrate.
b. Will you need to create an operational scenario to show how the UMAS integrates into the current paradigm of its intended field? For example, will you need to have a mock battle, or create a queuing backlog at a distribution plant or border crossing?

Conclusion

In this section, we described unique considerations of Unmanned Autonomous Systems. In particular, life cycle models that help structure cost estimates, existing cost estimation methods, product work breakdown structures, and parametric models. These led to a case study that described an Army Unmanned Vehicle and a recommended approach for estimating the per unit life cycle cost. We concluded by discussing two unique considerations of estimating the cost of the UMAS—levels of autonomy, test and evaluation, and demonstration—that have the potential to significantly influence the complexities involved with transitioning a UMAS into operation.

As the UMAS continue to be developed and deployed into operation we anticipate the maturity and accuracy of estimating their costs will similarly increase. At the moment, reliance on complete work breakdown structures, comparisons with historical data, and utilization of existing parametric cost models can provide a reliable estimation process that can be used to develop realistic cost targets.

References

ANSI/EIA. (2002). EIA-731.1 Systems engineering capability model.

ANSI/EIA. (2003). *Processes for engineering a system*. ANSI/EIA 632-2003. Philadelphia, PA: American National Standards Institute (ANSI)/Electronic Industries Association (EIA).

Blanchard, B., & Fabrycky, W. (2010). *Systems engineering and analysis* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Boehm, B. W. (2000). *Software cost estimation with Cocomo II.* Upper Saddle River, NJ: Prentice Hall.

Cherwonik, J. & Wehrley, A. (2003). *Unmanned aerial vehicle system acquisition cost research: Estimating methodology and database.* Washington, DC: Office of the Deputy Assistant of the Army for Cost and Economics.

Clark, B. K. (1997). *The effects of software process maturity on software development effort.* Unpublished Dissertation, USC Computer Science Department.

CMMI. (2002). Capability maturity model integration-CMMI-SE/SW/IPPD/SS, V1.1. Pittsburg, PA: Carnegie Mellon-Software Engineering Institute.

Cook, C. R., & Grasner, J. C. (2001). *Military airframe acquisition costs: The effects of lean manufacturing*. Project Air Force. Santa Monica, CA: RAND.

DoD. (2013). Unmanned systems integrated roadmap FY2013–2038. Washington, DC: Author.

DoD. (2014a, March). *DoD fiscal year (FY) 2015 budget estimates: Air Force justification book, Vol. 1.* Retrieved from http://www.saffm.hq.af.mil/shared/media/document/AFD-140310-041.pdf

DoD. (2014b, March). *DoD fiscal year (FY) 2015 budget estimates: Army justification book.* Retrieved from http://asafm.army.mil/Documents/OfficeDocuments/Budget/budgetmaterials/fy15/pforms//opa34.pdf

DoD. (2014c, March). *DoD fiscal year (FY) 2015 budget estimates: Navy*. Retrieved from http://www.finance.hq.navy.mil/fmb/15pres/OPN_BA_5-7_Book.pdf

Hagan, G. (2011, May 4). *Overview of the DoD systems acquisition process. DARPA Webinar.* Lecture conducted from DARPA.

ISO/IEC. (2002). ISO/IEC 15288:2002(E). Systems engineering—System life cycle.

ISO/IEC. (2012). ISO/IEC 15504:2003. Information technology—Process assessment, Parts 1–10.

Larson, E. W. (1952). *Project management: The managerial process* (5th ed.). E. W. Larson & C. F. Gray (Eds.). McGraw-Hill.

Lockheed Martin. (n.d.). SMSS [image]. Retrieved from http://www.lockheedmartin.com/us/products/smss.html

Markish, J. (2002). *Valuation techniques for commercial aircraft program design* (Master's thesis, Aeronautics and Astronautics Dept.). Cambridge, MA: MIT.

Martin, R., & Evans, D. (2000). Reducing costs in aircraft: The metals affordability initiative consortium. *JOM, 52*(3), 24–28.

Mills, M. E. (2014). *Product work break down structure for SMSS™* provided by Lockheed Martin Missiles and Fire Control.

SEER-H® Documentation Team: MC, WL, JT, KM. (2014). *SEER for hardware detailed reference—User's manual*. Galorath, Inc.

Spainhower, K. (2003). Life cycle framework. Retrieved June 10, 2014, from https://dap.dau.mil/aphome/das/Pages/Default.aspx

Valerdi, R. (2008). *The constructive systems engineering cost model (COSYSMO): Quantifying the costs of systems engineering effort in complex systems*. Saarbrucken, Germany: VDM Verlag Dr. Muller.

Valerdi, R., Merrill, J., & Maloney, P. (2013). *Cost metrics for unmanned aerial vehicles.* AIAA 16th Lighter-Than-Air Systems Technology Conference and Balloon Systems Conference.

Acknowledgments

This material is based upon work supported by the Naval Postgraduate School Acquisition Research Program under Grant No. N00244-15-1-0008. The views expressed in written materials or publications, and/or made by speakers, moderators, and presenters, do not necessarily reflect the official policies of the Naval Postgraduate School nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Panel 19. Organizing for Success: Managerial and Staffing Considerations

Thursday, May 5, 2016	
1:45 p.m. – 3:15 p.m.	**Chair: Reuben Pitts,** President, Lyceum Consulting, LLC ***Organization Analytics: Taking Cost-per-Dollar-Obligated (CPDO) Measures to the Next Level in Defense Contracting*** Timothy Reed, Principal Director, Beyond Optimal Strategic Solutions James Keller, Business Center Head, USMC Operations Analysis Division John Fallon, Professor, Villanova University ***Assessment of Navy Contract Management Processes*** Rene Rendon, Associate Professor, NPS ***Designing and Managing Successful International Joint Development Programs*** Andrew Philip Hunter, Senior Fellow and Director, Defense-Industrial Initiatives Group, CSIS Gregory Sanders, Deputy Director and Fellow, Defense-Industrial Initiatives Group, CSIS Samantha Cohen, Research Assistant, Defense-Industrial Initiatives Group, CSIS

Reuben Pitts—is the president of Lyceum Consulting. He joined the Naval Weapons Lab in Dahlgren, VA, in June 1968 after graduating from Mississippi State University with a BSME. His early career was spent in ordnance design and weapons systems. He subsequently served on the planning team to reintroduce the Navy to Wallops Island, VA, currently a multiple ship combat, over-the-water weapons testing lab for Surface Ship Combat Systems, fighter aircraft, and live missile firings. His outstanding service as the deployed science advisor to commander, U.S. Sixth Fleet, was recognized with the Navy's Superior Civilian Service (NSCS) Award and the Navy Science Assistance Program Science Advisor of the Year Award.

Pitts was selected to lead the technical analysis team in support of the formal JAG investigation of the downing of Iran Air Flight 655 by USS *Vincennes*, and he participated in subsequent briefings to CENTCOM, the Chairman of the Joint Chiefs, and the Secretary of Defense. As head of the Surface Ship Program Office and Aegis program manager, Pitts was awarded a second NSCS, the James Colvard Award, and the John Adolphus Dahlgren Award (Dahlgren's highest honor) for his achievements in the fields of science, engineering, and management. Anticipating the future course of combatant surface ships, Pitts co-founded the NSWCDD Advanced Computing Technology effort, which eventually became the Aegis/DARPA-sponsored High Performance Distributed Computing Program, the world's most advanced distributed real-time computing technology effort. That effort was the foundation for the Navy's current Open Architecture Initiative. In 2003, Pitts accepted responsibility as technical director for PEO Integrated Warfare Systems (IWS), the overall technical authority for the PEO. In September of that year, he was reassigned as the major program manager for Integrated Combat Systems in the PEO. In this position, he was the program manager for the Combat Systems and Training Systems for all U.S. Navy Surface Combatants, including aircraft carriers, cruisers, destroyers, frigates, amphibious ships, and auxiliaries. In July 2006, Pitts returned to NSWCDD to form and head the Warfare Systems Department. While in this position, he maintained his personal technical involvement as the certification official for Surface Navy Combat Systems. He also served as chair of the Combat System Configuration Control Board and chair of the Mission Readiness Review for Operation Burnt Frost, the killing of inoperative satellite USA 193.

Pitts has been a guest speaker, lecturer, and symposium panelist at many NAVSEA-level and DoD symposiums and conferences and at the Naval Postgraduate School, the Defense Systems Management College, and the National Defense University. For 19 years, Pitts was the sole certification authority of all Aegis Combat System computer programs for fleet use. He retired from the U.S. Civil Service in September 2008, with over 40 years of service to the Navy.

Organization Analytics: Taking Cost-per-Dollar-Obligated (CPDO) Measures to the Next Level in Defense Contracting

Timothy Reed—is the Principal Director at Beyond Optimal Strategic Solutions (BOSS), an acquisition consulting and training firm. He has served as a professor at the Naval Postgraduate School (NPS) and the Air Force Institute of Technology (AFIT). His 21-year USAF career included contracting assignments at F-22 and C-17 Program Offices and serving as the Director of Joint Contracting Command in Kirkuk, Iraq. He also served at the Pentagon, responsible for implementing strategic sourcing for the DoD. He earned a PhD in Strategic Management/Entrepreneurship from the University of Colorado and was awarded lifetime recognition as a Certified Purchasing Manager (CPM) with the Institute of Supply Management.

James Keller—has served the DoD in various capacities for over two decades. He currently heads the OAD Business Center and is responsible for contracting, budgeting and finance, and studies system management. He was deployed to Afghanistan from 2013–2014, receiving CG MEB-Afghanistan commendation for his contribution to the retrograde and redeployment mission. Over the last decade, his analysis of Marine Corps issues has been instrumental in major force restructurings of personnel and equipment. Before coming to the Marine Corps, Keller performed a variety of analyses for the Army, Navy, and NASA.

John Fallon—has approximately 20 years of acquisition, acquisition training, and consulting experience with the federal government and the private sector. He has supported both defense and civilian agencies, and his experience covers the full spectrum of contracting and program management, including market research, services, supplies, major weapons systems, inter-agency acquisition teams, policy, strategic sourcing, and teaching. As the Vice President of Acquisition and Training Solutions at Vysnova Partners, he is currently leading the GSA PMO support of the OFPP-sponsored Category Management initiative. Dr. Fallon also serves as an Adjunct Professor for Villanova University (MPA program) and University of Maryland University College (master's program in Supply Chain Management).

Abstract

Procurement efficiency measures were calculated for nine defense contracting organizations over three fiscal years. Cost per Dollar Obligated (CPDO) efficiency assessments were completed for the Air Force, Navy, Marine Corps, and non-military staff defense organizations. Trends were identified in various U.S. regions and in different military services. Efficiency measures were then compared to performance measures for cycle time (Procurement Administrative Lead Time [PALT]) and compliance (protests received and sustained). Comparison of these measures and CPDO provides insight into the relationships between cost efficiency and the quality and timeliness organization workload completion. In addition, the demographic makeup of organizations was captured to identify identified relationships between the performance measures and organization size, proportion of contracting officer warrants, percentage of military personnel, and average civilian pay level. This study provides acquisition leaders with actionable insight regarding organization efficiency, performance, and staff composition. An emerging typology is identified indicating different types of contracting organizations based on the characteristics of the portfolio they execute.

Introduction

How efficient is your contracting organization? How efficient should it be? How efficient are similar contracting organizations? Does paying a 3% assisted acquisition fee provide good value for your organization? How can workforce design improve efficiency, timeliness, and compliance? If you answered "I don't know," you are not alone. It is difficult or impossible for most contracting leaders to answer these questions today due to the absence of essential information. In this study, we present benchmark findings that provide information useful to answering these questions and to meeting these and other procurement challenges.

Acquisition workforce performance measurement and workload assessment have been areas of study for at least 70 years (Monczka & Carter, 1978; McCampell & Slaich, 1995). However, a review of the government organization literature indicates that the question of workload assessment and organization efficiency have been given significantly less attention than output measurement, and that output measurement has been conducted primarily with overly broad and inappropriate measures such as dollars obligated and actions completed. Further, the preponderance of the workforce modeling activity has focused on (1) measuring the size of the macro organization (impacts of retirement, accessions, etc.), (2) measuring the descriptive statistics or demographics of the workforce, and, to a lesser degree, (3) attempting to measure the capabilities of the organization vis-à-vis competency assessments (Lamm & Reed, 2009). While these assessments present leaders with important pieces of information, they are incapable of answering the critical questions: “How much work will we need to do?” And “how efficiently can we expect to accomplish quality work in our organization?”

In 2010, Reed found that workload measurement in DoD contracting organizations is either performed inconsistently or not at all. This study measures contracting workload, organization efficiency in completing work, and benchmark comparisons; and it identifies opportunities to improve organization performance based on the research findings. We utilize Cost per Dollar Obligated (CPDO) as the assessment model to baseline organization workload and serve as a comparison with other similar organizations.

In essence, CPDO identifies the cost that an organization incurs while conducting its mission. The operating costs incurred are then compared to the total work accomplished by the organization. The resulting ratio is the CPDO, or the cost for the organization to obligate (and de-obligate) each dollar.

We completed CPDO assessments on contracting organizations over three years, in nine Air Force, Navy, Marine Corps, and Defense Management Agencies. This multi-service assessment allows us to compare trends in organization efficiency across services and in various U.S. regions. We also are able to compare performance in different types of contracting organizations grouped by the complexity of work in that organization.

While the efficiency benchmarks alone represent useful information for contracting leaders, we also measured performance measures for cycle time (Procurement Administrative Lead Time [PALT]) and compliance (protests received and sustained). Comparison of these measures with CPDO provides insight into the relationships between cost efficiency and the quality and timeliness organization workload completion.

Finally, we analyzed the demographic makeup of the participating organizations and identified relationships between CPDO and proportion of contracting officer warrants and percentage of military personnel.

This study provides acquisition leaders with actionable insight regarding organization efficiency, performance, and staff composition.

Methodology

The methods utilized to obtain each of the variables of interest are presented below.

Cost per Dollar Obligated (CPDO)

CPDO is a measure of how efficiently a procurement organization accomplishes its mission. CPDO captures the cost of operating the organization, and standardizes it with the amount of work accomplished by the organization. Past research in CPDO has found a range in procurement organizations (in both industry and government) from .002 to .05 (McCampell & Slaich, 1995, p. 34). While these numbers are interesting points of comparison, we were unable to determine the methodology utilized in those studies. It appears that burdened organization costs were not utilized. In this study, we did use fully burdened labor costs, which result in higher CPDO. We believe this will result in more accurate indicators of organization efficiency.

CPDO Methodology

In order to conduct a CPDO analysis, operating cost information was identified and captured. First, an examination of available operating expense data was required. In this study, we used fully burdened GS rates provided by OPM to account for the organization mission cost calculation. We also utilized standard military manpower labor rates. Second, a listing of all staff positions occupied during each of the study years, and the GS level and step for that position, as well as the grade and length of service for each military member. Midpoint or organization average was used for step determination if required.

The second portion of the CPDO calculation is the amount of work accomplished by the organization. Historically, organizations report the net value of their obligations, that is, obligations less de-obligations (funds removed from contracts). This traditional process fails to recognize the work involved in the de-obligation process, nor the work involved in zero dollar contract actions. The absolute value of de-obligations typically ranges from 5–15% of an organizations obligation total. In order to ensure all work of the organization was better accounted for, we calculated the absolute value of all obligations and de-obligations, and utilized the sum of those absolute values to identify the amount of work accomplished in each fiscal year (FY) by each organization. We acknowledge that using the absolute value of de-obligation actions may provide disproportionate credit for those actions (e.g., a one million dollar de-obligation action likely requires less work than a one million dollar initial contract award). However, we identified a large number of zero dollar actions in each data set. These zero dollar actions are often associated with necessary post-award contract administration activities. Using the traditional workload methodology, organizations receive no work credit for these actions. We believe using the absolute value for de-obligation actions accounts for the work required to accomplish the large number of zero dollar actions as well as the work required to complete the de-obligation contract action. Obligation data was collected from organization contracting writing and reporting data system archives (non-Navy) and from actual obligation reports (Navy).

PALT Data Element Development and Evaluation

PALT data was extracted from the official contracting writing systems for each non-Navy organization, and from verified Navy data sets. PALT is the number of days it takes from acceptance of a purchase request/requirement to the award of the contract/issuance of the modification. PALT is reported as the duration or number of days the process takes. From a customer perspective, higher PALT numbers indicates it takes relatively longer to complete the process. Lower PALT numbers (in comparison) indicate the process took less time to complete. We acknowledge that there is variability in the way that PALT is tracked in different organizations. This is due in part to a lack of awareness by leaders of the usefulness of the measure, and subsequently a lack of awareness by staff of the importance of accurate data entry. Despite this variability, we believe PALT to be the most useful measure of contracting process time available.

We found the PALT classification systems in use varied by organization. Many recorded PALT in only two categories: those below the simplified acquisition threshold (SAT) of $150,000, and those above the SAT of $150,000. While some organizations measure PALT for multiple types of contract actions (e.g., orders off existing contracts, GSA schedule contracts, modification actions, etc.), all of the benchmark organizations measured PALT above and below SAT. As such, in this study, we were limited to the use of PALT either "Above" or "Below" the SAT. Capturing PALT in these two categories is useful as the contracting processes required for actions below the SAT is much more streamlined and can in most cases be accomplished in a timelier manner.

Protest Data Element Development and Evaluation

The second category of performance measures collected reflected the number of protests received and sustained. This information was provided by each organization. As protests are high visibility items that must be reported up the contracting and command chain, the documentation of protests is robust. In this study, each organization reported the number of protests that were filed either with the organization or with the GAO. Further, the organizations reported the number of those protests that were upheld, meaning the protest was recognized as valid and the organization was directed to take action. Based on interviews with senior leaders, protests received and upheld were identified as potential proxies for the quality of work accomplished by the organization, as well as the adherence to laws, regulation, and policy.

The data reported regarding the number of protests sustained yielded a much smaller range, with only a handful of protests upheld across the entire sample. The vast majority of protests are either dismissed or have some sort of corrective action taken in lieu of a final decision. The largest number of protests sustained in any organization in any FY was two. Having a protest sustained is clearly an indication of a need for attention in an organization. However, the small number of protests sustained in this sample made it difficult to utilize this measure. Many protests and contractor concerns are addressed via other corrective actions. Such actions are not currently tracked in a consistent manner. We believe standardizing the methodology for tracking corrective actions after protest to be worth consideration as a quality measure going forward.

Personnel Descriptive Data

Our senior leader interviews suggested that the type and mix of contracting personnel was an area of interest. To gain visibility into this area, several demographic personnel variables were added. Specifically, the *average GS grade* for each organization, the *total number of staff,* the *number of non-contracting personnel,* the *ratio of contracting officers to specialists,* and the *ratio of civilian to military personnel* were captured. These variables were compared to CPDO to determine whether any relationships exist. We report those findings in which we found a significant relationship.

Benchmark Organization Group Comparisons

This study was designed to identify benchmarks for contracting organizations. Through a combination of researcher colleagues and senior leader introductions, the following list of comparison organizations was identified:

- USMC 1
- USMC 2
- USAF 1
- USAF 2
- USAF 3
- A Civilian Defense Contracting Agency
- USN 1
- USN 2
- USN 3

CPDO Results

This section presents summary information for the benchmark organizations related to their CPDO ratios over time. In the first graph (Figure 1), a fairly consistent cluster of CPDO results is depicted, with most organizations achieving between .005 and .025. The notable exception is USMC 2, which ranges from .05 to .07.

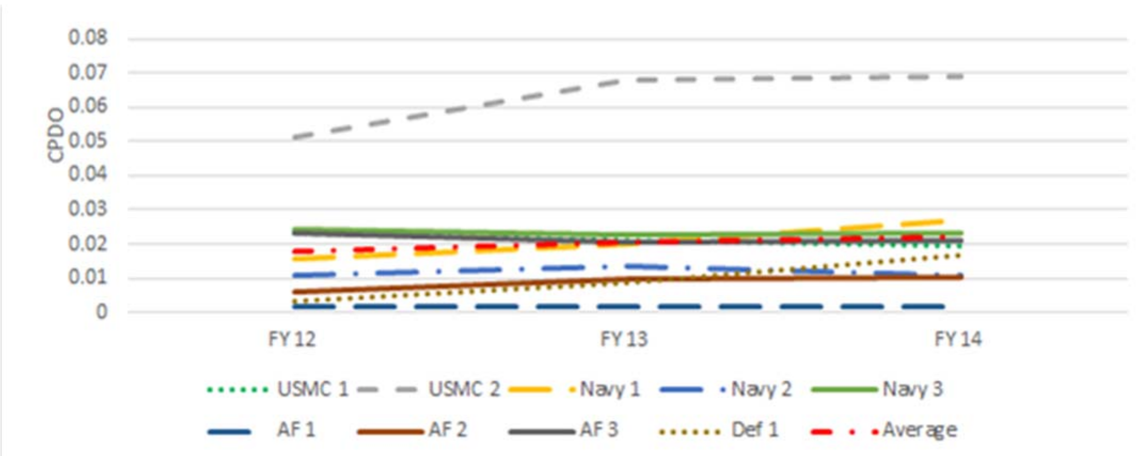


Figure 1. CPDO Results

We analyzed CPDO for each regional area represented by the benchmark organizations. The CPDO trend for all regions is up. The CPDO for organizations in the D.C. area (DC) and the rest of U.S. labor markets (ROUS) are plotted in Figure 2.

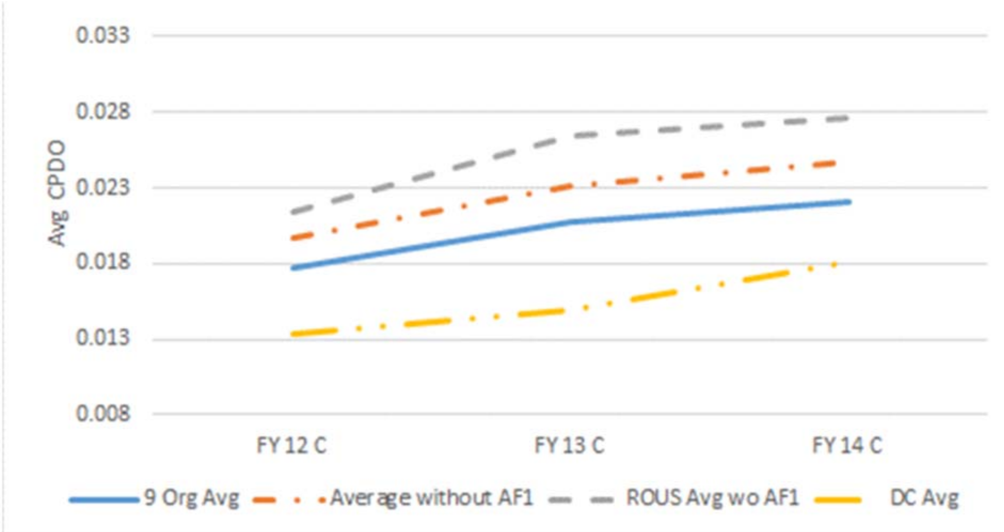


Figure 2. Average CPDO by Region

The study was designed to also analyze CPDO performance by service. Air Force CPDO averages .01, Navy .02, and USMC .045 (see Figure 3). While these results are from a small sample size, they identify differences in CPDO that warrant further examination in future research.

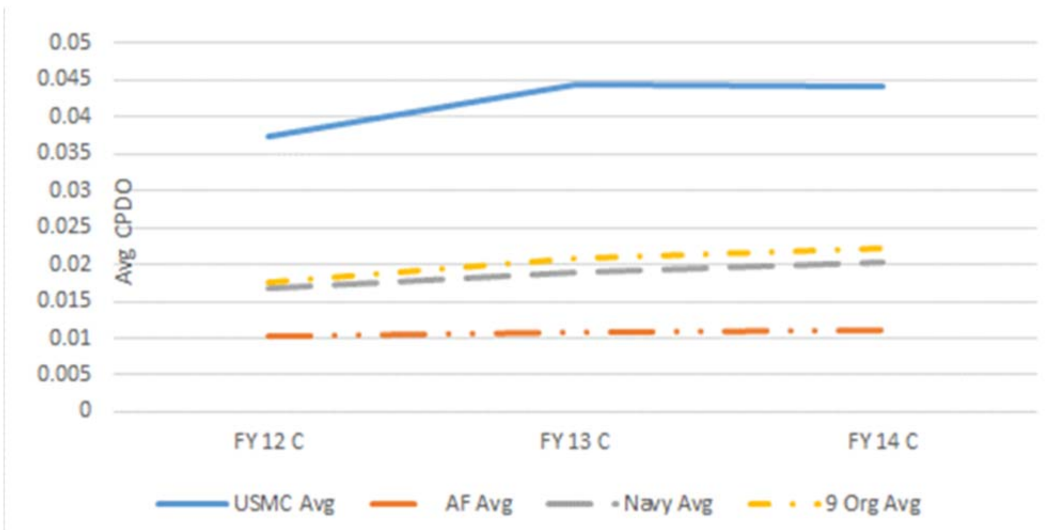


Figure 3. Average CPDO by Service

PALT

As discussed previously, PALT duration was identified by customers as the single most important performance measure. In this section, we present the benchmark comparisons for ASAT (over $150K) and BSAT (below $150K) contracting action PALT. The two charts that follow (Figures 4–5) show that USMC 1 and USMC 2 are the only two organizations with increasing PALT time durations in both BSAT and ASAT categories. All other organizations are either decreasing or flat.

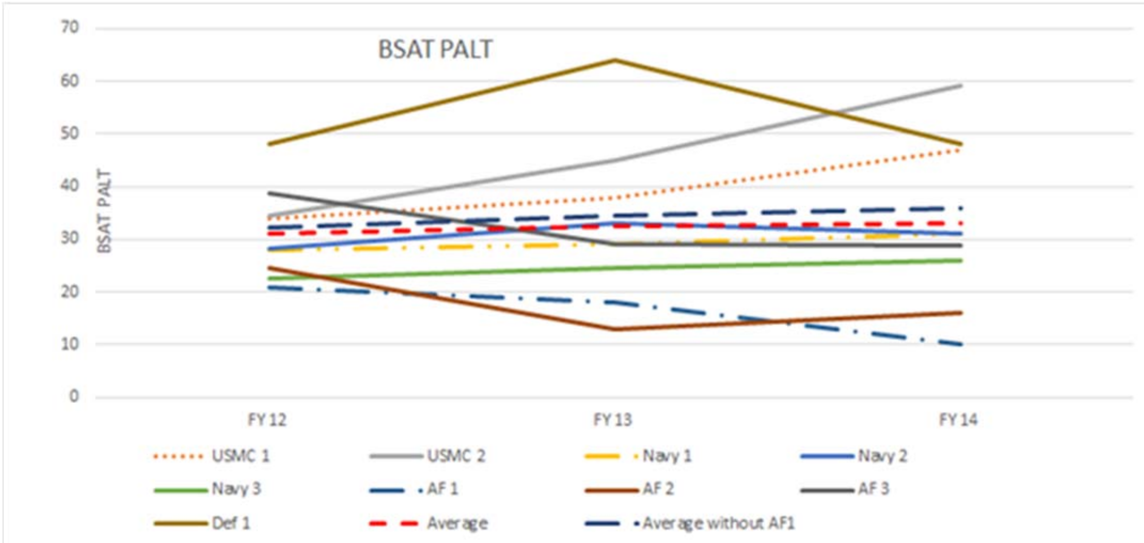


Figure 4. BSAT PALT

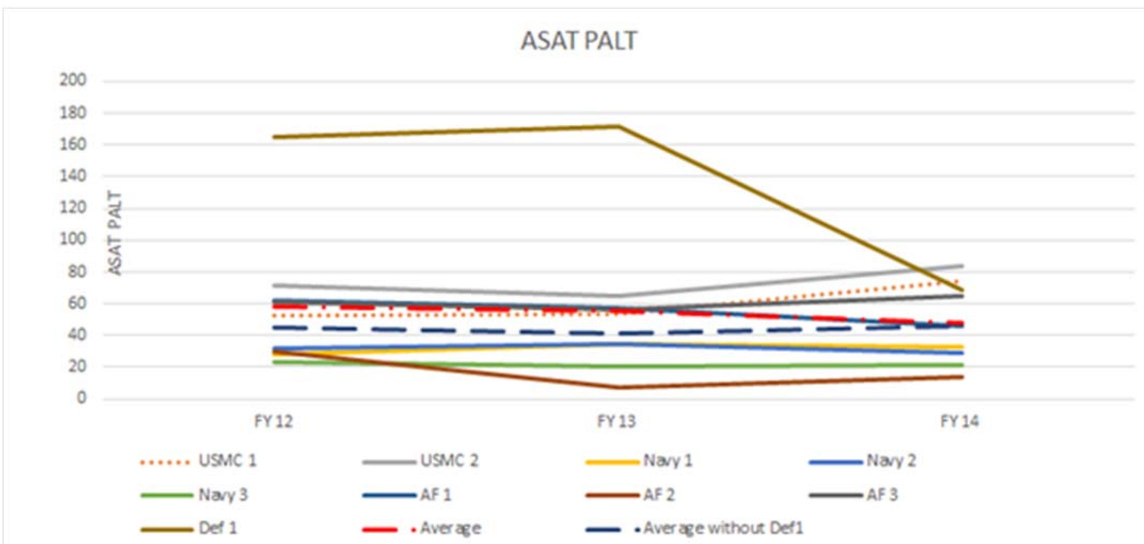


Figure 5. ASAT PALT

Relationship Between PALT and Other Variables

We analyzed the relationships between CPDO and PALT in order to identify correlations to provide insight into possible options for reducing PALT. In Figure 6, the analysis shows that BSAT PALT times rise as CPDO increases. This relationship may be a result of organizations dedicating new resources to high visibility ASAT requirements and staffing BSAT requirements with less-experienced, lower cost staff. Our analysis showed no consistent relationship between CPDO and ASAT PALT.

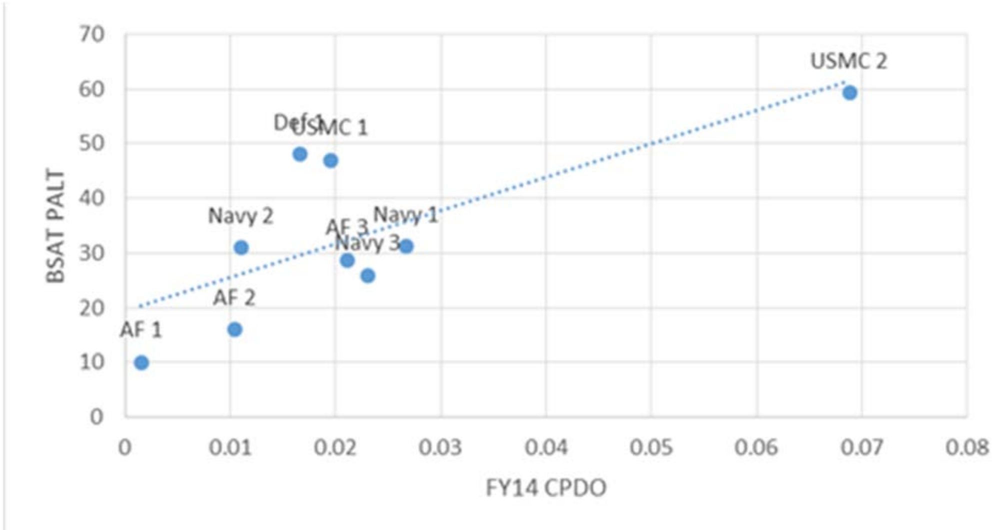


Figure 6. BSAT PALT by CPDO (FY14)

Statistical analysis of the relationship between CPDO and BSAT PALT for all benchmark locations over the three-year period confirms a significant relationship whereby BSAT PALT increases as CPDO increases.

$$PALT\ BSAT\ =\ 25.4\ +\ 341\ CPDO$$

$$(Significance\ Level\ =\ 98\%, df\ =\ 25) \quad (1)$$

Contracting Officer Warrants

We identified the number of core contracting personnel with contracting officer warrants at each benchmark organization. The roll of the contracting officer in completion of work is significant. The greater the number of contracting specialists assigned to each contracting officer, the more likely that there will be delays as contract documents queue waiting for contracting officer review. Contracting officer review delays extend PALT times and decrease customer satisfaction. We identified percentage of warranted contracting officers ranging from 24% to 91% of an organization's contracting staff.

Figure 7 depicts compelling relationships between higher percentages of warranted staff and lower CPDO. The graph shows the relationship between CPDO (along the horizontal axis) and contracting officer percentage (along the vertical axis). Our analysis found that CPDO decreases as the percentage of staff with warrants increase. A greater number of warrants results in the ability to complete contract actions in a more efficient manner, thus allowing the organization to accomplish more work with the resources allotted.

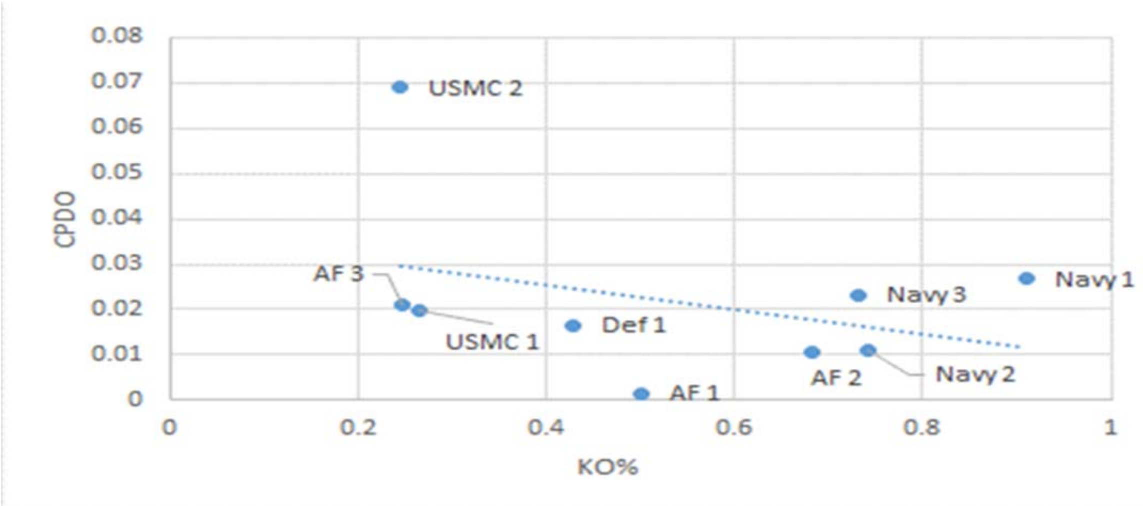


Figure 7. FY 14 CPDO and KO%

Statistical analysis of the relationship between CPDO and the percentage of contracting staff with warrants for all benchmark locations over the three-year period confirms a significant relationship whereby CPDO decreases as contracting officer percentage increases.

$$CPDO = 0.0337 + -0.0259\, Perc\ of\ contracting\ with\ warrants \qquad (2)$$

$$(Significance\ Level = 90\%, df = 25)$$

While not a focus of this paper, we also found a significant relationship between higher percentages of staff with warrants and reductions in both BSAT and ASAT PALT times. The relationship between increased warrant percentages CPDO and PALT times are noteworthy. Beyond the impact on processing time, we believe there is also a potential secondary impact of increasing the number of warrants in that it stimulates workforce development as contracting specialists strive to gain the knowledge and experience necessary to earn a warrant. In addition, contracting professionals may take more ownership of a contracting action when they know that they will be signing the document.

We recognize that limiting the number of warrants in a contracting organization is one strategy to mitigate risk. We suggest that an alternate way to mitigate risk is to develop a contracting workforce in which more professionals maximize their experience and knowledge as they pursue and earn warrants. Limiting the opportunity to obtain a warrant may have the unintended consequence of decreasing the motivation of specialists to maximize knowledge, and instead rely on the limited number of contracting officers to review and "grade" their work. Further, limited warrant opportunities may contribute to higher turnover as specialists see little chance of the goal of many contracting professionals—earning a warrant. In such a scenario, the best and brightest seeking such a goal will depart the organization and seek the opportunity elsewhere.

Protests

While protests are often identified as a potential measure of work quality, we question the use of protests as a performance measure. In this study, we did analyze the relationships between protests received and other variables. We found no significant relationship between changes in CPDO and the number of protests. The only significant relationship detected was between the total workload of an organization and the number of protests received. In Figure 8, the trend line show that protests rise as obligations increase in our benchmark sample. We identified a significant relationship in which an organization received an additional 3.24 protests for each billion absolute dollars obligated.

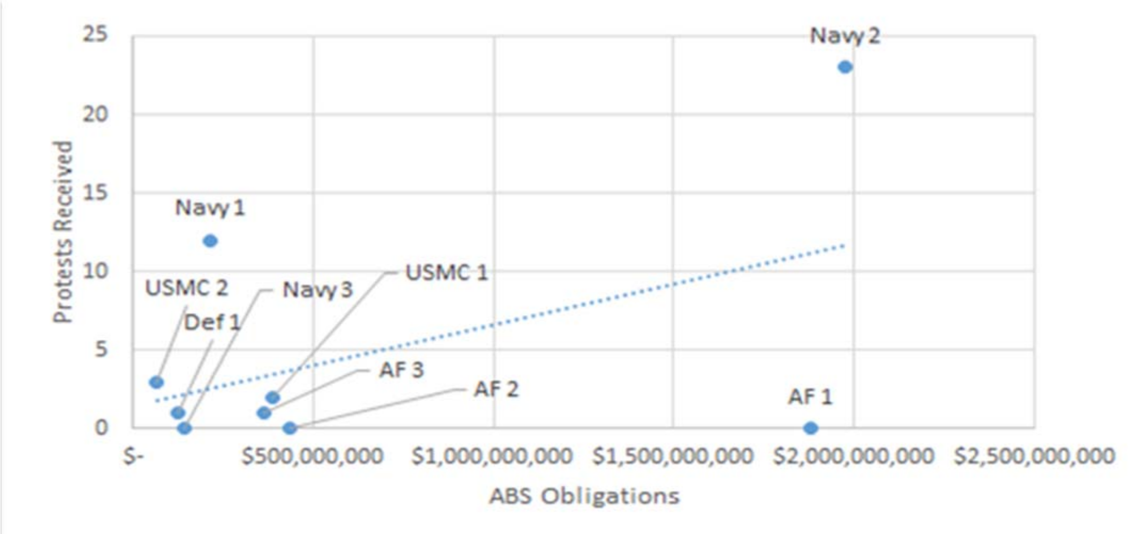


Figure 8. FY14 Protests and ABSO

$$Protests\ Received = 1.90 + 3.24\ ABS\ Obligations\ (\$B) \quad (3)$$

$$(Significance\ Level = 95.94\%, df = 25)$$

Civilian–Military Staff Mix

In this section, we explore the relationship between the civilian–military mix in a contracting organization and key performance measures. As discussed previously, the additional training and availability impact that having military personnel in a contracting organization has been frequently mentioned in our stakeholder interviews.

The following charts indicate that the higher the percentage of civilians in a contracting staff, the lower the CPDO (on the horizontal axis). This may indicate that organizations with lower percentages of military staff are able to focus more on contracting activities with less competition from military readiness demands.

Statistical analysis of the relationship between CPDO and the percentage of civilians on the contracting staff for all benchmark locations over the three-year period confirms a significant relationship whereby CPDO decreases as the civilian percentage increases.

$$CPDO = 0.0516 + -0.0372\ percent\ civ \quad (4)$$

$$(Significance\ Level = 95\%, df = 25)$$

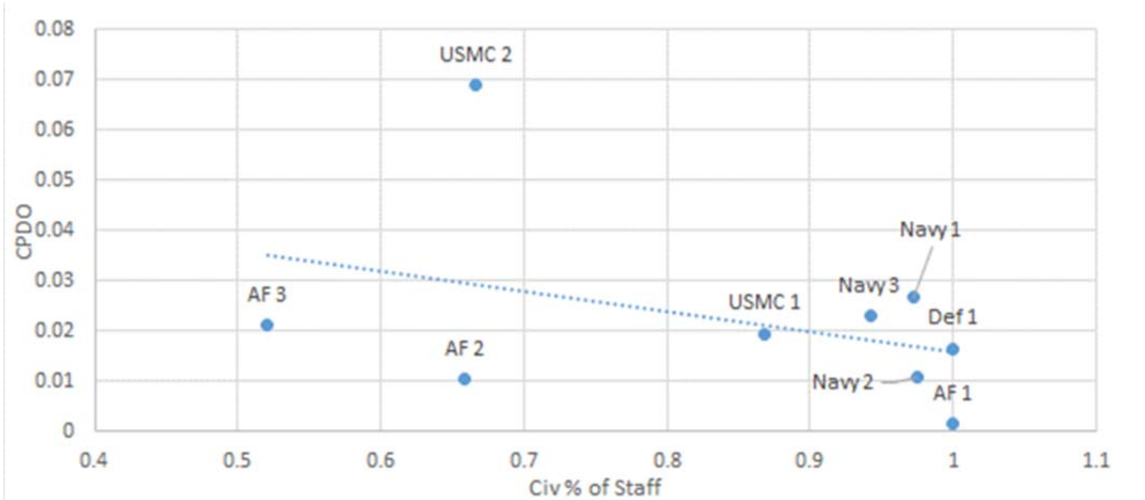


Figure 9. FY 14 CPDO and CIV%

The Emergence of a Contracting Organization Typology

As clusters of organizations began to emerge during our analysis, we noted that an organization portfolio typology facilitated comparison of organization performance to peer organizations with similar portfolios.

The two key characteristics that were utilized for this grouping were

- the percentage of actions accomplished by an organization that were below the SAT (actions lower than $150K)—the average of which was 74% for the benchmark group, and
- whether the median non-zero obligation action value was above or below the mean for the benchmark group ($54K).

Using these measures, we developed a 2x2 matrix and four potential contracting organization types. The distribution of our benchmark organizations is shown in Figure 10.

	Organization	Median value of contract action	Percentage of actions below the SAT
1	AF 1	$267K	46%
2	AF 2	$102K	71%
3	AF 3	$44K	79%
4	DEF 1	$24K	88%
5	USMC 1	$16K	90%
6	USMC 2	$12K	95%
7	Navy 1	$10K	73%
8	Navy 2	$22K	65%
9	Navy 3	$10K	69%
10	Average	$54K	74%

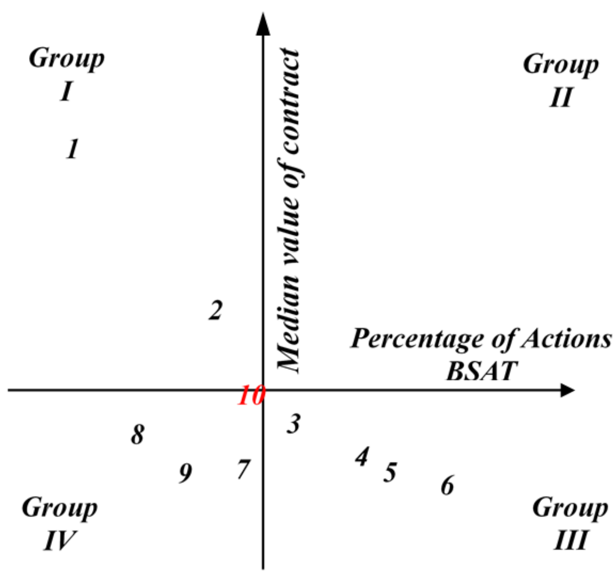


Figure 10. Peer Comparison Groups for Each Organization

Table 1. Peer Comparison Groups

Group I	Group II	Group III	Group IV
BSAT actions are less than 74% of all actions	BSAT actions are more than 74% of all actions	BSAT actions are more than 74% of all actions	BSAT actions are less than 74% of all actions
Median non-zero action value is higher than $54K	Median non-zero value is higher than $54K	Median non-zero value is less than $54K	Median non-zero value is less than $54K
These organizations do big dollar actions, and fewer BSAT actions than the average for the group. Requires special expertise to meet more complex actions	Big dollar actions, but still a higher level of BSAT actions than the rest of the group	Small dollar actions, and the vast majority of actions are BSAT. This organization should specialize in simplified acquisition procedures, organize to facilitate such actions, and implement warrants to support BSAT actions.	More of a mix of ASAT and BSAT actions than others in the sample. Requires a mix of contracting capability with a strong focus on the generally lower dollar levels for transactions. Of particular note, all the organizations in this group from our sample also used 1105 (procurement clerks) in their workforce staff.
AF 1, AF 2	No organizations were identified for this group in the benchmark study	AF 3, USMC 1, USMC 2, Def 1	Navy 1, Navy 2, Navy 3

The emerging typology allows for comparison of peer group organizations to within group averages on key performance measures. The average FY14 CPDO for each group is shown in Table 2.

Table 2. Benchmark Organization CPDO FY14

Organization	Group	CPDO FY 14
USMC1	3	.019
USMC2	3	.069
Navy 1	4	.027
Navy 2	4	.011
Navy 3	4	.023
AF 1	1	.002
AF 2	1	.010
AF 3	3	.021
DEF 1	3	.017
Average of Group 1 Peers	1	.006
Average of Group 3 Peers	3	.032
Average of Group 4 Peers	4	.020
Average of All	ALL	.022

Implications for Leaders and Practitioners

The contracting profession operates in a fast-paced, mission-critical environment and, as such, it is difficult to pause and consider changes and new ways of thinking. This has led to challenges and problems that remain largely the same over the past several decades. Whether it is PALT, resource constraints, or poorly written requirements, the retiring leaders of today are passing these same challenges to the millennials that will replace them. It is our collective opinion that the current environment is ripe for new analysis and thinking to better learn from one another to finally tackle and perhaps alleviate some of these decades-old challenges.

Comparing Contracting Organizations

Prior to illustrating the usefulness and applicability of the aforementioned study and related analysis, it is first useful to overcome the frequently held notion that each contracting organization is too unique for a comparative analysis. The claims that "My organization is not like the typical contracting shop," or "Our mission makes it impossible to benchmark our statistics with other organizations," or "What we do is so unique that I need an analysis independent of any other organization" have existed since the inception of the contracting profession and likely have never been as invalid as they are today. Significant efforts to streamline the profession regardless of the goods and services being procured have increased the similarity of the contracting profession across agencies and departments. For example, standardized contracting writing systems, government wide e-gov systems (e.g., FPDS, FBO, EPLS, etc.), heightened transparency, the increase in shared services, strategic sourcing, and, more recently, category management have all had a profound impact on making the contracting profession more uniform and a lot less "unique." Further validation that contracting organizations do lend themselves to a comparative analysis is offered in Table 1. This table uses the standard characteristics of median value of contract actions along with the transactions relative to the Simplified Acquisition Threshold, both common benchmarks in the contracting profession. The result of this analysis is that each of the nine participating organizations do indeed have peers when analyzed through the lens of these commonly accepted characteristics. In short, contracting organizations are not so unique that they cannot be compared to other, similar contracting organizations. Regardless of an agency or department's overall mission, there are enough common traits and characteristics to make a comparative analysis not only worthwhile but, in today's environment of data transparency and never ending budgetary challenges, essential.

CPDO Hypothesis, Insights, and Practical Applications

We posit that a comparative analysis is a worthwhile endeavor and that such comparisons offer practical application for contracting leaders.

BSAT PALT and CPDO

This particular comparison illustrated a direct correlation between Below the SAT PALT and CPDO. This resulting data offers that the higher the CPDO, the lengthier the BSAT PALT. At first glance, this seems to contradict conventional wisdom as one would logically assume that the higher CPDO, which may be driven by additional resources, would result in a shorter lead time (PALT), regardless of whether it was above or below the SAT. For BSAT, this analysis clearly illustrates a pattern that more resources does not equate to a decreased PALT for BSAT actions. In short, the old claims that "I need more resources if you want your PRs processed timely" are likely the wrong course of action, at least for BSAT contracting actions. Perhaps the additional resources were aimed towards ASAT contracting actions and the BSAT actions were secondary priorities for these commands with a higher CPDO thus leading to longer PALT durations.

Warranted Contracting Officer Percentage and CPDO

This particular comparison illustrated a direct correlation between the percentage of warranted contracting officers and CPDO. As the number of warranted contracting officers increased for each agency in the study, the respective CPDO of these agencies decreased. The implication here is straightforward: As warranted contracting officers are increased, the CPDO in that agency decreases. This has large implications for rightsizing staffing and how to approach warrant related policies, both important endeavors for contracting leaders. A larger number of warrants also implies that the related PALT should decrease, as there is an increase in the abilities of the organization to complete contract actions in a more efficient manner. In short, more warranted resources to complete actions translates into more work being accomplished with the resources allotted. This particular analysis and related findings offers a significant proposition that warrants further study as resource constraints and how to properly staff and right-size the workforce have been ongoing initiatives for decades in the contracting workforce with little to no agreement across agencies on how to move forward.

Admittedly, our analysis and research did not incorporate warrant levels (at or below SAP, certain dollar thresholds, etc.), the relationship between increased warrants and quality of work produced (measured by protests and/or other quality variables), impact on risk, and so on. These are all areas that demand further exploration. Conversely, examining the impact of increased warrant levels on employees ownership of their work given that they now sign the contracts, who receives a warrant (e.g., warrants are typically earned by the high performers in the organization), offers hypotheses that speak to the potentially positive outcomes related to workforce satisfaction in additional research.

Military Staff and CPDO

Many DoD contracting organizations have a blend of military and civilian staff. Like all things, this hybrid approach offers a myriad of pros and cons. For purposes of this discussion, the linkage between military and civilian staff and CPDO is evident in our study. The higher the percentage of civilians in a contracting staff, the lower the CPDO. This correlation/finding offers insight into the aforementioned importance of rightsizing an organization, thinking through the true costs of managing a hybrid organization, and trying to assess the proper mix of personnel.

Hypotheses stemming from this finding include that perhaps military readiness demands and other assigned duties detract from the position's primary focus of awarding contracts thereby requiring additional resources to backfill the military positions. Additionally, perhaps the constant turnover of military staff impacts training and organizational efficiencies thereby negatively impacting the agency's CPDO. Given that the organizations studied that offer nearly 100% civilian staff portray a range of PALT data, it is premature to add this critical variable into the discussion and further validates the need for additional research.

Using CPDO Moving Forward

This study and resulting analysis offers that CPDO can indeed be a useful tool in assisting leadership in how to properly structure contracting organizations as well as impact their efficiency. The implications and potential impact should not be taken lightly given that this particular workforce is responsible for executing the largest buying entity on the globe and doing so in an environment that offers little to no budgetary relief accompanied by unprecedented levels of scrutiny. While the various hypotheses beg for additional research, this study offers an encouraging and worthy starting point.

The final point to offer regarding CPDO is its rising importance in the current environment of shared services, fee-for-service organizations, federal-wide strategic sourcing and inter-agency agreements, and, most importantly, the OFPP sponsored category management initiative. As the government strives to "buy as one" and harness its collective bargaining power through centers of excellence and government-wide categories, leadership across the acquisition community (e.g., CAOs, CFOs, CIOs, Management Bureau leads, etc.) will all be keenly interested in how efficient the contracting organization is that is receiving government-wide funds. Prior to the DoD sending billions of dollars to the GSA for a category management initiative, an essential question that should be posed is what is the cost of the GSA's procurement activities compared to our own? While there are numerous, influencing variables that would inevitably find its way into this discussion, CPDO remains at the heart of the start of the conversation. Adding in PALT, warrants, and other variables mentioned above, the discussion becomes more sophisticated, leading to potentially sound, fact-based decisions that will inevitably produce not only a more efficient and effective workforce but, more broadly, savings to the taxpayer and a better use of the limited available dollars to support the warfighter.

Future Research

This study identified many opportunities for future research.

First, while the study sample does cover multiple services, the study sample size is small. The number of benchmark organizations should be increased to include

- additional Washington, DC, based organizations to confirm the lower CPDO identified for that area in this study,

- additional U.S. Army locations to complete the service comparison, and
- additional Group III organizations to further define the group of most significant interest to operational and base support contracting organizations.

Second, the significant differences between organizations on CPDO and other measures should be further studied. The differences may be attributable to the high percentage of "Below the SAT" transactions or other portfolio characteristics.

Third, the differences between each services' CPDO should be further assessed. Are there service policies or procedures that can be identified and leveraged by other organizations? Or are the differences driven primarily by portfolio type?

Fourth, the current study used a count of warrants to determine warrant percentage in each organization. Further study should be accomplished on the type and dollar level of warrants utilized by various organizations to provide a general roadmap of the most effective designation of warrants.

Fifth, the military–civilian mix in contracting requires more research—the benchmark sample indicates a significant relationship (e.g., 100% civilian organizations reduce CPDO from .051 to .014).

Finally, further analyze organization portfolio (percentage of actions that are task orders, delivery orders, full contracts, basic vehicles, etc.). A next level analysis of execution practices will provide insight into further optimizing CPDO and PALT.

This research provides insight into multiple uses of CPDO and other measures to optimize contract awards and meet the needs of procurement customers more effectively. Extending this research to a larger sample and with greater visibility into specific portfolio components will increase the precision of the findings and enhance the decision making of leaders throughout the contracting community.

References

Lamm, D., & Reed, T. (2009). *Demographics of the contracting workforce within the Army Contracting Command* (NPS-CM-09-140). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

McCampell, A., & Slaich, L. (1995). Purchasing efficiency and staffing benchmarks: A county government study. *Journal of Supply Chain Management, 31*(1), 29–36.

Monczka, R., & Carter, P. (1978). Measuring purchasing performance. *Management Review, 67*(6), 27–42.

Reed, T. (2010). *Army Contracting Command workforce model analysis* (NPS-CM-10-179). Monterey, CA: Naval Postgraduate School, Acquisition Research Program.

Appendix A. Acronyms

Acronyms

ASAT	Above the Simplified Acquisition Threshold (for this study, above $150,000)
BOSS	Beyond Optimal Strategic Solutions, the principal investigators for this study
BSAT	Below the Simplified Acquisition Threshold (for this study, below $150,000)

CPDO	Cost per Dollar Obligated, a measure of efficiency calculated by dividing the organization operating expense (cost) by the absolute value of obligations (work)
FY	Fiscal Year
GS	General Schedule, category of government civilian workforce
OPM	Office of Personnel Management
PALT	Procurement Administrative Lead Time, the duration of time required to accomplish a contracting action
PD2	Procurement Desktop Defense, the contract-writing system utilized by many contracting agencies to create and track contracting actions, also referred to as SPS
PR	Procurement Request, a form submitted by a requiring agency stating what needs to be purchased and providing documentation that funds are available
SAT	Simplified Acquisition Threshold, a threshold (for this study $150,000) below which streamlined, or simplified acquisition procedures are utilized to award contracts
SPS	Standard Procurement System, see PD2

Assessment of Navy Contract Management Processes

Rene G. Rendon—is an Associate Professor at the Graduate School of Business and Public Policy, NPS, where he teaches defense acquisition and contract management courses. He also serves as the Academic Associate for the MBA specialization in contract management. Prior to joining the NPS faculty, he served for over 20 years as an acquisition contracting officer in the United States Air Force. His career included assignments as a contracting officer for the Peacekeeper ICBM, Maverick Missile, and the F-22 Raptor. He was also a contracting squadron commander and the director of contracting for the Space Based Infrared Satellite program and the Evolved Expendable Launch Vehicle rocket program. Rene has published in the *Journal of Public Procurement*, the *Journal of Contract Management,* the *Journal of Purchasing and Supply Management*, and *the International Journal of Procurement Management*. [rgrendon@nps.edu]

Abstract

This research builds upon the emerging body of knowledge on contract management workforce competence and organizational process capability. In 2003, the Contract Management Maturity Model (CMMM) was first developed for the purpose of assessing Department of Defense (DoD) and defense contractor organizational contract management process capability. The CMMM has been previously applied at Air Force, Army, Navy, and defense contractor organizations. Specific to the Navy, assessments were conducted at three Navy contracting centers using the CMMM. These organizations included the Naval Air Systems Command (NAVAIR), Naval Sea Systems Command (NAVSEA), and the Naval Supply Systems Command (NAVSUP). The primary purpose of this paper is to summarize the assessment ratings, analyze the assessment results in terms of contract management process maturity, and discuss the implications of these assessment results for process improvement and knowledge management opportunities. This paper also provides insight on consistencies and trends from these assessment results to DoD contract management. Finally, this paper discusses these assessment results in an attempt to characterize the current state of practice of contract management within the U.S. Navy.

Background

In fiscal year (FY) 2015, the Department of Defense (DoD) awarded over $242 billion in contracts for mission-critical supplies and services. These contract obligations were executed through approximately two million contractual actions. Within the Navy, over $76 billion were obligated in the execution of over 220 thousand contractual actions (USA Spending, 2016). The amount of dollars obligated on contracts reflects the importance of the contract management function within the DoD and requires high levels of accountability, integrity, and transparency in its contracting processes. However, the Government Accountability Office (GAO) continues to identify DoD contract management as a high risk to the federal government due to the lack of skills and capabilities of the acquisition workforce, management and oversight of contracting processes and approaches, management of services acquisition, and need for improvement in operational contracting support (GAO, 2015). Additionally, the DoD inspector general (DoDIG) has identified deficiencies in the DoD agency's poor contract planning, contract administration, and contractor oversight (DoDIG, 2009, 2012, 2013, 2014).

The DoD's response to the GAO's high-risk rating and the DoDIG reported deficiencies include an increased hiring of contracting specialists and auditors, increased contracting training requirements, and an emphasis on individual competency assessments to identify contracting workforce skills and abilities (GAO, 2015). Additionally, the DoD has implemented a series of Better Buying Power initiatives outlining the steps needed to achieve better contracting results (Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics [OUSD[AT&L]), 2014). Thus, the DoD's approach to resolving its contracting deficiencies has been to focus only on increasing the contracting workforce and improving the competence of that workforce. What is missing from the DoD's response to its contracting deficiencies is an emphasis on organizational process maturity, specifically, contracting process capability. Auditability theory (Power, 1996, 2007; Rendon & Rendon, 2015) states that organizations also need capable processes and effective internal controls, in addition to workforce competence, to ensure mission success. Based on this author's experience, many of the DoD's contracting deficiencies are rooted more in the lack of organizational process capability, and less on the competence of the contracting workforce.

Research Scope and Objectives

This paper presents the results of process capability assessments for the U.S. Navy's contract management processes using the Contract Management Maturity Model (CMMM). The CMMM is used to assess an organization's contract management process capability and to develop a roadmap for implementing improvement initiatives for the contract management process. Using the Web-based survey assessment tool, the CMMM was applied to three Navy contracting agencies: Naval Air Systems Command (NAVAIR), Naval Sea Systems Command (NAVSEA), and the Naval Supply Systems Command (NAVSUP). The purpose of this paper is to summarize the assessment ratings, analyze the assessment results in terms of contract management process maturity, and discuss the implications of these assessment results for process improvement and knowledge management opportunities. The assessment results and related recommendations for contract management process improvement and knowledge management opportunities are proposed to the U.S. Navy for developing a road map for increasing contract management process capability. A thorough understanding of the Navy's current level of contract management process capability will help these organizations improve their procurement of defense-related supplies and services. This research also discusses the process assessment results by providing insight on consistencies and trends in an attempt to characterize the current state of practice of contract management within the U.S. Navy, as well as the DoD.

Research Method

This research is based on the application of the Contract Management Maturity Model (CMMM) for the assessment of organizational contract management processes. The CMMM was developed and validated in 2003 and subsequently applied to other defense contracting organizations (Garrett & Rendon, 2005; Rendon, 2003, 2008). The CMMM assessment tool is a Web-based survey comprised of 62 items related to each of the six contract management key process areas (approximately 10–11 items per key process area). See Appendix A for a description of the six contract management process areas The survey items use a Likert scale–option response with associated numerical values from 5 (Always) to 0 (I Don't Know). These options represent the organization's use of specific contract management best practices, as reflected in the acquisition and contract management literature. These best practices relate to contract management process strength, successful outcomes, management support, process integration, and process measurement. The numerical value associated with the responses to the CMMM survey items are then calculated to determine the process maturity level for each of the contract management processes. The CMMM designates process maturity levels ranging from Level 1 (Ad Hoc) to Level 5 (Optimized). See Appendix B for a description of each process maturity level.

The CMMM uses a purposeful sampling method designed to acquire data on organizational contract management processes. Purposeful sampling ensures that population samples are knowledgeable and informative about the phenomena being researched, thus increasing the utility of the information obtained from small samples (Creswell, 2003; McMillan & Schumacher, 2001). Thus, the survey is only deployed to warranted contracting officers and fully qualified contract specialists. The sampling in this research consisted of agency employees designated either as warranted contracting officers or as individuals that were considered fully qualified in the government contracting career field, in accordance with the Defense Acquisition Workforce Improvement Act (DAWIA). Warranted contracting officers are those individuals that have specific authority to enter into, administer, or terminate contracts and make related determinations and findings on behalf of the U.S. government (FAR, 2015). Full qualification in the contracting career field is interpreted to mean achievement of at least Level 2 certification in contracting under DAWIA. Level 2 certification requires completion of a baccalaureate degree with at least 24 semester hours in accounting, law, business, finance, contracts, purchasing, economics, industrial management, marketing, quantitative methods, and organization and management coursework; two years of contracting experience; and completion of the required contract training courses (DAWIA, 1990).

Results

The CMMM survey link was e-mailed to the directors of contracting for the specific agencies, and the link was then forwarded to the eligible contracting personnel. Reminder e-mails were sent approximately two weeks into the survey period. The survey instrument included the appropriate provisions for confidentiality and the protection of human subjects. Of the 369 eligible survey participants, 185 Navy contracting officers completed the survey, generating a response rate of approximately 50%.

Descriptive statistics were applied on the survey results, including a factor analysis to determine if the survey items closely correlated with questions designed to operationalize each of the contract management process areas. The factor analysis identified groupings of highly correlated survey items based on the survey responses. The results of the factor analysis indicated that the survey items related to each of the six contracting process areas loaded together (0.6 and above). (In factor analysis, factor loadings represent how much a factor explains a specific variable. Loadings can range from -1 to1. Loadings close to -1 or 1 indicate that the factor strongly affects the variable, either negatively or positively. Loadings close to zero indicate that the factor has a weak effect on the variable.) Based on the factor analysis, operationalized variables were created and used to perform reliability tests using Cronbach's α for each of the operationalized variables. As reflected in Table 1, the results of the reliability test indicated Cronbach's α value for each of the six key contracting process areas ranging from 0.91 to 0.94. These reliability coefficients are above 0.80, and thus, the survey instrument is considered to have high reliability and internal consistency (McMillan & Schumacher, 2001).

Table 1. Descriptive Statistics for the Contracting Process Area Scale Factors

Contracting process area scale factor	No. of items	*M* (*SD*)	Valid *N*	Cronbach's α
Procurement Planning	10	3.79 (.88)	185	.91
Solicitation Planning	10	3.74 (.87)	178	.92
Solicitation	10	3.61 (.93)	174	.92
Source Selection	11	3.85 (.90)	172	.93
Contract Administration	11	3.37 (1.03)	169	.94
Contract Closeout	10	2.46 (1.59)	168	.94

The Navy CMMM assessment results are reflected in Table 2, which lists the contract management process area, survey item number, and item process maturity enabler. Table 2 also shows the mean responses for each survey item, the standard deviation for each survey item, and the total number of responses for each survey item. The mean responses are based on the Likert scale's numerical value range from 5 (Always) to 1 (Never) and 0 (I Do not know) for each survey item in each contract management process area.

Table 2. U.S. Navy CMMM Assessment Results

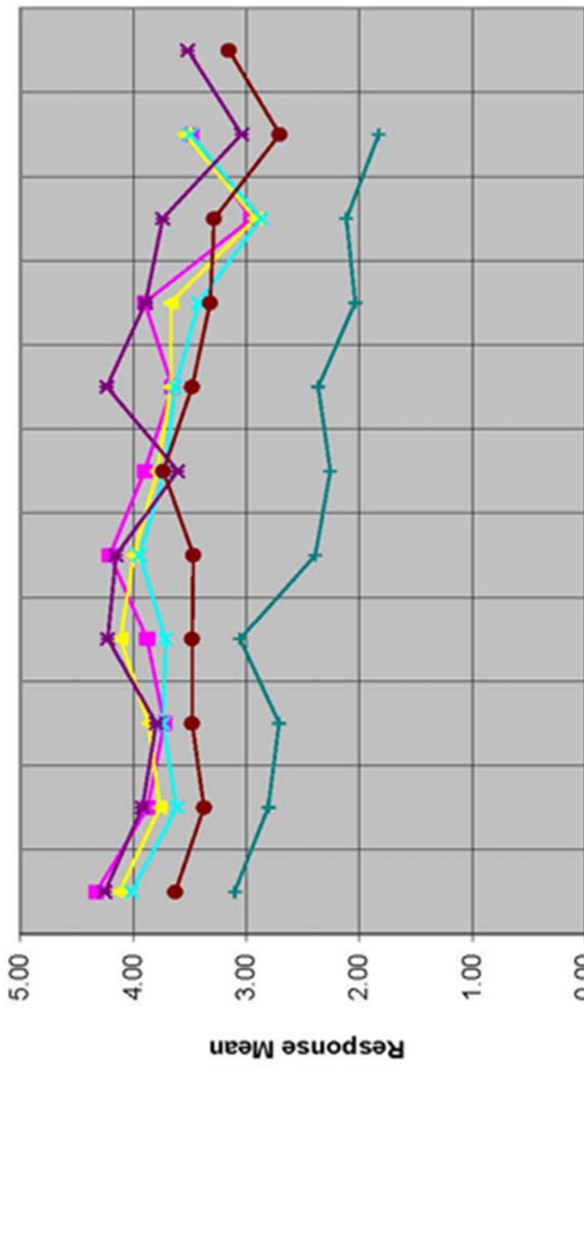
Key Process Area/Item Number and Description	Mean	SD	*n*
Procurement Planning			
1.1 Process Strength	4.32	1.04	187
1.2 Process Strength	3.87	1.28	187
1.3 Process Strength	3.72	1.13	187
1.4 Process Results	3.88	1.08	187
1.5 Management Support	4.21	1.00	187
1.6 Process Integration	3.90	1.13	187
1.7 Process Integration	3.65	1.21	187
1.8 Process Integration	3.90	1.12	187
1.9 Process Measurement	2.95	1.65	187
1.10 Process Measurement	3.49	1.15	187
Total	37.89		
Solicitation Planning			
2.1 Process Strength	4.12	1.09	180
2.2 Process Strength	3.76	1.31	180
2.3 Process Strength	3.87	1.17	180
2.4 Process Results	4.11	0.94	180
2.5 Management Support	3.99	1.03	180
2.6 Process Integration	3.79	1.07	180
2.7 Process Integration	3.67	1.14	180
2.8 Management Support	3.67	1.04	180
2.9 Process Measurement	2.92	1.65	180
2.10 Process Measurement	3.54	1.22	180
Total	37.44		
Solicitation			
3.1 Process Strength	4.01	1.22	176
3.2 Process Strength	3.61	1.43	176
3.3 Process Strength	3.74	1.29	176
3.4 Process Results	3.71	0.92	176
3.5 Management Support	3.94	1.03	176
3.6 Process Integration	3.72	1.14	176
3.7 Process Integration	3.63	1.12	176
3.8 Process Integration	3.42	1.11	176
3.9 Process Measurement	2.87	1.65	176
3.10 Process Measurement	3.49	1.20	176
Total	36.14		

Source Selection			
4.1 Process Strength	4.25	1.03	174
4.2 Process Strength	3.92	1.23	174
4.3 Process Strength	3.80	1.20	174
4.4 Process Results	4.23	1.04	174
4.5 Management Support	4.15	1.04	174
4.6 Process Results	3.60	1.17	174
4.7 Process Results	4.23	1.04	174
4.8 Process Integration	3.89	1.20	174
4.9 Process Integration	3.74	1.25	174
4.10 Process Measurement	3.04	1.71	174
4.11 Process Measurement	3.52	1.26	174
Total	42.37		
Contract Administration			
5.1 Process Strength	3.63	1.28	171
5.2 Process Strength	3.37	1.32	171
5.3 Process Strength	3.48	1.25	171
5.4 Process Results	3.48	1.16	171
5.5 Management Support	3.47	1.25	171
5.6 Process Integration	3.73	1.12	171
5.7 Process Integration	3.48	1.20	171
5.8 Process Integration	3.32	1.31	171
5.9 Process Integration	3.28	1.67	171
5.10 Process Measurement	2.70	1.66	171
5.11 Process Measurement	3.15	1.39	171
Total	37.10		
Contract Closeout			
6.1 Process Strength	3.10	1.82	170
6.2 Process Strength	2.80	1.89	170
6.3 Process Strength	2.71	1.86	170
6.4 Process Results	3.05	1.99	170
6.5 Management Support	2.39	1.82	170
6.6 Process Integration	2.26	1.87	170
6.7 Process Integration	2.36	1.86	170
6.8 Process Measurement	2.04	1.85	170
6.9 Process Measurement	2.11	1.81	170
6.10 Process Measurement	1.83	1.76	170
Total	24.65		

The survey item mean responses were totaled, and the resulting score was then converted to its associated process maturity level. Figure 1 reflects the process maturity level for each contract management process area based on the assessment results. Figure 2 reflects the comparison of survey item mean scores for each contract management process.

CONTRACT MANAGEMENT MATURITY MODEL©						
MATURITY LEVEL	PROCUREMENT PLANNING	SOLICITATION PLANNING	SOLICITATION	SOURCE SELECTION	CONTRACT ADMIN	CONTRACT CLOSEOUT
5 OPTIMIZED						
4 INTEGRATED						
3 STRUCTURED	N	N		N		
2 BASIC			N		N	N
1 AD HOC						

Figure 1. U.S. Navy CMMM Maturity Levels



	1	2	3	4	5	6	7	8	9	10	11
Procurement Planning	4.32	3.87	3.72	3.88	4.21	3.90	3.65	3.90	2.95	3.49	
Solicitation Planning	4.12	3.76	3.87	4.11	3.99	3.79	3.67	3.67	2.92	3.54	
Solicitation	4.01	3.61	3.74	3.71	3.94	3.72	3.63	3.42	2.87	3.49	
Source Selection	4.25	3.92	3.80	4.23	4.15	3.60	4.23	3.89	3.74	3.04	3.52
Contract Administration	3.63	3.37	3.48	3.48	3.47	3.73	3.48	3.32	3.28	2.70	3.15
Contract Closeout	3.10	2.80	2.71	3.05	2.39	2.26	2.36	2.04	2.11	1.83	

Figure 2. Comparison of Contract Management Process Survey Item Mean Scores

Discussion

The analysis of the CMMM assessment results can be discussed from the perspective of process capability maturity and process capability enablers. Process maturity is discussed first.

As reflected in Figure 1, the contracting process areas of Procurement Planning, Solicitation Planning, and Source Selection are rated at the Structured level of process maturity. This maturity level indicates that for these process area activities (see Appendix A) the processes are fully established, institutionalized, and mandated throughout the entire organization. These processes are supported by formal documentation and some processes may even be automated. Furthermore, the organization allows for the tailoring of these processes and documents in consideration for the unique aspects of each contract, such as contracting strategy, contract type, terms and conditions, dollar value, and type of requirement (product or service). Finally, senior organizational managers are involved in providing guidance, direction, and even approval of key process area strategy, decisions, and documents.

However, since these process areas are rated at only the Structured level, the assessment results also show that these processes are not fully integrated with other organizational processes that are part of the organization's contract management effort, such as financial management, schedule management, performance management, and technical management. Additionally, for these specific processes, the procurement team does not include representatives from other functional areas nor does it include the contract requirement end-user.

Also reflected in Figure 1, the contracting process areas of Solicitation, Contract Administration, and Contract Closeout are rated at the Basic level of process maturity. This indicates that for these process area activities (see Appendix A), some contract management processes have been established, but these processes are required only on selected contracts. Furthermore, there is no organizational policy establishing the consistent use of these processes and standards on all contracts awarded by the organization. Finally, although there may be some documentation of these processes and standards, not all processes are fully documented throughout the organization.

However, since these specific process areas are rated at the Basic level, the assessment results also show that these specific processes are not fully established, institutionalized, and mandated throughout the entire organization. Additionally, these processes are not supported by formal documentation nor are there any automated processes for these activities. Lastly, senior organizational managers are not involved in providing guidance, direction, or approval of key process area strategy, decisions, and documents.

As previously stated and reflected in Table 2, each CMMM survey item is associated with one of the five process capability enablers. These process capability enablers are Process Strength, Process Results, Management Support, Process Integration, and Process Measurement. As reflected in Table 2, the Navy's process areas with the highest scoring survey response means for Process Strength–associated survey items were in the process areas of Procurement Planning, Solicitation Planning, and Source Selection. These results indicate a stronger use of Process Strength best practices such as ensuring standardized, mandatory, and documented processes. Additionally, as reflected in Table 2, the Navy's process areas with the lowest scoring survey response means for Process Strength–associated survey items were in the process areas of Contract Administration and

Contract Closeout. These results indicate weaker use of Process Strength best practices in these specific contract management process areas.

As reflected in Table 2, the Navy's process areas with the highest scoring survey response means for Process Results–associated survey items were in the process areas of Source Selection. These results indicate a stronger use of Process Results best practices in ensuring appropriate evaluation standards and criteria and in maintaining integrity in the proposal evaluation process. Additionally, the Navy's process areas with the lowest scoring survey response means for Process Results–associated survey items were in the process areas of Contract Administration and Contract Closeout. These results indicate a weaker use of Process Results best practices in conducting surveillance of contractor performance, processing accurate and timely contractor payments, controlling contract changes, verifying final delivery, and obtaining seller's release of claims.

As reflected in Table 2, the Navy's process areas with the highest scoring survey response means for Management Support–associated survey items were in the key process areas of Procurement Planning and Source Selection. These results indicate a stronger use of Management Support best practices in ensuring that senior organizational management are involved in providing input and, if required, approval of Procurement Planning and Source Selection decisions and documents. Additionally, the Navy's key process areas with the lowest scoring survey response means for Management Support–associated survey items were in the process areas of Contract Administration and Contract Closeout. These results indicate a weaker use of Management Support best practices in ensuring that senior organizational management are involved in providing input and, if required, approval of Contract Administration and Contract Closeout–related decisions and documents.

As reflected in Table 2, the Navy's process areas with the highest scoring survey response means for Process Integration–associated survey items were in the process areas of Procurement Planning and Source Selection. These results indicate a stronger use of Process Integration best practices such as using integrated project teams and conducting an integrated assessment of contract type, risk management, and terms and conditions during Procurement Planning, and using integrated projects teams in the evaluation of proposals during contract Source Selection. Additionally, the Navy's process areas with the lowest scoring survey response means for Process Integration–associated survey items were in the process areas of Contract Administration and Contract Closeout. These results indicate a weaker use of Process Integration best practices such as integrating Contract Administration processes with other functional processes and using an integrated project team approach for monitoring and evaluating the contractor's performance and making related award fee and incentive fee determinations.

As reflected in Table 2, the Navy's process areas with the highest scoring survey response means for Process Measurement–associated survey items were in the process areas of Procurement Planning, Solicitation Planning, Solicitation, and Source Selection. These results indicate a stronger use of Process Measurement best practices such as adopting lessons learned and best practices for continuously improving the planning of procurements, issuing the procurement solicitation, evaluating contractor proposals, and awarding the contract. Additionally, the Navy's process areas with the lowest scoring survey response means for Process Measurement–associated survey items were in the process areas of Contract Administration and Contract Closeout. These results indicate a weaker use of Process Measurement best practices such as using efficiency and effectiveness metrics in administering the contract and closing out the contract. Additionally, these results also indicate a weaker use of practices such as adopting lessons learned and best practices for continuously improving the closing out of contracts and maintaining a lessons learned and best practices database for use in planning future procurements.

It is interesting to note that the CMMM summary-level survey response mean scores for the survey items related to each of the five process capability enablers show a clear distinction in the levels of the use of best practices. The relatively higher uses of best practices were identified in the pre-award process areas of Procurement Planning and Source Selection. The relatively lower uses of best practices were identified in the post-award phases of Contract Administration and Contract Closeout.

Process Improvement Initiatives

The true value of assessing an organization's contract management process capability is realized when the results are used in developing a road map for implementing contract management process improvement initiatives. The Navy was assessed at the Structured maturity level for Procurement Planning, Solicitation Planning, and Source Selection. In order for the Navy to progress to the Integrated maturity level, it should ensure these process areas are integrated with other organizational core processes, such as requirements management, financial management, schedule management, performance management, and risk management. The Procurement Planning process activities that need to be integrated with other organizational core processes include requirements analysis, acquisition planning, and market research. For the Solicitation Planning process, the activities include determining procurement method, developing evaluation strategy, and developing solicitation documents. The Navy should integrate Source Selection process activities such as evaluating proposals, applying evaluation criteria, negotiating contract terms, and selecting contractors. In addition to integrating these process areas with other organizational core processes, the Navy should also ensure that the procurement project's end-users and customers are included as integral members of the project procurement team and are engaged in providing input and recommendations for key contract management decisions and documents.

The Navy was assessed at the Basic maturity level for the Solicitation, Contract Administration, and Contract Closeout process areas. To progress to the Structured maturity level, the Navy should ensure that Contract Administration, Solicitation, and Contract Closeout processes are fully established, institutionalized, and mandated throughout the organization. Additionally, formal documentation should be developed for these process area activities. Also, senior management should be involved in providing guidance, direction, and even approval, when required, of key Solicitation, Contract Administration, and Contract Closeout strategies, decisions, related contract terms and conditions, and documents. The Solicitation process activities include advertising procurement opportunities, conducting solicitation and pre-proposal conferences, and amending solicitation documents as needed. The Contract Administration activities include monitoring and measuring contractor performance, managing the contract change process, and managing the contractor payment process. The Contract Closeout activities include verifying contract completion, verifying contract compliance, and making final payment. In addition to developing a road map for implementing contract management process improvement initiatives, the assessment results can also be used to identify training opportunities for increasing the process capability levels of the agency.

Implications for the DoD

The contracting processes and associated activities used in the Navy are the same processes and activities used in the Army, Air Force, and other DoD agencies. Therefore, these research findings provide insight into all DoD contract management. The results of the assessment of the Navy contracting processes reflect similar findings from an analysis of past DoDIG reports on contract management deficiencies. In their analysis of 149 DoDIG reports on contract management deficiencies, Hidaka and Owen (2015) found that 35.3% of the frequency of deficiencies was related to the Contract Administration process and 27.6% was related to the Procurement Planning process. Additionally, they found that 17.8% and 13.7% of the frequency of deficiencies were related to Solicitation Planning and Source Selection processes, respectively. Although the DoDIG investigations are focused on ensuring agencies are in compliance with contracting statutes and regulations, and not necessarily best practices, both the CMMM and DoDIG findings reflect a consistency in terms of weakness of contracting policies and procedures.

This consistency is also supported in Hidaka and Owen's (2015) findings that the DoDIG identified Control Environment as the internal control component associated with the majority (51.8%) of contracting deficiencies. The Control Environment internal control component is related to an organization's structure, authority, responsibility, and accountability. Additionally, Hidaka and Owen (2015) found that the Control Activities component was associated with 23.9% of the DoDIG-reported contracting deficiencies. The Control Activities internal control component is related to an organization's policies and procedures. As can be seen in the CMMM assessment results and Hidaka and Owen's findings, DoD contract management process capability is associated with its contracting internal controls. Both capable contracting processes and effective internal controls are needed to ensure auditability in DoD contract management (Rendon & Rendon, 2015).

Limitations of Findings

The CMMM is limited as an assessment model simply by the fact that it is based on qualitative survey data. Thus, the model is only as effective as the responses to the survey items. The CMMM should be used as an initial tool in assessing an organization's contract management process capability. The CMMM results should be validated with follow-up assessments, including personal interviews, procurement file audits, and reviews of procurement process documentation. Additionally, comparison of CMMM results with other procurement metrics such as procurement administrative lead-time, small-business awards, and the number of protested contract awards will also provide additional backup to the CMMM assessment.

Conclusion

This paper analyzed the results of contract management process maturity assessments conducted within the U.S. Navy. Although the CMMM assessment results indicated different contract management process maturity levels, ranging from Level 2 Basic to Level 3 Structured, for each contract management process area, some consistencies were identified. Generally, the assessment reflected higher maturity levels in the Procurement Planning, Solicitation Planning, and Source Selection process areas, while lower maturity levels were indicated in the Contract Administration and Contract Closeout process areas. These maturity levels reflect the extent of the implementation of contracting best practices in the areas of Process Strength, Process Results, Management Support, Process Integration, and Process Measurement. The assessment results identified opportunities for increasing contract management process maturity. The Navy assessment results also identified consistencies in DoD contract management process capability and internal control effectiveness. These consistencies include problem areas within the Procurement Planning and Contract Administration process areas. As the body of knowledge on government contract management process maturity continues to emerge, the use of maturity models will continue to gain wider acceptance as a tool for assessing organizational contract management process maturity and for providing a road map for implementing process improvement initiatives.

References

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, CA: Sage.

Defense Acquisition Workforce Improvement Act (DAWIA), 10 U.S.C. ch. 87 (1990).

DoD Inspector General (DoDIG). (2009). *Summary of DoD Office of Inspector General audits of acquisition and contract administration* (DoDIG Report No. D-2009-071). Washington, DC: DoD.

DoD Inspector General (DoDIG). (2012). *Award and administration of multiple award contracts at naval facilities engineering command specialty centers need improvement* (DoDIG-2013-007). Washington, DC: DoD.

DoD Inspector General (DoDIG). (2013). *Better processes needed to appropriately justify and document NAVSUP WSS, Philadelphia Site sole-source awards* (DoDIG-2013-034). Retrieved from http://www.dodig.mil/pubs/documents/DODIG-2013-034.pdf

DoD Inspector General (DoDIG). (2014). *Navy and Marine Corps have weak procurement processes for cost-reimbursement contract issuance and management* (DODIG-2014-092). Washington, DC: DoD.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2015). Retrieved from http://farsite.hill.af.mil/vffara.htm

GAO. (2015). *High-risk series: An update* (GAO-15-290). Washington, DC: Author.

Garrett, G. A., & Rendon, R. G. (2005). *Contract management organizational assessment tools*. McLean, VA: National Contract Management Association.

Hidaka, D., & Owen, J. L. (2015). *An analysis of internal controls for DoD contract management* (MBA thesis). Monterey, CA: Naval Postgraduate School.

McMillan, J. H., & Schumacher, S. (2001). *Research in education: A conceptual introduction.* New York, NY: Addison Wesley Longman.

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2014, September 19). *Better buying power 3.0* (White paper). Washington, DC: Author. Retrieved from http://bbp.dau.mil/docs/2_Better_Buying_Power_3_0%2819_September_2014%29.pdf

Power, M. (1996). Making things auditable. *Accounting, Organizations and Society, 21*(2), 289–315.

Power, M. (2007). *Organizations and auditability: A theory* (Vol. 9). Lancaster, England: Lancaster University.

Rendon, R. G. (2003). *A systematic approach to assessing organizational contract management maturity*. Unpublished doctoral dissertation, School of Business, Argosy University, Orange County, CA.

Rendon, R. G. (2008). Procurement process maturity: Key to performance measurement. *Journal of Public Procurement, 8*(2), 200–214.

Rendon, R. G., & Rendon, J. M. (2015). Auditability in public procurement: An analysis of internal controls and fraud vulnerability*. International Journal of Procurement Management, 8*(6), 710–730.

USA Spending. (2016) Data archives. Retrieved from https://www.usaspending.gov

Appendix A. Contract Management Processes

Procurement Planning: the process of identifying which organizational needs can be best met by procuring products or services outside the organization. This process involves determining whether to procure, how to procure, what to procure, how much to procure, and when to procure. Key process activities include conducting outsourcing analysis, determining and defining the procurement requirement, conducting market research, and developing preliminary budgets and schedules.

Solicitation Planning: the process of preparing the documents needed to support the solicitation. This process involves documenting program requirements and identifying potential sources.

Solicitation: the process of obtaining bids or proposals from prospective sellers on how organizational needs can be met.

Source Selection: the process of receiving bids or proposals and applying evaluation criteria to select a contractor.

Contract Administration: the process of ensuring that each contract party's performance meets contractual requirements.

Contract Closeout: the process of verifying that all administrative matters are concluded on a contract that is otherwise physically complete. This involves completing and settling the contract, including resolving any open items. Contract Closeout also includes contract termination.

Appendix B. Contract Management Maturity Levels

Level 1 Ad Hoc: Organizations at this maturity level do not have established organization-wide contract management processes. However, some established contract management processes do exist and are used within the organization, but these processes are applied only on an Ad Hoc and sporadic basis to various contracts. There is no rhyme or reason as to which contracts these processes are applied. Furthermore, there is informal documentation of contract management processes existing within the organization, but this documentation is used only on an Ad Hoc and sporadic basis on various contracts. Finally, organizational managers and contract management personnel are not held accountable for adhering to, or complying with, any basic contract management processes or standards.

Level 2 Basic: Organizations at this level of maturity have established some basic contract management processes and standards within the organization, but these processes are required only on selected complex, critical, or high-visibility contracts, such as contracts meeting certain dollar thresholds or contracts with certain customers. Some formal documentation has been developed for these established contract management processes and standards. Furthermore, the organization does not consider these contract management processes or standards established or institutionalized throughout the entire organization. Finally, at this maturity level, there is no organizational policy requiring the consistent use of these contract management processes and standards on contracts other than the required contracts.

Level 3 Structured: Organizations at this maturity level have contract management processes and standards that are fully established, institutionalized, and mandated throughout the entire organization. Formal documentation has been developed for these contract management processes and standards, and some processes may even be automated. Furthermore, since these contract management processes are mandated, the organization allows the tailoring of processes and documents in consideration for the unique aspects of each contract, such as contracting strategy, contract type, terms and conditions, dollar value, and type of requirement (product or service). Finally, senior organizational management is involved in providing guidance, direction, and even approval of key contracting strategy, decisions, related contract terms and conditions, and contract management documents.

Level 4 Integrated: Organizations at this level of maturity have contract management processes that are fully integrated with other organizational core processes such as financial management, schedule management, performance management, and systems engineering. In addition to representatives from other organizational functional offices, the contract's end-user customer is also an integral member of the buying or selling contracts team. Finally, the organization's management periodically uses metrics to measure various aspects of the contract management process and to make contracts-related decisions.

Level 5 Optimized: Organizations at this maturity level systematically use performance metrics to measure the quality and to evaluate the efficiency and effectiveness of the contract management processes. At this maturity level, continuous process improvement efforts are also implemented to improve the contract management processes. Furthermore, the organization has established programs for lessons learned and best practices in order to improve contract management processes, standards, and documentation. Finally, contract management process streamlining initiatives are implemented by the organization as part of its continuous process improvement program.

Designing and Managing Successful International Joint Development Programs

Andrew Philip Hunter—is a Senior Fellow in the International Security Program and Director of the Defense-Industrial Initiatives Group at the Center for Strategic and International Studies (CSIS). He focuses on issues affecting the industrial base, including emerging technologies, sequestration, acquisition policy, and industrial policy. From 2011 to November 2014, Hunter served as a Senior Executive in the Department of Defense (DoD). Appointed as Director of the Joint Rapid Acquisition Cell in 2013, his duties included fielding solutions to urgent operational needs and leading the work of the Warfighter Senior Integration Group to ensure timely action on critical issues of warfighter support. [ahunter@csis.org]

Gregory Sanders—is Deputy Director and Fellow with the Defense-Industrial Initiatives Group at CSIS, where his team analyzes data on U.S. government contract spending and other budget and acquisition issues. He derives knowledge from complex data collections via succinct and innovative tables, charts, and maps. His recent research focuses on contract spending by major government departments, contingency contracting in Iraq and Afghanistan, and European and Asian defense budgets. Sanders holds an MA in international studies and a BA in government and politics from the University of Denver, as well as a BS in computer science from the University of Maryland. [gsanders@csis.org]

Samantha Cohen—is a Research Assistant with the Defense-Industrial Initiatives Group at CSIS. Her work focuses on managing and analyzing data to identify relationships among policies, defense spending, and outcomes. Her recent research focuses on public opinion and defense spending in European countries. In 2015, she graduated from American University with a BS (honors) in economics. Additionally, she successfully completed the intensive French Language School program at Middlebury College and studied the economic and defense policies of the European Union and NATO with AU at the Université Catholique de Louvain in Brussels, Belgium. [scohen@csis.org]

Abstract

International joint development programs are important because of their potential to reduce costs and increase partnership benefits such as interoperability, economies of scale, and technical advancement. However, the performance of international joint development programs varies greatly. This paper compares the best practices of international joint development and domestic development programs through case-study analysis to identify the key variables that contribute to a program's eventual success or failure and to understand the elements that are crucial to managing these programs.

Introduction

The DoD recognizes the value of international joint development programs that include both research funding from and technology development with multiple countries. This is especially true in light of the Budget Control Act of 2011, which imposed caps on defense spending concurrent to European defense budget reductions. Additionally, the Defense Strategic Guidance issued in January 2012 commits the United States and the DoD to strengthening partnership with and cooperation in the global community by emphasizing pooling, sharing, and specializing capabilities with partner nations (DoD, 2012b). Furthermore, the *International Cooperation in Acquisition, Technology, and Logistics Handbook* states that when considering the pursuit of an international joint development program, the Milestone Decision Authority must consider whether a program executes "demonstrated best business practices, including a plan for effective, economical, and efficient management of the international cooperative program" (Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, 2012). While the value of international joint development programs is recognized, a theoretical basis for best practices in these programs is a paucity.

International joint development is not a novel idea. However, there is not yet consensus on what the design and management of successful international joint development programs looks like. The theoretical benefits of joint development projects include reduced costs, improved international cooperation, increased competition, and innovation. While unique combinations of these benefits drive each international program, most nations turn to international cooperation in defense acquisition to appease budget pressures and procure advanced programs that individual nations cannot financially afford. Utilizing the existing literature on best practices in both single nation and international joint development programs, this report investigates whether best practices have been actualized and what characteristics in the design and management of such programs equate to different outcomes.

Through evidence garnered from acquisition literature, the study team has established eight characteristics that are crucial to program outcomes. This report conducts an initial analysis of how the eight characteristics affect program outcomes. This interim report compares these characteristics over three initial cases, and three additional cases will be included in the full technical report. The three initial cases included in this interim report are the North Atlantic Treaty Organization (NATO) Alliance Ground Surveillance (AGS) program, the Joint Strike Fighter F-35 Lightning II (F-35) program, and the Lightweight 155mm Howitzer (M777) program.[1] For this analysis, the study team focused on defining the program characteristics that contribute to each program's challenges and successes. For the final technical report, the study team will build upon this research to develop a framework that will help guide future international cooperation in defense acquisition.

Methodology

To bolster the analysis, the study teem seeks to answer the research questions raised below. To achieve this, the study compares the defined best practices from acquisition literature with what the case studies have actualized by discussing what characteristics research has shown as crucial to international joint development program outcomes. The study team investigates these characteristics by interviewing program stakeholders from industry and government, as well as outside experts. Next, the study team uses the information gleaned from the interviews to assess the validity of the literature-derived hypotheses defined below. Lastly, the study team analyzes three cases to better understand the elements critical to managing and designing successful international joint development programs.

Research Questions and Hypotheses

In order to investigate the best practices of designing and managing international joint development programs in defense acquisition, the study team focused on two overarching research questions:

---

[1] The other three cases that will be analyzed in the final technical report for this study are the Medium Extended Air Defense System (MEADS), the Airbus A400M Atlas, and the Standard Missile-3 Block IIA program.

1. What are the characteristics of international joint development programs that result in positive or negative cost, scheduling, and end-product outcomes, such as a final product, interoperability, technical relevance, and development of existing defense industrial bases?[2]
2. How are best practices of international joint development programs in defense acquisition different from best practices of single-nation acquisition programs?

Additionally, four hypotheses are proposed to form a baseline for this analysis. These hypotheses were derived from a review of the existing literature on international joint development. This interim report will analyze these hypotheses to the extent feasible at this point in the research. The hypotheses are as follows:

1. The structure of cooperation in international joint development programs matters—the international joint development programs whose stakeholders cooperate only during the development or production phases will have less successful cost, scheduling, and end-product outcomes.
2. International joint development projects that are more grounded in security policies rather than economic efficiency interests are more likely to result in negative cost, scheduling, or end-product outcomes.
3. Countries that have cooperated in defense acquisition before have a higher chance of achieving positive cost, scheduling, and end-product outcomes.
4. Countries that are uniquely capable of producing complex acquisition programs benefit from working with smaller countries or industries who may have comparative advantages in certain technologies, but do not have the capacity to produce complex acquisition programs on their own.

Interviews With Program Stakeholders

The study team interviewed stakeholders from government and industry, as well as key leaders from research organizations to augment the information gleaned from the literature. The interviews focused on investigating which characteristics, out of the eight characteristics described in the section titled Case Study Analysis: Characteristics, each case manifested in addition to addressing the research questions and hypotheses. The interviews were accompanied by a Likert-scale survey to determine which characteristics the cases demonstrated. The survey results will be discussed in the section titled Case Study Analysis: Analysis. Figures 1–8 represent the results of the survey. Figures 1, 2, 4, 5, and 6 display the percentage of interviewees that voted for each response level. Figures 3, 7, and 8 report the percentage of interviewees that voted for each response level for two related questions. The questions are indicated on the y and x axes.

Case Study Analysis

While the full technical report for this study will go into greater detail on each case's history, this paper will touch upon key instances where there is evidence that the design and management of the program affected what actually happened, whether it be a success or a

---

[2] End-product outcomes are subjective to each case and successful end-product outcomes for each case will change depending on the purpose and goals of the program.

failure. When discussing the outcomes of the programs, the identified characteristics crucial to successfully designing and managing programs that are unique to international joint development programs will be analyzed to further investigate which characteristics are attributable to whether the program achieved its goals.

The NATO Alliance Ground Surveillance Program

The first inklings of the AGS program began in the early 1990s when NATO's Defense Capabilities Initiative (DCI) called for both higher standardization and interoperability of NATO alliance equipment and using cooperative development and production to realize the theoretical economic and technological benefits cooperation presents. Additionally, the United States' use of the Joint Surveillance Target Attack Radar System (JSTARS)[3] during Operation Desert Storm in 1991 accentuated the paramount role intelligence, surveillance, and reconnaissance (ISR) capabilities played in next-generation warfare (Chao, 2004). The original proposal for the AGS program was a "NATO-owned and operated core AGS capability, supplemented by interoperable national assets" (Chao, 2004, p. 5). Interviewees emphasized that when making decisions throughout the program, decision makers needed to ensure that there was a European face on the project, and that it was an inclusive NATO program.

It was not until 2009 that the 15 NATO partners signed a Program Memorandum of Understanding (PMOU) agreeing to the legal and budgetary framework for acquiring AGS (NATO, 2009). During the 14 years it took for the program to go from inception in 1995 to a PMOU in 2009, numerous factors were collectively responsible for the delayed beginnings of the program. In 2007, financial circumstances put pressure on defense budgets in Europe that lead NATO to buy off-the-shelf Global Hawk Block 40 RQ-4s with Multi-Platform Radar Technology Insertion Program (MP-RTIP) radars while the ground segment would be developed and procured by the European and Canadian partner nations (NATO, 2016).

The Joint Strike Fighter F-35 Program

The Joint Strike Fighter (JSF) program began in 1995 as the latest iteration of the Joint Advanced Strike Technology (JAST) program. JAST, which was initiated in 1993 following the bottom-up review of U.S. defense programs and policy, was originally designed to provide replacements for both the Navy's A-6 Intruder attack aircraft and the Air Force's F-16 Fighting Falcon multirole fighter. Two years later, an advanced short takeoff and vertical landing (ASTOVL) craft that was being developed by the Defense Advanced Research Projects Agency (DARPA) was congressionally directed to be incorporated into the JAST program, which would later be renamed the Joint Strike Fighter program.

Two years after Lockheed Martin was awarded the prime contract for the F-35 program in 2001, the Government Accountability Office (GAO) succinctly reiterated the purpose of the F-35 program: "The JSF program goals are to develop and field a family of stealthy, strike fighter aircraft for the Navy, Air Force, Marine Corps, and U.S. allies, with maximum commonality to minimize life-cycle costs" (Walker, 2003, p. 49). Furthermore, a major factor that influenced the international partner nations to join this program was to not only reap the anticipated operational and monetary benefits, but to develop a stronger

---

[3] Northrop Grumman's E-8 JSTARS is an aircraft designed to conduct ground surveillance, command and control, and battle management.

relationship with the United States and to play a role in future strategic and military collaboration.

In 2009, then-Under Secretary of Defense for Acquisition, Technology, and Logistics, Ashton Carter, issued an Acquisition Decision Memorandum (ADM), which led to the rescinding of Milestone B certification for JSF. This followed the release of the JET II report by DoD's Joint Estimating Team, which noted that JSF's system development and demonstration (SDD) phase would need an additional 30 months to complete, and JSF’s 2010 Nunn-McCurdy breach (Gertler, 2014, pp. 9, 29). Additionally, the F-35 Program Executive Officer commissioned a technical baseline review (TBR) that led the Secretary of Defense to announce that testing of the F-35A and F-35C would be de-coupled from testing of the F-35B (DoD, 2010, p. 4). The review noted that the F-35B was experiencing “significant testing problems” and placed the program on “the equivalent of a two-year probation” (Gertler, 2014, p. 31) that was lifted January 20, 2012 (DoD, 2011, p. 5).

According to the 2015 Director, Operational Test and Evaluation (DOTE) report, the greatest challenge the F-35 faces today is affordability (DoD, 2015). While cost baselines, schedule projections, and technical capabilities have consistently not been met, the program has reestablished its baselines and recently shows progress in punctuality. For instance, according to the 2015 DOTE report, the number of 2015 actual test flights[4] was only 7.4% below the scheduled amount. Compared to 2012, when only 34% of the planned flight tests had actually been executed, a 7.4 percentage point difference is a large improvement (DoD, 2012a). Cost performance is also improved since 2012.

The Lightweight 155mm Howitzer M777 Program

The M777 was designed to replace the M-198 Howitzer, previously used by both the U.S. Marine Corps (USMC) and the U.S. Army (USA), by introducing a lighter machine capable of higher fire speed and accuracy rates. M777’s request for proposal (RFP) stated that the platforms competing for the contract would first be presented at Yuma Proving Ground on April 25, 1996, and that the companies who were able to provide a platform that met the operational requirements detailed within the RFP would compete in a shoot-off and evaluation phase (U.S. Marine Corps and U.S. Army, 1999).

The team of Textron and VSEL won the Engineering, Manufacturing, and Development (EMD) contract in March of 1997. Textron dropped out in 1998 and VSEL experienced some challenges in adapting to the American systems engineering process, which led to an initial delay pushed as VSEL restructured pre-production systems engineering tasks, and resulted in a program cost growth of $43 million. This restructuring also generated a 21-month program delay, from December 1999 to September 2001 (Government Accountability Office [GAO], 2000, pp. 7–8).

In 1999, BAE Systems became the new M777 prime contractor when it acquired VSEL. BAE quickly began to encounter manufacturing challenges with the M777, many of which were driven by problems with titanium welding on the M777. Despite these challenges, M777 has found a larger market, suggesting a successful end-product outcome. In the initial conceptualization of the program, it was planned that the Marine Corps and Army would be the only groups to acquire the platform. However, in the past decade the

---

[4] As of November 2015

United States has utilized foreign military sales (FMS) to provide the M777 to allies across the globe following the M777's successful deployment in Afghanistan and Iraq.

Characteristics

In order to address the first research question,

> What are the characteristics of international joint development programs that result in positive or negative cost, scheduling, and end-product outcomes such as a final product, interoperability, technical relevance and development of existing defense industrial bases?,[5]

the study team identified eight characteristics that research and interviews with stakeholders have shown to be the most crucial and unique to impacting outcomes of international joint programs:

1. Integration
2. Number of Participating Countries
3. Decision Making
4. Commitment
5. Flexibility
6. Alignment of Operational Needs
7. Tradeoff between Leading-edge Technology and Cost
8. Workshare Distribution

The first characteristic is integration. As part of a Center for Strategic and International Studies (CSIS) project on complexity, Jeffrey A. Drezner, a senior policy researcher at RAND Corporation, discussed how organizational complexity is inherent in modern acquisition programs. Drezner (2009) stated that, "Organizational complexity addresses the structures and interactions of the government and industry organizations responsible for system design, development, production, and support" (p. 32). One complex aspect unique to international cooperation is the transnational partnerships that must be made for governments and industries to work together. Consequently, deeper layers of complexity exist: first, between governments; second, between government and industry; and third, between industries. In 2003, GAO published a report that argued "[t]he collaborative relationship between the customer and the product developer is essential to driving down operating and support costs" (p. 6). This study decided to focus on transnational relationships by analyzing the level of integration between the players involved.

The second characteristic is the number of participating countries. In 2012, defense economist Keith Hartley claimed that the number of partner nations in acquisition programs is associated with collaboration inefficiencies. Furthermore, the increasing number of partner nations adds additional layers of complexity.

The third characteristic is whether decision making throughout the program depended more on operational needs that could not be met by competing systems or on

[5] End-product outcomes are subjective to each case and successful end-product outcomes for each case will change depending on the purpose and goals of the program.

diplomatic and political needs. Major program decisions, such as those on requirements or contracting, are either based on operational needs, or political and diplomatic needs. Making decisions based on operational needs that could not be met by competing systems is more likely to achieve efficient costs and resource allocation, while decisions based on political or diplomatic demands might not reach the most cost-efficient outcomes (Hartley, 2012, p. 20).

The fourth characteristic is commitment. For programs to achieve the theoretical cost benefits of international joint development, partner nations need to be committed to the program. If one country defects, costs for the remaining countries will rise. Additionally, lack of commitment is a major warning sign that can lead to program failure.

The fifth characteristic is program requirements and the program's flexibility to respond to changing environments. The volatile technological and security environments of today require programs that can quickly change in response to emerging innovation and threats. Therefore, the management of programs must have the capacity to respond to changing environments without necessitating the termination of the program.

The sixth characteristic is the extent to which operational goals of partner nations involved align. Having multiple militaries working together could introduce varying operational goals. In order to produce a successful end-product, partner nations need to have compatible goals so that the program stays focused and partner nations are equally invested in acquiring the capability.

The seventh characteristic is whether the program was based on demand for leading-edge technology or based on demand for affordability. There is a tradeoff between achieving leading-edge technology and affordability structures, specifically economies of scale. Economies of scale exist when the scale of output increases to a point where the average per-unit costs of production begins to decrease. The exceptionally high cost of research and development (R&D) in modern defense acquisition is crucial to procuring technologically advanced capabilities. While the costs of R&D exhibit unremitting growth, the funds necessary to support R&D have shrunk in the United States and in U.S. partner nations. It is difficult to determine whether economies of scale will be achieved for a program from the outset. If a program decides to procure a system from scratch, it is not certain that the outcome will be successful enough to produce an adequate return on the initial investments made during R&D. This is increasingly risky if the program aims to procure leading-edge capabilities. Economies of scale are impossible to achieve before a final product has been developed and production has begun. It is uncertain whether a program based on leading-edge technology will reach levels of production that create cost-efficient output. International cooperation during development presents the opportunity to share costs of R&D over participants. From the outset, however, a program should elucidate whether the key mission is to achieve leading-edge technology or economies of scale.

The eighth characteristic captures how the program distributes workshare. To achieve cost-efficient outcomes, international programs present greater opportunity for competition based on comparative advantage. Competition is critical to achieving cost-efficiency because when there are many substitutable choices for consumers to choose from, suppliers will be forced to produce at the lowest cost possible since consumers will choose to buy the lowest-priced product. The international marketplace presents greater opportunity for competition among industries, which in turn supplies procurement at lower costs to the buyer. However, the international marketplace also introduces greater political and industrial-base variables into the equation. Costs are not typically the sole incentive for nations to participate in international cooperation in defense acquisition. Countries view strategic posture, trade policy, industrial gain, and technology transfer as spillover benefits to international cooperation. In some cases, these spillover benefits are more important to a country than cost-efficiency. Focusing on spillover benefits more than focusing on cost-efficiency will impact program outcomes and impact how countries work together.

Consequently, the last characteristic crucial and unique to impacting outcomes of international joint programs is whether the distribution of workshare was based on participating countries' comparative advantage or on political or industrial-base goals.

Analysis

Integration

For the integration characteristic, the study team asked the interviewees, "On a scale of one to six, rate the level of integration between government and industry, governments, and industries for each program." The responses are reported in Figure 1.

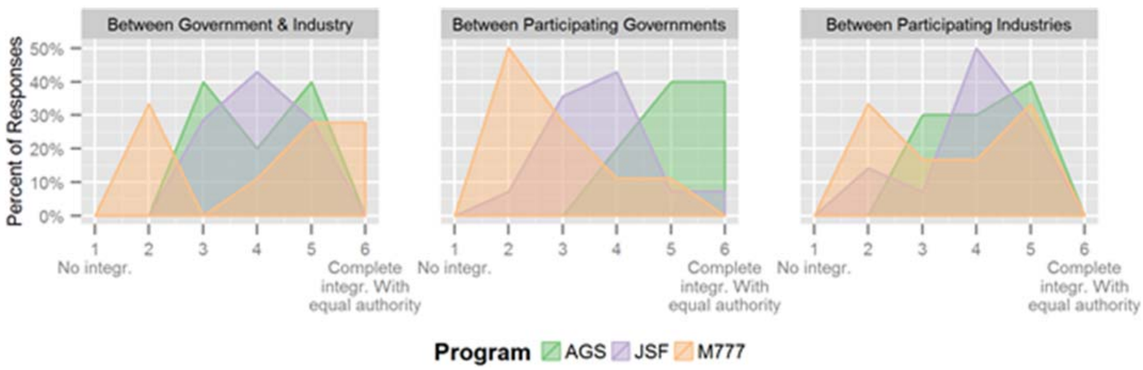


Figure 1. Extent of Integration

The AGS and the F-35 programs practiced similar levels of integration between government and industry, while the M777 program was reported as more integrated by the majority of survey respondents. For the M777 program, a higher level of integration between government and industry existed because the government made the final decisions while there was collocation of employees from both government and industry where the two participating bodies would consult before final decisions were made. This colocation increased during production. For the F-35 program, the contract was more of a top-down relationship in terms of decision making. The level of relationship between government and industry often depended on the company in question, but the government was in charge of the program for all companies. For AGS, the relationship between government and industry was very strong and positive in some instances, but in other cases caused problems for the program. The most notable and positive relationship was between NATO and Northrop Grumman (NG), who consulted regularly from the outset. For other companies, the integration between industry and government occurred domestically. Due to the fact that each country wanted to secure its own investments, issues arose because it was difficult to create work for every country based solely on investment levels. Without higher levels of transnational integration, certain countries viewed that the level of their industries' workshare was not worth the costs and defaulted from the program.

The reported level of integration between governments for the three programs varied greatly. For the F-35 program, the most frequently chosen level was "decision-making integration." The partnership between the governments involved in the program was not legally binding but did implement obligations that represented formal commitments between the partners. The decision-making mechanism between the partner nations was dominated by the United States, which had 80% voting power, while the other partners shared the remaining 20%. The UK, which was the first partner nation involved, however, could levy requirement alterations because of its tier-one status.[6] At the outset of the partnership, countries determined what was on their "must-have" and "nice-to-have" lists. Despite the unequal share of voting power and the eventual exclusion of most "nice-to-have" items, there have thus far been minimal defections and one interviewee argued that most partner nations were satisfied with at least one F-35 variant on an operational level. In addition, there were instances where non-U.S. partner nations teamed up to push for an operational requirement not initially planned by the United States. The mechanism encouraged the U.S. military branches and the partner nations to each coordinate their opinions before facing one another.

For the NATO-operated AGS program, the survey responses for integration between governments were both higher and less variable as can be seen in Figure 1. NATO's historic establishment as an intergovernmental organization bolstered AGS' achievement of high integration between governments. Decisions made on the AGS program were consensus-based with equal voting power for each partner nation. The study team did find, however, that formal voting arrangements did not always translate into how decision making works in practice. In a consensus-based voting mechanism, notionally every nation comes to the table with an equal stake in achieving their goals. In reality, there are the strong players and the followers during decision making. Typically, newly ascended or smaller partner nations fall into the latter category. When the larger contributing nations reached consensus, the other participating nations were generally quick to follow. Additionally, when there was disagreement between the larger players, delays could and did happen. Holdout partners could be outmaneuvered as long as the remaining nations all agreed, but escalation to direct contact between national leadership was sometimes necessary to resolve disagreement. While NATO as an organization has a strong institutional memory, throughout the first 15 years of the program there was not a standing office for joint acquisition. Instead, the designated NATO equivalent to a program office—the NATO AGS Management Agency (NAGSMA)—was not created until the PMOU was signed in 2009. This late organization standup cost the program the benefits of institutional memory because NAGSMA had not been present during the previous 14 years of the program.

For the M777 program, the United States established and built the program, and while the United States consulted with the other nations involved, the program was ultimately U.S.-led with unilateral decisions on requirements. International cooperation evolved when other nations decided that this program fulfilled their operational needs and joined the program. The partner-nation governments each had a representative collocated in the U.S.-based program office, but these foreign representatives were there for information gathering, rather than for sharing leadership.

For the integration between industries in the AGS program, the prime contractor, NG, controlled intra-industry relations. The AGS program was developed on the concept that cost share equaled workshare. Some experts argue that this concept contributed to the drawn-out, 14-year process of choosing the platform and signing the PMOU. Since European industry wanted access to U.S. technology, it was less desirable for the European

[6] Tier levels were made based on investment levels. The UK is the only tier-one partner nation and is the partner nation who has invested the most money after the United States.

partner nations to choose a platform manufactured by the United States because of U.S. hesitancy to share technology. This meant that, with a U.S. platform, there was not a sufficient ratio between cost and workshare. However, from the outset of the program, NG wanted to not only become the prime contractor for the AGS program, but to build a stronger relationship with European industry. To do this, NG built personal and professional relationships through consultation with a “we’re all in this together” attitude. Once the Global Hawk was chosen for the air segment, NG held responsibility for the intra-industry relations. From a U.S. perspective, any arrangement of industry cooperation was acceptable, as long as it lead to the best value.

The industries involved with the F-35 program had integrated decision-making processes; however, commercial tension between industries existed because of competition. Consequently, industries were less integrated during development and production. However, higher levels of integration during sustainment is anticipated, and tensions over competition appear likely to fade. One of the interviewees reported that, governments were more inclined to share information and work together than the industry partners.

For industry integration during the M777 program, one respondent marked the lowest value out of the three programs for this response. Interviewees, however, indicated that the prime contractor, BAE Systems, had regular consultation on decisions between the other industries involved. The prime and sub-contractors were more integrated here because BAE Systems controlled contracting with the sub-contractors. Regular consultations occurred between contractors when developing the system and throughout manufacturing. This was crucial for cutting-edge technologies. However, as the program matured, consultations happened less frequently.

Number of Participants

For the number of participating countries characteristic, the study team asked the interviewees, “On a scale of one to six, rate the extent to which the number of countries involved with the program impacted major decisions.” The results are displayed in Figure 2.

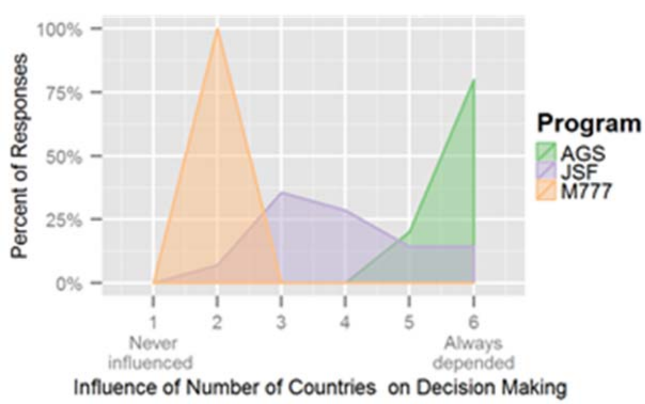


Figure 2. Number of Participating Countries

The number of countries participating in the AGS program varied from inception, and today 15 countries officially participate in the program.[7] In terms of participating countries, the AGS program has the highest number out of the cases analyzed for this study. Multiple interviewees discussed the program's core need to be an Alliance program with a European face. Furthermore, from the outset, the program wanted to include as many nations as possible while staying cost effective and maintaining operational requirements. Consequently, the number of nations involved always impacted the program, mirroring what the interviewees said. The core program characteristics of putting a European face on the program and including as many Alliance member-states as possible influenced many of the twists and turns the program took over its 20+ years of existence. Every decision made had to support inclusion and diversity at the same time that it satisfied each partner nation's investments and national interests; for instance, the United States originally offering JSTARS for the air segment faced political backlash. The United States then offered to use JSTARS radar technology on an Airbus aircraft. This caused further problems with technology gains desired by the EU and was ultimately an unsuccessful solution. Although the influence that the number of countries had on decision making throughout the program presented challenges, the large number of participating nations also kept the program moving forward. One expert from government noted that if there had been fewer partner nations, the program would have been more likely to fall apart because it would have lacked the broad political support within NATO to push the program forward.

Similarly, multiple interviewees from the F-35 program argued that the higher number of countries participating in the F-35 program prevented the program from being cancelled in the face of challenges. Unlike the AGS program, the experts interviewed responded very differently from each other on the characteristic describing the number of participating countries in the program. The study team has concluded that the varied responses for this characteristic can be attributed to different perspectives from different partner nations. Unlike the NATO-driven AGS program, the F-35 program is U.S.-centric. The United States is harder to integrate with for many of the participating nations because of the size and technical edge of the U.S. defense industrial base as well as technology transfer laws. Historically, the UK, Canada, and Australia have an easier time with this because of the long-term relationship that they have had with the United States and information sharing. Consequently, some partner nations feel that the sheer number of participants influenced decision making during the program, while other partner nations do not view the number of participants as a unique driver of decision making.

The interviewees for the M777 program unanimously chose "slight influence" for how the number of participants affected decision making during the program. The small number of partner nations coupled with the U.S.-centric program left little room for the number of partners to cause complications.

Decision Making

For the decision making characteristic, the study team first asked the interviewees, "On a scale of one to six, rate the extent to which decisions regarding the program were made depending on operational needs that could not be met by competing systems."

---

[7] Bulgaria, Czech Republic, Denmark, Estonia, Germany, Italy, Latvia, Lithuania, Luxembourg, Norway, Poland, Romania, Slovak Republic, Slovenia, and the United States.

Second, the study team asked, “On a scale of one to six, rate the extent to which decisions regarding the program were made depending on diplomatic or political needs.” The results are in Figure 3.

Figure 3. Decision Making

For the AGS program, one of the interviewees noted that decision making depended on different factors that changed over the different stages of the program. Another interviewee noted that, for some nations, decision making depended on operational needs in some instances, but considered industrial-base needs other times. AGS directly responded to NATO’s demand for both ISR capabilities and NATO Alliance equipment standardization and interoperability in the early 1990s. However, political factors for the AGS program were typically rated higher, which is not surprising given the inclusive alliance goals discussed in the Analysis: Number of Participants section. For some member states, acquiring the AGS system did not necessarily respond to their domestic strategic goals. Instead, these member states participated with the intention of either being good NATO partners or investing for the benefit of domestic industrial-base interests. For instance, governments who wanted their domestic constituent industries to benefit would be more likely to participate and contribute money. Additionally, the political and diplomatic pressures of periodic summits and major NATO events facilitated decision making during the program. When examining the timing of major decisions throughout the program and major summits or events, there is clear alignment. For example, the AGS procurement contract was signed at the 2012 Chicago Summit. These types of events accelerated key program milestones and decisions.

For the F-35 program, one interviewee was reluctant to rate “always depended” on operational needs, even though at the outset, the goals were to pursue, develop, and design based on the operational requirements of the predicted evolving security threats. As the program developed, additional countries joined through scheduled foreign military sales (FMS). FMS decisions were partially based on competing interests and best value rather than purely operational requirements. The responses rating the extent to which decision making depended on diplomatic or political needs for the F-35 program are more in the middle. One of the interviewees from a foreign government chose “depended more than occasionally,” because that country chose to participate in the F-35 program not only to reap the operational capabilities, but to also strengthen their interoperability with the United States and allied nations in the future. One interviewee discussed how operational requirements concerns drove decision making during the development stages of the program, while during and after the transition to follow-on development and production, political and diplomatic needs became more important. This could be attributed to the fact that the UK and United States cooperated from the outset, and as the program moved forward, other partner-nations joined, which added political concerns, national sovereignty requirements, and diplomatic interests.

The responses for the M777 program on whether decision making depended on operational needs conflict each other. While two interviewees rated decision making depending on operational needs as a six, “always depended,” one interviewee rated a one, “never depended.” For the United States, there were no competing systems at the time. The USMC and the USA jointly needed the lightweight and digitized firing system capabilities the M777 offered. The UK similarly had a demand for this technology that at the time had no competing systems. The Falklands War made it obvious that the UK’s land munitions lacked M777’s capabilities. The UK’s large stakes in this operational requirement made them the dominating industry when competing for the contract. The interviewees rated lower for the dependence of decision making on diplomatic or political needs for the M777 program. The reason why some of the survey respondents chose the third level, “depended occasionally,” is that the UK had already been developing this type of capability in response to their operational gaps during the Falklands War. Yet, due to the U.S. Arsenal Act, the United States had to establish a domestic supply chain. In the end, 70% of the program was made in the United States despite the UK’s effort at establishing the capacity to do so.

Commitment

For the commitment characteristic, the study team asked the interviewees, “On a scale of one to six, rate the extent to which commitments for the program stated in the contract or PMOU were binding.” No particular conclusions could be drawn from the M777 program for this characteristic. The responses rating the extent to which commitments are binding are displayed in Figure 4.

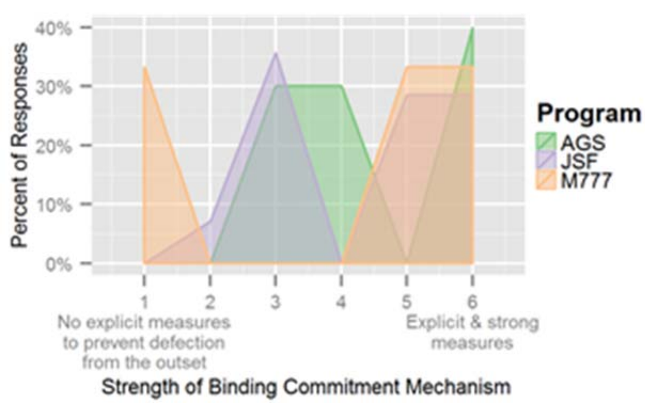


Figure 4. Commitment

There is evidence that these divergences can be imputed to differences in point of views from government and industry. For the AGS program, one interviewee expressed that NATO worked hard to put in disincentives for partner nations to quit. At the same time, another interviewee emphasized the importance of having disincentives, but not to the point that nations only stay in the program because the consequences of defecting are too harsh. It was equally important for the partner nations to benefit from participation as it was to prevent defection.

Responses to the commitment characteristic are also split for the F-35 program. When following-up with the interviewees, the study team found that one point of view argued that there are limited explicit measures to prevent defection. Instead, the partner nations made large investments that are ultimately a sunk cost if the country exits the program without procuring the capability. This situation acts as a measure to prevent defection but was not explicitly planned in the PMOU or contracts. Furthermore, countries who are benefiting from industrial spillovers have a disincentive to defect because exiting the program would mean losing industrial benefits. Another interviewee concluded that instead of being contractually binding, countries have moral, ethical, and political commitments to the program that act as measures to prevent defection.

Flexibility

For the flexibility characteristic, the study team asked the interviewees, "On a scale of one to six, rate the level of flexibility the program had in being able to change requirements in response to program updates such as an addition of participating countries or new developments." The results are displayed in Figure 5.

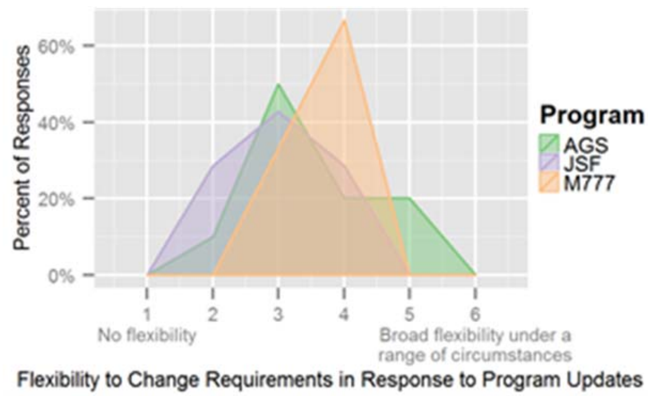


Figure 5. Flexibility

While the AGS program maintained its operational and political requirements made from the outset, the technical specifications on which these goals would be executed changed multiple times over 14 years of negotiations. These changes were made not necessarily because the program was flexible, but more so because the political nature of the program demanded it. To reiterate, the original mantra of the program was to include as many flags as possible. This level of complexity inherently faced political stalls, aggressive workshare negotiations, and affordability challenges. Whether or not these changes were "high hurdles" or "easy" depends on who is talking. Ultimately, the program had to be flexible, even if this meant 15 years of negotiations before the PMOU.

The interviewees for the F-35 program emphasized that the number of partner nations and increasing cost pressures created high hurdles for change. As a result of schedule delays, certain nations had to invest additional money to maintain their existing fleets on top of the money already invested in the F-35 program. These countries would have been more flexible and able to manage this situation if these risks were addressed from the outset. Daunting cost estimates were another barrier to flexibility. According to an interviewee, the vast majority of proposed changes were generally dropped after the proposing country saw the estimated cost increases. Additionally, bureaucracy imposed organizational constructs that controlled decision making within cost, schedule, and performance and did not work well with change.

From the discussion on the M777 program, there is evidence that the highest hurdles to flexibility existed between the services and not the international partners. Otherwise, there were a limited number of cases demanding a change to the program.

Operational Mission

For the operational mission characteristic, the study team asked the interviewees, "On a scale of one to six, rate the extent to which the countries involved with the program were compatible in operational requirements." The responses are displayed in Figure 6.

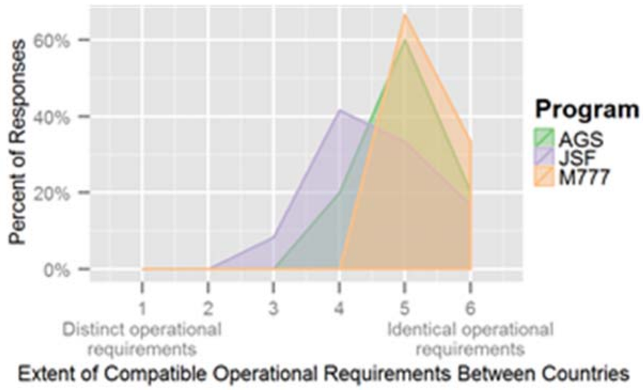


Figure 6. Alignment of Operational Needs

For the AGS program, the individual countries were less coherent on their core operational demands. Since this was a NATO program, however, the countries participating all agreed that the Alliance needed to satisfy the demand for ISR capabilities and greater interoperability and collaboration within NATO. The United States, for instance, already had access to this capability. Their interests were centered on helping NATO achieve this capability so that the United States would not be called upon to bolster strategic demands requiring ISR. Conversely, smaller nations could not achieve this capability on their own. Even if the need for leading-edge ISR capabilities is not in a nation's core strategy, having a more public access to ISR technology through NATO, benefits both their domestic and international security. Consequently, the participating partner nations do not perfectly align with their domestic interests, but, as a whole, the AGS program supports a common demand.

Similar to the AGS program, the partner nations of the JSF program did not have identical operational requirements, but because there are three variants of the program, major requirements for each country were met. The level of operational commonality between partner nations changes according to the country. Compared to the United States, the UK's level is more of a five or a six while Turkey's level is more a three or a four. Strategically, all partner nations are interested in interoperability, which acquiring the F-35 promotes. One interviewee suggested that the variance of responses is a result of changing operational requirements between times of peace and times of war.

Program Mission

For the program mission characteristic, the study team first asked the interviewees, "On a scale of one to six, evaluate the extent to which the mission of the program was based on the demand for leading-edge technology and a lower number of initial output." Second, the study team asked, "Evaluate the extent to which the program was based on the demand for developing low-cost economies of scale." The results are displayed in Figure 7.

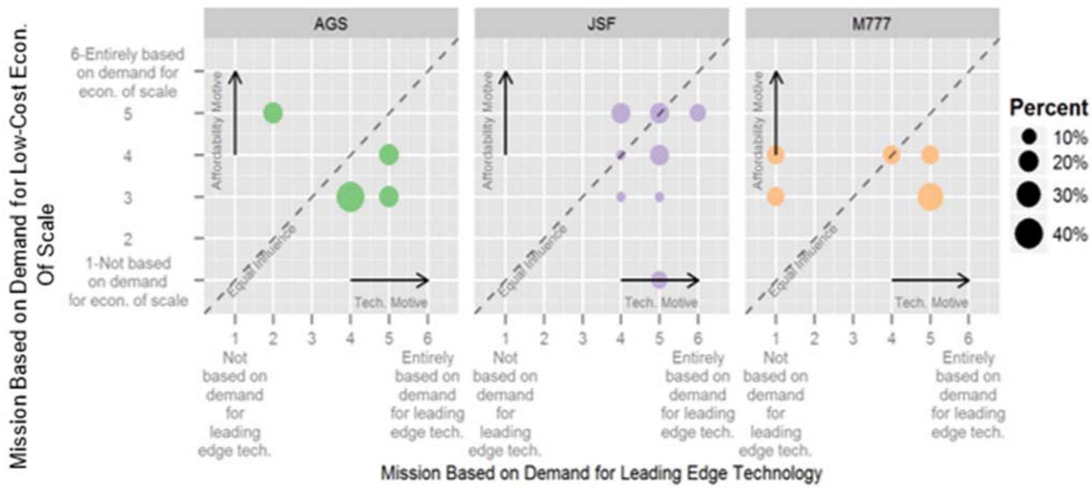


Figure 7. Tradeoff Between Leading-Edge Technology and Cost

The study team hypothesized that there is a tradeoff between these two underlying program goals. The survey responses, however, suggest that this is not the case. Or, rather, program directors did not consider them to be a tradeoff. For the AGS program, NATO was acquiring a leading-edge technology, but not necessarily developing it. The ISR capability NATO wanted to acquire already existed (JSTARS), just not as a platform that was NATO owned and operated. The United States tried its best to preserve and expand upon the leading-edge technology they had already developed with JSTARS to meet AGS requirements, despite understanding that there were other political dimensions it had to simultaneously account for. The underlying tradeoff was achieving new technology for Europe and keeping the program affordable. Before deciding on the Global Hawk RQ-4B Block 40 for the air segment, NG tried to work through the International Traffic in Arms Regulations (ITAR) process and make the more leading-edge capabilities exportable and deliverable to NATO. However, no new development solution was found that could satisfy export regulations, European participation requirements, nor the affordability requirements, which is why compromises continually had to be made. Financial circumstances in the 2000s caused the program to procure the RQ-4B Global Hawk Block 40 which was an already developed capacity. Even after the PMOU in 2009, AGS on its own does not achieve economies of scale. The additional USAF's procurement of RQ-4B Block 40, however, does create economies of scale, which helps AGS achieve higher cost efficiency.

What is notable about the responses for the F-35 program is that most respondents rated high levels for the program mission being based on both leading-edge technology and economies of scale. On the surface, the survey results reject the hypothesis that there is a tradeoff between leading-edge technology and low-cost economies of scale. The discussions with interviewees confirm, however, that this tradeoff still exists. Since the inception of JAST, which later became the F-35, the program was entirely based on acquiring a leading-edge fifth generation fighter. One interviewee, however, did note that the extent to which leading-edge technology prioritized over affordability depended on the service or partner nation. For example, while the USAF is buying the most F-35s and as a replacement fighter, the USN is using the F-35 to augment its current capabilities. Consequently, the USAF is more likely to rate the level of demand for leading-edge technology as a five or a six while the USN is more likely to rate it at a level four or five. Despite this difference, all the interviewees rated the program on the higher end of the scale of demand for leading-edge technology.

Affordability has been advertised from the outset of the F-35 program. Whether the program emphasized affordability to gloss the brochure sent to potential partner nations, or as a primary focus is unclear. Nothing about the leading-edge technology in the F-35 program is low cost. From the Nunn-McCurdy cost breach to the continuous LRIP, the program has consistently pushed prospects for economies of scale into the future. The program does, however, present opportunities for relatively low-cost production down the road if a global fleet in the four-digits is procured. With a large global fleet, low-cost sustainment will more likely be able to reap the benefits from global spare parts and supply chains, as well as economies of scale. Tagging affordability to this program from the outset was not realistic and could be considered a major source of the criticism the program has seen to date. One interviewee from government emphasized that the entire purpose of acquiring a fifth-generation fighter was technology. If the services wanted economies of scale, they could have procured more F-16s and F-18s for the Air Force and Carrier versions, respectively.

Workshare Distribution

For the workshare distribution characteristic, the study team first asked the interviewees, “On a scale of one to six, rate the extent to which the distribution of workshare was based on participating countries’ comparative advantage.” Second, the study team asked, “Rate the extent to which the distribution of workshare was based on political or industrial-base goals.” The responses are displayed in Figure 8.

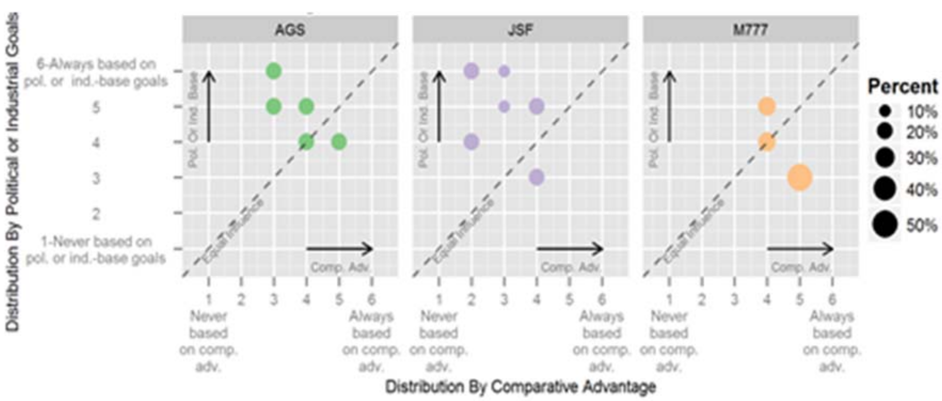


Figure 8. Basis of Workshare Distribution

The responses for the AGS program reflect the program’s blueprint. Interviewees rated the workshare level based on political or industrial-base goals generally higher than how they rated the workshare level based on comparative advantage. This is no surprise, given the political nature of the AGS program’s core goals. On the ground segment, stations were already available from the United States; however, this would have defeated the goal of high participation levels from the European partner nations. The program paid for a NATO ground station because it was ITAR-free and available for use in other European programs. While basing workshare distribution on comparative advantage is often times more economically efficient, the program would not have been able to exist without politically-based decision making. Spain, for example, withdrew from the program after not receiving enough industry participation. Although distributing workshare based on comparative advantage presents opportunities for cost efficiency, political and industrial base factors can be equally crucial in order to sustain program participation.

The workshare distribution for the F-35 program is proportionate to the level of investment a partner-nation contributes to the program. There would not have been factories built from scratch in Turkey where workforces were trained from ground zero if the workshare distribution was not constructed in this fashion. Additional political factors influenced workshare distribution as well. For example, supplementary wing production opened in the state of Georgia because that facility was facing an existential crisis from lack of work, not because producing wings in Georgia was the most efficient allocation of resources. Some countries were disappointed from the low-level of integration for technology-transfer, as technology-transfer laws significantly influenced the distribution of workshare. As one interviewee explained, if a country writes a piece of software, that software is property of the country, not the program. Because of this, technology-intensive production was automatically allocated to the United States.

The M777 program is a prime example of when political factors become more important than comparative advantage. Despite the fact that the UK had been developing the technology needed for this program, the USA and USMC were ordered by the Senate and House Appropriations Subcommittees on Defense in 1999 to develop a plan to utilize Rock Island Arsenal in producing various portions of M777 (U.S. Marine Corps and U.S. Army, 1999). This is yet another example of how monetary costs and benefits are not the only ways to measure efficiency with international cooperation. Furthermore, economic efficiency should not be the only measure of success here.

Initial Results

This interim report discusses what the study team can preliminarily conclude, while the full technical report of this study will provide a more comprehensive conclusion of the hypotheses. The final report will also develop a framework for designing and managing successful international joint development programs. In regards to the second research question,

> How are best practices of international joint development programs in defense acquisition different from best practices of single-nation acquisition programs?,

the study team found that both the single-nation and international defense acquisition programs of today face the different levels of complexities that Drezner (2009) recognized: organizational, environmental, and technical. The underlying difference between single-nation acquisition programs and international joint development programs is the high level of organizational complexity inherent in international cooperation. While modern single-nation defense programs face the complexities associated with integrating government and industry, international programs must also intermingle governments and international industries. To successfully overcome both the environmental and technical complexities of modern programs and manage the inherent organizational complexities of international programs, appropriate governance models that practice consolidation, flexibility, risk-management, and institutional memory are more likely to succeed in reaping the benefits international programs theoretically can achieve.

The following discussion considers the four hypotheses to the highest extent possible at this point in the research effort.[8]

The first hypothesis of this study is as follows:

> The structure of cooperation in international joint development programs matters: The international joint development programs whose stakeholders cooperate only during the development or production process will have less successful cost, scheduling, and end-product outcomes.

The countries participating in all three of the programs analyzed for this report participated during both the development and production phases. Consequently, the study team cannot confidently conclude whether the international joint development programs whose stakeholders cooperated only during the development or production process will have less successful program outcomes. However, the study team can use the results of the survey to extrapolate correlations between how successful programs were in achieving their goals and the level of integration between governments and industries. The interviewees from the M777 program, for example, rated higher levels of integration between government and industry than the interviewees from the F-35 and AGS programs. Looking at the history of the three programs, the M777 program executed the quickest contracting and signing of the PMOU. Additionally, unlike the F-35 program and the AGS program, the M777 program interviewees rated lower levels of integration between governments. The AGS and F-35 programs' interviewees similarly rated higher levels for these two characteristics. The AGS program took the longest to reach contracting and signing of the PMOU, while the JSF has experienced the most cost increases and schedule delays.

Second, the study team hypothesized,

> International joint development programs that are more grounded in security policies rather than economic efficiency interests are more likely to result in negative cost, scheduling, or end-product outcomes.

The AGS program's interviewees rated that the AGS program "often depended" or "always depended" on diplomatic or political needs, while the F-35 and M777 program interviewees never rated more than "often depended" for decision making based on diplomatic or political needs. Similarly, the AGS and F-35 program interviewees rated higher levels for workshare distribution being based on political or industrial-base goals than the interviewees for the M777 program. Program decision making and distribution of workshare are two areas where programs based decisions on costs and comparative advantage, or on political and industrial base goals that often reflect international and domestic security policies. While there were a high number of instances where the AGS program depended more on diplomatic or political needs, this does not always translate to negative outcomes. On the one hand, this could have contributed to the long period of time it took the program to reach a contract and PMOU. On the other hand, this was critical for the program to maintain its end-product goal of including as many NATO member-states as possible in order to put a European face on the program. The program successfully achieved its

[8] At this point in the research effort, the study team has only analyzed three out of the six case studies of this case-study analysis.

Alliance-based goals thanks to the program's ability to meet political factors reflecting different nation's security policies.

Third, the study team hypothesized,

> Countries who have cooperated in defense acquisition before have a higher chance of achieving positive cost, scheduling and end-product outcomes.

As previously discussed, the number of participating countries in a joint international program is a notable characteristic. While the study team has found that a higher number of partner nations supports a program's ability to move forward without cancellation, the number of countries involved often means that choices must be made in light of a diverse number of actors and while satisfying numerous investments, international interests, and domestic interests. Based on this hypothesis, the study team expected to see more positive outcomes with the AGS program based on the fact that the countries involved have historically worked together before through NATO. Information gleaned from the stakeholder interviews suggests that the office dedicated to integrating the program, NAGSMA, did not achieve the positive outcomes of a strong institutional memory because this office was not set up until the official PMOU was signed in 2009, 14 years after program inception. The F-35 program partially reaped benefits from having partner countries who have a history of cooperation in the past. Multiple interviewees noted how the United States, UK, Australia, and Canada were more fluidly integrated in the F-35 program because of their past experiences working together. However, it is hard to connect this to overall program outcomes because most participating nations are not a part of this construct.

Fourth, the study team hypothesized,

> Countries who are uniquely capable of producing complex acquisition programs benefit from working with smaller countries or industries who may have comparative advantages in certain technologies but do not have the capacity to produce complex acquisition programs on their own.

The research conducted thus far strongly supports the fourth hypothesis. For the AGS program, the United States had the ISR capabilities demanded by the program from the outset. They benefited, however, from working with the various-sized countries of NATO, as it ensured that the United States would not be the sole provider of support for NATO operations requiring ISR capabilities. Additionally, the United States benefitted from the partner nation's requirement to have a European face on the program because of the shared maintenance and lifecycle costs. Furthermore, all nations benefitted from international participation for the F-35 program because the leading-edge technology achieved by the program would not have been financially feasible by any one nation.

Concluding Thoughts

Globalizing the defense market at the research and development stages of acquisition poses crucial benefits for partner nations in light of budget pressures in the United States and Europe. In order to reap these benefits, international joint development programs must follow best practices. This interim report has identified eight characteristics critical to programs' capacity to achieve best practices. These eight characteristics impact the program's ability to achieve desired cost, scheduling, and end-product outcomes.

Based on the work conducted thus far, several of the characteristics are promising grounds from which best practices can be derived. For instance, interviewees across multiple programs agreed that including more member countries adds organizational complexity which can cause negative program outcomes such as schedule delays. Starting with a small group before reaching the PMOU may avoid some of these problems, although as AGS showed, a large group of countries does give a program momentum. Choosing the right partners is also important; as discussed in hypothesis three, institutional memory of past collaboration can contribute to better results. In other words, countries who are new to working together face greater challenges but simultaneously pave the way for future success. Additionally, many programs appear to be overestimating their ability to simultaneously pursue leading edge technology and economics of scale, with the latter often falling by the wayside. Multiple interviewees mentioned setting key parameters and anticipating technology transfer hurdles early in the process. However, adopting an attitude of humility about what joint development projects can actually achieve from a cost perspective may also be a critical first step.

The study team will continue analyzing three additional cases, the Standard Missile-3 Block IIA program, the A400M Atlas program, and the Medium-Extended Air Defense System program, in addition to those discussed in this report. This analysis will further develop the results of the hypotheses and bolster the framework for designing and managing international joint development programs in the future.

References

Chao, P. A. (2004). NATO AGS—*Finally ready to fly?* Center for Strategic and International Studies. Retrieved from http://csis.org/files/media/csis/pubs/0407_natoags.pdf

DoD. (2010). *Selected Acquisition Report (SAR) F-35.* Retrieved from http://fas.org/man/eprint/F-35-SAR.pdf

DoD. (2011). *Selected Acquisition Report (SAR) F-35.* Retrieved from http://www.defense-aerospace.com/dae/articles/communiques/F-35Dec11FinalSAR-3-29-2012.pdf

DoD. (2012a). *F-35 Joint Strike Fighter program*. Retrieved from http://www.dod.mil/pubs/foi/Reading_Room/Selected_Acquisition_Reports/F-35_December_2012_SAR.pdf

DoD. (2012b). *Sustaining U.S. global leadership: Priorities for the 21st century defense.* Retrieved from http://archive.defense.gov/news/Defense_Strategic_Guidance.pdf

DoD. (2015). *FY15 DOD programs: F-35 Joint Strike Fighter (JSF).* Retrieved from http://aviationweek.com/site-files/aviationweek.com/files/uploads/2016/01/DOT&E 2015 F-35 Annual Report.pdf

Drezner, J. A. (2009). Competition and innovation under complexity. In G. Ben-Ari & D. Berteau (Eds.), *Organizing for a complex world* (pp. 31–49). CSIS Press.

GAO. (2000). *Howitzer Program experiencing cost increases and schedule delays.* Retrieved from http://www.gao.gov/products/NSIAD-00-182

GAO. (2003). *Best practices: Setting requirements differently could reduce weapon systems' total ownership costs*. Retrieved from http://www.gao.gov/assets/160/157396.pdf

Gertler, J. (2014). F-*35 Joint Strike Fighter (JSF) program.* Retrieved from https://www.fas.org/sgp/crs/weapons/RL30563.pdf

Hartley, K. (2012). *White elephants? The political economy of multinational defence projects.* New Direction, The Foundation for European Reform. Retrieved from http://europeanreform.org/

North Atlantic Treaty Organization (NATO). (2009, September 25). NATO's Allied Ground Surveillance programme signature finalised. Retrieved from http://www.nato.int/cps/en/natolive/news_57711.htm

North Atlantic Treaty Organization (NATO). (2016). Alliance Ground Surveillance (AGS). Retrieved from http://www.nato.int/cps/en/natolive/topics_48892.htm

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. (2012). *International cooperatoin in acquisition, technology and logistics (IC in AT&L) handbook.* Retrieved from http://www.acq.osd.mil/ic/Links/IChandbook.pdf

U.S.Marine Corps and U.S. Army. (1999). *Lightweight 155mm Howitzer: Report to the Senate and House Appropriations Subcommittees on Defense*. Retrieved from www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA375183

Walker, D. M. (2003). *Defense acquisitions assessments of major weapon programs.* Retrieved from http://www.gao.gov/new.items/d03476.pdf

Panel 20. Analyzing and Incentivizing Performance Outcomes

Thursday, May 5, 2016	
3:30 p.m. – 5:00 p.m.	**Chair: Mark Deskins,** Director, Acquisition Career Management, ASN(RD&A) ***Big Data Analysis of Contractor Performance Information for Services Acquisition in DoD: A Proof of Concept*** Uday Apte, Professor, NPS Rene Rendon, Associate Professor, NPS Mike Dixon, Assistant Professor of Operations Management, Ivey Business School ***Knowledge, Experience, and Training in Incentive Contracting for the Department of Defense*** Kevin Carman, Dean, DAU Randall Gibson, Professor, DAU ***Learning in the Shadows: A Retrospective View*** Donna Kinnear-Seligman, Mission Assistance Program Analysis Manager, DAU

Mark Deskins—rejoined the ASN(RD&A) staff in May 2015 to serve as the Director, Acquisition Career Management (DACM), the Navy and Marine Corps' lead for the professional development and management of the DoN acquisition workforce. With over 25 years of professional experience, Deskins' spectrum of experience includes private industry and Navy field, headquarters, and program office assignments. From February 2009 to May 2015, Deskins was the Deputy Program Manager for Strategic and Theater Sealift (PMS 385), which included two ACAT programs: the Joint High Speed Vessel (JHSV) and the Mobile Landing Platform (MLP). During his tenure, both programs went from Milestone B to Initial Operating Capability, delivering five JHSVs and two MLPs. Prior to joining PMS 385, Deskins was the Chief of Staff for Team Ships and was responsible for front office operations and executive coordination and oversight of the vast array of issues related to surface ship acquisition, maintenance, modernization, and disposal. From 2004 to 2007, he was a member of the ASN(RD&A) staff and was responsible for executive oversight of issues related to surface ship, submarine, and carrier maintenance, modernization, and disposal.

From 2001 to 2007, Deskins was the Deputy Program Manager for Inactive Ships (PMS 333) in PEO (Ships), where he was responsible for the planning, programming, budgeting, and execution of the U.S. Navy's inactivation and disposal of conventionally powered surface ships. As the last stage of the acquisition life cycle, he oversaw the decommissioning of 20 ships and the disposal of 46 ships through foreign military sales (6), scrapping and recycling (12), fleet support through SINKEX (26), artificial reefing (1), and the donation of the aircraft carrier ex-Midway that operates today as a highly successful ship museum in San Diego.

Deskins' previous jobs include working five years as an industrial engineer in private industry in North Carolina and beginning his government career as an industrial engineer at the Indian Head Division of the Naval Surface Warfare Center (NSWC). Promotions brought him to the NSWC headquarters staff, where he worked in strategic planning and corporate operations. He was selected for NAVSEA's Commander's Development Program, where he held highly responsible positions working on combat systems, strategic and business planning at the Division and NAVSEA corporate level, and PEO Program Management assignments.

His awards include two Navy Superior Civilian Service Awards in 2004 and 2014, two Navy Meritorious Civilian Service Awards in 1994 and 2000, and several Special Act Awards and Outstanding Performance Awards. He is an acquisition professional, DAWIA Level III certified in Program Management, and has held memberships in the International Council on Systems Engineering and the Institute of Industrial Engineers. Deskins' education includes a bachelor's degree in industrial engineering from West Virginia University and a master's degree in technology management from the University of Maryland University College. His executive education includes Harvard Business School and the University of North Carolina.

Big Data Analysis of Contractor Performance Information for Services Acquisition in DoD: A Proof of Concept

Uday Apte—is a Distinguished Professor of Operations Management at the Graduate School of Business and Public Policy (GSBPP), Naval Postgraduate School, Monterey, CA. He also serves as the Associate Dean of Research and Development at GSBPP. Before joining NPS, Dr. Apte taught at the Wharton School, University of Pennsylvania, Philadelphia, and at the Cox School of Business, Southern Methodist University, Dallas. He is experienced in teaching a range of operations management and management science courses in the executive and full-time MBA programs. Prior to his career in academia, Dr. Apte worked for over 10 years in managing operations and information systems in the financial services and utility industries. Since then he has consulted with several major U.S. corporations and international organizations.

Dr. Apte's research interests include service operations management, supply chain management, technology management, and globalization of information-intensive services. He has completed over 10 sponsored research projects for the U.S. Department of Defense and has published over 60 articles, five of which have won awards from professional societies. His research articles have been published in prestigious journals including *Management Science*, *Interfaces, Production and Operations Management (POM), Journal of Operations Management (JOM), Decision Sciences Journal (DSJ), IIE Transactions,* and *MIS Quarterly*. He has co-authored two books, *Manufacturing Automation* and *Managing in the Information Economy*. Dr. Apte has served as a vice president of colleges at Production and Operations Management Society (POMS), as a founder and president of the POMS College of Service Operations, and as guest editor of POM journal. Currently he serves as the senior editor of POM and as associate editor of DSJ.

Dr. Apte holds a PhD in Decision Sciences from the Wharton School, University of Pennsylvania. His earlier academic background includes an MBA from the Asian Institute of Management, Manila, Philippines, and Bachelor of Technology (Chemical Engineering) from the Indian Institute of Technology, Bombay, India. [umapte@nps.edu]

Rene G. Rendon—is an Associate Professor at the Graduate School of Business and Public Policy, NPS, where he teaches defense acquisition and contract management courses. He also serves as the Academic Associate for the MBA specialization in contract management. Prior to joining the NPS faculty, he served for over 20 years as an acquisition contracting officer in the United States Air Force. His career included assignments as a contracting officer for the Peacekeeper ICBM, Maverick Missile, and the F-22 Raptor. He was also a contracting squadron commander and the director of contracting for the Space Based Infrared Satellite program and the Evolved Expendable Launch Vehicle rocket program. Rene has published in the *Journal of Public Procurement*, the *Journal of Contract Management*, the *Journal of Purchasing and Supply Management*, and the *International Journal of Procurement Management*. [rgrendon@nps.edu]

Michael Dixon—is an Assistant Professor of Operations Management, Ivey Business School. He is interested in service operations management and specifically the effect of operational decisions and service design on customer experience. Dr. Dixon conducts management research from an analytical perspective and analyzes large archival data sets to model customer behavior, finding ways to represent the behavioral aspects of service design in quantitative and analytical ways. Specifically, he is interesting in understanding the role of experiential sequence on participant perception. In past research, he analyzed season subscription ticket sales for a large performing arts venue and showed that season subscription re-purchases are influenced by event sequences from previous seasons, leading to managerial direction for event planners to properly schedule performances across a season to maximize experiential effects. He received the Liskin Teaching Award from the Naval Postgraduate School in 2014 and 2015.

Dr. Dixon is a co-editor of the Production and Operations Management Society (POMS) Chronicle newsletter. He has publications in the *Journal of Operations Management* and articles under review at the *Production and Operations Management*, *Journal of Service Management*, and *Journal of Service Research*. He has served as a reviewer for *Management Science, Production and Operations*

Management, *Journal of Operations Management*, *Journal of Services Research*, *Service Science*, and the *Decision Sciences Journal of Innovative Education*. [mdixon@ivey.uwo.ca]

Abstract

This paper explores the use of Big Data analytic techniques to explore and analyze large datasets that are used to capture information about DoD services acquisitions. We describe the burgeoning field of Big Data analytics, how it is used in the private sector, and how it could potentially be used in acquisition research. We test the application of Big Data analytic techniques by applying them to a dataset of CPARS (Contractor Performance Assessment Reporting System) ratings of acquired services, and we create predictive models that explore the causes of failed services contracts using three analytic techniques: logistic regression, decision tree analysis, and neural networks. The report concludes with recommendations for using Big Data analytic techniques in acquisition.

Introduction

In April 2015, the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) issued his implementation guidance for Better Buying Power (BBP) 3.0 with the theme Achieving Dominant Capabilities through Technical Excellence and Innovation. The purpose of the BBP 3.0 acquisition initiative is to strengthen the Department of Defense's (DoD's) efforts in innovation and technical excellence while also continuing the DoD's efforts to improve efficiency and productivity (USD[AT&L], 2015). One of the major components of BBP 3.0 is its emphasis on improving the tradecraft in acquisition of services. The implementation guidance focuses on strengthening the contract management function for installation level services, improving requirements definition in the services acquisition process, and improving the effectiveness and productivity of contracted engineering and technical services.

It is not surprising that the USD(AT&L) has focused on improving services acquisition in the DoD. Services contracting specifically, and contract management generally, have been identified as a "high risk" by the Government Accountability Office (GAO). Since 1992, the GAO has found that the DoD lacks an adequate number of trained acquisition and contract oversight personnel, uses ill-suited contract arrangements, and lacks a strategic approach for acquiring services (GAO, 2015). Additionally, the GAO has reported that the DoD lacks adequate data needed to inform its decision-making on services acquisition and contract management. The GAO has also stated that the DoD lacks established metrics to assess its progress in improving services acquisition, and that the DoD should leverage its acquisition data by developing baselines to identify trends, thereby enabling it to develop measurable goals and gain more insight into whether its initiatives are improving services acquisition.

The purpose of this research is to explore how the DoD can leverage acquisition data, specifically contractor performance information, in identifying drivers of success in services acquisition. Through the use of exploratory descriptive and predictive statistical models, we describe and uncover the drivers of low and high contractor performance scores. In uncovering and describing these drivers, we develop recommendations for cost-effective management of services acquisition. Furthermore, we perform additional statistical analysis to determine if there is any relationship between contractor performance assessment factors (quality, schedule, cost, business relations, and management of key personnel), service type, contract type, level of competition, and contract dollar value. In researching the relationships among these variables, we perform predictive-modeling-based statistical methodology appropriate for Big Data including predictive regression modeling, decision-tree analysis, and neural-network analysis to determine which variables—

contractor performance data, contract characteristics, and management approach—can be considered the drivers of success for services acquisition.

This research report is organized into five sections. This introductory section is followed by the second section which reviews our past research in services acquisition with a focus on investigations into contractor performance information and drivers of success in services acquisition. The third section provides a primer on the use of Big Data analytics and selected Big Data analysis tools. The fourth section provides the results of our Big Data analysis on contractor performance information and its relationship to drivers of success in services acquisition. We complete the report in the fifth section with conclusions and recommendations for using Big Data analysis in investigating success drivers in services acquisition.

Past Research

We have addressed the need for research in the increasingly important area of services acquisition by undertaking six sponsored research projects over the past several years. The first two research projects (Apte et al., 2006; Apte & Rendon, 2007) were exploratory in nature, aimed at understanding the types of services being acquired, the associated rates of growth in services acquisition, and the major challenges and opportunities present in the service supply chain.

The next two research projects were survey-based empirical studies aimed at developing a high-level understanding of how services acquisition is currently being managed at a wide range of Army, Navy, and Air Force installations (Apte, Apte, & Rendon, 2008, 2009). The analysis of survey data indicated that the current state of services acquisition management suffers from several deficiencies, including deficit billet and manning levels (which are further aggravated by insufficient training and the inexperience of acquisition personnel) and the lack of strong project-team and life-cycle approaches. Our research (Apte, Apte, & Rendon, 2010) also analyzed and compared the results of the primary data collected in two previous empirical studies involving Army, Navy, and Air Force contracting organizations so as to develop a more thorough and comprehensive understanding of how services acquisition is being managed within individual military departments.

As a result of these research projects dealing with the service supply chain in the DoD, we have developed a comprehensive, high-level understanding of services acquisition in the DoD, have identified several specific deficiencies, and have proposed a number of concrete recommendations for performance improvement.

In our research, we analyzed 715 Army Mission Installation Contracting Command (MICC) service contracts found in the PPIRS database. These contracts were specifically for professional and administrative, maintenance and repair, utilities and housekeeping, and automated data processing and telecommunication services (Hart, Stover, & Wilhite, 2013). The results of our analysis of contract variables and contract success (Rendon, Apte, & Dixon, 2014) are summarized as follows.

- Utilities and housekeeping services had the highest failure rate of all the product service codes analyzed. The reasons for contract failure included business relations and management of key personnel.
- Contracts with a dollar value from $50 million to $1 billion had the highest failure rate of all the contract categories. This group's most common reason for failing was cost control.

- Contracts awarded competitively had the highest failure rate when compared to the other two forms of competition available. The reasons that most often resulted in a contract failure were in the areas of schedule and cost control.
- Contracts structured as a combination contract type had the highest failure rate when compared to the other five types of available contracts.

In this past research (Rendon et al., 2014), we further analyzed our contract data to determine whether any of the variables had a significant relationship with contract success by specifically looking at the contract failure rates. We used the chi-square test (Fisher's exact test) to test if the actual failure rates are significantly different than what would be expected if the total contract failure rate was applied to each variable. The results of the chi-square test identified that Contractual Amounts and Contract Type were our only statistically significant variables.

We also looked at the relationships between percentage of filled 1102 billets and failure rates, and between workload dollars per filled billet and failure rates, and made some interesting observations. We saw that as the percentage of 1102 filled billets increased, the contract failure rate decreased. This would seem intuitive, that as the workforce increases, the contract success rate would also increase, since there would be sufficient resources to manage the contracting process.

In our most recent research (Rendon, Apte, & Dixon, 2015), using the original data set of 715 Army service contracts (Hart et al., 2013), we analyzed the narrative section of the CPARS (Contractor Performance Assessment Reporting System) reports to determine alignment with the objective assessment ratings (Black, Henley, & Clute, 2014). Based on interviews, we also analyzed the value added, not only of the narrative section, but also of the usefulness of the CPARS as a contractor assessment tool. Our focus was to recommend improvements to the CPARS contractor performance information documentation process. The results of our analysis of CPARS narratives and interviews, reported earlier in Black, Henley, and Clute (2014), are summarized as follows.

- The contracting professionals are doing a better job at providing beneficial CPARS data in the narrative when the contract is unsuccessful versus when it is successful.
- The contracting professionals were slightly better at matching the narrative sentiment to the objective scores in unsuccessful contracts than in successful contracts.
- The results of the interviews found that the CPARS database is still often not reliable, robust, or comprehensive enough. The interviews also reflected that unsuccessful contracts tend to have more reliable, robust, and comprehensive past performance information available in their CPARS reports. The interviewees also stated that the information found in the PPIRS database sometimes contains information in the narrative that is either contradictory or does not quite match up with the objective ratings.

In our current research, we use exploratory descriptive and predictive statistical models to describe and uncover the drivers of low and high contractor performance ratings. Additionally, we perform statistical analysis to determine if there is any relationship between CPARS factors and contract variables, as reflected in Figure 2. In researching the relationships among these variables, we perform predictive-modeling-based statistical methodology appropriate for Big Data including predictive regression modeling, decision-tree analysis, and neural-network analysis to determine which variables—CPARS factors, contract variables, characteristics, and management approach—can be considered as the drivers of success for services acquisition. The next section of this report provides a primer on the use of Big Data analytics and the various types of Big Data analysis tools.

Big Data Analysis

The term *Big Data* is fairly new in modern business nomenclature. It refers to the massive influx of data that has been and is currently being collected in the digital and Internet era. In some estimates, 90% of the data that is currently being stored on computers and servers around the world was collected in just the past two years (Baesens, 2014, p. 1). Other authors (Mayer-Schoenberger & Cukier, 2013, p. 28) cite that in the year 2000, only one quarter of the world's data was digitized; the remainder was on paper and other analog media. However, by 2013, 98% of all data was digital.

The flood of data comes primarily from the digitization of processes, interactions, and communications brought about by digital innovations such as internet-consumerism, mobile technology, and social networking (Mayer-Schoenberger & Cukier, 2013). In addition, data storage capacity is becoming ever cheaper, making it easier to keep data indefinitely. The term *datafication* refers to turning aspects of life that, in the past, have never been quantified into data that can be analyzed; for example, GPS coordinates are being recorded in mobile transactions or photos, photo images are being "datafied" to find face matches by Facebook, and words and sentences from Twitter status updates are being analyzed for content and sentiment using various text analysis techniques.

The term *Big Data* is used to discuss how to store, manage, and—perhaps most importantly—analyze these large stocks of data. Specifically, Big Data analytics refers to the ability to make distinct observations from large amounts of data that might not be able to be inferred from smaller amounts (Mayer-Schoenberger & Cukier, 2013). According to these authors, Big Data analytics differ from traditional statistics in three important ways. First, sample sizes are much bigger, approaching at times the size of an entire population. Traditionally, statisticians use small, unbiased samples to make inferences about larger populations, which has worked well for simple questions. Complicated sampling techniques have to be deployed for more complex, layered questions in order to make inference about specific sub-groups of a population. Second, Big Data analytics have to settle with unclean data. Finally, Big Data analytics leads to correlational explanations and not causational, that is, the results of Big Data analytics can only be interpreted as correlational relationships between variables.

The new term *data science* refers to the skillset needed to make sense of Big Data (see Schutt & O'Neil, 2013). A data scientist is made up of equal parts computer scientist, statistician, mathematician, and graphic designer, with capabilities to pull and combine datasets; manipulate, clean, and analyze data; and communicate aggregate results in a meaningful way. Data scientists are found across multiple sectors, including journalism, academia, information technology, banking, insurance, sports, and government.

Big Data is used by computer scientists that feed computers volumes of data with hopes that computers can make inferences on the probability of intuitive analytics that, in the past, have proven very difficult to teach to a computer. The success of the IBM Watson project provides evidence that Big Data analytics can outperform the world's most clever trivia masters. Big data analytic techniques are being used to generate algorithms for computer learning, search engines, and risk management.

The focus of this paper is to describe, as a proof of concept, how Big Data analytic techniques could be used to further the understanding of successes and failures of the DoD and other federal service contracts. Using the CPARS data previously described, we consider the range of analytics that could be used to expand the research and practice of service acquisitions.

Typical Form of Big Data

Datasets used for Big Data analytics are usually formed by taking multiple measurements of multiple cases. Data is organized in rows and columns. Data in the same row are all from the same case or observation, and the columns have the same measurement or variable for all cases. Typically, a dataset's size is described by the number of cases and its number of variables. One of the variables is an identification number that is unique for that individual case. There may also be other identification variables that can be used to describe the case's membership to some other category; for example, the zip code, state, unit, etc. Identification variables can be used to extract data from other sources, adding to the number of variables available for analytic modeling.

Analytical modeling is a term that describes various methods that specifically quantify relationships between variables using past data as an indicator of how relationships form and how they might exist in the future. In predictive analytics, analysts create models that attempt to explain relationships between a specific target variable (sometimes called a dependent variable) and any number of input or independent variables. Analytic modeling has two important tasks: to predict outcomes of future cases, and to quantify relationships between inputs and target variables. These two tasks are not always congruent; at times a model might be very good at predicting future cases while at the same time present a challenge in interpreting relationships found in the data.

In most cases, target variables are either continuous across a large scale (e.g., dollars, time, or distance) or categorical with just two categories, that is, binomial (e.g., defaulting on a loan, failing an assessment, or repurchasing of a product). Binomial target variables take the form of either "yes" or "no." Less common, but still available, is predictive modeling with categorical target variables with more than just two categories.

Predictive modeling uses probability and statistics to estimate relationships between variables. In traditional statistics, a sample of cases is used to make these estimations and the model is used to infer something about a larger or future population. Using larger samples sizes found in Big Data allows the analyst to compare a model's ability to predict and describe relationships with existing data; analysts will randomly select a percentage of cases to be withheld during the model building phases. After a model is proposed, an analyst will "validate" the model using the withheld dataset to see how it would perform using existing data. Having a "validation" dataset adds to the ability to use the model outside the sample that is used to create it.

Predictive analytic models, estimated using Big Data, can provide a good indication of how target variables can be predicted using other measurements of a case. Predictive models are used widely in situations in which there is a complex set of variables, some of which might be correlated to a target variable for part of the time. Take, for example credit scoring in which lending companies will use a predictive model to assess the risk that a borrower might default on a loan (binomial target variable). Creating models using data from past lenders, a portion of which defaulted, credit issuers can make decisions about whom to offer credit. The model might show that people who are young and have little income are at high risk of default. However, the quantifiable relationships that make up the model are entirely correlational and cannot be said to cause default; that is, being young with low income does not cause default. We stress this important point that predictive models are correlational and should not be used to describe causes of target variables.

Decision Tree Analysis

Decision tree analysis is a predictive analytics technique that attempts to identify and isolate portions of a dataset that seem to act in similar ways in regard to a target variable. Target variables can be binary, nominal, or continuous. The purpose of a decision tree analysis is to propose a set of rules that can be used to estimate or predict a target variable.

To begin decision tree analysis, the methodology first identifies the independent variable that most discriminates the target variable, that is, the one in which a separation will lead to the most divergent prediction of the target variable. This is done by considering what the typical target variable will be if the data is divided at points within the range of values of all the independent variables. Most software that conducts decision tree analysis will algorithmically consider all division across all independent variables, giving each divergent scores using one of various methods. The independent variable with the highest divergent score is usually chosen to be the first "branch" in the decision tree. The division of the data results in "nodes" that are further divided by other variables in the same manner, resulting in a tree in which the "root" is on the top and the "branches" go down. The final "nodes" are called "leaves" and give a prediction of the target variable for data that fits within the path that leads to it. What results is a fan-shaped visual depiction of simple decision-based models that can be used to predict the target variable. In addition to providing a prediction model, decision tree models also provide a good interpretation of how different values of independent variables impact a target variable.

Typically, the more branches in a tree, the better a model can predict target variables in a training dataset; analysts typically have to set rules about when to stop branching within the training dataset. However, it is often the case that only a few branches are appropriate for validation data. To combat overfitting, an analyst can "trim" the branches of the tree back to only those that contribute to the prediction of the target variable of the validation data.

Logistic Regression

The next method we discuss is modeling a binomial decision variable using regression techniques. Linear regression is taught in most college-level statistics courses. In traditional regression, an analyst will estimate a model predicting a continuous target variable using any number of both continuous and discrete independent variables. In decision tree analysis, the "model" resulted in a visual tree diagram that can be used to interpret and predict outcomes of cases; in regression the result of the modeling is a mathematical equation that can consider values of new case in order to predict the target. Traditional regression analysis is considered "linear" because the resulting mathematical model is in the form of a linear equation representing a line, or a multi-dimensional surface, that has slope and intercept. The equation of traditional linear regression analysis takes the following form:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots\ b_nx_n \qquad (1)$$

In Equation 1, the x_{n} are the values of each of the independent variables and collectively the equation can be used to predict the value of a target variable, $\hat{y}$. The "slope" portion of the equations are called "coefficients" and can be used to formally and explicitly describe relationships between independent and target variables. In the previous equation, the $b_1, b_2, \dots b_n$ are the coefficients that are estimated for each of the independent variables. The coefficients are "estimates" in the same way that the average of a sample is an estimate of the average of an entire population. Through independent hypothesis testing, a p value for each coefficient is calculated that can be used by analysts to determine if a coefficient significantly influences estimation of the target variable (recall that a low p value means that a coefficient is significant).

The traditional linear regression assumes that the target variable is continuous (e.g., temperature, weight, dollars) across a scale. When a target variable is binary (e.g., defaulting on a loan, failing an assessment, or repurchasing of a product), analysts use an extension of traditional linear regression called logistic regression. In logistic regression, the target variable takes on the binary form of zeros and ones, that is, the analyst assigns one of the two options to take the value of 1 and the other to take the value of 0. In traditional regression, the estimated model can be used to predict the actual values of the continuous target variable; in logistic regression, the equation will instead predict the *probability* that the case will take the value of 1 (instead of 0). The equation for a logistic regression takes the following form:

$$Prob(y = 1 \mid x_1, x_2, \ldots x_{n1}) = \frac{1}{1+e^{-(b_0+b_1x_1+b_2x_2+\cdots b_nx_n)}} \quad (2)$$

Equation 2 reads that the probability that the target variable y is equal to 1 given a set of independent variables $(x_1, x_2, \ldots x_{n1})$ is equal to the fraction that has 1 as the numerator and $1 + e^{-(b_0+b_1x_1+b_2x_2+\cdots b_nx_n)}$ as the denominator. The form of the fraction ensures that the probability will be between 0 and 1 and the exponential function allows the traditional linear equation $(b_0 + b_1x_1 + b_2x_2 + \cdots b_nx_n)$ to be represented linearly even if the target variable is binomial. Using the past data, software packages use an algorithm called "maximum likelihood" to find the value of the coefficients that best fit the past data to the equation form.

Typically the interpretation of the coefficients $b_1, b_2, \ldots b_n)$ are converted into "odds" or more precisely into "log odds." Odds are the ratio of probabilities; for binomial variables, odds can be represented as follows:

$$Odds\ (y = 1) = \frac{Prob(y=1)}{Prob(y=0)} \quad (3)$$

Since we are dealing with binomial variables, this can be rewritten as follows:

$$Odds\ (y = 1) = \frac{Prob(y=1)}{1-Prob(y=1)} \quad (4)$$

Reformulating the previous regression equation model in terms of odds, we get the following:

$$\ln\left(\frac{P(y=1)|(x_1,x_2,\ldots x_{n1})}{P(y=0)|(x_1,x_2,\ldots x_{n1})}\right) = b_0 + b_1x_1 + b_2x_2 + \cdots b_nx_n \quad (5)$$

The right-hand side of the reformulated equation now mimics the linear regression equation and is now linear in term of log odds. This reformulation is called a "logit transformation." In order to interpret the coefficients from a logistic regression, an analyst would typically calculate the exponent of the coefficient e^{b_n} and interpret it in terms of the original probability equation. For example if the exponent variable is above 1, say 1.8, you would say that the probability that the target variable would take the value of 1 will increase by 80% for every unit increase in the independent variable. If the exponent variable is below 1, say .80, you would say that the probability that the target variable will decrease by 20% for every unit increase in the independent variable.

Just like in decision tree analysis, regression models can be "overfit" by including too many non-generalizable independent variables. In addition, analysts using regression methodologies need to be aware that when independent variables are highly correlated with one another, the interpretation of the model is called into question (this problem is called multicollinearity). Deciding which variables to use in a model is typically done in one of two ways: (1) independent variables are chosen based on preconceived or theoretical understanding, or of their relationship with the target variable; or (2) independent variables are considered algorithmically to determine their individual contribution to an overall model. This algorithmic consideration of independent variables is typically known as “step-wise” regression and consists of calculating the “goodness-of-fit” for models with differing combination of possible independent variables. The model that can explain the most amount of the variation of the target variable with the least amount of independent variables is usually chosen because of its “parsimonious” appeal, that is, its ability to explain with little complication.

Neural Networks

The final type of data analytics technique that we evaluate in this research is neural networks. Neural networks gets its name from neural pathways and connection in brains; the way ideas, thoughts, and facts are connected together in a dense web of connections within the brain. These pathways often have nodes that act as connectors between disparate paths. In neural networking with Big Data, algorithms are deployed to uncover layers of connecting nodes between different independent variables in order to better predict the target variable.

Neural networks essentially involves creating a series of regressions to uncover hidden connecting nodes which are in turn used as input for additional regressions to find deeper connecting layers, eventually leading to a regression model of a prediction of a target variable. In short, it is a series of regression models uncovering latent connecting layers of data that can, in turn, be used to better predict target variables. Analysts can control the level of connecting layers and which independent variables to use in the initial phases. The end result is a prediction model that can be verified using an independent validation dataset. The logical structure of a neural networks model with a single hidden layer is shown in Figure 1.

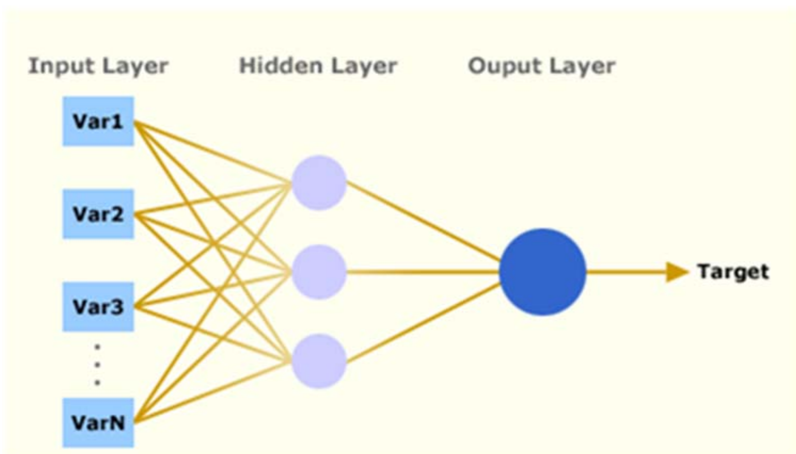


Figure 1. Logical Structure of Neural Networks Model

As we discussed in the previous section, regression techniques generally force analysts to create “linear” models, but using neural networks, analysts are able to model complex, nonlinear relationships using the intermediate layer nodes. The hidden layer nodes are able to handle the complexity of conditional (if/then) modeling that is not possible using traditional regression techniques.

Neural networks tend to work well with large datasets for which the analyst has very little preconceived theoretical model in mind. The results of a neural networks model are extremely difficult to interpret, and, as such, it is used primarily as a prediction modeling technique as opposed to a descriptive or explanatory technique. Typically, the analyst is unable to describe the explicit connection between independent and dependent variables due to the complexities of the intermediate nodes.

Concluding Remarks

In addition to these methods, Big Data analysts are also concerned with topics such as missing data, data transformations, and model validations. Model validation will be addressed in subsequent discussions about training and validation datasets. Data transformations is a topic that is too broad for this paper, typically makes interpretation of results very challenging, and often leads to "overfitting" of the data. Missing data is often approached by "imputing" a value for data that is missing based on the mean or modes of the variable. In some cases, an analyst will infer a missing value based on a regression type formula with the missing value as the target variable. In our subsequent analysis, we imputed a small amount of missing data by replacing missing values with the mean value.

Big Data Analysis in Acquisition Research

A. Data Collection and Preparation

As mentioned earlier, the contract data used in our research was collected with the assistance of our graduate students (Hart et al., 2013). We searched the PPIRS database to identify Army Mission Installation Contracting Command (MICC) services (non-systems) contracts for the period 1996–2013. This search yielded 14,395 contracts in total. The data was then refined to include only those contracts associated with the following product/service codes:

- R: Professional, Administrative, and Management Support Services
- J: Maintenance, Repair, and Rebuilding of Equipment Services
- S: Utilities and Housekeeping Services
- D: Automatic Data Processing and Telecommunications Services

Based on the filtering for the previously mentioned service contracts, we identified 5,621 contracts. We then further filtered this database to include only contracts from the following Army MICC field directorate offices (FDOs) contracting organizations:

- MICC Region Fort Eustis
- MICC Region Fort Knox
- MICC Region Fort Hood
- MICC Region Fort Bragg
- MICC Region Fort Sam Houston

This data filtering resulted in 715 service contracts that were used in conducting our analysis, as seen in Table 1.

Table 1. Database Breakdown
(Hart et al., 2013)

	Total Contracts
Total Army MICC Non-System Contracts	14395
Less: Non R, J, S, D Service Contracts	8774
Total R, J, S, D Service Contracts	5621
Less: R, J, S, D Service Contracts at other MICC	4906
R, J, S, D Service Contracts at MICC FDO Eustis, Knox, Hood, Bragg, Sam Houston	715
Fort Eustis	238
Fort Knox	119
Fort Hood	114
Fort Bragg	55
Fort Sam Houston	189

For each contract, data was collected on specific contract variables (type of service, contract dollar value, level of competition, contract type) and specific contractor assessment ratings (quality of product/service, schedule, cost control, business relations, management of key personnel, and utilization of small business). Determining a contract to be successful or unsuccessful was made based on whether the contractor received a marginal or unsatisfactory rating in any of the CPARS assessment areas (quality of product/service, schedule, cost control, business relations, management of key personnel, or utilization of small business). The contractor receiving a marginal or unsatisfactory rating in any one of these assessment areas results in the determination of the contract as unsuccessful. It should be noted that the data collected from the PPIRS database was sanitized by removing identifiable data such as contract number, contractor name, DUNS number, and place of performance.

In addition to the contractor performance information accessed from the PPIRS-RC database, we also collected MICC region organization demographic data (annual workload in dollars, annual workload in actions, number of 1102 billets authorized, and percent of 1102 billets filled; Hart et al., 2013). This data was also analyzed to determine if these organizational demographics were related to contract success.

During our research we were able to receive access to PPIRS query tool that allows users to look up CPARS records individually. Unfortunately, we were not able to gain access to the CPARS databases with PPIRS directly; instead, we were required to pull records one at a time in order to conduct research. As previously described, our research team was able to pull 715 CPARS records (cases). While this is not a “Big Data” dataset, we believe that the actual CPARS dataset stored in PPIRS in its entirety is indeed Big Data. To our knowledge, there has been little to no research into this dataset. Therefore, in this paper we propose several techniques that could be used to gain information from the Big Data that is being recorded and stored by the federal acquisition community.

Because our dataset is fairly small in Big Data terms, the results of our analysis should not be construed as being conclusive or indicative of general trends. However, if we are able to gain access to more or all of the CPARS records, the same analytics that we explore in the remainder of this paper can be used to gain a rich understanding of the dynamic and complex relationships between contracting attributes and CPARS scores. We intend to petition the gatekeepers of the CPARS records to make available the entire dataset so as to go forward with improved analytics.

In the following sections, we focus on three predictive modeling techniques: decision tree analysis, logistic regression, and neural networks. Each of these techniques has unique strengths to help researchers understand underlying relationships. All three are predictive modeling techniques that create models to predict a target variable. In our case, we use the CPARS data that we had collected for the previous studies; we use as a target variable a binomial indication of contract failure as previously described (a contract with either a marginal or unsatisfactory rating in any of the CPARS assessment areas.) As possible input variables we use the following variables:

- MICC
- Contract Start Month
- Contract Start Day
- Contract Start Year
- Contract End Month
- Contract End Day
- Contract End Year
- Fiscal Year of Contract
- Duration in days
- Contract Type: RJSD
- Awarded Dollar Value
- Current Dollar Value (at time of CPARS)
- Basis of Award
- Type of Contract (FFP, CPFF, CPAF, etc.)
- Annual Workload of Contracting Office (Dollars)
- Annual Workload of Contracting Office (actions)
- # of 1102 Billets Filled by Contracting Office
- % of 1102 Billets Filled by Contracting Office
- Workload ($) by Filled Billet
- Workload (actions) by Filled Billet

All analysis done in the following section was conducted using SAS Enterprise Miner, a leading software for Big Data analysis.

The first step in conducting any of the three types of analysis is to divide the original dataset into two datasets, the first being called a “training” dataset and the second called a “validation” dataset. The training dataset is used to create the analytical model, while the validation data is used to determine if the model is “overfit,” that is, if the model is too dependent on the training dataset to be applicable to other data. The validation data then "validates" the model that was created using the training dataset. Overfitting is a problem if the model is going to be used to predict target variables from observations outside what was used in the training dataset. In our case, we specified that 80% of the 715 cases be used for training the model and 20% be used to validate the model. The same cases were used to train and validate in all three techniques subsequently described.

Proof of Concept—Decision Tree Analysis

As discussed earlier, decision tree analysis is a predictive analytics technique that attempts to identify and isolate portions of a dataset that seem to act in similar ways in regard to a target variable. Figure 2 shows a decision tree we identified using SAS Enterprise Miner software for the binary target variable "unsuccessful contract." At the highest node, we see that 2.98% of the training dataset contracts were unsuccessful (1 = unsuccessful, 0 = successful) and 3.45% of the validation data. The first division is by the continuous variable called "Awarded Dollar Value"; those contracts that were less than $90,698,261 in awarded dollar value (ADV) had a much smaller failure rates (1.95% in training dataset and 3.05% in validation) compared to those that had higher awarded dollar value (12.07% and 7.14%).

The thickness of the line in the chart displays where the majority of the data lie; 512 cases in the training dataset had less than $90.6 million ADV while only 58 cases had more than $90.6 million ADV. Because there are so few cases with ADV greater than $90.6 million, there is little reason to further divide this section; however, if more data were available, the decision tree could be much more complex.

For those contracts with ADV less than $90.6 million, the next division is the "Workload (Actions) by Filled Billet." The contracting offices with less than 74.5 workload actions by filled billets had much lower failure rates (0.99% training, 3.7% validation) than that for offices with higher workload actions by filled billets (5.66% training, 0% validation). This would suggest that contracting offices that are understaffed or overworked tend to have larger number of contracts with low CPARS scores. However, take note that the validation dataset does not follow the same direction as the training dataset, suggesting that the model is overfit. Having a model that is overfit this early in a decision tree model is a symptom of having a small initial sample size.

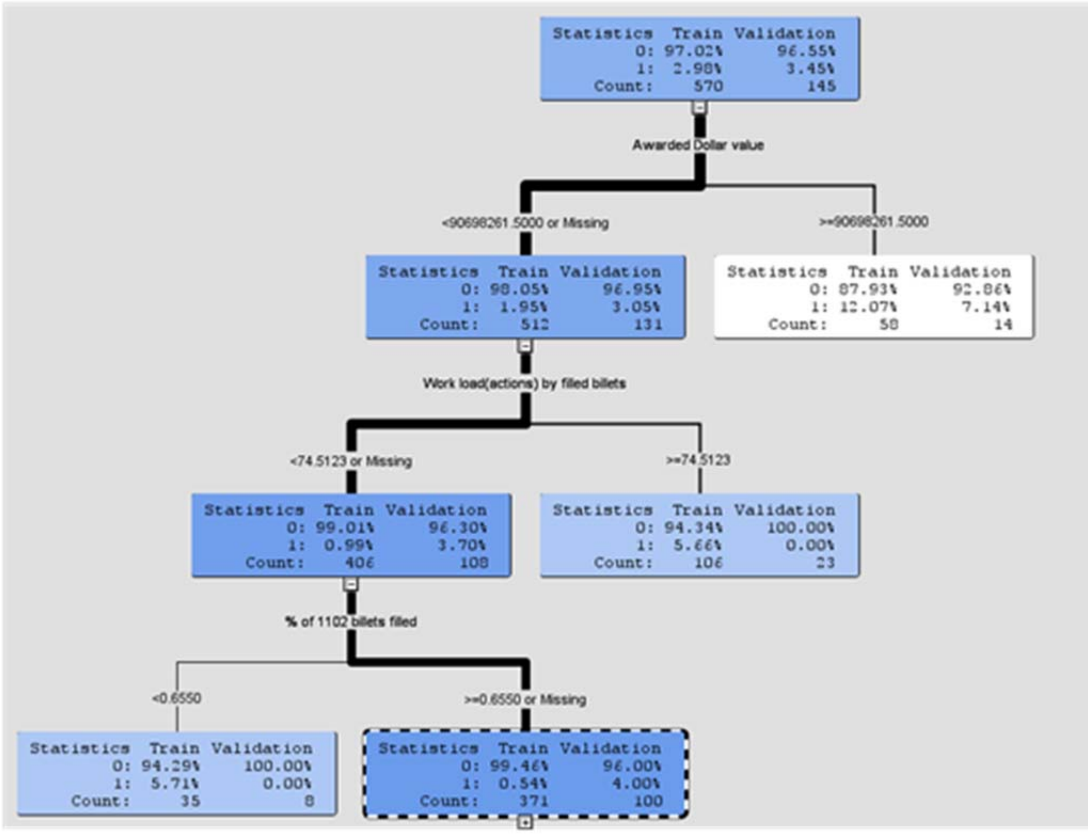


Figure 2. Decision Tree Analysis for "Unsuccessful Contract"

The final division happens with those contracts that are both less than $90.6 million ADV and from contracting offices with less than 74.4 workload actions per filled billet. The division shows that the offices that have less than 65.5% of their 1102 billets filled have a larger failure rate (5.71% and 0%) compared to those with a higher percentage of 1102 billets filled (0.54% and 4%). This suggests that contracting offices that are unable to fill their billets are likely to have higher rate of failed contracts.

Training Versus Validating

The decision tree presented in Figure 2 shows how the training dataset could best be divided into groups based on the independent variables. The resulting divisions make groups that are the most divergent in terms of the percentage of the binary target variable "unsuccessful contracts." Unfortunately, the "validation" dataset does not always follow the divergent nature of the training dataset, and, as a result, it appears that this analysis is overfit. If a model is overfit, it is less useful to generalize to other observations. However, overfit models can be useful in interpreting past data. In our case, the dataset is relatively small and therefore it is not necessarily very representative of any large set of contracts. Consequently, it is difficult to make any definitive or generalizable observations. However, the purpose of this research is to assess how Big Data analytics can be used to gain better understanding the success of contracts and that purpose has been well served with this proof of concept study.

Proof of Concept—Logistic Regression

As described in the logistic regression section, we performed the regression analysis using a step-wise regression methodology. In this method, a regression was estimated first with no independent variables; that is, with only an intercept. Next, a model was estimated with an intercept and only one variable that could explain the most variability in the target variable. Next, a model with an intercept and two top variables was estimated. This process was continued until all the independent variables had been included in the analysis. At the conclusion of the modeling, the software program displays which of the models explains most of the variability in the target variable with the least amount of independent variables. The results are shown in Table 2.

Table 2. Results of Stepwise Logistic Regression

Parameter	*Estimate*	*p value*	$e^{(Estimate)}$
Intercept	-12.213	<.0001	0
Work load actions by filled billet	0.0129	0.0117	1.013
Type of Contract – CPAF	8.8507	<.0001	6979
Type of Contract – CPAF & CPFF	-3.2748	0.9986	0.038
Type of Contract – CPFF	9.2498	<.0001	10402
Type of Contract – CPFF FFP	37.0026	0.9954	1.7×10^{16}
Type of Contract – CPIF	-3.3486	0.9978	0.035
Type of Contract – FFP	7.8061	.	2455
Type of Contract - Other	-3.7514	0.9970	0.0264

	Training	*Validation*
Average Squared Error	0.0266	0.0290
Misclassification Rate	0.0281	0.0276

The numbers in the “Estimate” column are the estimated coefficients for the regression equation previously described. A p value less than 0.05 is typically considered significant. The final column is the exponent of the estimate; these are easier to interpret since the original coefficient is in terms of log odds. This model reveals that two main characteristics of the contract tend to do a fairly good job of classifying failures (see the misclassification rate for training and validation datasets around 2.8%). Introducing additional variables to this model did not significantly improve the estimates.

The variable “workload action by filled billets” is the number of work actions that the entire office did divided by the number of filled billets that a contracting office had during the time period. The calculation provides an average number of actions worked for each billet filled. The logistic regression results show that an increase of one more worked action per filled billet would increase the odds of a failed contract by 1.013 or 1.3%. That means that increased workload of 10 actions per billet would be 13% more likely to have a failed contract. This variable was also a significant indicator of failure in the decision tree analysis.

The type of contract is also a significant indicator of CPARS failures in our dataset. The variable "Type of Contract" is a categorical variable with multiple different categories, as follows:

- CPFF Cost Plus Fixed Fee
- CPAF Cost Plus Award Fee
- CPIF Cost Plus Incentive Fee
- FFP Firm Fixed Price
- Other Other types of contracts

Using categorical variables in regression requires analysts to construct "dummy variables" for each category that take binary values 0 or 1. A dummy variable is created for all categories except for one category which is referred to as the "base case." The coefficients for the regression models should be interpreted in terms of the base case. In our example, the base case is FFP contract. The interpretation of the coefficients for these variables is as follows: CPAF contracts are 6,979 times more likely to have CPARS failures than the FFP contracts in our dataset. CPFF contracts are 10,402 times more likely to have failed CPARS than the FFP contracts. All other categories of contracts are not significantly different from the FFP contracts. Interestingly, these findings were not uncovered in either the decision tree analysis or the previous research we did with this dataset.

Proof of Concept—Neural Networks

In our earlier introduction of the neural networks technique, we stated that this technique tends to work best using very large data sets. In addition, we stated that the modeling of neural networks is primarily only useful for prediction with no meaningful ability to describe or explain relationships between independent and target variables. Instead, neural networks modeling is described in terms of its ability to correctly predict cases in the validation dataset.

Given that our dataset was rather small (only 512 cases in the training dataset), the results of neural network modeling were not much better than those for the logistic regression modeling. We found that by using a simple neural network model with only one layer of hidden nodes, we could create a model that would mimic both the average squared error and the misclassification rates found on Table 2 reporting on the previously mentioned logistic regression model. Our conclusion is that because our dataset was limited in size, a more complex modeling technique such as neural networks did not improve the prediction capacity. Hence, it would be better for an analyst to stay with the logistic regression model which is easier to interpret. However, if a large dataset were available, the neural networks modeling could have been useful for risk prediction.

Conclusions and Recommendations

Conclusions

In the previous section, we applied three Big Data analysis techniques—decision tree, logistics regression, and neural networks—to the CPARS data as proof of concept. As discussed earlier, we found that the following four variables exhibit the largest impact on the success/failure rates of contracts:

- Type of Contract (FFP, CPFF, CPAF, etc.)
- Awarded Dollar Value
- Workload (Actions) by Filled Billets

- % of 1102 Billets Filled by Contracting Office

As noted earlier, the size of the CPARS dataset that was available and used in this research was rather small, and as a result the previously mentioned conclusions cannot be unequivocally considered as being definitive. However, based on the results of our prior research and on work experience of one of the researchers as a contracting officer, we have every reason to believe the previously listed variables play important roles in affecting the success/failure rates of contracts.

Regarding the applicability and use of three Big Data analysis techniques tested in this research, we found that the first two techniques are scalable in a sense that although they are ideally suited for analyzing large datasets, they are also useful for analyzing datasets of limited size. In contrast, the neural networks technique is not likely to be particularly useful unless the dataset being analyzed is large in size.

Recommendations for Big Data Analysis Techniques in Acquisition

The current DoD acquisition community uses a number of disparate databases that capture specific acquisition and contracting data. Some databases consist of structured data while others consist of unstructured data (Rendon & Snider, 2014). Structured data are typically comprised of program data and contract data that can be mined through data mining techniques. For example, FPDS-NG provides pre-award summary data of contracts awarded by federal executive agencies. This database provides contract specific data such as contracting agency, contractor, type of contractor, federal supply class or service code, contract type, level of competition, contract dollar value, and so on. Additionally, the DoD's Selected Acquisition Report (SAR) provides post-award information to Congress such as cost, schedule, and performance data for major acquisition programs. The SAR reports are generally submitted on an annual basis and reflect changes from the previous report such as cost variances, changes in procurement quantities and changes in earned value management (EVM) metrics. Other sources of acquisition data include the Federal Business Opportunities (FEDBIZOPPS) website that contains contract solicitations (e.g., requests for proposals), industry conferences notices, and contract award notifications. Another source of acquisition data, specifically contractor performance data, is the already discussed Past Performance Information Retrieval System (PPIRS) that contains the contractor performance report cards known as the Contractor Performance Assessment Reports (CPARS).

The previously mentioned databases provide both pre-award (inputs) and post-award (outputs) sources of acquisition data. The optimum use of Big Data analysis would be to apply Big Data analysis techniques to both input and output acquisition data to explore any relationships between acquisition inputs and outputs. We propose the following recommendations for these types of Big Data analysis techniques in defense acquisition, as reflected in Figure 3.

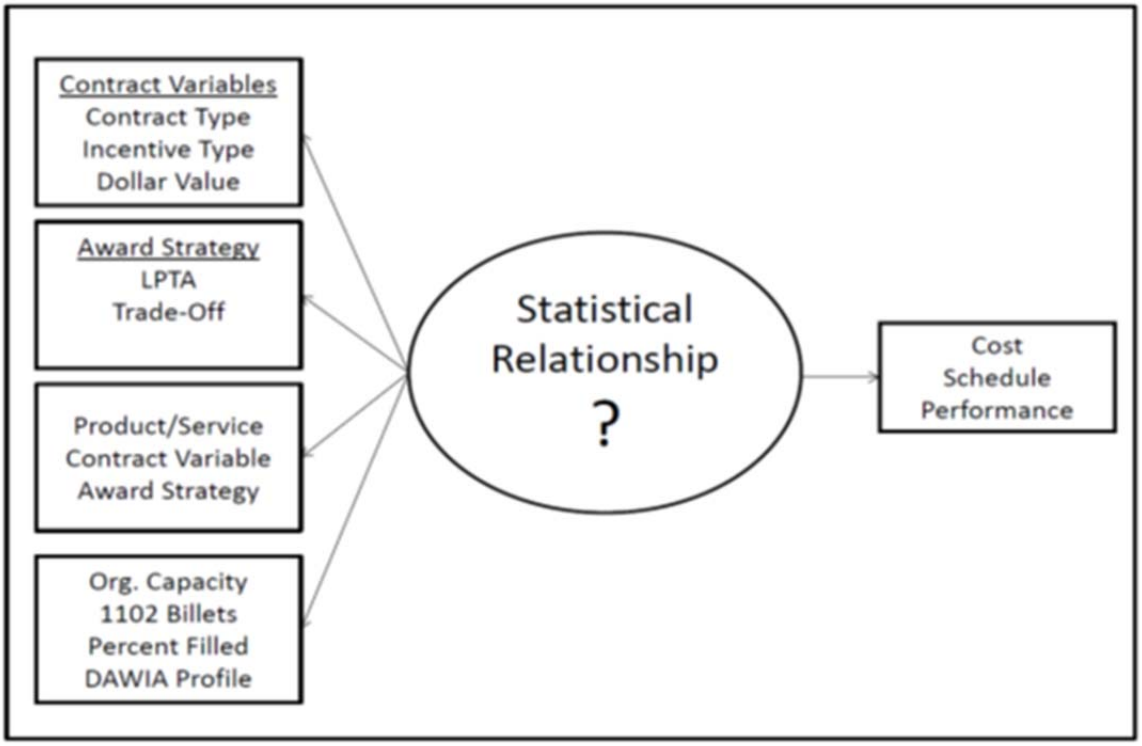


Figure 3. Proposed Recommendations

1. Analysis of specific contract variables and related contract cost, schedule, and performance outcomes. This Big Data analysis would look at the specific contract variables of contract type, incentive type, and contract dollar value and the resulting cost, schedule, and performance outputs of the contract. The purpose is to determine if contract type (fixed priced or cost reimbursement), incentive type (objective incentive such as FPI or CPI, subjective incentives such as award fee or award term), or dollar value is statistically related to the contract final cost, schedule, and performance results. This would require access and integration of the FPDS-NG, SAR, and PPIRS databases. The findings of this type of analysis would be beneficial in selecting contract type and incentive types on future contracts.
2. Analysis of specific contract award strategy variables and related contract cost, schedule, and performance outcomes. This Big Data analysis would look at the specific contract award strategy of price-based awards (such as lowest priced, technically acceptable) and tradeoff based awards (such as performance price tradeoff) and the resulting cost, schedule, and performance outputs of the contract. The purpose of this analysis is to determine if contract award strategy is statistically related to the contract final cost, schedule, and performance results. This would require access and integration of FEDBIZZOPPS database of solicitations, contract source selection files, SAR, and PPIRS databases. The findings of this type of analysis would be beneficial in selecting contract award strategies on future contracts.
3. Analysis of specific product/service codes, specific contract variables, contract award strategy variables and related contract cost, schedule, and performance outcomes. This Big Data analysis would look at the different

products and services procured by the DoD by product/service codes, as well as by contract type, contract award strategy and the resulting cost, schedule, and performance outputs of the contract. The purpose of this analysis is to determine if specific types of products or services are associated with specific contract variables and contract award strategy and if there is a statistical relationship with the contract final cost, schedule, and performance results. This would require access and integration of FEDBIZZOPPS database of solicitations, contract source selection files, SAR, and PPIRS databases. The findings of this type of analysis would be beneficial in selecting contract variables and contract award strategies on future procurement of specific products and services.

4. Analysis of organizational contracting capacity and related contract cost, schedule, and performance outcomes. Organizational contracting capacity includes metrics such as number of contracting (1102 and military equivalent) billets, percent of filled contracting billets, and number of DAWIA certified contracting personnel. This analysis would explore the relationship between the organization's capacity to contract (reflected in number and percent filled billets and DAWIA profile) and the organization's resulting cost, schedule, and performance outputs of its awarded contracts. The challenge in this Big Data analysis application is getting access to the organization's contracting capacity metrics. These metrics are not necessarily maintained by organizations, or may only be maintained at the higher headquarter levels. The benefit in conducting this Big Data analysis would be to see the relationship between contracting workforce (in terms of numbers and competence level) and contract performance.

References

Apte, A., Apte, U., & Rendon, R. (2008). M*anaging the services supply chain in the Department of Defense: An empirical study of current management practices* (NPS-AM-08-137). Retrieved from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Apte, A., Apte, U., & Rendon, R. (2009). *Managing the services supply chain in the Department of Defense: Empirical study of the current management practices in the Army* (NPS-AM-09-136). Retrieved from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Apte, U., Apte, A., & Rendon, R. (2010). *Services supply chain in the Department of Defense: A comparison and analysis of management practices in Army, Navy, and Air Force* (NPS-CM-10-161). Retrieved from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Apte, U., Ferrer, G., Lewis, I., & Rendon, R. (2006). *Managing the service supply chain in the U.S. Department of Defense: Opportunities and challenges* (NPS-AM-06-032). Retrieved from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Apte, U., & Rendon, R. (2007). *Managing the service supply chain in the US Department of Defense: Implications for the program management infrastructure* (NPS-PM-07-126). Retrieved from Naval Postgraduate School, Acquisition Research Program website: http://www.acquisitionresearch.net

Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications.* Hoboken, NJ: John Wiley & Sons.

Black, S., Henley, J., & Clute, M. (2014, May). *Determining the value of Contractor Performance Assessment Reporting System (CPARS) narratives for the acquisition process* (NPS-CM-14-022). Monterey, CA: Naval Postgraduate School.

Cleland, D. I. (1986). Project stakeholder management. *Project Management Journal, 17*(4), 36–44.

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review, 14*(1), 57–74.

El-Gohary, N. M., Osman, H., & El-Diraby, T. E. (2006). Stakeholder management for public private partnerships. *International Journal of Project Management, 24,* 595–604.

Federal Acquisition Regulation (FAR), 48 C.F.R. ch. 1 (2015).

Fitzsimmons, J. A., & Fitzsimmons, M. J. (2006). *Service management: Operations, strategy, and information technology* (5th ed.). New York, NY: McGraw-Hill.

Freeman, R. E. (1984). *Strategic management: A stakeholder approach*. Boston, MA: Pitman.

GAO. (2002). Best practices: *Taking a strategic approach could improve DoD's acquisition of services* (GAO-02-230). Washington, DC: Author.

GAO. (2009). *Defense acquisitions: Actions needed to ensure value for service contracts* (GAO-09-643T). Washington, DC: Author.

GAO. (2015). *High risk: An update* (GAO-15-290). Washington, DC: Author.

Hagan, P., Spede, J., & Sutton, T. (2012, December). *Defining and measuring the success of services contracts in the United States Navy* (NPS-CM-12-201). Monterey, CA: Naval Postgraduate School.

Hart, J., Stover, A., & Wilhite, T. (2013, November). *Management levers that drive services contracting success* (NPS-CM-13-122). Monterey, CA: Naval Postgraduate School.

Lee, L., & Dobler, D. W. (1971). *Purchasing and materials management: Text and cases.* New York, NY: McGraw-Hill.

Luo, Y. (2002). Contract, cooperation, and performance in international joint ventures. *Strategic Management Journal, 23*(10), 903–19.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.

Moe, T. M. (1984). The new economics of organization*. American Journal of Political Science,* 739–77.

Nash, R. C., Schooner, S. L., O'Brien-Debakey, K. R., & Edwards, V. J. (2007).*The government contracts reference book: A comprehensive guide to the language of procurement*. Washington, DC: George Washington University.

Rendon, R. G. (2010, June). *Assessment of Army Contracting Command's contract management processes* (NPS-CM-10-154). Monterey, CA: Naval Postgraduate School.

Rendon, R. G., Apte, U., & Dixon, M. (2014, February). *Services supply chain in the Department of Defense: Drivers of success in services acquisition* (NPS-CM-14-001). Monterey, CA: Naval Postgraduate School.

Rendon, R. G., Apte, U., & Dixon, M. (2015, March). *Services acquisition in the Department of Defense: Analysis of operational and performance data to identify drivers of success* (NPS-CM-14-190). Monterey, CA: Naval Postgraduate School.

Rendon, R. G., & Snider, K. F. (Eds.). (2008). *Management of defense acquisition projects.* Reston, VA: American Institute of Aeronautics and Astronautics.

Rendon, R. G., & Snider, K. F. (2014). Retrieving what’s already there: Archival data for research in defense acquisition. In J. Soeters, P. M. Shields, & S. Rietjens (Eds.), *Routledge handbook of research methods in military studies* (pp. 78–91). New York, NY: Routledge.

Seifert, A. J., & Ermoshkin, I. K. (2010). *Analysis of Government Accountability Office and Department of Defense Inspector General reports and commercial sources on service contracts* (MBA professional report). Monterey, CA: Naval Postgraduate School.

Schutt, R., & O'Neil, C. (2013). *Doing data science: Straight talk from the frontline.* Sebastopol, CA: O'Reilly Media.

Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]). (2010, June 28). Bet*ter buying power: Mandate for restoring affordability in defense spending* [Memorandum]. Washington, DC: Author.

Williamson, O. E. (1979). Transaction-cost economics: The governance of contractual relations. *The Journal of Law and Economics, 22*(3), 233–61.

Knowledge, Experience, and Training in Incentive Contracting for the Department of Defense

Kevin R. Carman—Dean of the DAU West Region, has over 40 years of program and contract management experience with major systems acquisition. As a qualified Executive Coach, he has extensive experience in the senior level program management and contract curriculum. A retired USN Captain with 30 years of service, Dr. Carman holds a BA, University of Massachusetts, in Mathematics; an MBA, National University; an MS, Acquisition and Contract Management, Naval Postgraduate School, Monterey, CA; an MA, National Strategic Studies, Naval War College; and a Doctor of Business Administration in Strategic Management from Alliant International University. He is a Certified Professional Contracts Manager and an NCMA Fellow. [kevin.carman@dau.mil]

Randall Gibson—is a Professor of Contracts Management at DAU West Region. Gibson has over 30 years of experience in contract compliance and pricing issues. A contract specialist on Surface Ship Weapons Directorate at NAVSEA and cost and pricing analysis for the Department of Defense Secretariat, he holds a bachelor's and master's in business administration from Marshall University and a Juris Doctorate from The Ohio State University. He is a Certified Public Accountant and a licensed attorney in Ohio and the District of Columbia. [randall.gibson@dau.mil]

Abstract

This paper assesses the incentive and award fee contracting training and experiences gained by the Department of Defense (DoD) Acquisition Workforce (DAW) through the implementation of various incentive arrangements to influence more favorable performance outcomes. The researchers developed a model to measure current and expected gaps between training, experience and knowledge. Additionally, the model correlates training and experience with knowledge and correlates training and experience with performance outcomes. The researchers used a survey instrument that included 30 questions to capture observations and assessments of the DAW who attended advanced classes between the period of 2013 and 2015. The survey provided key data to determine the presence of any noticeable gaps between required and actual levels of training, experience and knowledge.

Introduction

Part 16 of the Federal Acquisition Regulation (FAR) describes a wide variety of contract types that may be used for the acquisition of goods and services using appropriated funds in the Federal Government. The contract types primarily vary according to (1) responsibilities assumed by the contractor for costs of performance and (2) profit incentives for achieving or exceeding specified standards or goals. The contract types are grouped into two broad categories: "fixed-price" and "cost-reimbursement." The contract types range from "firm-fixed-price," where the contractor has full responsibility for performance costs and profits, to "cost-plus-fixed-fee," where the contractor is allocated minimal responsibility for performance costs with the fee (or profit) fixed by the terms of the contract. In between these two extremes are various "incentive" contract types, where the contractor's responsibility for the performance costs and profit are adjusted depending on the actual results of specific uncertainties identified at the time of contract award. Tables 1 and 2 present a summary of the FAR descriptions for three primary fixed-price type contracts and three cost-reimbursement type contracts.

Table 1. Summary of Fixed Price Contract Types

	Firm-Fixed-Price (FFP)	Fixed-Price Incentive Firm Target (FPIF)	Fixed-Price Award-Fee (FPAF)
Use When . . .	The requirement is well-defined. •Contractors are experienced in meeting it. •Market conditions are stable. •Financial risks are otherwise insignificant.	A ceiling price can be established that covers the most probable risks inherent in the nature of the work. The proposed profit sharing formula would motivate the contractor to control costs and to meet other objectives.	Judgmental standards can be fairly applied by the fee determining official. The potential fee is large enough to both: •Provide a meaningful incentive. •Justify related administrative burdens.
Contractor is Obliged to:	Provide an acceptable deliverable at the time, place and price specified in the contract.	Provide an acceptable deliverable at the time and place specified in the contract at or below the ceiling price.	Perform at the time, place, and the price fixed in the contract.
Contractor Incentive *(other than maximizing goodwill)*	Generally realizes an additional dollar of profit for every dollar that costs are reduced.	Realizes profit on cost by completing work below the ceiling price. May earn higher profit by incurring costs below the target cost or by meeting objective performance targets.	Generally realizes an additional dollar of profit for every dollar that costs are reduced; earns an additional fee for satisfying the performance standards.

Background

In September 2010, the Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) in his Better Buying Power (BBP) initiative memos clearly expressed, as a matter of policy, the importance of properly choosing contract types as a way of "aligning the incentives of the government and contractor" (Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics [OUSD(AT&L)], 2010). While the September 2010 memo *BBP 1.0* emphasized the increased use of Fixed-Price Incentive Firm Target type contracts, BBP 2.0 refined the guidance by emphasizing the use of the "appropriate contract vehicle for the product or services being acquired" (e.g., "one size does not fit all") and also "focus[ed] on improving the training of management and contracting personnel in the appropriate use of all contract types" (OUSD[AT&L], 2012). More recently, in BBP 3.0, there was even more refined guidance: "Employ appropriate contract types, but increase the use of incentive type contracts" (OUSD[AT&L], 2015a).

Table 2. FAR Summary of Cost Plus Contract Types

	Cost-Plus-Incentive-Fee (CPIF)	Cost-Plus-Award-Fee (CPAF)	Cost-Plus-Fixed-Fee (CPFF)
Use When ...	An objective relationship can be established between the fee and such measures of performance as actual costs, delivery dates, performance benchmarks, and the like.	Objective incentive targets are not feasible for critical aspects of performance. Judgmental standards can be fairly applied. Potential fee would provide a meaningful incentive.	Relating fee to performance (e.g., to actual costs) would be unworkable or of marginal utility.
Contractor is Obliged to:	Make a good faith effort to meet the Government's needs within the estimated cost in the Contract, Part I the Schedule, Section B Supplies or services and prices/costs.		
Contractor Incentive ***(other than maximizing goodwill)***	Realizes a higher fee by completing the work at a lower cost and/or by meeting other objective performance targets.	Realizes a higher fee by meeting judgmental performance standards.	Realizes a higher rate of return (i.e., fee divided by total cost) as total cost decreases.

In a report entitled *Performance of the Defense Acquisition System, 2015 Annual Report,* dated September 16, 2015, the OUSD(AT&L) concluded “that incentive contracts (cost-plus-incentive-fee and fixed-price-incentive) control cost, price, and schedule as well as, or better than, other types—and with generally lower yet fair margins” (OUSD[AT&L], 2015b).

Prior to the September 2010 BBP memo cited above, the Government Accountability Office (GAO; 2005) reported that the Department of Defense (DoD) had paid billions in award fees regardless of acquisition outcomes. The GAO (2005) audits stated the award fee criteria were subjective and faulty, resulting in fees awarded for marginal performance.

Even though a number of corrective actions were implemented, they still failed to address a key area of training and experience in the application of incentive arrangements. The report by the GAO stated the guidance on Award Fees has led to better practices but is not consistently applied for the DoD. The audit report also addressed the following topics: programs that paid fees without holding contractors accountable for achieving desired acquisition outcomes, such as meeting cost and schedule goals and delivering desired capabilities, and programs that paid contractors a significant portion of the available fee for what award fee plans describe as “acceptable, average, expected, good, or satisfactory" performance when the purpose of these fees is to motivate excellent performance.

The shortfalls of incentive contract arrangements are well documented in two GAO studies (GAO, 2005, 2009) and other literature. They address the areas of cost control, schedule, management, and technical performance. Generally, contractors who fail to meet incentive-related goals also frequently fail to meet other terms and conditions of the contract, which if structured properly, would carry performance penalties. Some of the shortcomings seen in the application of incentive contracts led to several clarifications from senior leadership in the DoD, including additional amplification of what’s important in the BBP. All too often, the contractors continue to earn high fee levels on late or cost over runs because the criteria tends to be overly subjective and the outcomes are not always well written with clear and equivocal objective outcomes. This could be a result of a knowledge shortfall, ineffective training, and/or inexperience. However, no research has been conducted to look more closely at the root cause of the problem in incentive contracting. This study hypothesized that knowledge, experience, and training could be contributing to the root cause, impeding the appropriate use of contract types to include incentive type arrangements.

This research began with the development of a Research Model that accounts for the variables that contribute to incentive arrangement failures. The research pursuit is intended to help the DoD better understand the level of experience, knowledge and training required for the effective application of incentive contract arrangements. The researchers leveraged the Kirkpatrick Learning Model (Figure 1), specifically Levels II–Learning, III–Behavior, and IV–Results.

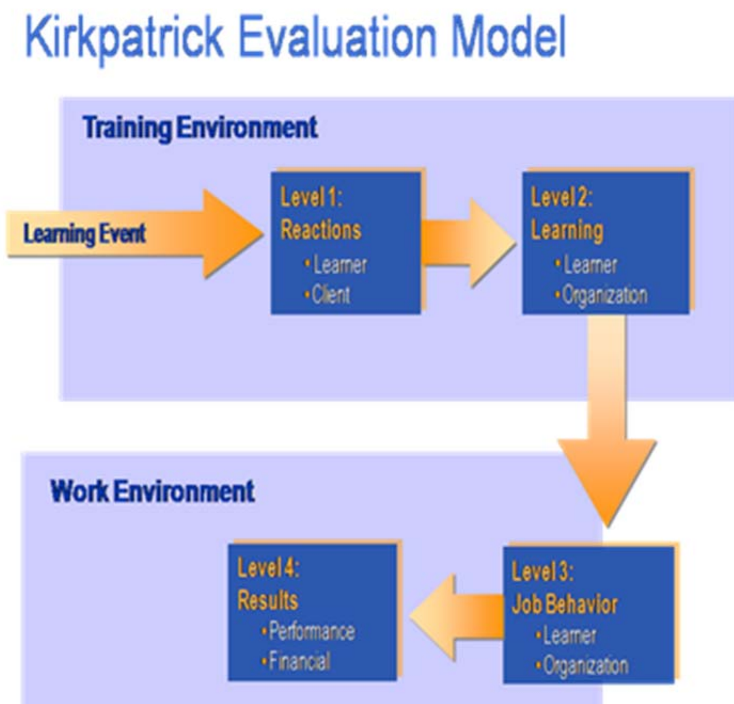


Figure 1. Kirkpatrick Evaluation Model

Kirkpatrick's four levels of evaluation consist of

- Step 1: **Reaction**—How well did the learners like the learning process?
- Step 2: **Learning**—What did they learn? (the extent to which the learners gain knowledge and skills)
- Step 3: **Behavior**—What changes in job performance resulted from the learning process? (capability to perform the newly learned skills while on the job)
- Step 4: **Results**—What are the tangible results of the learning process in terms of reduced cost, improved quality, increased production, efficiency, etc.?

Research suggests that as much as 90% of training resources are spent on the design, development, and delivery of training events, yet only yield 15% on-the-job application (Brinkerhoff, 2006). The learning reinforcement that occurs after the training event actually produces the highest level of learning effectiveness, followed by activities that occur before the learning event.

Research Model

Training Effectiveness Model

The Training Effectiveness Model (Figure 2), would assess if Knowledge, Training, and Experience gaps exist among Contracting Officers, Contract Specialists, Acquisition Professionals, Program Managers, and Deputy Program Managers. The primary focus of this research is to better understand the experiences (H1A and H2A) gained through the management of various incentive arrangements and determine if gaps exist in experience levels, and whether these gaps have a potential causal relationship with programmatic performance. Training received through the DAU and other training (H1B and H2B) were used to determine the presence of any substantiated gaps and their influence on performance outcomes. A gap analysis would confirm the disparity between current and required levels of knowledge, training, and experience to achieve desired performance outcomes performance. For Hypothesis 2, a correlation was used to determine the strength of the relationship between experience and training and its potential causal relationship to knowledge. A gap analysis and correlation was also conducted to determine the strength of knowledge supported by training and experience.

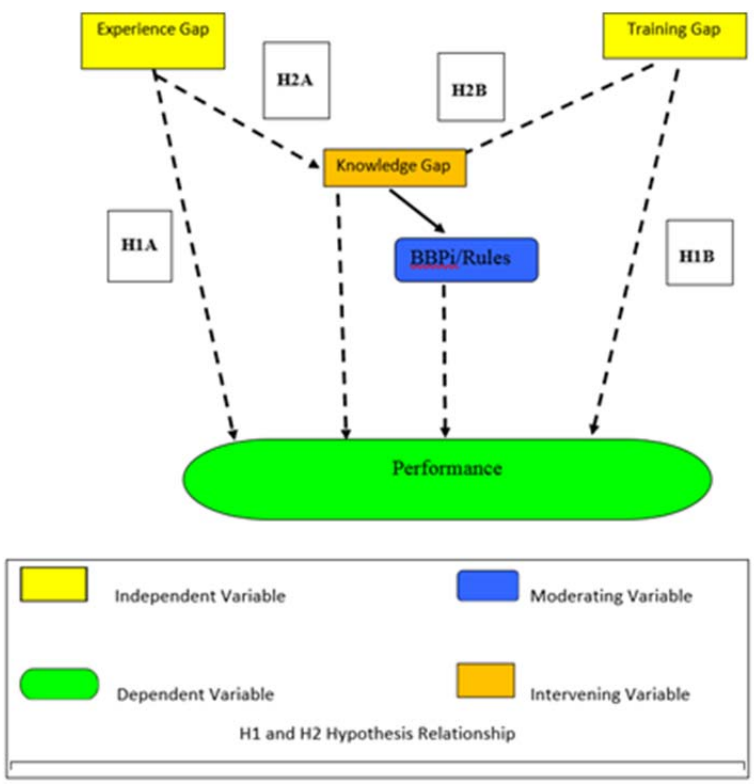


Figure 2. Knowledge, Experience, and Training Effectiveness Model

A *Knowledge Gap* is defined as the difference between what Acquisition professionals from the Program Management (PM) and Contracts Management (CM) communities observed as the current knowledge levels and the ideal or desired knowledge level. A Knowledge gap was measured by the absolute difference of the arithmetic mean of the observed or current knowledge level and the required knowledge level in the context of incentive and award fee contracts.

A *Training Gap* is defined as the difference between what the Acquisition professionals from the PM and CM communities observed as current training level and the ideal or desired training level. A Training gap is measured by the absolute difference of the arithmetic mean of the observed or current training level and the desired or required training level in the context of incentive and award fee contracts.

An *Experience Gap* is defined as the difference between what the Acquisition professionals from the PM and CM communities observed as their current experience levels and the ideal or desired experience level. An Experience gap is measured by the absolute difference of the arithmetic mean of the observed or current experience level and the desired or required experience level in the context of incentive and award fee contracts.

Performance is the overall effectiveness of the acquisition professional as perceived by the Program Managers and Contracting Officer groups. Performance is defined by the eight generic attributes (DoD, 2014) that program managers and contracting officers encounter with incentive and award fee contracts. The eight attributes are cost growth, program schedule, technical requirements, user requirements, technical issues, program risk, cost control, and contractor performance. The validation of the hypotheses of this research depended in part on the correlation analysis (at a minimally acceptable level of significance of at least .05 or $p < 0.05$) between the gaps or intervening variables and the incentive contracting performance. It is important to stress that a correlation coefficient of any magnitude or sign, regardless of statistical significance, does not imply causation (Emory & Cooper, 1991). However, the study explored the independent variables that might in some way influence performance outcomes. Consistent with the Kirkpatrick Evaluation Model, the DoD Acquisition Workforce survey volunteers provided useful information about themselves and their peers. The data centered on

1. the Effectiveness of classroom training received by the respondent and its impact on job performance of incentive contracting (Level 2)
2. job performance levels achieved, expected and targeted with respect to incentive and award fee contracting (Level 3)
3. performance effects of perceived gaps in knowledge, training, and experience with respect to incentive and award fee contracting (Level 4)

Research Methodology

The intended population for this research included Defense Acquisition Workforce Improvement Act (DAWIA) students who completed the Program Management (PM) Office Course (PMT 352) and Contracting (CON) for Decision Making (CON 360) from each of the military services, DoD agencies and included support contractors between the period of October 2013 and December 2015. Two control groups were surveyed: a Program Management group consisting of Program Managers, Deputy Program Managers, and function acquisition leads; and the Contract Management group consisting of Procuring Contracting Officers, Administrative Contracting Officers, Contract Specialists, and contract support administrators/staff. A statistical analysis was performed to answer the following research questions:

Research Question 1

What are the relationships among training gap, experience gap, and performance attributes that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field?

- ***Hypothesis 1:*** There are gaps in training and experience that affect performance attributes that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field.
    - *Hypothesis 1A:* There is a reliable relationship among experience gaps and performance attributes among contracts and program managers.
    - *Hypothesis 1B:* There is a reliable relationship among training gaps and performance attributes among contracts and program managers.

Research Question 2

What are the relationships between training gaps, experience gaps, and knowledge gaps that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field?

- ***Hypothesis 2:*** There are differences in the relationships between training gaps, experience gaps, and knowledge gaps that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field.
    - *Hypothesis 2A*: There is a reliable relationship among experience gaps and knowledge gaps among contracts and program managers.
    - *Hypothesis 2B:* There is a reliable relationship among training gaps and knowledge gaps among contracts and program managers.

The researchers used a web-based survey consisting of 30 questions; 1,194 individuals from the program management track responded, and 946 individuals from the contracting track responded. A Beta Testing of the survey instrument was conducted with the DAU West Contracting and PM Department faculty prior to the release of the survey. The researcher performed a Cronbach's alpha coefficients assessment that resulted in a very high reliability value of 0.936, $p < .05$.

Data Analysis

This research measured the gaps and correlation relationship among knowledge, training, experience, and performance gap that exist in incentive and award fee contracts. This was accomplished by assessing the students' current level of training, knowledge and the experience required to support incentive contract arrangements, compared to the level of knowledge, training, and experience required to achieve expected outcomes. Data from two population samples (i.e., PMs and CMs) identified the gaps.

This research pursuit was based on empirical data in order to assess knowledge, training, and experience gaps that could be influencing the success or failure of incentive and award fee contract arrangements. This accompanying analysis required a systematic method that accurately described the relationships among the independent, dependent and intervening variables measured.

Descriptive correlation research was the most suitable choice to determine accuracy (Isaac & Michael, 1977). The variables of this study consisted of two independent variables, one intervening variable, and one dependent variable, as follows:

<u>Independent Variables</u>

- Observed vs. Required/Expected Incentive Contracts Experience Gaps
- Observed vs. Required/Expected Incentive Contracts Training Gaps

<u>Intervening Variables</u>

- Observed vs. Required/Expected Incentive Contract Knowledge Gaps

<u>Dependent Variables</u>

- Performance Attributes

Hypothesis 1**:** There are gaps in training and experience that affect performance attributes that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field.

Table 3 provides the descriptive statistics of the knowledge, training, and experience gaps. It also shows the mean, median, mode, and standard deviation of each variable. The gaps for both the PM and CM were moderately strong at 3.8 to 4.3 on a 6-point Likert scale. The PMs and CMs indicated noticeable gaps in training, experience, and knowledge.

Table 3. Descriptive Statistics for CM/PM Knowledge, Training, and Experience Gaps

	CM Know Gap	CM Train Gap	CM Exp Gap	PM Know Gap	PM Train Gap	PM Exp Gap
N	424	424	424	424	424	424
Mean	3.7783	3.8113	4.1061	3.9458	3.9033	4.3302
Median	4.0000	4.0000	5.0000	4.0000	4.0000	5.0000
Mode	6.00	6.00	6.00	6.00	6.00	6.00
Std. Deviation	1.77396	1.82503	1.88952	1.67710	1.72249	1.63861
Variance	3.147	3.331	3.570	2.813	2.967	2.685
Minimum	1.00	1.00	1.00	1.00	1.00	1.00
Maximum	6.00	6.00	6.00	6.00	6.00	6.00

Table 4 shows the Pearson's r analysis of the independent variable, training gap, and the dependent variable of performance by PMs. All nine performance attributes had a moderately low correlation that ranged from .359 to .441, $p < .000$. All of these correlations had a significant confidence level at the 0.01 level.

Table 4. Results of the Pearson's r Correlation Between Program Managers Training Gap and Performance Attributes

Variables	N	Pearson's r	Sig.
Cost Growth – Training Gap	506	.441	.000
Program Schedule – Training Gap	506	.407	.000
Tech Requirements – Training Gap	506	.399	.000
User Requirements– Training Gap	506	.383	.000
Technical Issues – Training Gap	506	.373	.000
Program Risk – Training Gap	506	.359	.000
Cost Control – Training Gap	506	.401	.000
Contractor Performance – Training Gap	506	.381	.000

Note. *Correlation is significant at the 0.01 level (two-tailed)

Table 5 shows the Pearson's r analysis of the independent variable, training gap, and the dependent variable of performance by CMs. Cost growth indicated a very strong correlation of r = .934, p < .000. The remaining seven performance attributes had a moderately high correlation ranging from .592 to .717, p < .000. All of these correlations had a significant confidence level at the 0.01 level.

Table 5. Results of the Pearson's r Correlation Between Contracts Managers Training Gap and Performance Attributes

Variables	N	Pearson's r	Sig.
Cost Growth – Training Gap	372	.934	.000
Program Schedule – Training Gap	372	.690	.000
Tech Requirements – Training Gap	372	.694	.000
User Requirements– Training Gap	372	.622	.000
Technical Issues – Training Gap	372	.669	.000
Program Risk – Training Gap	372	.592	.000
Cost Control – Training Gap	372	.717	.000
Contractor Performance – Training Gap	372	.659	.000

Note. *Correlation is significant at the 0.01 level (two-tailed)

Table 6 shows the Pearson's r analysis of the independent variable, experience gap, and the dependent variable of performance by PMs. Cost growth indicated a moderately strong correlation r = .728, p < .000. The remaining seven performance attributes had a moderate correlation ranging from .466 to .584, p < .000. All of these correlations had a significance at the 0.01 level.

Table 6. Results of the Pearson's r Correlation Between Program Managers Experience Gap and Performance Attributes

Variables	N	Pearson's r	Sig.
Cost Growth – Experience Gap	506	.728	.000
Program Schedule – Experience Gap	506	.532	.000
Tech Requirements – Experience Gap	506	.511	.000
User Requirements– Experience Gap	506	.466	.000
Technical Issues –Experience Gap	506	.483	.000
Program Risk – Experience Gap	506	.549	.000
Cost Control – Experience Gap	506	.584	.000
Contractor Performance –Experience Gap	506	.546	.000

Note. *Correlation is significant at the 0.01 level (two-tailed)

Table 7 shows the Pearson's r analysis of the independent variable, experience gap, and the dependent variable of performance by CMs. Program Risk indicated a strong correlation of r = .911, p < .000. The remaining seven performance attributes had a moderately strong correlation ranging from .782 to .894, p < .000. All of these correlations had a significance at the 0.01 level.

Table 7. Results of the Pearson's r Correlation Between Contracts Managers Experience Gap and Performance Attributes

Variables	N	Pearson's r	Sig.
Cost Growth – Experience Gap	372	.874	.000
Program Schedule – Experience Gap	372	.873	.000
Tech Requirements – Experience Gap	372	.851	.000
User Requirements– Experience Gap	372	.823	.000
Technical Issues –Experience Gap	372	.854	.000
Program Risk – Experience Gap	372	.911	.000
Cost Control – Experience Gap	372	.894	.000
Contractor Performance – Experience Gap	372	.782	.000

Note. *Correlation is significant at the 0.01 level (two-tailed)

Table 8 shows a mean gap analysis of current and should be performance levels of incentive and award fee contracts by the CM respondents. The scale measures the various levels of understanding of outcomes that can be achieved through incentive arrangement. On a Likert scale of 1–6, 1 represents knowing the characteristics of an incentive arrangement, 2 represents understanding the benefits, 3 represents applying the mechanics, 4 represents analyzing the risks, 5 represents evaluating outcomes, and 6 represents creating and developing suitable contracts.

Table 8. Results of a Mean Gap Analysis Between Contract Managers Current and Should Be Performance Level

Variables	N	Current Mean	Current SD	Should be Mean	Should be SD	Mean Gap
Program Manager	304	3.0968	1.7832	4.1184	1.7723	1.0216
Deputy Program Manager	301	3.0370	1.7899	4.0864	1.7318	1.0494
Contracting Officer	327	4.0185	2.0340	5.1804	1.5412	1.1619
Contract Specialist	369	3.3425	1.9019	4.7015	1.7621	1.3590

The CM respondents identified a gap greater than 1.0 for all four groups. The greatest gaps were contract specialists, with a 1.3590, and contracting officers, with a 1.1619. The contracting officers could believe that among the contracting officers and contracting specialists, performance levels should result in more meaningful performance outcomes.

Table 9 shows a mean gap analysis of current and should be performance levels of incentive and award fee contracts by the PM respondents. The PM respondents identified a gap greater than 1.0 for each of the groups. The greatest gaps were deputy program managers, with a 2.0384, and contract specialists, with a 1.4396. Program managers could believe that among deputy program managers and contract specialists, performance levels should result in more meaningful performance outcomes.

Table 9. Results of a Mean Gap Analysis Between Program Managers Current and Should-Be Performance Level

Variables	N	Current Mean	Current SD	Should be Mean	Should be SD	Mean Gap
Program Manager	386	3.0200	1.7777	4.1036	1.7756	1.0836
Deputy Program Manager	383	2.0112	1.5380	4.0496	1.7351	2.0384
Contracting Officer	343	3.9510	2.0372	5.1590	1.5382	1.2080
Contract Specialist	327	3.2416	1.9019	4.6812	1.7604	1.4396

The results of the multiple regression analysis are shown in Table 10. The analysis was conducted with the independent variable of CM training means and dependent variables of performance attributes. The constant determines the overall effect of the variables on the performance attributes. The multiplier R-value was 0.935, p < 0.000; and the multiple R-squared was 0.872, suggesting that the independent variable, CM training, explains the 93.52% of the variances in performance attributes. Multiple regression analysis was used to evaluate the strength of the relationship among the independent and dependent variables. The regression model summary and coefficients implies that CM training had a strong positive correlation with performance attributes.

Table 10. Multiple Regression of CM Training as a Function of the Performance Attributes

Dependent Variables	*B*	*SEB*	Beta	*t*	*p*<
(Constant)	-2.821	0.155		-18.166	0.000
Cost Growth	1.490	0.058	0.937	25.775	0.000
Program Schedule	-0.015	0.054	-0.010	-0.276	0.783
Tech Requirements	-0.015	0.067	-0.010	-0.230	0.818
User Requirements	0.051	0.051	0.036	1.011	0.312
Technical Issues	-0.015	0.054	-0.010	-0.281	0.780
Program Risk	-0.083	0.052	-0.058	-1.591	0.113
Cost Control	0.110	0.053	0.076	2.050	0.041
Contractor Performance	-0.049	0.046	-0.033	-0.043	0.298

Independent Variable: CM Training Mean — Multiple $R = 0.935$ R-squared = 0.872 $p < 0.000$

The results of the multiple regression analysis are shown in Table 11. The analysis was conducted using the independent variable of CM experience means and dependent variables of performance attributes. The constant determines the overall effect of the variables on the performance attributes. The multiplier R-value was 0.920, p < 0.000; and the multiple R-squared was 0.844, suggesting that the independent variable, PM experience, explains the 92.0% of the variances in performance attributes. Multiple regression analysis was used to evaluate the strength of the relationship among the independent and dependent variables. The regression model summary and coefficients can be interpreted that PM experience had a strong positive correlation with performance attributes.

Table 11. Multiple Regression of CM Experience as a Function of the Performance Attributes

Dependent Variables	*B*	*SEB*	Beta	*t*	*p*<
(Constant)	0.236	0.096		2.459	0.014
Cost Growth	0.238	0.081	0.233	2.960	0.003
Program Schedule	0.074	0.056	0.075	1.312	0.190
Tech Requirements	0.013	0.071	0.013	0.189	0.850
User Requirements	-0.015	0.050	-0.015	-0.293	0.770
Technical Issues	0.070	0.049	0.071	1.441	0.150
Program Risk	0.369	0.056	0.361	6.536	0.000
Cost Control	0.290	0.050	0.285	5.815	0.000
Contractor Performance	-0.062	0.053	-0.064	-1.186	0.236

Independent Variable: CM Experience Mean   Multiple $R = 0.920$ R-squared = 0.844 $p < 0.000$

The results of the multiple regression analysis are shown in Table 12. The analysis was conducted using the independent variable of PM training means and dependent variables of performance attributes. The constant determines the overall effect of the variables on the performance attributes. The multiplier R-value was 0.450, p < 0.000; and the multiple R-squared was 0.203, suggesting that the independent variable, PM training, explains the 45.0% of the variances in performance attributes. Multiple regression analysis was used to evaluate the strength of the relationship among the independent and dependent variables. The regression model summary and coefficients imply that PM training has a weak positive correlation with performance attributes.

Table 12. Multiple Regression of PM Training as a Function of the Performance Attributes

Dependent Variables	*B*	*SEB*	Beta	*t*	*p*<
(Constant)	0.818	0.231		3.533	0.000
Cost Growth	0.586	0.168	0.446	3.488	0.001
Program Schedule	0.159	0.162	0.125	0.983	0.326
Tech Requirements	0.066	0.196	0.052	0.336	0.737
User Requirements	0.067	0.159	0.052	0.422	0.673
Technical Issues	-0.226	0.159	-0.173	-1.418	0.157
Program Risk	-0.184	0.147	-0.145	-1.252	0.211
Cost Control	0.072	0.144	0.056	0.499	0.618
Contractor Performance	0.029	0.132	0.023	0.223	0.824

Independent Variable: PM Training Mean — Multiple $R = 0.450$ R-squared = 0.203 $p < 0.000$

The results of the multiple regression analysis are shown in Table 13. The analysis was conducted considering the independent variable of PM experience means and dependent variables of performance attributes. The constant determines the overall effect of the variables on the performance attributes. The multiplier R-value was 0.730, $p < 0.000$; and the multiple R-squared was 0.534, suggesting that the independent variable, PM experience, explains the 73.0% of the variances in performance attributes. Multiple regression analysis was used to evaluate the strength of the relationship among the independent and dependent variables. The regression model summary and coefficients imply that PM experience has a moderately positive correlation with performance attributes.

Table 13. Multiple Regression of PM Experience as a Function of the Performance Attributes

Dependent Variables	B	SEB	Beta	t	$p<$
(Constant)	-0.543	0.296		-1.833	0.068
Cost Growth	0.975	0.095	0.701	10.274	0.000
Program Schedule	-0.097	0.103	-0.071	-0.942	0.347
Tech Requirements	0.014	0.119	0.011	0.116	0.908
User Requirements	0.059	0.087	0.048	0.675	0.500
Technical Issues	-0.006	0.086	-0.004	-0.066	0.948
Program Risk	0.073	0.107	0.053	0.689	0.491
Cost Control	0.060	0.094	0.044	0.636	0.525
Contractor Performance	-0.049	0.090	-0.036	-0.549	0.583

Independent Variable: PM Experience Mean — Multiple $R = 0.730$ — R-squared $= 0.534$ — $p < 0.000$

***Hypothesis 2*: There are differences in the relationships between training gaps, experience gaps, and knowledge gaps that contribute to incentive contract arrangements among Defense Acquisition Workforce members in the contracting and program field.**

- *Hypothesis 2A*: There is a reliable relationship among experience gaps and knowledge gaps among contracts and program managers.
- *Hypothesis 2B*: There is a reliable relationship among training gaps and knowledge gaps among contracts and program managers.

A Pearson's Correlation (i.e., linear correlation) analysis was conducted between the training gaps, experience gaps, and the knowledge gaps and is recorded in Table 14. For CM knowledge gap and training gap, a strong correlation of $r = 0.898, p < 0.000$ is supported. CM knowledge gap and CM experience gap have a strong correlation of $r = 0.893, p < 0.000$. For program managers (PM) knowledge gap and training gap, a strong correlation of $r = 0.843, p < 0.000$ is supported. PM knowledge gap and CM experience gap have a strong moderate correlation of $r = 0.767, p < 0.000$. Pearson's r correlation results for CM and PM training, experience, and knowledge gaps are presented in Table 14.

Table 14. Results of the Pearson's r Correlation Between Training Gap, Experience Gap, and Knowledge Gap

Variables	N	Pearson's r	Sig.
CM Knowledge Gap - CM Training Gap	445	.898	.000
CM Knowledge Gap - CM Experience Gap	445	.893	.000
PM Knowledge Gap - PM Training Gap	369	.843	.000
PM Knowledge Gap - PM Experience Gap	369	.767	.000

Note. *Correlation is significant at the 0.05 level (two-tailed)

The multiple R-squared was 0.898 and .893, suggesting that the two independent variables, experience gap and training gaps, explain 89.8% and 89.3% of the variances in knowledge gap, respectfully.

Multiple regression analysis was used to evaluate the strength of the relationship among the independent and dependent variables. The regression model summary and coefficients can be interpreted such that knowledge gap and experience gap have a strong positive correlation with knowledge gap.

Table 15 shows the multiple regression of knowledge gap as a function of CM training gap and CM experience gap. There is a strong correlation between the variable of $r = .925, p < 0.000$. This implies that when the CM training gap aligns with experience gaps, knowledge gap will be optimized. Hypothesis H2is supported for CM Training gaps, CM Experience gaps and Knowledge gaps.

Table 15. Multiple Regression of Knowledge Gap as a Function of Training Gap and Experience Gap for the Contracting Group

Variable	*B*	*SEB*	Beta	*t*	*p*<
(Constant)	.160	.089		1.789	.074
CM Training Gap	.430	.035	.458	12.251	.000
CM Experience Gap	.490	.037	.497	13.316	.000
Dependent Variable: Knowledge Gap	Multiple $R = 0.925$		R-squared = 0.855		$p < 0.000$

Note. *Correlation is significant at the 0.05 level (two-tailed)

Table 16 shows the multiple regression of knowledge gap as a function of PM training gap and PM experience gap. There is a moderately strong correlation between the variable of $r = .785, p < 0.000$. This implies that when the PM training gap aligns with experience gap, knowledge gap will be moderately optimized. Hypothesis H2 is supported for PM Training Gaps, PM Experience gaps and Knowledge gaps.

Table 16. Multiple Regression of Knowledge Gap as a Function of Training Gap and Experience Gap for the Program Management Group

Variable	*B*	*SEB*	Beta	*t*	*p<*
(Constant)	.913	.161		5.663	.000
PM Training Gap	.554	.066	.458	8.389	.000
PM Experience Gap	.334	.064	.497	5.189	.000
Dependent Variable: Knowledge Gap	Multiple $R = 0.785$		R-squared = 0.616		$p < 0.000$

Note. *Correlation is significant at the 0.05 level (two-tailed)

Research Findings

Summary

With respect to incentive contracting arrangements,

- Moderate (self-assessed) gaps exist in knowledge, training and experience.
- The experience gap, vice the training gap, is perceived to be more closely related to performance outcomes, and Contract Managers see training to be more closely related than do Program Managers.
- Training and experience are highly correlated to knowledge.
- The current state of skill-sets is more about the mechanics, while the desired state is more about creating and developing suitable contracts.
- There are only a few significant differences between Contract Managers and Program Managers regarding assessments of knowledge, training, and experience.

Discussion

Both Program Manager and Contract Manager Respondents reported moderate knowledge, training, and experience gaps amongst their organization personnel who implement incentive contract arrangements.

Knowledge associated with contractor incentives, and associated incentive contracting approaches and techniques, are necessary elements in addressing performance obstacles. Our research indicates that both Program Managers and Contract Managers perceive a strong relationship between both the observed *training gap* and *experience gap* to the observed *knowledge* gap.

Our research suggests that *experience*, rather than *training*, is more closely identified with performance issues using incentive arrangements. More specifically, with a noted exception, respondents did not strongly relate the observed *training* gap as impacting skill-sets needed to address acquisition obstacles (e.g., “Unexpected Cost Growth,” “Changes in Program Schedules,” “Changes in Technical Requirements,” “Changes in User Requirements,” Technical Issues,” and “Program Risk”). The exception was Contract Manager Respondents who more strongly identified the relationship between *training* gaps and meeting performance challenges but reached the strongest only for the cost-related acquisition obstacles (e.g., “Unexpected Cost Growth” and “Cost Control”).

Our survey asked Contract Managers (Contracting Officers and Contract Specialists) and Program Managers (Program Managers and Deputy Program Managers) to self-assess their “current” and “should be” levels of performance in working with incentive contract arrangements, thereby allowing visibility of any perceived gaps. Our research also asked Contract Managers and Program Managers to assess each other.

Using a 6-point Likert scale, Contract Managers self-assessed themselves near the middle of the range and indicated they needed to increase performance levels, significantly from the current state. Program Manager assessments of Contract Manager performance level were very closely aligned with those of the Contract Manager self-assessments previously mentioned and corroborated what the research confirmed.

The Program Manager self-assessments were very similar to both the Contract Managers’ self-assessments and the Contract Managers’ assessments of Program Managers except when it came to the assessment of Deputy Program Managers, where there was more than a 2-point gap (Current = 2.0 or “understands benefits,” while Should-be = 4.05 or “analysis of risks”).

As expected, a strong relationship was noted between the observed training and experience gaps and knowledge gaps, suggesting that reducing training/experience gaps could also reduce any knowledge gaps.

Qualitative Data

The respondents provided over 6,000 comments, and the following is a representative sample of the strengths and/or shortcomings of the application of incentive contracts.

Knowledge

- “Contractor outcomes—profit/fee—were higher than they should have been because personnel routinely failed to hold contractors to the criteria found in the Award or Incentive plans.”
- “I think people don’t want to use contract types they don’t fully understand.”
- “I strongly believe that incentive contracts are necessary at my command and have recommended them after being brought on. It was acknowledged that this is the best contract type in the interest of our program, but overly complicated and burdensome given the lack of training of our staff.”
- “KO was unable to write the outcomes to meet my program outcomes. I was pushed to use FFP or CPFF or IDIQ arrangements because KO did not fully understand the formation of award fee contracts.”
- “Not sure all of my people are good at thinking through how different incentive structures will cause the contractor to behave. Incentives other than cost are particularly tricky and I generally shy away from them because I sense that the contractor will ‘outfox‘ us and we will end up regretting the structure down the road.”

Training

- “People had the training, but did not understand how to use it in their duties.”
- “Most time is spent trying to apply mechanics of type rather than truly implementing meaningful measures.”

- “The biggest problem is developing meaningful criteria. Services acquisition are very hard to ‘incentivize.’ The wording of the meaningful criteria is the bigger problem than coming up with 10, 20, 30 percent incentive. That is simple math.”
- “Because we don‘t have good command level training for how to administer these overly complex contracts, everyone struggles to efficiently do their work.”
- “Lack of training is the biggest problem.”
- “We do not receive enough training to understand these concepts enough to execute effectively.”

Experience

- “Experience is the main driver of shortfalls. Not everybody is in situations where they are using CPAF, FPIF, or CPIF contracts regularly.”
- “This lack of experience, understanding and training makes it very difficult to effectively utilize these contract types.”
- “Nobody seems to know what to do, if and when we use it and/or the required information is not passed to the field.
- “Experience seems to drive individuals to contracts that they are familiar with.”
- “Lack of experience with multiple contract types can cause under-performance.”
- “The lack of experience outs (place) the government at a disadvantage in execution of these types of contracts.”
- “There’s just no substitute for experience, not IQ, not education. With experience comes intuition, and it’s intuition that recognizes what flies and what doesn’t. Oftentimes, people are thrust into programs and projects for which they lack a basis for making informed decisions about the future.”

Recommendations

1. Ensure incentive contract development training be reinvigorated for PMs and CMs to fully understand the fundamental principles and benefits of incentive contracts as well as provide a rigorous setting where they can practice designing incentive contracts and applying them.
2. Identify the lessons learned from successful and unsuccessful incentive arrangements within the past 10 years and make them available across the defense acquisition workforce; incorporate these to the greatest extent possible into existing courseware.
3. Produce an incentive and award fee guidebook that addresses the performance attributes identified in this study and address the question “How do I write an incentive arrangement?” to include incorporating performance attributes that lead to achievement of the PM’s goals and outcomes.
4. Ensure the appropriate acquisition workforce qualification competencies (or any variation thereof) incorporate the key standards that address incentive and award fee contract proficiencies that are assignment specific.

5. Recommend that the Functional Integrated Product Team (FIPT) explore a wide range of competencies that are specifically tuned to the implementation of incentive arrangements for PMs and CMs in assignment specific positions.
6. Conduct a follow on study to address the relationship between knowledge, performance, and the applicable regulations.

References

Brinkerhoff, R. O. (2006). *Telling training's story: Evaluation made simple, credible, and effective.* San Francisco, CA: Berrett-Koehler.

DoD. (2014). *2014 study of program managers training and experience.* Retrieved from https://acc.dau.mil/adl/en-US/712361/file/78162/OSD%20PM%20STUDY_brief%20for%20140821%20Hot%20Topic%20Trng%20Forum_v10.pdf

Emory, W. C., & Cooper, D. R. (1991). *Business research methods.* Homewood, IL: Richard D. Irwin.

GAO. (2005, December). *DOD has paid billions in award and incentive fees regardless of acquisition outcomes* (GAO-06-66). Retrieved from http://www.gao.gov

GAO. (2009, May). *Guidance on award fees has led to better practices but is not consistently applied* (GAO-09-630). Retrieved from http://www.gao.gov

Issac, S., & Michael, W. B. (1977). *Handbook on research and evaluation*. San Diego, CA: Edits.

Kirkpatrick Evaluation Model. (2016, March). The Kirkpatrick Model of Training Evaluation. Retrieved from http://www.astdla.org/Resources/Pictures/Kirkpatrick%20Evaluation%20Model.png

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2010, September 14). *Better buying power: Guidance for obtaining greater efficiency and productivity in Defense spending* [Memorandum]. Retrieved from http://bbp.dau.mil/docs/Better%20Buying%20Power--Guidance%20for%20Obtaining%20Greater%20Efficiency%20and%20Productivity%20in%20Defense%20Spending.pdf

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2012). *Better buying power 2.0.* Retrieved from http://bbp.dau.mil/docs/BBP%202%200%20Memo%20to%20the%20Workforce%20(13%20Nov)%20Final.pdf

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics. (OUSD[AT&L]). (2015a, April 09). *Better buying power 3.0*. Retrieved from http://bbp.dau.mil/docs/BBP3.0Implementation Guidance MemorandumforRelease.pdf

Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics (OUSD[AT&L]). (2015b, September 16). *Performance of the defense acquisition system, 2015 Annual Report.* Retrieved from http://bbp.dau.mil/docs/Performance-of-Defense-Acquisition-System-2015.pdf

Learning in the Shadows: A Retrospective View

Donna J. Kinnear-Seligman—is a Mission Assistance Program Analysis Manager at the Defense Acquisition University West Region. She has over 20 years of experience with developing and managing complex business knowledge applications, performing comprehensive system analyses, and conducting extensive research. Her research topics previously presented during the NPS Acquisition Research Symposium include the following:

- Challenges in Meeting OSD's Obligation & Expenditure Rate Goals (*ARJ*, October 2013)
- OSD's Obligations and Expenditure Rate Goals (*AT&L*, August 2013)
- Learning Organizations (*ARJ*, April 2013)
- Human Capital Accelerators (*AT&L*, March 2011)
- It's Time to Take the Chill Out of Cost Containment (*ARJ,* April 2010)

Seligman holds an MBA (emphasis in Program Management) from the University of Management and Technology and a BS in Information Decision Systems from San Diego State University. [donna.seligman@dau.mil]

Abstract

How does the Acquisition Workforce train and create experience for the next generation? This research addressed how the Defense Acquisition University (DAU) develops its emerging leaders among its support faculty and staff. Through a retrospective view of the emerging leadership program in particular, the author investigated the value and impact of the program details and the resulting effect it had for the Emerging Leader Program (ELP) graduates. ELP graduates were invited to participate in a survey to quantify their ELP experiences and express any concomitant value for the DAU. Focusing on the various ELP activities coupled with Better Buying Power's emphasis on professional development to reinforce the importance of improving workforce professionalism, this paper assessed the outcomes of the DAU's ELP over the course of six years. Originally, the ELP was intended to create a development pathway. Has it?

Issue

Like any Human Capital development program, is the investment worth it? After completing the Emerging Leader Program (ELP) at the Defense Acquisition University (DAU), were graduates able to influence leadership (a key indicator of leadership performance) with their new skill sets? Additionally, how many graduates actually achieved advancement or became more competitive for various leadership positions? The DAU has conducted a total of six year-long ELP sessions and by the end of FY16 will have graduated 58 emerging leaders.

Results

This research confirmed the ELP's effectiveness and identified the impact along with its efficacy through a variety of metrics. The research results also quantified the specific learning experiences that surfaced. What was the single most influential factor for ELP graduates that gave them enough momentum to move forward as a future influencer/leader? Beyond what the ELP graduates actually experienced in the way of workplace influence, they also addressed how they felt about the program overall. The results of this research can provide very useful insights for future ELP candidates as well as the ELP program itself.

Figure 1 displays how ELP graduates rated the ELP program activities. Those activities earning the highest survey ratings are circled. The participants found self-assessments—such as the DDI 360 Leadership Mirror® (360), Myers-Briggs Type Indicator

Assessment (MBTI), etc.—effective, largely because they allowed for self-awareness and self-managed change according to the qualitative comments. Additionally, the soft skill unique training and shadow experiences appeared to be influential program components. The soft skills workshops and associated assessments provided an experiential platform as well as some intellectual muscle for emerging leaders as they developed even further.

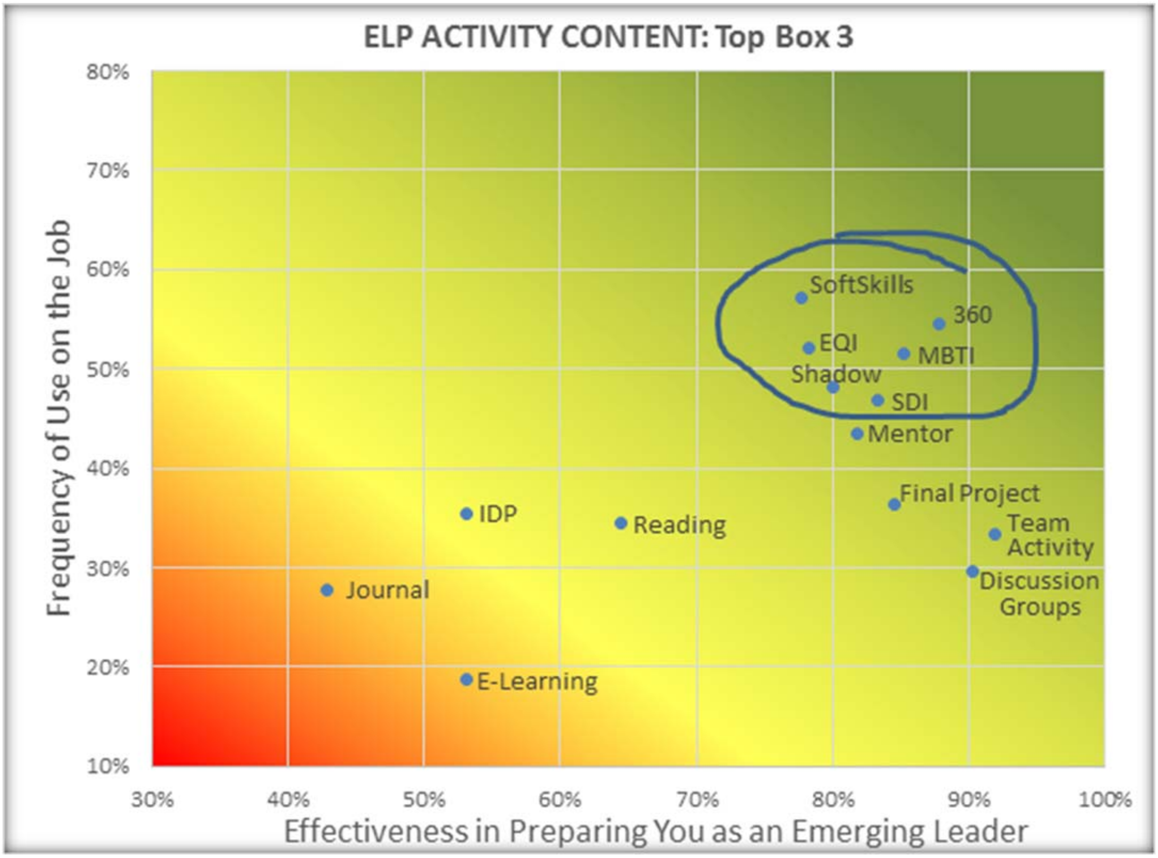


Figure 1. Emerging Leader Program (ELP) Activity Content

Background

The DAU conducts an internal climate survey every two years. Based on the required improvements, climate survey improvement teams are normally instituted to address suggestions and take any necessary action as a result. In 2009, the DAU's Climate Survey Improvement (CSI) team noted a trend of lower staff satisfaction in comparison to faculty. After a closer look, DAU staff had identified the need for "more recognition" and the ability to "influence decisions" affecting their contributions in their respective workplaces. DAU staff appeared less optimistic about their future opportunities at the DAU. The CSI team recommended that the DAU develop a "Future (Emerging) Leader Program" (Seligman, 2009).

Recognizing that Gen-X and Millennials bring a change in leadership responsibility mindsets while having less experience at the same time, DAU leadership decided to pilot an "Emerging Leadership Program" in 2011. After the pilot, the staff participants generally felt the program helped them bridge the opportunity gap they previously experienced. Today, the DAU's ELP Charter and Introduction states the program will provide "experience and knowledge that fosters professional and personal growth … and … prepares select DAU employees for positions of increasing responsibility" (Fowler, 2015). Since its inception, the DAU has continued its investment in the ELP over the last six years to further develop DAU staff members.

This research is the first effort that measured organizational learning outcomes by evaluating/assessing the perceived effectiveness of constituent ELP activities and participant comments after graduation.

Emerging Leadership Program Specifics

Participation in the ELP is competitive. If accepted, ELP selectees are exposed to a wide range of leadership competencies. During a year-long program, ELP participants meet once a month virtually, and twice face-to-face to discuss them. The two face-to-face meetings are reserved for the first and last meetings. The virtual forums help pace the participants through the various ELP program activities. The program is designed to strengthen seven core competencies:

1. Customer Service
2. Communications Skills
3. Interpersonal Skills
4. Flexibility/Adaptability
5. Problem Solving
6. Developing Others & Continuous Learning
7. Integrity & Honesty

Below are 14 ELP activities along with a brief description for each:

- **Discussion Groups:** ELP participants meet virtually to cover program activities and check on ELP progress. The first discussion group is face-to-face to help establish a trusting relationship among the ELP participants and set a foundation for the team building ahead. Several of these meetings include visits/roundtables with the DAU's most senior leadership.
- **Myers-Briggs Type Indicator Assessment® (MBTI®):** Self-assessment that accounts for 16 distinctive personality types. Through a self-assessment survey, participants identify their own preference. During the workshop, they learn different ways of leveraging their personality preferences with those of others.
- **Strength Deployment Inventory (SDI):** A self-assessment that helps participants better understand a wide range of their own behavioral preferences and better understand how they may react when faced with conflict and turbulence. ELP participants also discuss how their own SDI relates to those within their DAU leadership.
- **Individual Development Plan (IDP):** The ELP IDP is a more extensive version of the DAU IDP. ELP participants develop extensive short term goals and identify specific developmental assignment prospects. The IDP is then coordinated with supervisors and the principle ELP Coordinator.
- **Journal:** Journals are not required or submitted for completion, but rather encouraged as a tool to capture thoughts and experiences during the ELP since they support discussion groups and the ELP student's final project.
- **DDI 360 Leadership Mirror® (360):** The 360 is a web-based multi-rater feedback system which maintains confidential, anonymous feedback from the

ELP's peers, managers, and anyone who may be reporting to them (formally or informally). This assessment report allows participants to uncover any blind spots and make adjustments.

- **Team Activity:** The ELP participants are assigned a topic to present. The presentation is delivered as part of the capstone activity and represents another opportunity to practice some of the skills acquired during the program.
- **Emotional Quotient Inventory (EQI):** This self-assessment is designed to measure emotional-social intelligence. The report produces an overall EQ score as well as separate scores for Intrapersonal (self-awareness and self-expression), Interpersonal (social awareness and relationships), and Stress Management (emotional management).
- **E-Learning Curriculum (e-Learning):** This e-Learning curriculum is supported by a suite of courses located in the DAU's intranet though "Skillport." The curriculum is mapped to the competencies identified for the ELP participant. Completion of one course is required each month.
- **Reading Report:** The readings are selected from the ELP reading list and require a written report. The report summarizes the key learning points and practical leadership application of the content.
- **Soft Skill Workshops:** In this series of workshops, the ELP students learn strategies they may not experience in their on-the-job training. Students participate in workshops such as Crucial Conversations®, Crucial Confrontations®, Influencer Training™, or Crucial Accountability®, all taught by certified Vital Smarts trainers. Some ELP groups have also experienced Leading at the Speed of Trust® by Franklin Covey.
- **Mentor:** ELP participants propose a leadership mentor for approval by the ELP Coordinator. ELP participants record mentor observations in a journal and create a one page summary report of the journal entries.
- **Shadow:** The "shadow" is an on-the-job assignment of "job shadowing" a supervisor who has a job related to the staff's current position or one they plan to seek. The shadow focuses on management styles and interaction with employees and/or customers.
- **Final Project:** Each ELP participant submits a final individual project which summarizes the participant's experiences over the year. The final project includes lessons-learned, journal reflections, notes from the group discussions, details of the job shadow experience, and how the ELP participant plans to apply the newly learned skills back on the job. A copy is also sent to the ELP Coordinator after the end of the Capstone.

Methodology

The researcher used a survey consisting of 11 quantitative and qualitative questions sent to a population of 58 DAU ELP graduates. In order to make comparisons between different views, most demographics were pre-loaded to lessen the burden for respondents to answer several upfront demographics. Figure 2 depicts the survey which leveraged a matrix style format and cell groupings to shorten the appearance of the instrument to minimize survey length.

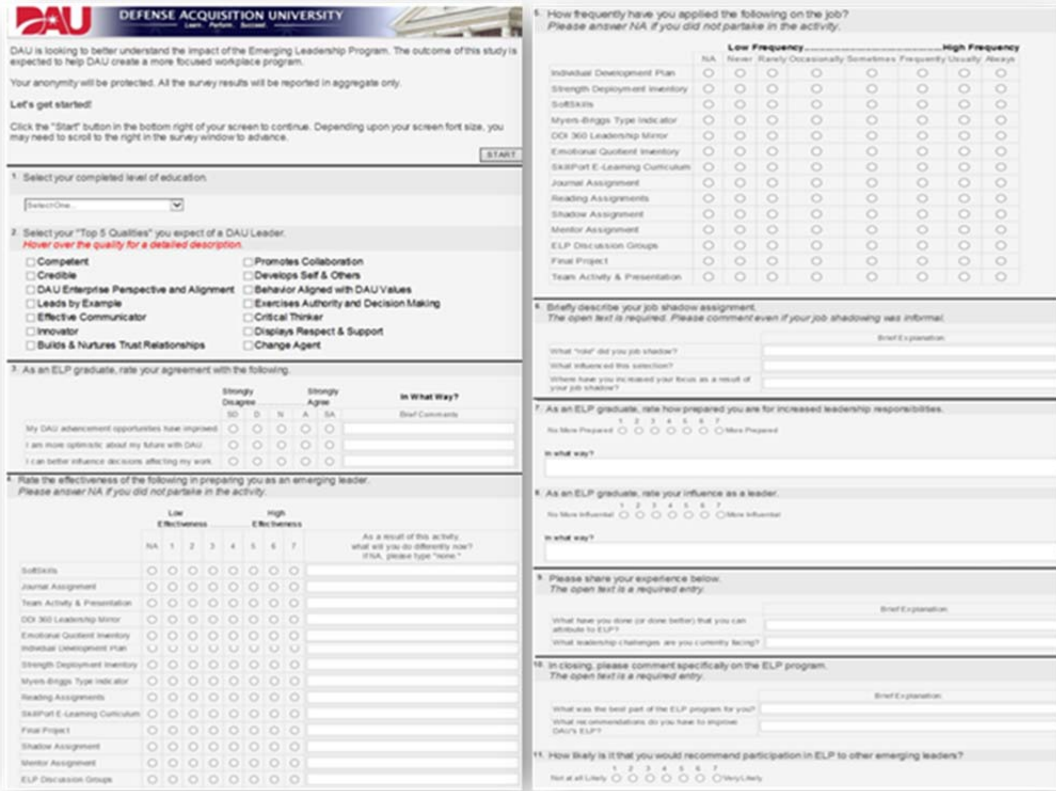


Figure 2. ELP Questionnaire

The survey questions were designed to measure the perceived effectiveness of 14 ELP activities and how frequently the ELP graduates were using their newly found skills. As a past graduate of the ELP program, the researcher included anecdotal observations relative to the 14 activities to underscore a greater (or lesser) value of each. Because the need for a staff leadership program was initially tied to climate survey results, the survey instrument also included three DAU climate survey questions for comparison to benchmarked data.

The results were collected in a survey tool and exported to Excel, as seen in Figure 3. The researcher used custom visual basic formulas to build summary arrays to display the results by respondent groupings. Below is an image of the “one world” summary sheet. The details will be explained in Findings—the next section.

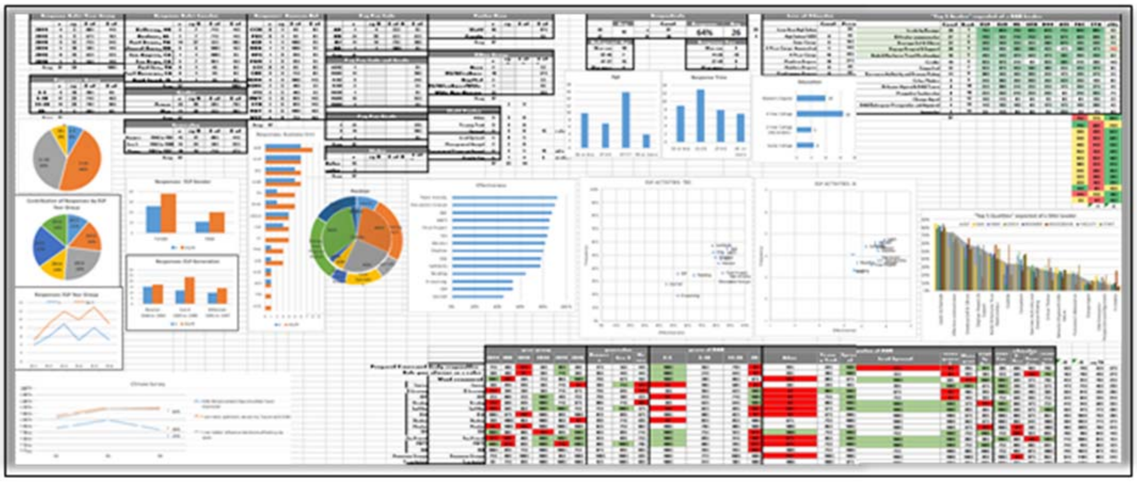


Figure 3. Summary Spreadsheet

Findings

The findings indicated a positive learning experience overall from the ELP. Respondents were constructive and suggested changes that the DAU should consider to improve the effectiveness of the activities that would lead to more favorable learning outcomes. Despite the fact that the ELP graduates' responses were similar to related questions in the climate survey, the ELP qualitative feedback indicated that the DAU should expand leadership opportunities for emerging leaders. The ELP graduates felt they needed to apply their newly found leadership skills more frequently and needed more opportunities.

Demographics

The participants from the DAU's ELP were well represented in the research study. The total response rate was 64%. Figure 4 displays each ELP year group and their representative percent of the total responses. With the exception of the initial ELP pilot, class size each year has ranged from nine to 13 students.

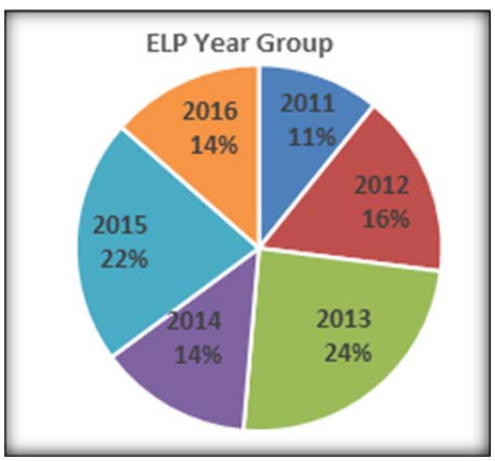


Figure 4. Contribution by ELP Year Group

Figure 5 shows how the two groups varied when comparing the respondents (n) to the total number invited (N) by year group, the gap (non-responders). Response rates also varied by generation:

- Boomers: 88%
- Gen-X: 50%
- Millennials: 71%

As Figure 6 shows, the Gen-X group was the largest non-responding group in this assessment. The lower response rate among Gen-X could be explained Gen-X's mistrust of technology (Erickson, 2008).

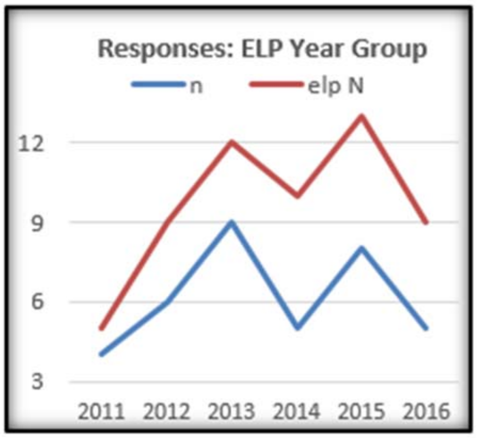


Figure 5. ELP Year Group Response Gap

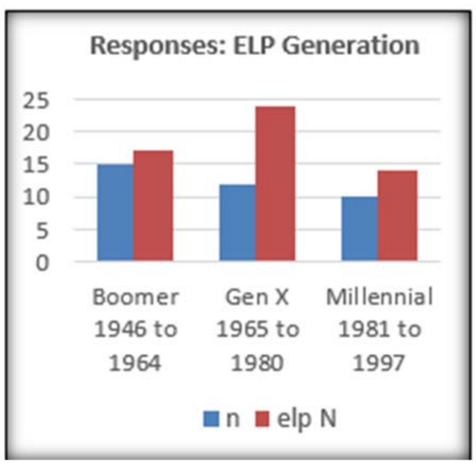


Figure 6. Responses by ELP Generation

For the purpose of this research, respondents were categorized as Line Staff (admin, training techs, and specialists), Mid-Level Staff (lead specialists, management analysts, and management program analysts), or Senior Staff (designated deputy personnel or senior supervisors). DAU staff who responded to the questionnaire came from diverse educational backgrounds. Almost a third held a master's degree, and almost half held a bachelor's degree. The remaining respondents either obtained a two-year degree or are actively seeking college credit. Additional stratification of the ELP responses will be conducted in supplementary research analysis to determine if there is any significant modulation within the ranking of the ELP activities among any particular respondent group.

Assessing the 14 ELP Activities

Fourteen diverse learning activities were assessed using a top box three (TB3) methodology (i.e., totaling the responses of 5, 6, and 7 on a Likert Like scale from 1–7 and then dividing by the total respondents). Figure 7 displays the learning effectiveness of each activity (a larger view is shown in Figure 1). The effectiveness of each activity could also be influenced by their frequency of use (e.g., daily, weekly, monthly, etc.). The scatterplot shows the respondents' aggregated average of the ELP attributes they rated for both effectiveness and frequency.

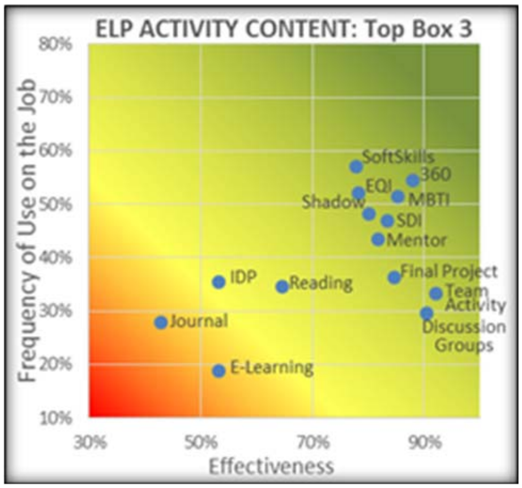


Figure 7. ELP Activities Rated

Combining both components could suggest a tight coupling (or not) between the two. For example, any particular activity with a high frequency and high effectiveness might start to wane (or not) if exercised less frequently. The researcher decided for the purpose of this study to keep "effectiveness" as the more influential attribute. Follow-on research could validate any changes in effectiveness for those components used more (or less) frequently over time.

Representative Comments

The following quotes are representative of the respondents' qualitative comments among the 14 ELP activities. They are listed in order of the TB 3 rating for effectiveness. Specifically, ELP graduates were asked, "What will you do differently now?"

Team Activity (TB: 92%) "I will …

… dive into projects that affect DAU as a whole."

… plan my own goals and keep others accountable to the project goals."

… stay connected with my ELP project team as a professional network."

… work better with teams to get full participation whether I'm a lead or a team member."

Author's Note: We learn from others and learn by doing … finding ongoing team opportunities can pay huge dividends.

Discussion Groups (TB: 90%) "I will …

... participate more and share my ideas even when they are different."

… provide candid feedback to other members."

… get to know my couterparts from across the university."

… appreciate the power of collaboration and 'bring it' when it applies."

Author's Note: Collaboration is important for innovation and creative tension … both of which can strengthen relationships and achieve more synergism.

360 (TB: 88%) "I will …

… work on continuous improvement and be more self aware in areas where I need to improve."

… improve on my effectiveness with communication skills and delegate."

… engage and work on perceptions on how I am seen by my leadership."

… continue to reflect—the 360 assessment was a profound learning moment about myself as a leader."

Author's Note: Assessing blind spots requires the perspective of others … but working past them is up to us and requires an honest commitment.

MBTI (TB: 85%) "I will …

… make more effective interactions with differing and similar personality types."

… consider adjustments needed with other personality types on how they work and interpret information."

… continue to apply MBTI and learn to better support my self-leadership."

… be more aware of other co-workers preferences so as to reduce conflict and increase group cohesion."

Author's Note: Armed with more knowledge about ourselves helps us break through our own mental models … we can learn even more by recognizing our filters.

Final Project (TB: 85%) "I will …

… continue to reflect and look back at projects."

… remember this—it was rewarding to see it come together as value added."

… look for IPT participation opportunities … good vibes on my presentation."

… do this again … enjoyed the process of pulling together a final project."

Author's Note: The experience of working a project can test our mental muscle … participating in "special projects" can also further strengthen learning.

SDI (TB: 83%) "I will …

… adjust my behavior and approach to conflict situations to be more effective."

… strategically approach conflict with leaders armed with knowledge of how my SDI compares to my leadership."

… be more aware of my stress reactions and make an effort to better deal with daily work stressors … especially when working on teams."

… apply immediately! An informative tool and profound moment to learning about myself as a leader."

Author's Note: Hesitation to engage could be the result of a blurred lens on "how to" approach or "how to engage." … The SDI offers insights to help clear the lens and better understand our proclivities.

Mentor (TB: 82%) "I will …

… continue weekly mentor vector checks … I have too much to work on but perspective is invaluable … learning to engage."

… find time to take on this challenge."

… reference back to my mentor when/if a situation warrants."

… utilize several suggested methods and reflect back on my mentor's insights."

Author's Note: It can be difficult to find a mentor and difficult to find time for mentoring … the right mentor–mentee chemistry is an ideal formulary.

Shadow (TB: 80%) "I will …

… request more stretch opportunities … research and/or support deep dives."

… ask for a specific project to lead now that I have had a shadow so I can exercise my new skills … I can contribute and lead from below."

… approach others with more confidence … a transformational experience."

… model my own approach from what I saw demonstrated … so much learning in the shadow!"

… look for more shadow opportunities … even if informal, this is an outstanding networking opportunity and learning opportunity."

… emulate some traits I observed as well as avoid some."

Author's Note: It's important to professionally stretch and build self-confidence … the trick is to monitor, fuse, translate, and eventually emulate from a master leader's shadow.

EQI (78%) "I will …

… be more patient with others."

… handle daily situations differently."

… show more empathy … a profound learning moment about myself as a leader."

… treat all colleagues as humans, regardless of position or status."

Author's Note: The first step is to understand yourself … make yourself smarter… then set a personal goal and look for ways to be more effective.

Reading (65%) "I will …

… work on developing an ongoing professional reading habit."

… look for another good leadership book."

… try to use and remember the skills I read … great recommended reading!"

… read additional books to continue to learn new leadership approaches and techniques."

Author's Note: Read more … the key is to pick the right reading that has the best return. Leaders should be avid readers.

e-Learning (53%) "I will …

… look for continued e-curriculum—it was free and it was helpful."

… use the new skills to deal with others in the circumstances I read about."

… do more e-leaning beyond the 'requirements' to improve my self-knowledge and skill-knowledge."

… keep soft skills on my reading list."

Author's Note: E-learning can be convenient but isn't always very "learning immersive" as live participation is. To make it more effective is all about the "approach" to learning.

IDP (53%) "I will …

… better assist subordinates with setting up their IDPs now that I have done one."

… continue IDP from a holistic perspective."

… spend more time discussing options outside of DAU's programs that supplement the ELP learning."

… plan my opportunities—'need' vs. 'want' because I can't do everything."

… continue to plan as I take my development seriously, but in many ways I think we check the box on this."

Author's Note: IDPs are very useful as long as they are realistic professional goals, thoughtful, and aligned with a strategic plan.

Journal (45%) "I will …

… look for guidance on how to journal or what to record."

… update my journal."

Author's Note: Reflection through journaling is most effective when integrated with other learning activities.

Representative Comments in Review

The ELP graduates' "I will …" statements reinforced how the ELP activities motivated them to apply what they learned. The comments also emphasized how this program can have a lasting impact. For example, most of the graduates spoke of seeking additional opportunities to participate in projects, IPTs, informal mentoring, and focus groups to practice and enhance their new skills. Recognizing the need to continuously learn new skills, ELP graduates inferred the importance of keeping their learning apertures wide open for additional opportunities.

Three Questions Asked to Target an Assessment of Impact

We generally measure what matters, but how do we know if we got it right with the ELP? It could be a question of benchmarking against the DAU's internal climate survey (collected on a Likert Like scale of 5). Figure 8 displays TB2 and BB2. Like TB3, TB2 totals the responses of 4 and 5 and then divides by the total respondents; likewise, BB2 totals the responses of 1 and 2 and then divides by the total respondents. Three DAU Climate Survey questions were considered in this research for measuring impact and how the ELP graduates responded post-graduation from the year-long program. ELP graduates offered candid comments—some very thought-provoking.

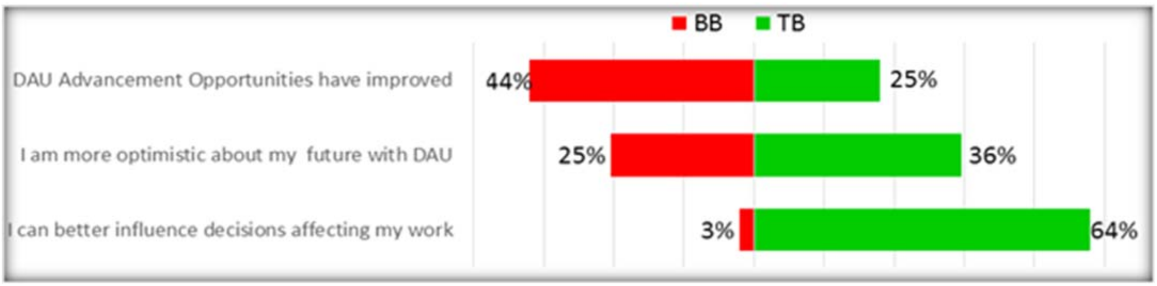


Figure 8. DAU Climate Survey Questions

DAU Advancement Opportunities & ELP Optimism

Bottom Box [red]: Opportunities for advancement are perceived dismal at best. The ELP, as the name suggests, should help develop leaders. But what does the ELP graduate emerge to? Respondents pointed out that advancement is limited for staff members. Because of the small departments organized around specific technical skills, they felt the opportunities for formal leadership positions appear to be rare. Initial discussions at the program outset also make it clear that there are few advancement opportunities at the DAU. A few respondents felt they were even encouraged to "move on" if they were seeking a leadership position—and they did. Over 10% of the graduates have left the DAU for advancement positions.

Top Box [green]: Who is this select group of ELP graduates who felt there are improvements in advancement opportunities when the facts do not necessarily support this ranking? Leadership and supervisory positions appear to be diminishing with recent efforts to "align leadership to at least ten per working group." Some of the respondents said they are "waiting it out" or "hanging in there" for leaders to retire or "move on." This group appears to be optimistic. They said "optimists get more done" and "attitude is everything," inferring that it could change how others see them as well. This group also does not see a clear path for advancement (i.e., "it isn't evident") but they seem to learn the power of influence as a surrogate leadership strength. Several respondents noted an equally rewarding experience through a cross departmental and highly collaborative project where they felt they markedly influenced the outcome.

Influencing Decisions at Work

Bottom Box [red]: There were very few comments in this group due to the small response rate in the bottom box. After getting to know the other ELP participants, one respondent remarked that their influence largely depended on the daily duties of the staff member. Another said, "I'm just a worker bee …" Perhaps seeing influence as an inherent feature of a particular position is holding some ELP graduates back who could otherwise be influential. One ELP graduate expressed a certain disappointment and said “almost everyone here at ‘x’ has been through the program” and none were “selected” because they were potential leaders.

Top Box [green]: The ELP graduates scoring in the TB box appeared to be more “comfortable” with everything from communication to making decisions when empowered to do so. They noted their communication skills had improved and they even understood themselves better. It wasn’t that they were the “loudest” or even the “most critical,” but they had developed their “voice” along with an increased ability (and responsibility) to use it and do what was needed for the team or project to succeed. The ELP graduates also noted the usefulness of creative tension and taking risks to share perspectives when it can help the project. One respondent said “accept it—there is potential to lead from below.” Another respondent said “my boss has confidence in me to perform my duties above my duty description.” This group of ELP graduates saw leadership differently and felt “I could influence decisions about my work … and now what I have learned is helping me to be a better asset to DAU.” In some cases, the ELP has provided the additional skills to act more assertively when they have an opinion they believe would make a difference.

Impact … Leadership … Isn’t It All About Influence?

ELP graduates said, “with the right skill and attitude, you can influence and win confidence to make an impact.” Research suggests that a Leader without Influence is not a very effective Leader. Others say, “An Influencer is a Leader. … They challenge processes.” In that context, ELP graduates who learn to influence despite certain hurdles can lead up and across, and ultimately achieve some of the same outcomes as leading down. The depiction shown in Figure 9 was adapted from a leadership blog (Rockwell, 2015) and captures another interpretation of influence.

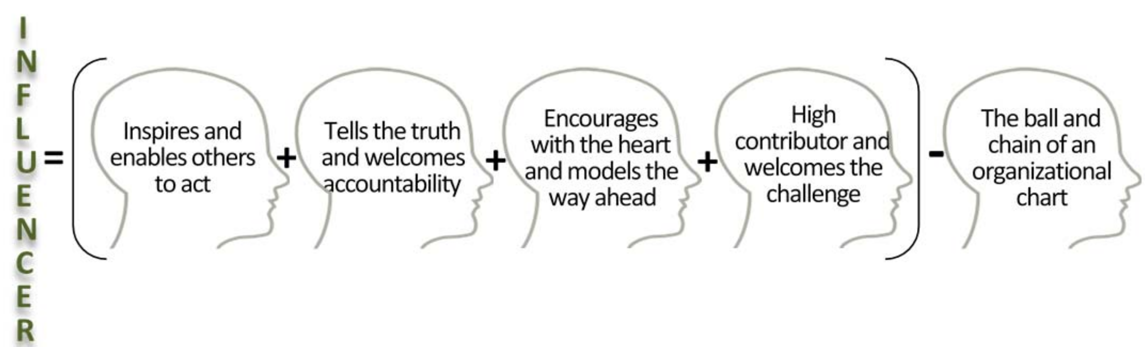


Figure 9. An Interpretation of Influence
(Rockwell, 2015)

Conclusion

After conducting a retrospective review of the ELP from its graduates, the DAU is in a better position to make several adjustments to the program and keep it relevant, challenging, and a platform to further develop staff. The ELP components that the respondents found very useful don’t need much tweaking. The ones that showed less value, including course readings, IDP, Journal, and e-learning could all use a boost. Either selecting more useful reading, allocating more time for the readings, and/or using the readings as a precursor for group discussions might ignite more interest and reinforce its value. Demonstrating a stronger connection between the IDP and ELP graduates' developmental guide path will help raise its importance. Maintaining a journal can be laborious and similar in context to maintaining a diary. Since it was the least effective ELP component, it might be better to withdraw it from the program in the context of other ELP demands. While e-learning is so pervasive these days, the learning drawback for ELPs could be affected by its asynchronous method of learning, that is, absent discussions among peers. Professional gains afforded by each of the ELP components varied, though appeared to have a closer grouping. Shadow assignments, self-assessments, and the soft skills all were rated as skillsets used more frequently, but what needs to happen to make all these more effective? Some of the respondents felt cohorts needed to continue after graduation to continue the focus on their development. TB3 81% of the ELP graduates felt the ELP prepared them for increased leadership responsibilities and that they had more useful tools to influence decisions. More importantly, they are ready to take a leap with the proviso that "Leaders make mistakes, too, but it's how they communicate and take accountability that makes them stronger leaders and real influencers." As one respondent so aptly said, if you aspire to be a "leader," take a hard look at how you "influence" and begin to develop those skills. As a fellow ELP graduate, I frequently practice how to influence my way ahead, and I also have a much deeper understanding of the powerful techniques at my disposal thanks to the ELP.

References

Fowler, D. (2015, October 19). *The emerging leader program guide*. Fort Belvoir, VA: Defense Acquisition University. Retrieved from https://daunet.dau.mil/Personnel/SitePages/Emerging%20Leader%20Program.aspx

Erickson, T. (2008, May 10). 10 reasons Gen Xers are unhappy at work. *Harvard Business Review*. Retrieved from https://hbr.org/2008/05/10-reasons-gen-xers-are-unhapp

Rockwell, D. (2015, December 29). *The exponential leadership goal for 2016* (C. Commons, Producer, & WorldPress.com). Retrieved from Leadership Freak: The Exponential Leadership Goal for 2016.

Seligman, D. (2009). *Climate survey IPT: Staff*. Fort Belvoir,VA: Defense Acquisition University.

Panel 21. Methods for Improving Cost Estimates for Defense Acquisition Projects

Thursday, May 5, 2016	
3:30 p.m. – 5:00 p.m.	**Chair: Major General Casey Blake, USAF,** Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force (Acquisition) ***The Role of Inflation and Price Escalation Adjustments in Properly Estimating Program Costs: F-35 Case Study*** Stanley Horowitz, Assistant Division Director, Institute for Defense Analyses Bruce Harmon, Research Staff Member, Institute for Defense Analyses ***The Impact of Learning Curve Model Selection and Criteria for Cost Estimation Accuracy in the DoD*** Candice Honious, Student, Air Force Institute of Technology Brandon Johnson, Student, Air Force Institute of Technology John Elshaw, Assistant Professor of Systems Engineering, Air Force Institute of Technology Adedeji Badiru, Dean, Graduate School of Engineering and Management, Air Force Institute of Technology ***Cost and Price Collaboration*** Venkat Rao, Professor, Defense Acquisition University David Holm, Director, Cost and Systems Analysis, TACOM LCMC Patrick Watkins, Chief, Stryker/Armaments Pricing Group, Army Contracting Command

Major General Casey D. Blake, USAF—is the Deputy Assistant Secretary for Contracting, Office of the Assistant Secretary of the Air Force for Acquisition, Washington, DC. He is responsible for all aspects of contracting relating to the acquisition of weapon systems, logistics, and operational and contingency support for the Air Force.

Prior to this assignment, he served as the Commander, Air Force Installation Contracting Agency (AFICA), Office of the Assistant Secretary of the Air Force for Acquisition, Wright-Patterson AFB, OH.

Maj Gen Blake entered the Air Force in 1984 through the ROTC program at The Citadel, Charleston, SC. He has served as an Information Management Officer, Air Force Mission Director on a NATO fighter wing, Warranted Procuring Contracting Officer for major defense acquisition programs, and Contracts Chief at both an air logistics center and product center. He commanded a contracting squadron at the base and major command specialized levels, served as both a Deputy Mission Support Group Commander and Defense Contract Management Agency Commander, has deployed as a Contingency Contracting Officer in support of Operation Iraqi Freedom, and served as the Senior Contracting Official–Afghanistan during Operation Enduring Freedom.

Maj Gen Blake has also served in several assignments as an acquisition contracting staff officer at a major command headquarters. He is a joint qualified officer, having served at the Office of the Secretary of Defense. During that time, he deployed to Iraq for two months in support of the Joint Forces Command Enabling Force Study.

The Role of Inflation and Price Escalation Adjustments in Properly Estimating Program Costs: F-35 Case Study

Stanley A. Horowitz—is Assistant Director of the Cost Analysis and Research Division at the Institute for Defense Analyses (IDA) in Alexandria, VA. Much of his work involves analysis of the Defense personnel compensation and management policies and the cost, measurement, and enhancement of readiness. Recently he has also been studying the use of inflation indexes in the DoD. He has directed studies of Reserve Component readiness, Reserve costing, Reserve training, and Reserve volunteerism. In 2015, he received the Andrew J. Goodpaster Award for Excellence in Research from IDA. Horowitz was trained as an economist at MIT and the University of Chicago. [shorowit@ida.org]

Bruce R. Harmon—works for the Institute for Defense Analyses, where he has been a professional research staff member for over 25 years. Harmon has extensive experience in modeling the costs and schedules of various aerospace systems, as well as in analyses of other acquisition issues. He is a PhD candidate in economics at American University, Washington, DC. [bharmon@ida.org]

Abstract

Applying price indexes presents a challenge in estimating the costs of new defense systems. An inappropriate price index can introduce errors in both development of cost estimating relationships (CERs) and in development of out-year budgets. In this paper we apply two sets of price indexes to the F-35 Joint Strike Fighter procurement program, both to help cost analysts understand the impacts of different price indexes and to provide guidance in their choice.

We approach this problem via hedonic price indexes derived from CERs. These indexes isolate changes in price due to factors other than changes in quality over time. We develop a "Baseline" CER model using data on historical tactical aircraft programs available at the F 35's late-2001 Milestone B decision. Comparisons are made between the Baseline model estimates, F-35 program office estimates, and estimates using cost models employing more conventional approaches to inflation adjustment. We find that the Baseline hedonic model provides estimates close to actual F-35 costs. As the hedonic index is directly estimated only for the historic period, we develop a procedure to project inflation rates based on historical hedonic index values.

Introduction

Background

The application of price indexes presents a substantial challenge in estimating the costs of new defense systems. The problem is twofold. First, the analyst must use a price index when normalizing historical cost data to a common point in time (where the normalized costs are referred to as "base year" [BY] dollars in defense acquisitions or, more generally, "real" dollars), so that these data can be used to help estimate the costs of future systems. Second, as budget requirements for future acquisitions are in "then-year" (TY) dollars (or more generally, "nominal" dollars), BY dollar estimates must be escalated to TY dollars using a price index. Using an inappropriate price index can introduce errors in both of these steps. In this paper, we apply two sets of price indexes to a cost estimating problem—the F-35 Joint Strike Fighter (JSF) procurement program. The purpose is to help cost analysts and others involved in the acquisition process understand the impacts of different price indexes and to provide guidance in their choice.

In general, price indexes isolate changes in price due to factors other than quality changes. These changes can be categorized into changes due to general inflation, changes in the overall price level in the economy (subsequently often just called "inflation"), and real price growth—price changes for a particular class of products relative to inflation. The combination of inflation and real price growth constitute price escalation—overall change in the price of a specified, constant quality, good, or service.

The point of departure for this work is the analysis of escalation indexes presented in Harmon, Levine, and Horowitz (2014; hereafter "D-5112"). The overall goal of that research was to identify a price index that is better than current indexes at meeting the Department of Defense's (DoD) need for a sound basis for cost estimation. In particular, we explored an alternative "hedonic" approach for calculating price indexes for tactical aircraft. In this analysis, we used updates to the hedonic model presented in D-5112 in the F-35 example.

The F-35 Cost Estimating Problem

The F-35 program has experienced significant program cost growth since its October 2001 Milestone (MS) B decision that initiated Engineering and Manufacturing Development (EMD). A substantial portion of this cost growth has been in its unit recurring flyaway (URF) cost, with much of this attributed to the incorrect application of price indexes (Arnold et al., 2010). Given the tactical aircraft focus of the Institute for Defense Analyses' (IDA) previous hedonic models, the F-35 makes for a suitable case study.

We used information available at MS B to develop models for exploring the effects of escalation adjustments on estimated F-35 URF costs. The resulting estimated costs can then be compared to several benchmarks, including cost estimates produced by the JSF program office (JPO) and observed URF costs for F-35s procured from 2007 to 2013. From this exercise, we draw lessons for future cost estimating practice.

Hedonic Price Index Models for Tactical Aircraft

Introduction

In this chapter, we review past work on hedonic price index models and present updates developed specifically for the F-35 cost estimation problem. The estimation of the hedonic indexes for tactical aircraft builds upon tools that cost estimators have used for years. The basic setup is

nominal system unit price = f(year, quality variables, other control variables)

The hedonic index application has commonalities with cost estimating relationships (CERs), which also model system costs as a function of quality variables, and cost/quantity relationships (primarily learning), which are control variables in the hedonic model. The hedonic index estimation differs from past cost estimating practice in that the price index is estimated simultaneously with other model parameters and the dependent variable is expressed in TY (nominal) dollars. In CER development, adjustments needed to normalize historical cost data to BY dollars used as the dependent variable are often performed using a general deflator based on an index of overall inflation, such as that published in the *National Defense Budget Estimates* "Green Book" by the Office of the Under Secretary of Defense (Comptroller; OUSD[C]).[1] For commodities such as tactical aircraft, a given observed price may reflect both inflation and relative price changes, including those due to

[1] The *National Defense Budget Estimates* is commonly referred to as "the Green Book." It is a reference source for data associated within current DoD budget estimates.

variation in the quantity purchased. Typically normalization to a common quantity (e.g., first unit or 100th unit)[2] is performed using BY dollars prior to CER estimation. Thus, another unique aspect of our modeling is the simultaneous estimation of CER and learning curve parameters, as well as production rate effects.

The hedonic analysis described in D-5112 used the direct time-dummy variable approach formulated by Triplett, an early developer of hedonic analysis (Triplett, 2006). The update to the earlier analyses also used this approach, along with the same set of explanatory variables (presented in Table 1). Five quality variables describe the aircraft, two quantity variables capture the cost effects of learning and production rate, and the time-dummy variables identify each fiscal year in which the aircraft were procured. The hedonic index is defined by the expression $b_t^{D_t}$,where D_t is a 1/0 dummy variable with a value of 1 for fiscal year t, and b_t is the estimated index for that year. BY dollars are calculated as $BY\ dollars_t = \frac{TY\ dollars_t}{b_t^{D_t}}$. In the application of the Green Book index, the index (where the base year value equals 1) replaces the $b_t^{D_t}$ expression in calculating BY dollars.[3]

Table 1. Explanatory Variables

Quality variables
Empty weight in pounds
Maximum speed in knots
Advanced materials as percentage of structure weight
Dummy variable for fifth-generation aircraft[a]
Dummy variable for Short Take-Off and Vertical Landing (STOVL) aircraft[b]
Quantity variables
Cumulative production
Lot size (number of aircraft produced in a year)
Time-dummy variables

[a] Fifth-generation aircraft are characterized by stealth, internal weapons carriage, avionics with information fusion, and support of net-centric operations. In the D-5112 sample, the F-22 and F-35 A/B/C were classified as fifth-generation aircraft; in the update, we added the F-117.

[b] The A/V-8B and F-35B, aircraft with STOVL capability needed for operations from small aircraft carriers and short unimproved airfields.

The database used in regression estimation contains pooled cross-section and time-series data, often called "panel data" in the econometrics literature, where each panel is an aircraft program. The cost metric of interest is the unit recurring flyaway cost (URF). In D-5112, the time series included 40 fiscal years (FYs 1973–2013), with 2012 as the base year; the cross-sections (panels) consisted of the 11 aircraft programs' original designs plus derivatives of these designs from series or block changes. In model estimation, the quality changes associated with the series/block changes are captured in the changes in empty weight over time. Production rate effects were calculated by estimating the annual fixed cost

[2] Although unit prices are also sensitive to production rate, this typically has not been taken into account.

[3] If the values for the Green Book escalation index were the same as the hedonic price index, all other model parameters would also be the same.

for each program.[4] Learning spillovers due to commonality between the EA 18G and F/A 18E/F and between the F-35 variants were included in the model.[5] We also accounted for loss of learning due to series/block changes.[6]

Updating Hedonic Price Index Models for Tactical Aircraft

For the current analyses, we made multiple changes to the previous work, including several versions of the model meant to capture different aspects of the F-35 cost estimating problem. Our primary focus is on the "Baseline" F-35 model; the intent was to use the vintage of information available for the MS B (October 2001) cost estimate. As the FY 2002 budget materials were released earlier in 2001, we used data through FY 2002. Eliminating the newer data means that we dropped the EA-18G from the data sample along with the three F-35 variants (F-35A, F-35B, and F-35C); also, the F-22A program is truncated. This left the F-22A as the sole fifth-generation aircraft with only two data points (2001 and 2002). In order to include another fifth-generation aircraft, we added the F-117A[7] to the updated sample.

In addition to the original series aircraft, derivative follow-on aircraft were relevant for the F-14A (F-14A+ and F-14B), F-15A (F-15C, F-15C MSIP, and F-15E), F-16A (F 16C Blocks 25/30/50), F/A-18A (F/A-18C and F/A-18C Night Attack), and A/V-8B (A/V-8B Night Attack and A/V-8B Radar).[8] As these derivative aircraft were produced serially, they were included in the same panel as the original design. We use 2002 as the BY price index; this was also the BY for the F-35 MS B estimates and the associated URF goal.

In addition to the Baseline model, we estimated other model variations to address different aspects of the F-35 cost estimating problem. The Green Book model replaces the statistically estimated hedonic index with the procurement budget index published in the FY 2002 *National Defense Budget Estimates*. This would be more typical of the approach used in CER estimation. All hedonic model variations follow the "Full CER Hedonic Model" approach from D-5112. We also estimated a "Full Information" model, using complete actual data through 2013. The purpose of that model is to provide a close comparison with the model included in D-5112.[9] A slight modification of this model excludes the F-35—the "Full Information less F-35" variation provides hedonic index values through 2013 without using any information from F-35 program cost experience. Unlike in the D-5112 and Full Information models, the Baseline model does not generate price index values from 2003 to

---

[4] Fixed costs for each program were estimated as a function of the estimated maximum variable costs.

[5] Learning spillovers are captured by estimating parameters that assign some portion of the cumulative quantity across related aircraft.

[6] This is accounted for by parameters that decrement cumulative quantity at each block change.

[7] Stealth technology is the prime feature of fifth-generation aircraft and the F-117. The F-117 differs from newer examples of fifth-generation aircraft in having less sophisticated electronic systems.

[8] Military aircraft are described by Mission-Design-Series (MDS). For the F-14A, for example, the mission is fighter (F), the design is 14, and the original series is A. The aircraft in column headings of Table 1 are new designs, with the exception of the F/A-18E, which was a major change from the previous F/A-18s. The three F-35 variants are being built for different missions and produced in parallel.

[9] The model in D-5112 only used data through 2012 and did not include the F-117A.

2013; instead, a methodology is presented in which model results are extrapolated to produce estimated index values through 2013.

Model Estimation and Results

This section presents regression results for the different model variations. Comparisons are shown between these models and the Full CER Hedonic Model described in D-5112. As the functional form of the models is the same, we do not repeat the detailed exposition presented in D-5112—instead, we highlight the differences in the regression results.

We estimate the model parameters using maximum likelihood estimation. The models are fit using the nonlinear optimization package within Microsoft Excel. The distribution of errors is assumed to be multiplicative/lognormal—this is analogous to estimating a log-log regression using linear regression.

Table 2 presents key regression metrics and parameter estimates for the five models.

Table 2. Comparison of Regression Results

	FY 1973–FY 2002		FY 1973–FY 2013		
Metric	**Baseline**	**Green Book**	**D-5112**	**Full Information**	**Full Information Less F-35**
Price index used	Hedonic	Green Book	Hedonic	Hedonic	Hedonic
Number of data points	117	117	150	159	143
Parameters estimated	41	11	55	54	53
Adjusted R^2	.97	.84	.97	.97	.97
Standard error	.09	.20	.09	.09	.09
Quality coefficients					
Empty Weight[a]	0.78	0.75	0.83	0.84	0.81
Maximum Speed[a]	0.29	0.08	0.30	0.28	0.26
Advanced Materials[b]	1.95	1.86	1.67	1.63	1.77
Fifth-Generation[b]	1.24	1.44	1.11	1.16	1.14
STOVL Capability[b]	1.00	1.00	1.10	1.05	1.00
1st unit cost (T1), FY02$M					
F-14A	240	119	271	261	261
F-15A	196	94	218	207	209
F-16A	97	50	109	104	104
F/A-18A	140	73	158	153	153
F-117A	187	128	189[c]	192	192
A/V-8B	81	49	94	88	87
F/A-18E	197	101	219	210	213
F-22A	370	212	368	367	365
F-35A	235[c]	144[c]	233	234	233[c]
F-35B	246[c]	154[c]	267	259	246[c]
F-35C	278[c]	169[c]	276	277	277[c]
Learning curve slope	84.5%	88.1%	83.9%	84.1%	84.1%
Escalation growth rate: 73–02	7.4%	4.5%	7.6%	7.5%	7.5%
Escalation growth rate: 02–13	N/A	2.1%[d]	3.6%	3.5%	3.2%

[a] The coefficients on these variables enter the model in the form x^b.
[b] The coefficients on these variables enter the model in the form b^x.
[c] Out-of-sample estimates.
[d] Extrapolated from projections in the FY 2002 Green Book.

The regression fits for the models in which a hedonic index is estimated are comparable. Restricting the index to that prescribed in the 2002 Green Book results in a significantly worse model fit. The learning curve slopes are similar for the hedonic models, but the slope is substantially shallower for the Green Book model (88% vs. 84%)—again, this is consistent with the embedded underestimation of escalation when normalizing the data to constant year dollars. Systematically lower constant dollar costs in the earlier years mean that the estimated learning effect is blunted. The steeper learning slope is also consistent with values of fighter/attack aircraft learning curve coefficients estimated using labor hour costs in previous studies (Resetar, Rogers, & Hess, 1991; Younossi, Kennedy, & Graser, 2001; Harmon, 2010).

Coefficients on weight, speed, and materials composition are relatively stable across the models and are consistent with those reported in past CER studies (Resetar et al., 1991; Younossi et al., 2001; Harmon, 2010; Harmon, Nelson, & Arnold, 1991). Unit prices increase with weight, maximum speed, and more advanced materials. The one exception is the speed variable in the Green Book model—as the aircraft with the highest maximum speeds (the F-15 and F-14) appear early in the sample, the underestimates of aircraft inflation associated with the model tend to bias its parameter estimate downward. Estimates for the fifth-generation and STOVL aircraft effects change some when the F 117 is introduced into the sample. The fifth-generation factor increases from 1.11 to 1.16, while the STOVL factor decreases from 1.10 to 1.05. When the F-35 is excluded from the regression, the STOVL factor goes to 1.00—this reflects the influence of the F 35B (which is a fifth-generation STOVL aircraft), with the A/V-8B the only other STOVL aircraft in the sample.[10] The range of fifth-generation premiums for the hedonic models is generally consistent with values from an earlier IDA paper on the cost of stealth (Nelson et al., 2001), although the 1.24 factor for the Baseline model is somewhat higher than expected. The 1.44 factor estimated with the Green Book model is clearly too high—the bias is a mirror image of the maximum speed coefficient, where underestimated escalation and newer fifth-generation aircraft interact. Thus, if there is a relationship between time and the values of the quality variables, a systematic bias in the price escalation used will result in a related bias in the coefficients on the quality variables. Also note that the analogous cost drivers in the historical studies are usually estimated using labor hour data, eliminating the possibility of bias from price escalation.

Estimated first unit variable costs (T1s) for each initial Mission-Design-Series (MDS) (usually the “A” series) are calculated using the quality coefficients, the regression intercept, and the values of the quality variables for each MDS. Table 2 (on page 5) shows the T1s for all relevant MDS, including “out-of-sample” cases in which the MDS was not used in model estimation. These cases are the F-35 variants, with the exception of the F-117A, which was not used in estimating the D-5112 model. For the models using the hedonic indexes, the out-of-sample estimates were close to the values calculated using the models that included those MDS. The exception is the F-35B, where the more complex STOVL capabilities were not well captured in the models not using the F-35 data. Even in this case, the out-of-sample F-35B T1s are only around 5% lower than the estimates from the other hedonic models. The T1s from the Green Book models are all substantially lower than those from the hedonic models. This is consistent with the shallower learning curve for the Green Book model, where the real prices of the initial lots are systematically underestimated because of biased

[10] This does not mean that STOVL capabilities are free in the model; holding all else equal, STOVL aircraft will tend to be heavier and have more advanced materials than a conventional aircraft. Also note that in model estimation, the coefficient on the STOVL dummy was restricted to ≥ 1.00.

escalation. Figure 1 shows the escalation indexes for a selection of the regression models.[11] Also included for comparison is the latest (FY 2015) Green Book index.

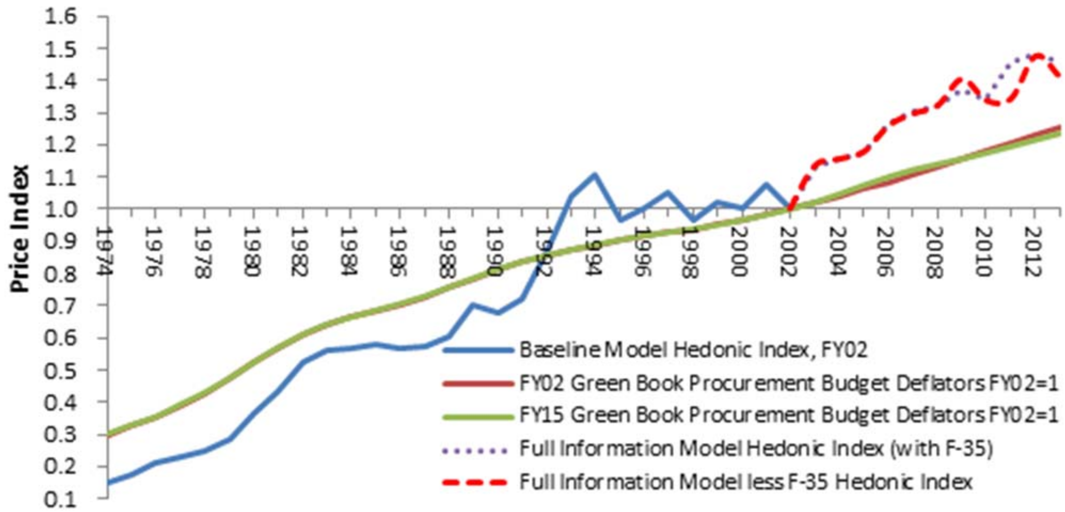


Figure 1. Comparison of Price Indexes

These indexes are portrayed in the price growth rates shown in Table 2. Of most interest for the F-35 estimating exercise are the Baseline and Green Book models. The other models are included for comparison purposes as well as to provide escalation estimates through 2013. There is no 2002–2013 escalation associated with the Baseline model; one of the goals of our analyses is to suggest a methodology for extrapolating forward growth rates from the Baseline model hedonic index. Also note how little the Green Book inflation changed from the FY 2002 forecasts (including extrapolations from FY 2007 to FY 2013) through the actuals reflected in the latest FY 2015 values.

Normalizing the data using the Green Book index results in a constant-dollar cost data set and associated model that systematically underestimates costs in the earlier years and overestimates costs in the later years. In addition to introducing bias in the quality parameters, using the Green Book index also results in a shallower learning curve. This behavior is not evident in the Baseline model. It is clear in both the distortion of the parameter estimates and the systematic errors in estimating the actual data that a naïve application of price indexes can be problematic.

[11] The published FY 2002 Green Book deflators only include projections through FY 2007. Beyond FY 2007, we use the 2.1% inflation rate evident in the FY 2004 to FY 2007 projections.

F-35 Cost Estimating Applications

Introduction

We compare F-35 URF estimates generated by the Baseline and Green Book models against three sets of benchmarks. They include

- MS B program cost estimates and subsequent cost estimates associated with the 2009 "Nunn-McCurdy" unit cost breach,[12] in *BY 2002* dollars;
- Actual TY dollar budget values for the 2008–2013 fiscal year lots; and
- The latest program cost estimate as reported in the December 2013 selected acquisition report (SAR), reported in TY dollars.

To do this, we use the Baseline and Green Book models to produce BY 2002 cost estimates for each scenario. For comparisons with the TY actuals and estimates we use either an index calculated from the historical hedonic index ("projected hedonic index") or the Green Book index. The BY 2002 estimate comparisons demonstrate the effect of different price indexes on the structure of the CER model, while the TY dollar estimates also show the effect of the different indexes in projecting BY estimates forward.

F-35 MS B and Nunn-McCurdy Breach Estimates

MS B estimates are the initial benchmarks used for budgeting and for calculating program cost growth. As both models take into account production rate and learning, they can produce an analog of the MS B estimate using the quantities and production schedule associated with the October 2001 program. The IDA model estimates in this application do not carry explicit assumptions regarding future (post-2002) escalation—they are in BY 2002 dollars as directly calculated by the model. Figure 2 shows comparisons between the MS B URF estimates (all F-35 variants combined) and those generated by the Baseline and Green Book models using MS B input values.

The estimates from the two models converge as a result of the shallower learning slope of the Green Book model. Both models produce estimates above the program MS B URF estimate. However, they are substantially below the 2009 SAR estimates that triggered the Nunn-McCurdy breach. Many elements of F-35 cost growth are not captured in the above model estimates. Data from Arnold et al. (2010) allow us to isolate and deconstruct the URF portion of the cost growth.[13]

---

[12] A Nunn-McCurdy unit cost breach (10 U.S. Code § 2433a, "Critical cost growth in major defense acquisition programs") occurs when cost growth in program or acquisition unit costs surpasses 15%.

[13] The 2009 F-35 Nunn-McCurdy breach was driven by cost growth in EMD and nonrecurring procurement as well as by URF.

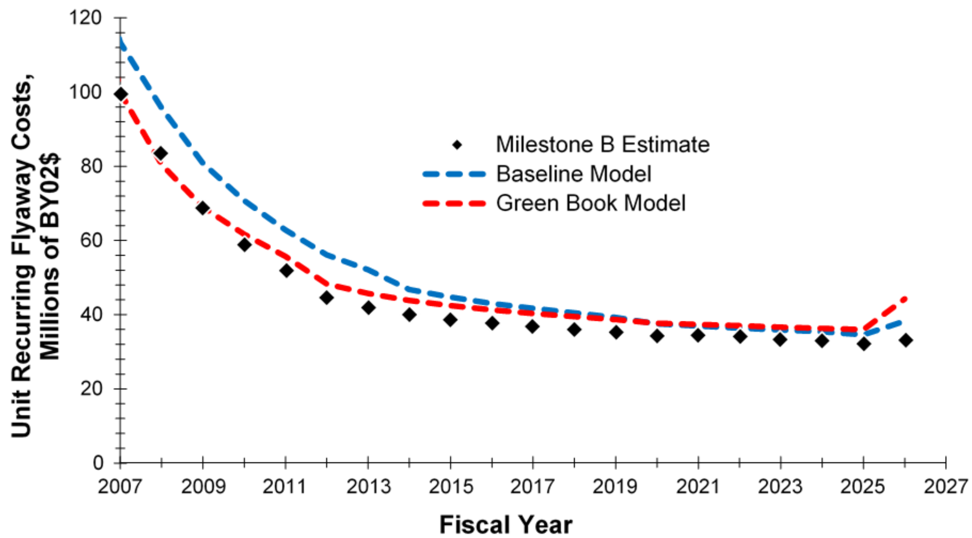


Figure 2. Comparisons of Milestone B and Model Estimates for All F-35 Variants

Weight growth in all F-35 variants was a driver of cost growth between MS B paper designs and the current designs reflecting the aircraft as produced. Almost all weight growth attributable to redesign was evident by the 2009 Nunn-McCurdy breach and reflected in the production lots.[14] As empty weight is an input to the models, the weight growth must be taken into account when comparing model outputs to the MS B estimates and subsequent cost growth. Another change affecting cost model application is the decrease in commonality between variants (F-35A/F-35B/F-35C) since MS B. Current commonality is reflected in the "spillover" parameter affecting learning across variants estimated as part of the Full Information model. The cost effects of commonality have been estimated by the JSF program using a detailed assignment of the learning quantities depending on common component applications. As we cannot reproduce such a detailed analysis, we make use of the spillover parameter instead—for the MS B estimate we increase its value to reflect higher commonality assumed at that point.

Table 3 shows the MS B URF estimate, a buildup of cost growth drivers to the 2009 estimate as derived from Arnold et al., and comparisons with the model estimates. Model estimates presented include calculations with MS B inputs, and with inputs reflecting contemporary values for empty weight and commonality (learning spillovers).

[14] We used the latest available weight status to characterize the F-35 variants as procured. These values were fixed across the procurement lots and do not include any weight growth margin.

Table 3. F-35 Program Growth Track From Milestone B to 2009 SAR and Model Estimate Comparisons

Metric	F-35 Program URF Cost, in Millions of BY 2002$			
	Cost Growth Increment	Cumulative Cost Growth	Baseline Model	Green Book Model
MS B Estimate		40.7		
Major Subcontractor Fee	1.5	42.2		
Change in Materials Manufacturing Efficiency	3.0	45.2	47.3[a]	44.6[a]
Design-Negated Affordability and Production Efficiency Plans	3.0	48.2		
Aircraft Weight Growth	3.0	51.2	52.1[b]	48.4[b]
Change in Buy Profiles (2009 SAR)	2.5	53.7		
Escalation Rates (2009 SAR Estimate)	7.0	60.7		

[a] MS B weight and commonality
[b] Contemporary weight and commonality

We orient the model outputs in the table to reflect how they relate to the cost-growth elements from the MS B estimates. Elements that represent underestimates based on a departure from business as usual (i.e., the historical database) are included above the model estimates calculated with the MS B weight and commonality assumptions. The estimates reflecting updated weight and commonality are in line with cost growth through the Aircraft Weight Growth row. Not accounted for in this application of the IDA model estimates are cost increases due to buy profile changes (a reduction in quantities and a stretch-out of the procurement schedule) and a misapplication of escalation rates for future costs.[15] The last cost-growth element is informative of our research question. Instead of using contractor-specific labor rate escalation, the JPO used OUSD(C) Green Book inflation when converting constant dollar estimates to TY dollar estimates.

From Arnold et al. (2010):

> However, at the time of Milestone B, the Defense Contract Management Agency (DCMA) and Lockheed Martin had already agreed to a Forward Pricing Rate Agreement (FPRA) that increased rates more than the OUSD(C) escalation indices ... therefore, the fully burdened labor rates turned out to be significantly higher than those used in the JPO Milestone B [estimate]. (p. 12)

The preferred methodology reflected in the 2009 JPO cost estimate is to escalate estimated constant year costs to TY dollars using escalation rates appropriate to the different cost elements. The OUSD(C) index is then used to de-escalate the TY dollars to BY dollars, which are, in turn, reported in the SARs and used as a basis for cost-growth calculations. This correction of the original methodology is responsible for the $7 million unit cost growth due to escalation rates shown in Table 3. Analogous steps are not reflected in the BY 2002 model estimates in Table 3; thus, the constant year model estimates presented for comparison are conceptually similar to the JPO's MS B estimates, reflecting the same

---

[15] Both of these effects are addressed in the later benchmark comparisons.

error.[16] In the next sections, we focus on model-generated TY estimates in the context of more up-to-date F-35 estimates.

F-35 Actual Budget Values

This section compares model-generated estimates with actual historical costs. The emphasis is on the results from the Baseline model. The budget experience is taken from Navy and Air Force President's Budget (PB) Justification Books, "Exhibit P-5, Cost Analysis" sheets. In collecting these data, we used the values in the latest PB in which they appeared; for example, for the FY 2013 lot, we used data presented in the FY 2015 PB submission. For this exercise, we used the unadjusted TY URF values.

For the Baseline model, we developed the projected hedonic index to generate TY estimates through FY 2013. We also included results for the Green Book model, where the FY 2002 Green Book index (including extrapolations through FY 2013) is applied. We used the hedonic indexes generated by the Full Information and Full Information Less F-35 models for comparison purposes. For model inputs, we used contemporary values for the quality variables and the procurement profiles reflected in the budget data.

The projected hedonic index is based on the relationship between the FY 2002 Green Book and Baseline hedonic indexes; it has the advantage of using only information through 2002 while taking into account the systematically higher escalation rates associated with the hedonic indexes vs. the Green Book rates.

To calculate the projected hedonic index, we first define the relationship between the Green Book index and the hedonic index using data through 2002 as estimated by the Baseline model. Given the year-to-year volatility of the hedonic index, we do this by comparing 10-year compounded annual growth rates. These data are shown in Figure 3.

[16] Although it would be possible to capture the 2009 procurement profile and escalation application effects in the modelling exercise, we only address these issues in the context of more up-to-date cost data.

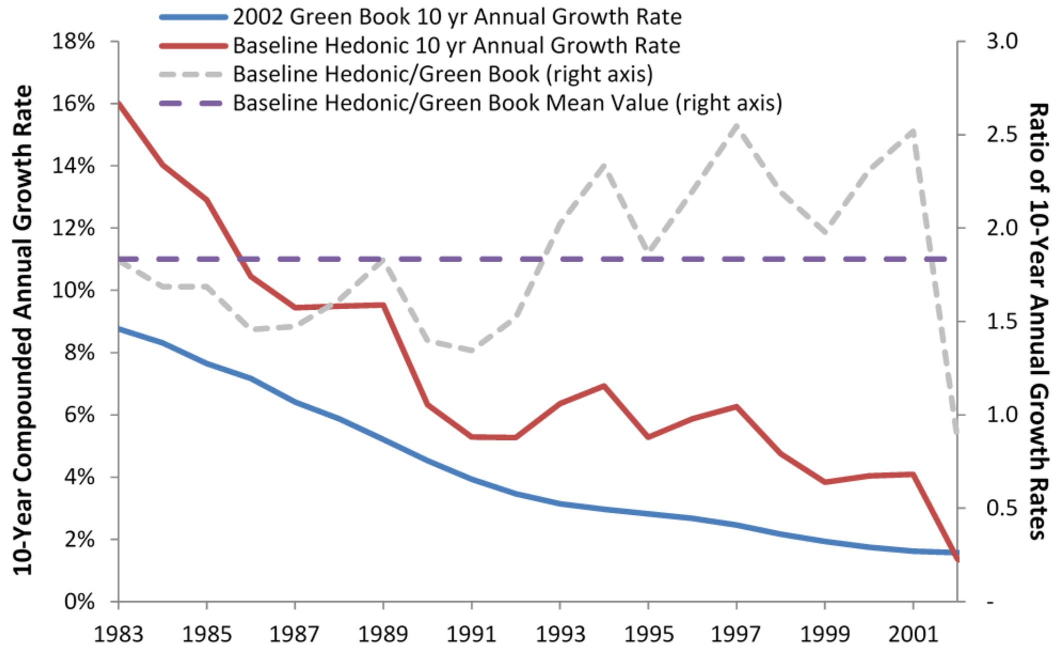


Figure 3. Comparison of Baseline Hedonic and Green Book Index Growth Rates, 1983–2002

Examination of the data shows that the hedonic and Green Book indexes relate to one another most consistently through a multiplicative factor vice an additive adjustment. We use the calculated average ratio (mean value) of 1.83, shown in the figure as a conversion factor on the 2003–2013 Green Book values, to arrive at the projected hedonic index. This is shown along with the other indexes in Figure 4.

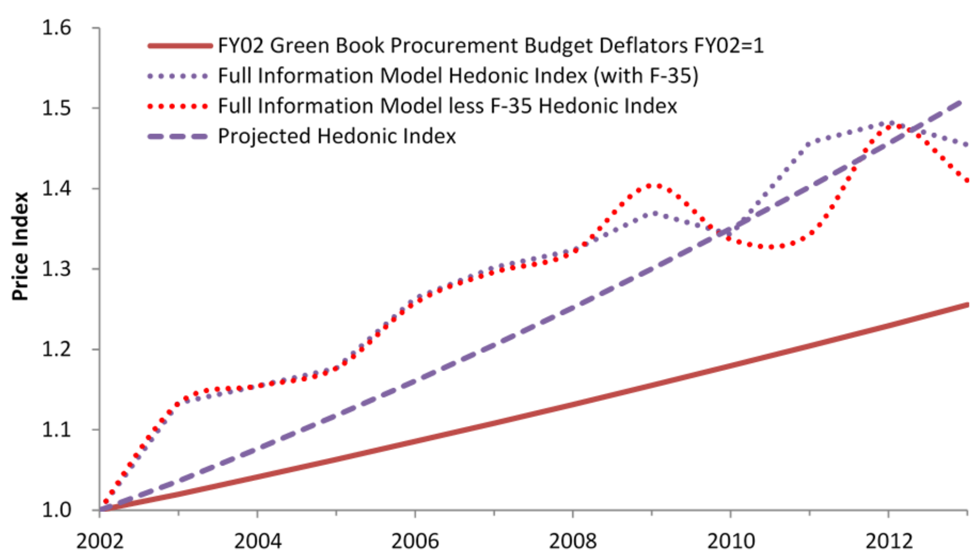


Figure 4. Comparison of Hedonic and Green Book Indexes, 2002–2013

Figure 5 compares the URF estimates associated with the two models and three escalation index assumptions with the budget actuals.

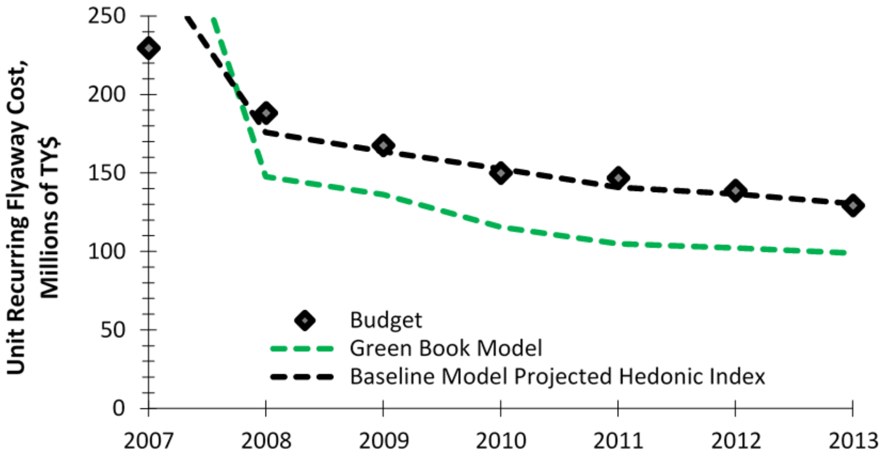


Figure 5. Comparison of Model Estimates With Budget Actuals, All F-35 Variants

Table 4 compares the estimated URF costs with the budget actuals calculated for the 2007–2013 budget years, broken out by variants.

Table 4. Comparison of Estimates of 2007–2013 URF Costs, Millions of TY$

Variants	Actual Budget	Baseline Model, Projected Hedonic Index	Green Book Model and Index
All Variants	149	147	115
F-35A	139	137	110
F-35B	160	152	121
F-35C	167	175	124

The results show that the Baseline model estimates when projected forward using the hedonic index come close to the actual budget values for 2007–2013; estimates depending on the Green Book index consistently underestimate the budget URF costs. However, the Baseline model tends to miss the costs for the individual variants, with the F-35B underestimated and the F-35C overestimated. This result is consistent with the differences in parameter estimates between the Baseline and Full Information models, which are, in turn, a result of the more complex STOVL implementation of the F-35B relative to the A/V-8B that is not completely captured in weight differences.

F-35 2013 SAR/PB 2015 Estimates

This section takes a somewhat different approach to the F-35 estimating problem. The question we want to answer is: what scaling of the FY 2015 Green Book index results in the closest fit to the latest JPO estimates? While the previous F-35 estimating exercises took the data available in 2002 as given, in this case we assume contemporary data for escalation projections. To address this question, we only use the Baseline model with the projected hedonic index as presented above. For 2014 onwards, we scale the FY 2015 Green Book index by a multiplier analogous to the factor used to calculate the projected hedonic index. The multiplier is determined by scaling the Green Book index such that the model-estimated totals for 2014–2037 are the same as those reported in the SAR. The resulting factor is 1.75—comparing directly with the 1.83 factor used to calculate the projected hedonic index. This analysis is shown graphically in Figure 6.

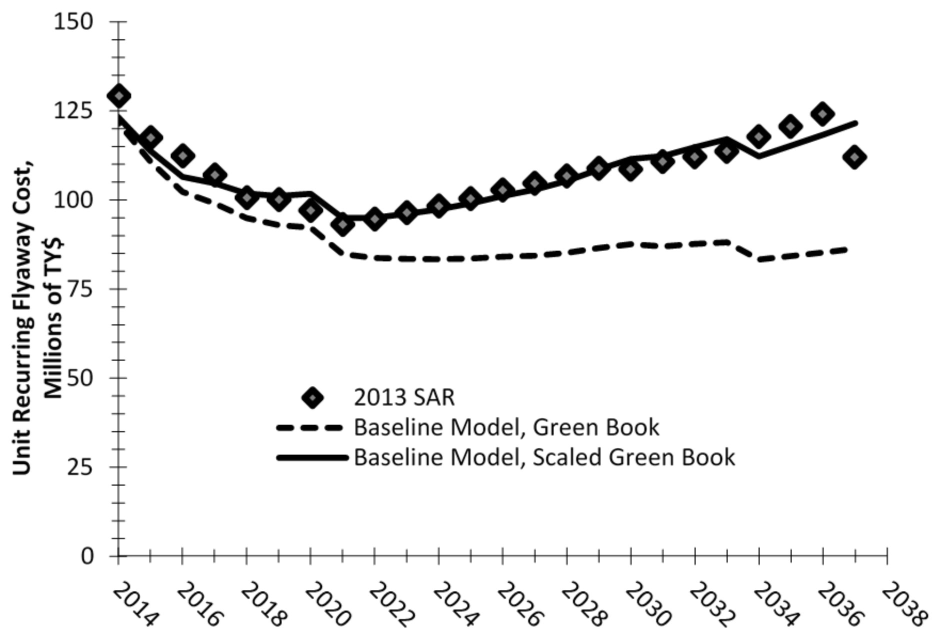


Figure 6. Comparison of Model Estimates With the 2013 SAR Estimates, All F-35 Variants

If the estimates are projected using the unadjusted Green Book index, the 2014–2037 URF estimate is $88 million vs. $106 million reported in the SAR. This shows the impact of the different indexes on projected costs, isolated from their influence on defining the CER model.

Summary and Conclusions

This paper describes different approaches to estimating expected price growth in defense system costs. The comparison of cost estimates based on escalation predictions derived from hedonic modeling with F-35 budget actuals through FY 2013 is particularly interesting. Although the model inputs reflect the latest F-35 aircraft characteristics and program parameters, in terms of the structure of the model and escalation projections, the models are defined by the information that was available at MS B. As the hedonic index is directly estimated only for the historic period, we apply a methodology to project forward escalation rates associated with the hedonic index. This example shows the close correlation between the Baseline hedonic model estimates and the budget actuals. The lower estimates from the Green Book model are due to two factors: the underestimates of escalation from FY 2002 to FY 2013 and biases introduced into the model parameters because of underestimates of escalation in the historical period.

Looking out to FY 2037, we find that projecting escalation using our approach closely mimics the more detailed buildup of input-specific escalation rates used by the JPO. This is in contrast to projections using Green Book escalation, which result in an $18 million underestimate in unit costs.

We demonstrate the effect of different escalation methodologies using top-level CER models. Cost analysts usually build up their estimates from a more detailed level. However, issues regarding the proper application of price indexes, for both normalizing historical data and making projections, are equally valid in more typical cost estimating environments. For example, rates of price growth for raw material inputs, propulsion systems, electronic components, and labor inputs are likely to be different from that of general inflation. In our last example, we calculated overall escalation rates implied in the JPO estimates for the rest of the F-35 program; we found these escalation rates to be consistent with those projected using values from the historical hedonic price index.

The main point is not the superiority of hedonic development of escalation indexes. Rather, it is that cost analysts should be attentive to possible differences between inflation and escalation and the implications of using inflation as a proxy for escalation when it is not a good one.

References

Arnold, S. A., Byun, J. S., Cloud, H. A., Gallo, A. O., Gonwa, M. W., Harmon, B. R., & Bronson, P. F. (2010). *WSARA 2009: Joint Strike Fighter root cause analysis* (IDA Paper P-4612). Alexandria, VA: Institute for Defense Analyses.

Harmon, B. R. (2010). (U) *Cost estimating techniques for tactical aircraft manufacturing labor* (IDA Paper P-4490). Alexandria, VA: Institute for Defense Analyses. (Unclassified/PI/LR).

Harmon, B. R., Levine, D. B., & Horowitz, S. A. (2014*). Inflation adjustments for defense acquisition* (IDA Document D-5112). Alexandria, VA: Institute for Defense Analyses.

Harmon, B. R., Nelson, J. R., & Arnold, S. A. (1991). *Unit cost implications of new materials: Preliminary analyses of airframe experience* (Revised) (IDA Document D-908-REV). Alexandria, VA: Institute for Defense Analyses.

Nelson, R., Harmon, B. R., Bontz, R. J., & Devers, W. C. (2001). (U) *Cost estimating for modern combat aircraft: Adjusting existing databases and methods to include low-observable cost considerations* (IDA Paper P-3528). Alexandria, VA: Institute for Defense Analyses. (Secret/PI/LR).

Office of the Under Secretary of Defense (Comptroller). (2015, March). *National defense budget estimates for FY 2016* ("The Green Book"). Washington, DC: Author.

Resetar, S. A., Rogers, J. C., & Hess, R. W. (1991). *Advanced airframe structural Materials: A primer and cost estimating methodology* (R-4016-AF). Santa Monica, CA: RAND.

Triplett, J. E. (2006). *Handbook on hedonic indexes and quality adjustments in price indexes: Special application to information technology products*. Paris, France: Organisation for Economic Co-operation and Development.

Younossi, O., Kennedy, M., & Graser, J. C. (2001). *Military airframe costs: The effects of advanced materials and manufacturing processes* (MR-1370-AF). Santa Monica, CA: RAND.

Acknowledgments

Thank you to Sarah K. Burns and David M. Tate for performing a technical review of this document, and to Steve Miller, Carol Moore, and Garret Summers of the Office of the Secretary of Defense, Cost Assessment and Program Evaluation for inspiring and commenting on the analysis.

The Impact of Learning Curve Model Selection and Criteria for Cost Estimation Accuracy in the DoD

Candice M. Honious—is the EMD Cost Analyst for the Combat Rescue Helicopter program at Wright-Patterson Air Force Base in Dayton, OH. She received her master's degree in cost analysis from the Air Force Institute of Technology and a bachelor's and master's degree in accountancy from Wright State University in Dayton, OH. [candice.honious@us.af.mil]

Captain Brandon J. Johnson—is a Weapons Systems Cost Analysist at Wright-Patterson Air Force Base in Dayton, OH. He received his master's degree in cost analysis from the Air Force Institute of Technology, as well as a Master of Business Administration from Oklahoma State University and a bachelor's degree from the United States Air Force Academy. [brandon.johnson@afit.edu]

John J. Elshaw—is an Assistant Professor of Systems Engineering in the Department of Systems Engineering and Management at the Air Force Institute of Technology. He holds a BS in accounting from The University of Akron, an MBA from Regis University, and a PhD in management with a specialization in organizational behavior and human resource management from Purdue University. [john.elshaw@afit.edu]

Adedeji B. Badiru—is a Professor of Systems Engineering at the Air Force Institute of Technology. He is a registered Professional Engineer (PE), a certified Project Management Professional (PMP), and a Fellow of the Institute of Industrial Engineers. [adedeji.badiru@afit.edu]

Abstract

The first part of this manuscript examines the impact of configuration changes to the learning curve when implemented during production. This research is a study on the impact to the learning curve slope when production is continuous but a configuration change occurs. Analysis discovered the learning curve slope after a configuration change is different from the stable learning curve slope pre-configuration change. The newly configured units were statistically different from previous units. This supports that the new configuration should be estimated with a new learning curve equation. The research also discovered the post-configuration slope is always steeper than the stable learning slope. Secondly, this research investigates flattening effects at tail of production. Analysis compares the conventional and contemporary learning curve models in order to determine if there is a more accurate learning model. Results in this are inconclusive. Examining models that incorporate automation was important, as technology and machinery play a larger role in production. Conventional models appear to be most accurate, although a trend for all models appeared. The trend supports that the conventional curve was accurate early in production and the contemporary models were more accurate later in production.

Introduction

The Budget Control Act of 2011 subjected the Department of Defense (DoD) to a more fiscally constrained and financially conscious environment than ever before, juxtaposed with a demand for new aircraft programs of almost every type. As an increasing number of programs are terminated, with budget overruns being a contributing factor, managers at every level in the DoD are expected to ensure the Department's shrinking budget is being used in the most effective way. The increased scrutiny adds greater emphasis on the accuracy of program office cost estimates given that an approved program cost estimate supports every major aircraft acquisition program funded by the Department.

The current state of the DoD includes shrinking budgets and large funding cuts for acquisition programs. An extra emphasis on scrutiny of accurate cost estimates is the result of the current cuts and budget issues. There is a new standard for Financial Managers and Program Managers who have to support and maintain a cost estimate like no time before.

Having a balanced budget is a concern for the DoD, and the budget depends on the cost estimate. Gone are the days when the DoD had ever-growing budgets where the fiscal mentality was to spend. The fiscal mentality now involves saving and receiving as much value as possible for every budgeted dollar.

In order to obtain reliable cost estimates, cost estimating models and tools within the DoD present the opportunity for an evaluation on their accuracy. The current learning curve methods within the DoD's cost estimating procedures are from the 1930s. As automation and robotics increasingly replace human touch-labor in the production process, a model that is 80 years old and assumes constant learning may no longer be appropriate for accurate learning curve estimates. Robotics and automation do not learn, and they are inevitably a part of future production. New learning curve methods that consider automated production should be examined as a possible tool for cost estimators to utilize. The modern learning curve methods could be a useful tool for obtaining better cost estimates within the DoD. The purpose of this research ultimately is to investigate new learning curve methods, develop the learning curve theory within the DoD, and pursue a more accurate cost estimation model.

A vital input to the cost estimate for a production program is the assumed learning curve slope for the program. The learning curve often depicts the learning phenomenon that occurs in manufacturing. Learning is defined as a constant percentage reduction of the required touch labor hours (or costs) to produce an individual unit as the quantity of units produced doubles (Yelle, 1979, p. 302)—as the number of units produced doubles, the number of hours required to produce a single unit decreases by the learning curve rate. Learning is also defined as both the conceptual and the physical learning of a physical process (Watkins, 2001, p. 18). The learning curve for a program is generally considered stable once the program is substantially into production because the manufacturer and laborers have produced enough units to learn the most efficient production process. However, intuitively and through past research, it is known that learning is disrupted by changes in production, and only the production of additional units can recover the lost learning (Watkins, 2001, p. 18). It is critical to capture the change in the learning rate due to production modifications to better estimate DoD program costs.

This comparative analysis study will examine whether different learning curve models are more accurate than Wright's Learning Curve model (the status quo) when comparing actual values to predicted. The current DoD learning curve methodology does not take into account available information and factors that contribute to learning. The point of emphasis for this research and the issue that needs to be resolved is that DoD agencies need to estimate more accurately. Prior research on this subject shows that the learning curve methods have room for further development. There may be an opportunity to incorporate alternate learning curve models and more DoD programs into this area of research. Research found that an important factor (incompressibility) was not explicitly researched or known. Towill and Cherrington defined incompressibility as the percentage of the learning process that is automated (Towill, 1990). Robotics and automation are not going away and will likely play a larger role in the future. Research on what that factor actually equals or how it relates to different airframes could be critical for obtaining a more precise model. Using integration, assembly, and checkout (IA&CO) processes instead of complete touch labor processes should provide an analysis that is more insightful and potentially leads to a more accurate model. IA&CO are specific work that occurs during production.

The idea of learning in a production environment is well established. T.P. Wright first published the learning curve phenomenon in early 1936. Wright observed that in a manufacturing environment, as the cumulative quantity of units produced doubled, the cumulative average cost decreased at a constant rate (Wright, 1936, pp. 124–125; Yelle, 1979, p. 302). During World War II, government contractors investigated the usefulness of the learning curve concept to predict labor hours and cost requirements for aircraft and ship construction projects. The private sector went on to adopt the learning curve theory into practice shortly thereafter. Although learning curve theory has evolved and has been referred to by different names in the decades following Wright's report, including the experience curve, the progress curve, and the improvement curve, Wright's model remains one of the models most widely used by manufacturers to predict labor hours and costs (Yelle, 1979, pp. 303–304; Badiru, 1992, p. 176.).

Wright's original findings postulated a constant learning environment; however, researchers have not ignored the idea that constant learning may not exist on a continual basis in a manufacturing environment. In fact, the ideas of regressed and lost learning have been widely studied. Research studies support that a break in production creates an environment of relearning because the labor resources have stopped working, at least on the same project, and will be less efficient at manufacturing when production restarts (Anderlohr, 1969, pp. 16–17).

In addition to production breaks, instances also exist when a major configuration change occurs during production and disrupts the learning process. In this situation, the new configuration is immediately incorporated into the next units on the production line; the units already produced are retrofitted at a later time. Intuitively, the units with the configuration change should initially have a different learning rate than the units without a change because the manufacturers must learn how to incorporate the change into the production process. However, because the learning rate for the new configuration is unknown, DoD program offices generally do not treat the reconfigured units with a different learning rate. As a result, the program often experiences substantially different hours/costs for the newly configured production units than their original learning curve projected. A configuration change in a production program does necessitate learning for the contractor, and the impact to learning attributable to the configuration change should be understood by all levels of the DoD acquisition community. Wright (1936) understood this limitation to the learning curve theory application even in the infancy of the idea:

> The tremendous cost of changes introduced into a production order during construction is too well known to require emphasis. This cost is involved, not only in shop delays, but in the engineering expense of re-designing. It is appreciated that in a rapidly moving art such as aviation, changes are more or less inherent. … In using the curve developed in this paper, it should be recognized that the factors derived are based on the assumption *that no major changes will be introduced during construction*. (Wright, 1936, p. 124)

One of the first and most recognizable learning curve formulas is y= 〖ax〗 ^b. This is referred to as the Unit Learning Curve Model,

$$y = ax^b \quad (1)$$

where

y = the estimated production hours or cost

a = the production hours of the theoretical first unit

x = is the unit produced

b = is a factor of the learning = log R (learning rate)/log2

The Air Force methodology on learning curves and guidance in their application is found in Chapter 8 of the *Air Force Cost Analysis Handbook* (AFCAH) and Chapter 17 of the DoD Basic Cost Estimating Guidebook (BCE). These two guides focus on two theories: unit theory and cumulative average theory. The unit theory, Equation 1, predicts a specific unit cost. Cumulative average theory focuses on the average of all units produced up to a certain point in the production process. The cumulative average and unit theory have been the standard in manufacturing. However, research has shown other models may provide a more accurate predictor of cost.

Study 1: Production Break and Lost Learning

Current DoD program office cost estimating assumes a stable rate of learning once a program is substantially into production. However, intuitively, a configuration change introduced into the production line will initially disrupt the learning effect. This study will research two main questions to address the implications when a configuration change occurs during production:

1. Is there an impact to the learning curve slope when a configuration change is introduced to the production line? Specifically,
    a. What is the learning curve slope for each new configuration;
    b. Are the production segments for each configuration significantly different; and
    c. What is the difference between the hours predicted based on the prior configuration and actual hours for each segment?
2. How many units of the newly configured aircraft are produced before the contractor recovers the stable learning rate?

The first research question leads to a single testable hypothesis to determine if the mean amount of labor hours prior to a configuration change is the same as the mean amount of labor hours subsequent to a production change?

Hypothesis 1:

H_0: Mean labor hours $_{\text{prior to configuration change}}$ = Mean labor hours $_{\text{post configuration change}}$

H_a: Mean labor hours $_{\text{prior to configuration change}}$ ≠ Mean labor hours $_{\text{post configuration change}}$

If the analysis results fail to reject the null hypothesis, this would indicate that the data points come from the same population and a configuration change did not have a significant impact to the learning during production. If the analysis rejects the null hypothesis, this would indicate the opposite, that the data points representing different configurations come from different populations and that a configuration change did have a significant impact to the learning during production. If the results support rejecting the null hypothesis, using the prior learning curve equation is inappropriate to predict the hours of the new configuration because the units come from different populations. The second research question does not require a hypothesis test.

Production Break and Lost Learning

As the learning curve theory has evolved, researchers and practitioners have investigated the impact to the learning rate when other than constant production exists. George Anderlohr (1969) is credited with developing a model to determine the additional hours/costs that result from a break in production. Anderlohr (1969) defined a production break as "the time lapse between completion of a contract for the manufacture of certain units of equipment and the commencement of a follow-on order for identical units" (p. 16). A break in production results in increased hours and costs because the laborers are no longer performing their tasks on a constant repetitive basis and the laborers become less efficient (have a loss of learning) during the production break timeframe (Anderlohr, 1969, pp. 16–17).

Studies have also discovered that lost learning can be a result of forgetting at times other than during a production break (which is considered scheduled forgetting). Two other instances when forgetting can occur are (1) at random due to the inability to continue work (e.g., machine breakdowns), and (2) based on a natural process (e.g., aging workforce; Badiru, 1995, p. 780). Badiru goes on to conclude that "whenever interruption occurs in the learning process, it results in some forgetting." The amount of forgetting is a function of both the length of disruption and the initial performance level (Badiru, 1995, p. 780).

Additional Work Theory

A similar circumstance to the production break theory that has a similar result is the idea of new learning, when manufacturing is interrupted with a major configuration change to the production unit. When the unit being manufactured is changed, the laborers must adjust their processes to learn how to correctly produce the newly configured unit. Historically, adjusting the learning curve to account for the impact due to configuration changes is referred to as splicing or splitting the curve, although little research has been done in this area with empirical data. The theory of splitting the curve provides rationale to split the curve between units of different configurations (pre- and post-configuration change) because the latest production unit usually provides the greatest estimate for future production units (Dahlhaus & Roj, 1967, p. 16).

Sample Data

There were three limiting conditions the data had to satisfy to be included in this study: (1) at least one identified configuration change must come into the production line during production, (2) all units must be produced on the same production line, and (3) the program must be "substantially" into production. For the purposes of this analysis, substantially into production is defined as those units considered by the program office to be representative of stable production and exclude any units identified as developmental or pre-production.

After excluding programs that did not meet the research conditions, there were four data sets available at the time of the analysis, including one joint service and three Air Force aircraft programs. Due to the proprietary nature of the production data, the program names are not disclosed and will be identified as Programs A, B, C, and D. In addition, three classes of aircraft are represented in this study: Unmanned Air Vehicle, Cargo, and Fighter aircraft.

Analysis Methods and Results

To test the research hypothesis, the data will be split into separate segments at each identified configuration change to identify if the segments are statistically similar based on the mean or median labor hour values. Using the touch labor hours, the learning rate before an identified configuration change and the learning rate after the change will be calculated to address the remaining areas of the first research question. Both calculations will use Crawford's unit theory equation $y = ax^b$; because the data is available in units, a unit analysis is appropriate. In addition, to avoid the smoothing effect and the obfuscation of unit variation a cumulative unit curve can create, the unit learning curve method will provide the most explanatory results of the two methods (unit and cumulative average) for the intent of this study (ICEAA Module 7, 2013, p. 14).

The slope will be calculated each time an identified configuration change occurs and not at other instances, even if a pattern change is evident in the scatter-plot of the data. The learning curve equation of a segment will forecast the touch labor requirements of the successive production segment. The forecasted hours of an identified configuration change will be compared to the actual hours of the same configuration to calculate the difference and the percentage difference.

To answer the second research question, an analysis will determine the number of aircraft produced after a configuration change until the prime contractor was able to return to a stable learning rate. This will be accomplished by removing one production unit at a time (in sequential order beginning with the first unit of the segment) and calculating the learning curve slope of the remaining units until the stable rate of the prior segment is achieved. An overall commonality is not expected because every program, every contractor, and the associated production process are different. Instead, the results are informational and may support contract negotiation efforts with more insight into post-configuration change production.

Table 1 includes the slope calculations for each program for each segment identified. Configuration A is always the initial configuration, prior to any changes. Based on this summary, the slope never remained the same after a configuration change. Program C and Configuration C of Program B were excluded from the analysis at this point based on the inappropriateness of the analysis method displayed through scatter-plots in both unit and log space.

Table 1. Segment Learning Curve Slope Values

		Configuration					
		A	B	C	D	E	F
Program	A	63.26%	49.84%	-	-	-	-
	B	87.02%	84.33%	-	-	-	-
	D	91.96%	73.15%	59.04%	68.56%	75.24%	82.49%

In nearly every case involved in this study, the segmented data are statistically different when compared to an adjacent segment, which addresses the issue in the first research question (Configurations A and B of Program B were not found to be statistically different). There is a change to the learning curve slope each time a configuration change is introduced, and in every case examined except one, the median labor hours (which are partially a function of the learning curve slope) for the different configurations are statistically different. These findings suggest that using the prior learning curve equation is inappropriate to predict the hours of the new configuration because the units come from different populations.

Further addressing the first set of questions, the learning curve regression equation for each segment is used to predict the touch labor hours for each unit in the following segment. The total predicted hours for each segment are compared to the total actual hours of the segment, and the results are shown as a difference in hours as well as a percent difference for comparison between the programs. A negative value indicates the estimate was lower than the actuals. Table 2 details the results of the predicted and actual hour comparisons.

Table 2. Learning Curve Equation Prediction vs. Actuals Summary

	Predicted Hours	Actual Hours	Difference	% Difference
Program A				
	Predicted Hours	Actual Hours	Difference	% Difference
A predicting B	11,336,756.40	11,371,252.00	(34,495.60)	-0.30%
Program B				
	Predicted Hours	Actual Hours	Difference	% Difference
A predicting B*	229,114.62	295,348.35	(66,233.73)	-22.43%
Program D				
	Predicted Hours	Actual Hours	Difference	% Difference
A predicting B	1,014,525.48	986,331.30	28,194.18	2.86%
B predicting C	490,909.41	531,988.54	(41,079.13)	-7.72%
C predicting D	339,726.00	368,921.32	(29,195.31)	-7.91%
D predicting E	678,070.58	698,789.63	(20,719.06)	-2.96%
D predicting F	397,530.17	542,429.97	(144,899.80)	-26.71%

*Configuration B not considered a statistically significant change from configuration A

Given that this portion of the analysis only includes three programs, and two of the programs only compare two segments, there are too few data points to develop any meaningful CER or factor. However, the results are still impactful because for each of the seven segment comparisons, no fewer than 20,000 hours were the difference between the predicted and actuals, which equates to millions of dollars per segment misestimated (generally underestimated) in a cost estimate. Underestimation requires the program office to find dollars not currently in its budget, and overestimation temporarily ties up funding that can be used for other purposes.

In reality, a contractor will submit a tech-refresh proposal to the program office to account for the configuration change, but will estimate the unit costs based on an extrapolation of its stable learning curve because the new slope is unknown. In every program analyzed in this study, the learning curve slope becomes much steeper after the configuration change (when compared to the initial stable slope), and a extrapolation of the stable curve will create a higher per unit cost than the contractor would actually experience with the steeper learning curve. This phenomenon is explored in the next section, which analyzes Program A to answer the second research question of how many production units are manufactured before the contractor returns to its stable learning rate.

Program A was selected for analysis in addressing the second research question because Program A has a large sample size in total and within each segment. In addition, only one configuration change came into the production line, so this program provides a consistent sample to analyze. The stable slope for Program A is 63.26% as determined by the units in Configuration A (units 41 to 71). Table 3 summarizes the slopes for Configuration B beginning with units 72 to 124 and removing one unit at a time from the beginning of the segment until the stable slope was reestablished.

Table 3. Program A Stable Slope Analysis Summary

First Unit	Slope	Units to Stabilize
72	49.84%	
73	50.69%	1
74	51.34%	2
75	51.95%	3
76	52.48%	4
77	52.85%	5
78	52.83%	6
79	52.81%	7
80	53.21%	8
81	53.44%	9
82	53.80%	10
83	54.36%	11
84	54.85%	12
85	55.25%	13
86	56.33%	14
87	57.39%	15
88	59.18%	16
89	60.52%	17
90	62.03%	18
91	63.60%	19
92	64.36%	
93	64.30%	
94	64.52%	
95	63.51%	
96	62.06%	
97	60.54%	

The stable learning rate is achieved with the production of unit 91, which is 19 units after the configuration change came into the production line. While every program will stabilize at different production rates, the important point in this analysis is that after the configuration change is introduced, the contractor learns much quicker on the units after the configuration change than the stable flatter learning rate pre-change. The units immediately following the stabilized rate (92 to 97) are included in the table to show that the contractor does not continue to learn for all units after the stabilized rate is achieved; rather, the contractor's learning rate stays around the stabilized rate. While this analysis is for only one program and cannot be generalized for all programs, the prior analysis did show that for each program, the contractor learned at a much steeper rate following the configuration change. These results provide evidence to support a position other than the contractor extrapolating the prior stable learning curve in a tech-refresh proposal before a configuration change is introduced.

Conclusions of Research

The hypothesis testing indicated a statistically significant difference in the median production touch labor hours in the pre-configuration change and post-configuration change aircraft for every pair of data segments analyzed, except for one. Comparing the median values may equate to a statistically significant difference in the learning curve slopes for those data segments because the impact to the learning curve slope is evaluated through the touch labor hours of the data points, as they are partially a function of the slope value.

The data point analysis to address the stable learning curve research question produced interesting results. The analysis did show a pattern that post-configuration change, the contractor initially learns at a much faster rate and the learning rate decreases with each subsequent unit until the stable learning rate is again achieved. The learning rate did appear to stabilize at this point and did not continue to decrease.

While sample size was limited to a few programs, the results of this study may imply two things. First, that a majority of the time there is an impact to the learning curve slope whenever a configuration change is introduced during production. Second, that the contractor is able to learn to incorporate the change much more quickly than its stable learning rate for the entire aircraft.

Significance of Research

The results of this research indicate there may be a significant impact to the learning curve slope when a configuration change is introduced during production, even if the program is substantially into production, as were the programs included in this analysis. The findings suggest more research in this area is important for two reasons. First, if more programs are examined, additional data points may lead to the development of a CER or factor to adjust a stable learning curve, which would be a useful tool for cost estimators given the ever-changing acquisition environment. Second, because the learning curve slope is such a crucial factor in production contract negotiations, empirical evidence strengthens the DoD's position of what the contractor's expected learning curve should be—which this study found is not the same as the extrapolation of the contractor's stable learning curve.

An initial estimate that does not anticipate any configuration changes will underestimate unit production hours or costs required for the newly configured unit. If the DoD negotiates a contract based on an extrapolation of the contractor's stable rate, these results provide evidence that the stable rate will overestimate the production requirements; this analysis showed the contractor learns at a steeper rate after a configuration change. The initial underestimating, coupled with the contractor's overestimation, will result in the program office requesting millions of dollars, possibly in excess, per configuration change.

Learning curve theory advises the use of the most recent or most representative production methods to predict the follow-on articles. While this is intuitive and proven to result in better estimates, program offices cannot disregard the prior units. If program offices track the configuration change information and the resulting impacts, the DoD may be in a better position to estimate costs and negotiate production contracts.

Study 2: A Comparative Study of Learning Curve Models

History shows that there is a flattening effect near the end of production runs, learning does not remain constant in aircraft production, and machinery is becoming more involved in the production process. There is evidence to support a hypothesis that a different model may be more accurate than Wright's model. Prior research establishes the foundation for further research into additional types of aircraft. There is evidence to support a hypothesis that a different model may be more accurate than Wright's model. Previous research found that the contemporary models are more accurate than Wright's model given an incompressibility factor (M) that is somewhere between 0.0 and 0.1. M is a number between zero and one where zero indicates a completely manual process and one indicates a fully automated process. Wright's model is the most accurate predictor of cost if M is assumed to be greater than 0.1. Specifically, further research and analysis using program integration, assembly, and checkout. Additional research on the impressibility factor may indicate a model that is more applicable to DoD methodology. For the purpose of this study, the contemporary models examined are the DeJong and S-Curve Models. The following investigative questions are the basis for study 2:

1. How does the application of learning curve models using program integration, assembly, and checkout data affect learning curve models that incorporate an incompressibility factor?
2. How sensitive are IA&CO data to the incompressibility range?
3. Which learning curve model is the most accurate at predicting cost or time?

4. How can the individual airframe work codes prove beneficial for predicting cost or time?

The questions lead to the following hypotheses:

H1: One of the compared models will have Mean Absolute Percent Error significantly different from the others.

H2: One of the modern learning curve models will be significantly more accurate than Wright's model in predicting costs or hours.

H3: The S-Curve model with IA&CO will have a significantly lower MAPE than Wright's and DeJong's Learning Curve Models.

H4: The incompressibility factor will have a significant influence on the accuracy of the DeJong and S-Curve Models.

The first hypothesis' null (H_0) is $\mu_1 = \mu_2 = \mu_3$. This means that the MAPE (lower MAPE is better) for each learning curve model is the same. The alternative hypothesis(H_a) is that one of the model's MAPE is statistically different. A rejection of the null hypothesis in favor of the alternative hypothesis supports significant finding. The significant finding means that testing each contemporary learning curve model against Wright's model is the next phase. The second hypothesis (H2) has a null $\mu_1 = \mu_i$, where i will equal models 2 and 3 (DeJong's and the S-Curve). The H_a is that the contemporary learning curve models will have a lower MAPE than the conventional model$\mu_1 > \mu_i$. H3's null hypothesis is $\mu_2 = \mu_3$. The H_a is that $\mu_2 > \mu_3$, meaning that the S-Curve will have the lowest MAPE and thus be the most accurate predictor of cost or hours. The last hypothesis is that small changes in the incompressibility factor will have a large influence on the MAPE of each model.

Relevant Learning Curve Research

Relevant research has highlighted an important point in why military programs have not adapted a contemporary learning curve model:

> Because of the regularity of production in military programs, organizational forgetting, and spillovers of production experience are less apparent. If forgetting is present, it may be very difficult to identify (e. g., data could be consistent with either a 20 percent learning rate or a 25 percent learning rate with 5 percent forgetting). And, in most cases there are not many model variants, so spillovers are not important. (Benkard, 1999, p. 4)

The newest fighter weapon in the U.S. military arsenal will be the F-35. The F-35 has three variants, and the Pentagon plans to spend over $390 billion on these aircraft (Luce, 2014). Five percent of $390 billion attributed to learning/forgetting processes is still a staggering number. The point is that there is room for improvement. Many of the fighter aircraft in use today have had multiple models. The F-15 had models A-E, and the F-16 had models A-F. The DoD can use the hypothesized 5% forgetting to save millions of taxpayers' dollars. The accurate estimates result in less spending, or savings that could go into other taxpayer needs or public works. The estimates enable the DoD to truly forecast budget and spending levels.

One must consider the question of what does the future actually look like in regards to machinery and automation? Are people still going to be relevant for production? In the Defense acquisitions realm, the basis is that with low purchase quantities for state of the art machines will not rely on technological or machine dominated production. This idea really comes down to the machine verses machines argument. Asking whether a robot will take the jobs of humans is key. Experts say yes and no. In the past, machines were used to replace manual labor that was intensive and repetitive. According to a study by the Bank of

America, robots are likely to be performing 45% of manufacturing tasks by 2025, versus 10% today (Madigan, 2011). The price of a computer, a robot, a chip, etc. is falling, and it is speculated that it will fall even more in the future. However, the jobs that require human interaction are least likely to be replaced by a robot. Maybe DoD acquisitions are in the clear and 100% of learning is still realizable for learning curve methodology. Experts agree that the future does include a significant presence of machinery because the prices of robotics and computers are decreasing while the cost of human employees are increasing (Aeppel, 2012).

Contemporary Learning Theory

A contemporary variation of learning curve models is DeJong's Learning Formula. This formula is a modification of Wright's model, and it takes into account the constant, M. M is the incompressibility factor, which is a constant between zero (fully manual operation) and one (fully automated or machine dominated operation; Badiru et al., 2013). Equation 2 highlights DeJong's Learning Model.

$$t_n = t_1\{M + (1 - M)n^{-b} \qquad (2)$$

where

t_n= the cumulative average time after producing n units

t_1 = time required to produce theoretical first unit

n = cumulative unit number

b = log R/log 2 (learning index)

M = incompressibility factor (a constant)

A machine based production process would result in no learning, and thus an M value of one. It the basis of this thesis and belief that aircraft production, complex in nature, has an M value close zero because aircraft production is a highly manual process. Thus, M would be closer to zero for IA&CO. M does not have a specific value. This research will focus on the best M value for the particular aircraft production. A potential weakness of the DeJong model is that it does not take into account previous units produced as much as the S-Curve model does.

The S-Curve Model takes into account both previous units produced and the incompressibility factor. Figure 1 shows the effects of learning over time as hypothesized from the S-Curve Model. The linear nature of Wright's original learning curve model has been in question for many years (Everest, 1988). The Rand Corporation first sought to explain the progression of the learning curve used to estimate costs for both military and civilian airframes. The report attempted to describe the relationship between units (quantity) and costs and, ultimately, whether the relationship was linear on a log scale. The results of the Rand Corporation found that a convex curve may provide more accuracy if producing a large number of units (Asher, 1956). The results found that a convex model provides less error if there is a need for large extrapolation. Essentially, an estimation of significantly more units in production instead of fewer provides less error. For units where large extrapolation was required, a non-linear model was more appropriate. The S-Curve model, convex in nature, presents a shape of learning. The S-Curve, when plotted using a log scale relationship, follows an S function. The experience over time (attempts at learning) may exhibit the S-Curve (Everest, 1988).

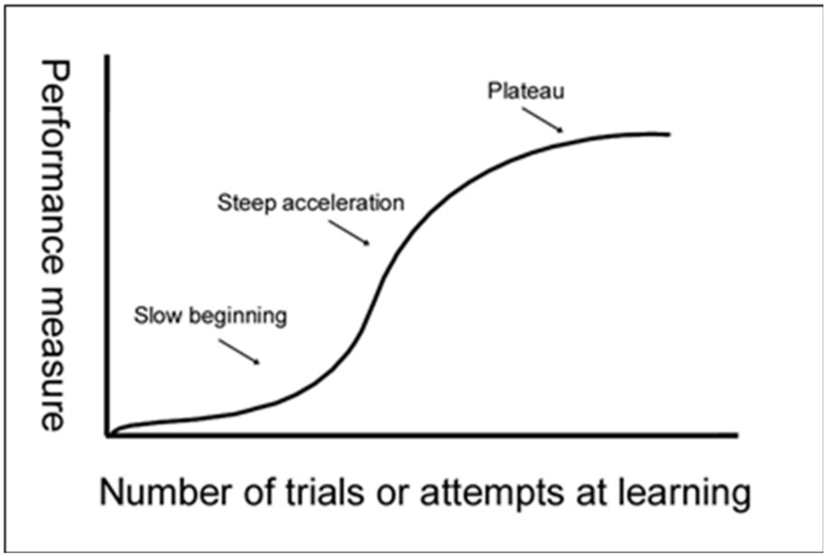


Figure 1. S-Curve Learning Model

Initially, there is a slow beginning as a worker learns the production process. The newness of the product is a characteristic of the slow initial beginning. New tools, methods, shortages of parts, reworks, and the challenge of developing a cohesive production team are all potential contributors to the slow beginning. The fact that the initial stage in production deals changes from tooling to even workers contributes to the gradual start (Badiru, 1992). From there, learning and familiarity of tools, methods, and workers occur. The learning enables a steep acceleration of production. Production improvement occurs with attempts on the process, or learning by doing. An example from the literature is aircraft production. Aircraft production that includes workers and tools that are more efficient leads to an assembly process that is also more efficient. The efficiencies found result in less time to complete an aircraft (Asher, 1956). However, the improvement and efficiencies eventually begin to fade. The plateau at the trailing edge of the curve is the *slope of diminishing returns* where the curve begins to flatten out, or in many occurrences at the end of production cycles, there is a "tailup" (Everest, 1988). After time, inefficiencies can occur: forgetting, experienced workers focusing on new projects, failure to repair worn tooling at the normal rate, increase of machine disassembly, lack of key materials (safety stock), and workers taking more time to prolong their employment (Everest, 1988). The S-Curve equation is shown in Equation 3:

$$t_n = t_1 + M(n + B)^{-b} \tag{3}$$

where

t_n= the cumulative average time after producing n units

t_1 = time required to produce theoretical first unit

n = cumulative unit number

b = log R/log 2 (learning index)

M = incompressibility factor (a constant)

B = equivalent experience units (a constant)

From this equation and Figure 1, the forgetting concept is evident. The S-Curve portrays that with time, some inefficiencies will occur. Use of the S-Curve and DeJong

Models may provide more precision to learning curve estimation and enable higher accuracy within DoD cost estimating because they include influences that were previously unaccounted for.

Data

The Air Force Life Cycle Management Center Cost Staff (AFLCMC/FCZ) at Wright-Patterson Air Force Base (WPAFB) provided learning curve data for 17 Major Defense Acquisition Programs (MDAPs). A MDAP is classified as a major program that exceeds a certain dollar threshold. There are 80 MDAPs in the DoD as of 2014. The numbers have decreased slightly over the years. The data files consist of average Learning Curve Reports of Annual Unit Cost (AUC) in addition to the MDAPs estimate methods using Wright's conventional learning curve model. Only one program provided was broken into the specific work codes that include the needed data (IA&CO). When comparing models based on airframe's integration, assembly, and checkout data, the assumption of incompressibility close to zero is acceptable due to the highly individualized process completed by humans.

A description of labor categories Integration, Assembly, and Checkout highlights that not all definitions are synonymous amongst manufacturers. The manufacturers largely consider what is involved in each category as proprietary information. For the purpose of this study, Final Test Integration, Electrical and Mechanical Assembly, Test/Integration, Composites (all locations), and Quality Control are considered IA&CO. Final Test Integration includes the direct labor for the final integration and test, which includes final assembly, system burn-in, payload integration and interface, autopilot checks, taxi tests, range tests and first flight support. Electrical Assembly is the direct labor required to assemble electronic components. Mechanical Assembly is the direct labor required to build servos for the aircraft, to build landing gear, build starter/alternators, to perform rework, and high-time maintenance on those components. Test/Integration is the direct labor for new build electronics, field repairs, integrating avionics, and testing them at the system level. Composites manufacturing is the direct labor required to lay up, cure, and finish components such as the fuselage, wings, tails, and landing gear. Quality Control is the direct labor required to provide inspection of electronics and mechanical components and assemblies, document discrepancies, and resolve problem areas. Of note, these labor categories involve mainly direct labor performed by humans where the learning process is observable. All of the data includes Test Support, Machine Shop, Program Support and Design. These work codes are not repetitive in nature like IA&CO.

Analysis Methods and Results

The data includes actual costs as well as predicted costs using one of the learning curve models. Once calculation of the predicted costs is complete, the error is simply the difference between the actuals costs and the predicted costs. To provide a comparison, a difference calculation in the absolute value and absolute value percent error are the means of analysis. The next step is to perform an analysis to test the hypotheses. ANOVA or the Kruskal-Wallis test will provide the basis for comparing the percent errors. The tests will produce an F-statistic (a test statistic) that falls within a Chi-distribution and a p-value. This comparative study will produce results based on a 95% confidence level (an $\propto$ of 0.05). If the P-value is less than 0.05, rejection of the null hypothesis in favor of the alternative hypothesis will occur. Rejecting the null hypothesis for this study will represent that there is a 95% chance that the tested populations are different. The conditions for ANOVA are as follows: the samples must be from a random selection of the population, normally distributed, and population variances must be equal. If the conditions for ANOVA fail to meet the needed criteria, the Kruskal-Wallis test (non-parametric equivalent to ANOVA) will be the test to determine if multiple samples arise from the same distribution and have the same parameters (“Kruksal-Wallis Excel,” n.d.). The ANOVA or Kruskal-Wallis f-test provides insight into the first hypothesis. The Kruskal-Wallis test is beneficial since the one-way ANOVA is usually robust based on the assumptions for ANOVA. The Kruskal-Wallis test becomes useful in particular when group samples strongly deviate from normal (sample size is small and unequal and data are not symmetric) and variances are different (potential outliers exist). The assumptions for the Kruskal-Wallis test are that no assumptions are made about the underlying distribution; however, assume that all groups have a distribution with the same shape, and no population parameters are estimated (no confidence intervals in the data; Zaiontz, 2015). If the F-statistic is significant, then rejection of the null hypothesis in favor of the alternative that at least one of the sample means is different is the outcome.

The t-Test for two samples test will evaluate the second hypothesis that one or more of the models is a better fit to the data than Wright’s Model. The control for this comparison is Wright’s Learning Curve Model. Since Wright’s Learning Curve Model (WLC) is the control for this study, a comparison to the other model’s MAPEs is the method. If the assumption for equal variance is not met, the t-Test for two samples assuming unequal variances will be used. The next analysis that corresponds to H3 will be testing which model is most accurate given significant results for more than one model from H2. Once again, the paired difference t-test is the next step. A paired difference experiment uses a probability distribution when comparing two sample means and produces a t-statistic that falls within a student-t distribution that can either reject or fail to reject the null hypothesis depending on the desired confidence level. Lastly, H4 will require reiterations of the tests in order to determine a good estimate for incompressibility factor based on the airframes. This method will include IA&CO and then all of the data in order to provide a comparison.

Of note, the reader may question why the means cannot provide the basis for the analysis. This lies in the variation of the means. If the coefficient of variation (standard deviation as a percentage of the mean) is low, the mean may be a good predictor of the better model. However, as a rule of thumb, if the coefficient of variation (CV) is greater than 15%, the mean indicates a looser distribution. Most analysts would likely prefer a tighter distribution with less variability. In practice, a low CV (say, 5%) would indicate that the average (mean) of the cost data is a useful description of the data set. On the other hand, if the CV is much higher (say, greater than 15%), there should be a cost driver in the data set that causes the cost to vary. The CVs for the analysis will provide insight into the dispersion of the data points. CVs for the analysis all exceeded the threshold and resulted in the mean not being a good predictor. Table 4 highlights the MAPE analysis for M = 0.05. Results were inconclusive as to which model was more accurate.

Table 4. MAPE Analysis M = 0.05

WLC	DeJong	S-Curve
4.11%	3.00%	2.64%

A description of the results for the Air Force Program, based on the assumption of low incompressibility values of 0.05 and 0.1, highlights the effects of learning. The results changed between these values and indicated that the S-Curve model may be a more accurate predictor at 0.05, but WLC is more accurate at 0.10. After 0.10, WLC becomes significantly less error prone for both IA&CO and the analysis with all of the work codes. WLC is a better predictor of cost when the incompressibility is 0.10 and higher. Analysis on all of the data points, not just IA&CO, showed no difference between the models at an M of 0.05. However increasing the M value results in WLC becoming an increasingly more accurate predictor of cost. The following figures highlight the effects of M and the Absolute Percent Errors.

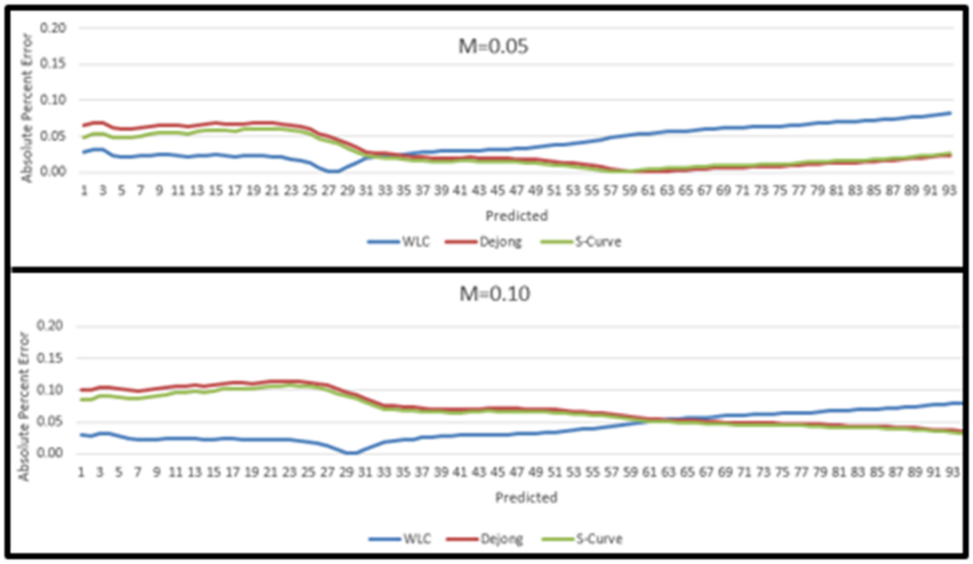


Figure 2. IA&CO APE Trends

When plotting the Absolute Percent Errors for all of the data points, a trend similar to the analysis using IA&CO was evident. WLC starts as a more accurate predictor of cost and then becomes less and less accurate, whereas the DeJong and S-Curve models become increasingly more accurate predictors. Figure 3 shows the results of APE for all of the data.

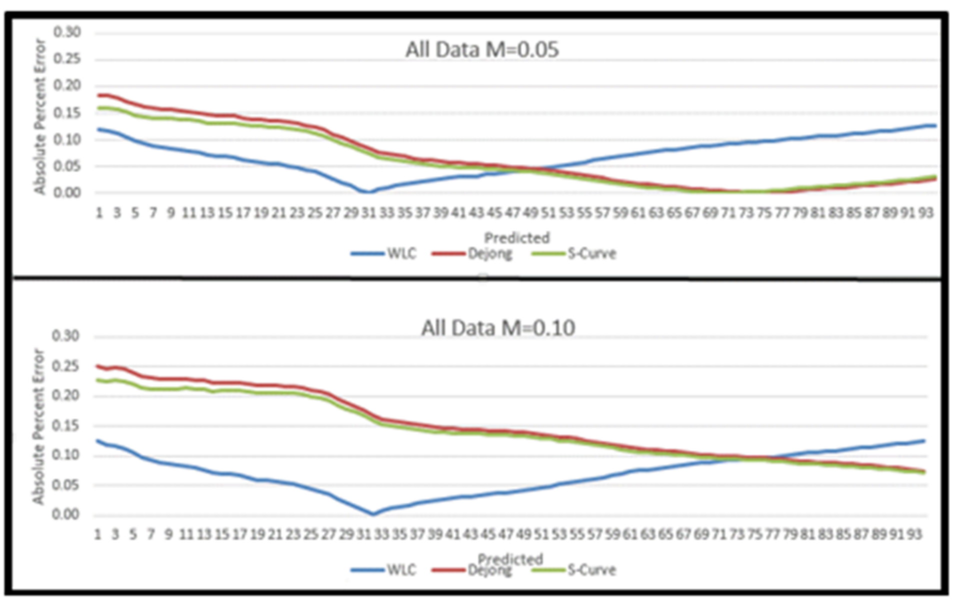


Figure 3. All Data APE Trends

Conclusions of Research

In summary, results support the first hypothesis that there is a significant difference between the models. Results are inconclusive as to whether any models are significantly more accurate than Wright's model. Between an incompressibility of 0 to 0.1, DeJong and S-Curve models were more accurate (less error prone). Nevertheless, at an incompressibility of 0.10 and beyond, Wright's model is most accurate. The third hypothesis was also inconclusive as to which model is most accurate at an incompressibility of 0.05.

Both DeJong and S-Curve models were more accurate than WLC, but there was no difference between the two. Finally, incompressibility was highly influential as hypothesized. The results of the findings lead to questioning why, for the program chosen, incompressibility would become increasingly more error prone when more automation is present. In addition, the findings put into question how the DoD can draw a conclusion about the application of contemporary learning curve models in acquisitions and specifically cost estimation. Absolute Percent Error figures highlight that WLC is accurate initially and eventually becomes increasingly less accurate. The opposite, S-Curve and DeJong Models are not as accurate initially, but become increasingly accurate over trials. The MAPE analysis averages all of the errors. If the data set included more units, results may trend towards results in favor of the contemporary models. That answer is based on the visual trend from the APE figures. Of interest when the incompressibility factor is 0.10, the models portray that 90% of learning is obtainable. Because the data set was small, changes in incompressibility may not be as evident to the significance of the comparative study.

The findings also put into question how the DoD can draw a conclusion about the application of contemporary learning curve models in acquisitions and specifically cost estimation. If the production cycle is long and many trials will be realized, there is potential that the contemporary models may capture a more accurate picture of learning. Aircraft production may provide starkly different results from a missile production run where more units are produced over time. The results support that there is potential for a more accurate model. However, it may not be in the realm of aircraft production. Aircraft production may include some automation. It is not implausible that aircraft production is 95% manual and supports an M factor of 0.05. The contemporary models may support a more automated process such as a production line much like the automobile industry. Prior studies and subject matter expert opinion support that aircraft production is manual. However, there is a belief that more automation will be present in the future.

Significance of Research

Results from the analysis show that there is reason to believe Wright's Learning Curve may not be the best method for estimating costs. By extrapolating from actuals, the method for Wright's model may not incorporate enough of the variability of learning. The results provide evidence that Wright's Model is accurate initially, but with attempts at learning (trials) the amount of error increases. The comparative analysis on learning curve models provides a standalone analysis of program actuals. The conclusions from this study are that there is potential for a more accurate cost-estimating model and that the conventional learning curve models become increasingly less error prone over trials. The DeJong and S-Curve models show promise as a way to improve DoD cost estimating.

The results of the research do not support all of the hypotheses. Results did confirm that the incompressibility factor was highly influential for both the S-Curve and DeJong models. The results of the comparison changed drastically with a small change in the incompressibility factor. The DeJong and S-Curve models were both more accurate than WLC, but there was no difference between the two. This finding makes it challenging to simplify the results given the uncertainty of incompressibility. The influence of machinery in longer production cycles is a valid assumption for the future. The influence of automation in this comparative study was evident by the absolute percent error graphs.

References

Anderlohr, G. (1969, September). Determining the cost of production breaks. *Management Review,* 16–19.

Asher, H. (1956). Cost-quantity relationships in the airframe industry. *Project Rand*. Retrieved from http://www.rand.org/content/dam/rand/pubs/reports/2007/R291.pdf

Badiru, A. B. (1992). Computational survey of univariate and multivariate learning curve models. *IEEE Transactions on Engineering Management, 39*(2), 176–188. http://doi.org/10.1109/17.141275

Badiru, A. B. (1995). Multivariate analysis of the effect of learning and forgetting on product quality. *International Journal of Production Research, 33*(3), 777–794.

Daulhaus, F. J., & Roj, J. S. (1967). *Learning curve methodology for cost analysts*. Washington, DC: Department of the Army.

Department of the Air Force. (2007). *Air Force cost estimating handbook.* Washington, DC: Author.

Everest, J. D. (1988). *Measuring losses of learning due to breaks in production* (Master's thesis). Monterey, CA: Naval Postgraduate School.

International Cost Estimating and Analysis Association. (2013). *Cost estimating body of knowledge training, Unit III, Module 7. Learning curve analysis: How to account for cost improvement.*

Kruksal-Wallis Excel. (n.d.). Retrieved December 7, 2015, from https://www.youtube.com/watch?v=KmVTkVnWFuc

Towill, D. R. (1990). Forecasting learning curves. *International Journal of Forecasting, 6*(1), 25–38. http://doi.org/10.1016/0169-2070(90)90095-S

Watkins, P. N. (2001). The persistence of learning and acquisition strategies. *Acquisition Review Quarterly*, 15–30.

Wright, T. P. (1936, February). Factors affecting the cost of *airplanes. Journal of Aeronautical Sciences, 3,* 122–128.

Yelle, L. E. (1979). The learning curve: Historical review and comprehensive survey. *Decision Sciences*, *10*, 302–328.

Zaiontz, C. (2015). Real statistics using Excel. Retrieved from http://www.real-statistics.com/

Cost and Price Collaboration

Venkat V. Rao—serves as Professor of Acquisition Management and currently teaches acquisition and program management courses for the Defense Acquisition University (DAU). Rao is part of the faculty of the Midwest Region, which is responsible for providing professional education and training to approximately 30,000 defense department personnel in the 12 states within the Midwest. Rao is also the Site Lead for the DAU-Midwest Sterling Heights office and has broad program management and product development experience in the technology industry spanning 30 years. [venkat.rao@dau.mil]

David Holm—CCEA, is the Director of Cost and Systems Analysis, United States Army TACOM Life Cycle Management Command. [david.a.holm6.civ@mail.mil]

Patrick Watkins—is currently the Pricing Chief for the Stryker/Armaments/R&D Pricing Group, Army Contracting Command–Warren. He has 37 years of contracting experience with about half of that career spent in contract pricing. His specialty is negotiation and pricing of combat vehicles and support services. Over his career, Watkins has also developed or participated in the development of cost models used in the analysis of acquisition program costs, especially for forecasting cost changes due to competitions and dual sourcing. [patrick.n.watkins4.civ@mail.mil]

Abstract

This paper examines how collaboration between the cost analysis and price analysis functions can achieve program efficiencies. Collaboration is defined along two dimensions: (1) as interactions between analysts from the two functions and (2) as exchange of relevant data and information between the functions. This study facilitated an exchange of financial data common to all programs, namely, the Cost and Software Data Reports and Price Negotiation Memorandums, between the two functions. Interviews with subject-matter experts provided the basis of estimates of the effects of the information exchange on the respective functions. The effects estimated were improvements in price negotiations by price analysts and improvements in program cost estimates by cost analysts. In addition to the common reports outlined above, the paper also identified other sources of information used by price and cost analysts to perform their functions and realize improvements. Based on the results, the paper proposes an information environment to systemically improve and institutionalize collaboration.

Introduction

This paper examines the benefits of collaboration, defined as information sharing, between price analysts and cost analysts. Both groups support a program's acquisition activities with financial analysis. Price analysts typically reside in acquisition contracting commands, and cost analysts reside in either System Command cost analysis organizations, on Program Executive Organization staff, or in business management offices. Both groups access financial information for different purposes: Price analysts support contracting actions, evaluate contractor proposals, and develop government positions to support negotiations to arrive at a final contract price. Cost analysts, on the other hand, develop life cycle cost estimates that are used to support program budgets, develop user requirement documents, and support program offices in their tradeoff and affordability analysis. Both groups need accurate cost information to meet their goals and have adopted various techniques and access multiple sources to obtain that information. Price and cost analysts sometimes collaborate and share information, but this collaboration is sporadic and based on existing practices, known data sources, and individual relationships.

This paper examines the state of current collaboration between price and cost analysts in four Army ground vehicle programs, captures the resultant benefits to both functional disciplines, and recommends business process improvements to promote further collaboration.

Background

From 1992–2003, the Cost and Software Data Reports (CSDRs) were rarely collected on Army ground vehicle programs. In 2004, however, after more than a decade, a renewed emphasis was placed on contractors' contractually providing the CSDRs to the cost analysis community. The CSDRs report the actual recurring and non-recurring costs incurred by the contractor on the contract. By contrast, the Price Negotiation Memorandums (PNMs) are internal documents developed by price analysts that analyze contractor proposals' costs and price, document the government position, and record the final negotiated price. Both the CSDRs and the PNMs include detailed costs for labor, material, and overhead, but the CSDRs also provide costs by a standardized detailed work breakdown structure. Currently, this information is not exchanged between the pricing and cost analysis communities on a routine basis.

This paper examines the impacts of exchanging this information between cost and price analysts for four Army programs. In addition to the CSDRs and PNMs, the analysts were asked to report additional sources of price and cost information that supported their activities and improved their results. The additional sources include information from external organizations like the Defense Contract Management Agency (DCMA) and the Defense Contract Audit Agency (DCAA) and can vary by program to include the contract data requirements lists (CDRLs). As subject matter experts, price and cost analysts in the programs estimated the impacts of this information exchange by measuring the percentage improvements in cost-estimating and price-analysis outputs.

Project Description

Methodology

Ideally, we would have research data available on those programs that displayed high collaboration between cost analysts and price analysts and those that did not display such collaboration. We might then be able to compare results from those programs to quantify the benefits of collaboration. However, such program data is not generally available. Further, most programs in the ground vehicle community are well into the second or third cycle of system enhancements on very mature platforms, so early program data, such as for the Technology Maturation and Risk Reduction (TMRR) phase and the Engineering and Manufacturing Development (EMD) phase, are not available. Given the paucity of historical data, we developed the methodology described below which relies on questionnaires administered to subject matter experts in both cost analysis and price analysis to determine the value of collaboration to each discipline. While this methodology is less precise than one using accurate, matched historical data, it does provide enough information on the value of collaboration to inform potential changes in work practices.

To help inform the participants and to obtain some insight from them into the benefits of the collaborative process, we selected four Army ground vehicle programs and documents to exchange. For each program, we matched a cost analyst volunteer and a price analyst volunteer, each a subject matter expert with at least one year of experience on that particular program and with several years of experience in their respective disciplines. The selected participants were well versed in the details of their particular programs and had participated in at least one full budget cycle or one full contract negotiation cycle, respectively. The two principle sets of documents common to all of the programs were CSDR data and the PNMs. We had the cost analysts provide the Cost Performance Reports

(DD 1921) and the Functional Cost Hour Reports (DD 1921-1) to their price analyst counterparts. We had the price analysts provide the PNMs to their cost analyst counterparts. Participants were able to ask questions and discuss data with their counterparts prior to completing the questionnaires. We then administered the questionnaire to gauge the value of collaborating using these documents. The questionnaires also included open-ended questions about other practices that might foster collaboration, which proved informative.

The CSDRs are submitted by the contractor to the Government after completion of key events in a program (e.g., Critical Design Review, Prototype delivery, etc.) and at or near major decision points in the program lifecycle. It is at these points that contractor proposals are reviewed in preparation for the next contract award. Hence, the CSDRs are expected to inform the price analysis and proposal review with actual costs from the most recent contractor effort, thereby improving the negotiation position of the price analysts for the next contract award.

CSDR

The CSDR may consist of up to seven different reports. The details of the cost-reporting requirement (including, for example, the types of reports required, reporting structure, frequency, due dates, etc.) are communicated to the contractor in the request for proposal through the DD 2794 CSDR Plan. The CSDR Plan identifies which of the seven report types the contractor will be required to submit under that contract. The following are the seven different report types and short descriptions of the information they contain:

1. Contractor Work Breakdown Structure (CWBS) Dictionary: Provides a detailed description and definition of both technical and cost content for each Work Breakdown Structure (WBS) reporting element from the CSDR plan.
2. Cost Data Summary Report (DD 1921): A contract-level report that lists all WBS elements from the DD 2794 CSDR Plan. Provides a breakout of non-recurring and recurring costs incurred to date and estimated costs at contract completion. Provides a breakout of overhead costs (General & Administrative, Management Reserve, Profit/Fee, etc.) and quantities completed to date and estimated at contract completion.
3. Functional Cost and Hour Report (DD 1921-1): Provides a detailed breakout of all resource data (labor hours, labor dollars, material dollars, overhead dollars) across four functions (Engineering, Tooling, Quality Control, and Manufacturing) for each identified WBS element on the DD 2794 CSDR Plan. Costs are identified as non-recurring or recurring. They are further identified as incurred to date or estimates at contract completion. Reported costs do not include overhead costs from DD 1921 (General & Administrative, Management Reserve, Profit/Fee, etc.). The report also includes direct-reporting subcontractors' costs.
4. Progress Curve Report (DD 1921-2): Provides a detailed breakout of all resource data (labor hours, labor dollars, material dollars, overhead dollars) across four functions (Engineering, Tooling, Quality Control, and Manufacturing) for specified WBS elements on the DD 2794 CSDR Plan. Costs are direct recurring costs incurred to date and hours incurred to date. Costs are also segregated by unit or lot to develop learning curves and to project future units. Reported costs do not include summary element costs from DD 1921 (General & Administrative, Management Reserve, Profit/Fee, etc.). The report also includes direct-reporting of subcontractors' costs.

5. Contractor Business Data Report (DD1921-3): Annual report designed to facilitate overhead cost analysis at a specific contractor's site. Includes specific overhead information on all Major Defense Acquisition Programs, government contracts, and other government and commercial business. Also includes actual direct and indirect cost data on Government contracts and proposed direct and indirect cost data for future fiscal years.
6. Contractor Sustainment Functional Cost Report (DD 1921-5): This report is similar to DD 1921-1 except that its focus is on sustainment activities. Provides a detailed breakout of all resource data (labor hours, labor dollars, material dollars, overhead dollars) across four functions (Engineering, Program Management, Maintenance Operations, Materials) for each identified WBS element on the DD 2794 CSDR Plan. Costs are identified as non-recurring and recurring. They are further identified as either incurred to date or as estimates at contract completion. Reported costs do not include overhead costs from DD 1921 (General & Administrative, Management Reserve, Profit/Fee, etc.). Includes direct-reporting subcontractors' costs.
7. Software Resource Data Report (SRDR): Provides information for selected WBS elements on the DD 2794 CSDR Plan on software size, effort, and schedule.

PNM

The PNM contains several sections of interest to the cost analyst, including a reference to a DCAA Audit if one was performed, a reference to the Program Management Technical Evaluation of the proposed labor and material costs, a reference to an technical evaluation by the DCMA if one was performed, and a cost element summary for each of the contract deliverables and the total contract. The cost element summary includes many details of interest to cost analysts, such as the following:

- Material costs with part number detail on the contractor's proposal and an explanation of the U.S. Government analysis
- Additional detail on other part numbers, non-Bill of Material, and other material costs is also included in the discussion on material costs
- Total Labor Hours by Work Breakdown Structure description and cost center for the contract deliverables followed by the U.S. Government analysis
- Details on other direct costs
- Direct Labor Rates by rate and skill band
- Indirect Rates for Manufacturing, Engineering, G&A, Material Acquisition, and Material Handling and other overhead costs with an analysis of the overhead pool
- Facilities Cost of Capital (FCCM)
- An analysis of the profit based on the risk and other factors impacting the contract
- Government Furnished Material used by the contractor to prepare the proposal

The PNMs are documents reflecting the negotiated cost for labor and material by component in an awarded contract. Hence, the PNM is expected to inform the cost analyst in generating more accurate program budget estimates for the next funding cycle.

As stated earlier, the CSDRs are expected to inform the price analysis and proposal review with actual costs from the most recent contractor effort, thus improving the negotiation position of the price analysis for the next contract award.

These two premises were captured in a questionnaire with a section for the cost analyst and a section for the price analyst.

The price analyst section of the questionnaire asked a series of questions on the improvement in the negotiated price of a contract due to the availability of the CSDRs from the previous phase. For example, the questionnaire asked for improvement in the negotiated price for the Engineering and Manufacturing Development (EMD) Phase contract when the CSDRs from the Technology Maturation and Risk Reduction (TMRR) phase were made available. Similarly, improvements in the negotiated price of the Low Rate Initial Production (LRIP) Contract were sought given the CSDRs from the EMD phase. Finally, improvements in the Full Rate Production (FRP) contract were recorded given the CSDRs from the LRIP phase. Additional questions on the benefits of the detail in the CSDRs were also sought, along with additional reports or data that the price analyst used in determining their final negotiated position.

The cost analyst section of the questionnaire similarly focused on the improvement in the cost estimating given the PNM for the TMRR phase. Analysts also estimated the improvement in program estimates given the PNM for the EMD phase followed by improvements given the PNMs for the LRIP and the FRP contracts. Cost analysts were also asked to discuss other sources of information besides the PNMs that improved their cost-estimating efforts.

The questionnaire was reviewed by the Director, Cost and Systems Analysis and Pricing Chief for the Stryker program, and the Acquisition Contracting Command leadership.

Cost and price analysts responsible for four Army ground vehicle programs were asked to complete the questionnaire. A discussion was conducted with each team of price and cost analysts by program to document and clarify the responses. The discussions also captured the benefits of other sources of information that analysts used, such as Earned Value Management (EVM) Reports, Defense Contract Management Agency (DCMA) analysis of Forward Pricing Rate Proposals and Agreements (FPRP & FPRA), and the Defense Contract Audit Agency (DCAA) analysis of actual costs reported by contractors.

The results of the survey along with the analyst discussions are summarized in the paper under the Results section of the paper.

Selected Programs

Four Army ground vehicle programs were selected for this study.

- **Stryker**: A family of eight-wheeled armored fighting vehicles.
- **M88**: M88 and its variants are one of the largest armored recovery vehicles in use and include the M88, M88A1, and the M88A2 Hercules.
- **Paladin Integrated Management (PIM) Program**: An artillery vehicle delivering the M109A7 self-propelled howitzer.
- **Heavy Tactical Vehicles (HTV):** Program for Combat Support and Combat Service Support.

Four teams of price and cost analysts with responsibility for the above programs supported this study and paper.

Summary Results: Percentage Improvement Due to Collaboration

Tables 1 and 2 summarize the results recorded by the analysts and are followed by a discussion of each program's questionnaire responses.

Table 1 records the improvement that the CSDR and other reports provided the price analyst in supporting their price negotiation position. Each column records the percentage improvement estimated by the analyst in the government's price analysis of the contractor's proposal for the next phase due to the availability of the actual costs reported via the CSDR.

For example, the Stryker price analyst, in reviewing the CSDR submitted during the TMRR phase, was able to realize an estimated 5% improvement in the negotiated price for the EMD contract. Similarly, data from the CSDR received during the EMD phase supported an improvement in the government position of between 5% and 10%, while the CSDR from the LRIP phase supported a 25% improvement in the final position for the FRP contract.

A more detailed discussion by program follows the summary results tables.

Table 1. Price Analyst Input Percentage Improvement

Program	TMRR CSDR to EMD Contract	EMD CSDR to LRIP Contract	LRIP CSDR to FRP Contract
Stryker	5%	>5 ≤10%	>25%
M88	< 5%	>5 ≤10%	>5 ≤10%
PIM	< 5%	>5 ≤10%	>5 ≤10%
HTV	NA	NA	>0 ≤5%

Table 2 records the improvement in the cost estimates going forward that the cost analysts estimated due to the availability of the PNM from the last contract. For example, given the PNM for the TMRR contract, the cost analyst was able to improve the program cost estimates going forward by 10% to 15%. Similarly, given the price negotiation memo of the EMD contract, the cost analyst was able to improve the program cost estimates by 10% to 15%, the PNM of the LRIP contract enabled a 5% to 10% improvement, and the PNM for the FRP contract again showed a 5% to 10% improvement in program cost estimates going forward.

Table 2. Cost Analyst Input Percentage Improvement

Program	TMRR PNM to Inform Program Cost Estimates	EMD PNM to Inform Program Cost Estimates	LRIP PNM to Inform Program Cost Estimates	FRP PNM to Inform Program Cost Estimates
Stryker	>10 ≤15%	>10 ≤15%	>5 ≤10%	>5 ≤10%
M88	<5%	>15 ≤ 20%	>20 ≤ 25%	>20 ≤ 25%
PIM	>5 ≤10%	>15 ≤20%	>15 ≤20%	>5 ≤10%
HTV	>25%	>20 ≤25%	>15 ≤20%	>15 ≤20%

Price and Cost Discussion by Program

Stryker Program Discussion

The Stryker program is unique in that collaboration has been a longstanding practice. Price and cost analysts were collocated, which led to collaboration by physical proximity and resulted in improved cost and price outcomes. Observation of this effect led to program leadership exploring the effects of collaboration and has been an additional impetus for this paper.

Cost analysts reported the use of the following information in addition to the PNMs that supported their cost and budgeting efforts. Current Contractor Rates enabled the estimation of current costs. The Contractor Cost Proposals (CCPs) provided a window into expected costs that supported forecasting. The Bills of Material (BOM) enabled analysis of system configurations and have been reported as a significant benefit. Contract Scope helped in understanding the relationship of proposal costs to work effort and the time frame of the contract. Cost analysts also used CDRL reports provided as part of the contract deliverables, while access to Army supply systems provided visibility to parts costs.

The data from these sources was used to respond to requests for the Initial Government Cost Estimates (IGCEs) and also supported the Basis of Estimates for program costs. The sections of the PNM on Other Direct Costs (ODCs) and Indirect Rates were beneficial in developing a more complete cost estimate.

Cost analysts also stated that while access to reports was important, discussions with subject matter experts and price analysts for additional insights on current negotiations and contract structure helped develop future estimates. The structure of future contracts also helped the cost analysts appropriately bucket future expenses into the appropriate categories. Price analysts also provided a future Period of Performance (POP) which helped identify the fiscal year in which funding would be required to determine the plan for funds obligation.

Price analysts for the Stryker program reported that in addition to the CSDRs, the standard CDRL A007 that provided actual hours and cost expended was available to them. The Stryker program also had unique CDRLs; one example is the CDRL 0005 Parts Receipt Report which included part number detail, unit costs and quantities, and average unit costs over a rolling 12-month period that was especially useful in evaluating contract costs. These actual costs were beneficial in preparing for the next contract negotiations.

The PNMs from prior contracts and back up detail were also beneficial in developing the government position. DCAA Audit reports and DCMA Forward Pricing Rates were useful in analyzing current data and preparing for future contracts. Price analysts also received and used internal technical reviews from engineers regarding labor hours and types and quantities of material to prepare government positions and validate contractor proposals.

The CSDR 1921 was beneficial for a top level price analysis, but detailed analysis for configuration specific labor hours, dollars, and material dollars required the 1921-1. The detail also supported an analysis at the individual element level rather than at a total price level to determine the major changes from the awarded contract to the new proposal.

The IGCE from the cost analysis group also supported the evaluation of new proposals.

The Stryker cost analysts estimated greater percentage improvement in the early phases, around 10–15%, when programs' costs are less precise and when the negotiated contracts and contractor proposals can assist the analyst in developing more effective and accurate cost estimates.

Price analysts did not experience significant benefit—only 5% to 10%—in the early phases of programs. One possible explanation is that engineering effort can vary significantly from the early technology development in the TMRR phase to engineering development and integration effort in the EMD phase. As such, analyzing proposals based on past technical efforts is highly dependent on the scope of the engineering effort, which can vary significantly. In the early program phases, configurations for engineering models and prototypes are not firm, thus limiting the use of information from early prototypes.

However, price analysts reported greater benefit in the production phase—about 25%—going from the LRIP to FRP since actual costs from production are available for negotiating future production costs along with projected learning curve reductions. The production and post-production phases also benefited in the negotiations for program support costs as parts usage and cost information became available.

In summary, both price and cost analysts leverage information from multiple sources in addition to the PNMs and CSDRs to support cost estimates and price analysis effectively and have realized measurable and meaningful benefits from this collaboration. The mechanism for collaboration is primarily a human interface with access to reports along with limited automation of data.

M88 Program Discussion

For the M88 program, in addition to the PNMs, the cost analyst used the Forward Rate Pricing Proposal (FPRP) and FPRA to support cost analysis along with detailed data showing the Base and Overhead costs that the contractor assumed when the FPRPs/FPRAs were formulated. In addition to the PNMs and FPRAs, Actual Incurred Cost Reports (from the DCAA), Purchase Orders for selected parts (also from the DCAA), Hours per Vehicle (HPV) Reports (supplied by the DCMA and the contractor), and cost data collected from contractors via tailored CDRLs (i.e., Systems Technical Services (STS) Monthly Cost Reports) also proved beneficial. EVMS and CSDR data from other programs were also used.

Specific sections of the PNM, such as the Cost Element Summary, helped to break out the Base, Overhead, ODC, G&A, FCCM, and Profit at the negotiated price. This helped to compare estimates and assumptions with actual prices. It provided a method to compare PNM to PNM to understand and determine the cause for price increases, which helped explain the differences between actual and estimated costs. On the M88 program, the negotiated prices in the BOM of the PNM were important to identify the largest cost-driving components, and to track changes from PNM to PNM, which in some instances helped identify the change in part sourcing from Contractor Furnished Material to GFM. The Labor Hours by WBS provided a useful means to verify estimates and to track negotiated labor rates over time (from PNM to PNM), but Hours per Vehicle (HPV) Reports from the DCMA and the contractor provided better estimates. The Other Direct Costs (ODC) breakdown between interdivision, subcontractor, travel and other miscellaneous categories helped identify costs/scope that sometimes slip through the cracks of estimates. The Direct Labor Rates in a PNM were useful in some circumstances for generating IGCEs or tracking costs from PNM to PNM, but generally the additional detail was not accessed as frequently. The Indirect Rates were useful in some circumstances for identifying changes from PNM to PNM and understanding the variations to rates in the FPRPs/FPRAs, and helped in formulating IGCEs, but the level of detail is not typically applicable. The Pool Analysis in the PNM was marginally beneficial, but helped to point to other sources that the Acquisition Contracting Center used, such as documentation from formal audits to negotiate the various Pool Rates. It was also useful to account for Material Handling Overhead that the contractor adds to the GFM, and the PNM was helpful in identifying the specific additions that were applicable. The analyst also concluded that additional analysis into the various costs contained within different overhead pools would provide insights into how the contractor splits Direct and Indirect costs.

The price analysts did not rely on information from cost analysis and primarily used information from contractor and subcontractor proposals (current and past), prior PNMs, Request for Quotes (RFQ), the Government Supply System, and DCAA Audit Reports. The availability of CSDRs is a recent development, and it is expected to be an important source of information for price analysts. In addition, technical evaluations from the PMO, the DCMA, and the ACC, along with historical labor costs and material purchase orders from the contractor, were useful sources of data for the price analyst. Industry forecasts such as Global Insights on price escalations and learning curve calculators also supported the price analysis.

A review of the CSDRs by the price analyst led to the conclusion that the level of detail in the CSDR between recurring and non-recurring hours would improve the analysis of the contractor's proposal. An analysis of the reported costs and negotiated price on contracts currently under execution would improve the government's negotiation position for the next contract. Inflation estimates used by the cost analysts would provide better estimates than the OMB estimates currently used by price analysts which do not capture DoD pricing as effectively. The CSDRs used to develop Program Objectives Memorandums may also provide better insight into Indirect Costs for negotiations.

In summary, both price and cost analysts relied on quality information from independent sources such as the DCMA, DCAA, past negotiations, and contractor proposals and actual costs from different sources to deliver effective results. However, in this case, the review suggests that a systemic access to actual costs and the corresponding detail from the CSDRs can lead to greater insights and better government positions. The estimates on improvements by the cost analysts suggest that program cost fidelity increases from the EMD phase to the LRIP and FRP phases combined with access to negotiated costs from concluded contracts is supporting improved program cost estimates. The price analysis function suggests that access to current and traditional sources of information such as the DCMA and DCAA is providing improvement, while access to cost analyst sources such as the CSDRs might result in additional improvements.

PIM Program Discussion

In addition to the BOM and labor hours from the PNMs, the cost analyst accessed the FPRA from the DCMA. Additional sources included Earned Value information from the contractor's system and actual costs from the System Technical Services (STS) contracts.

The PNM includes the total Material Price and total Material Overhead by Contract Line Item which supports a top level analysis of material costs. The PNM also assisted detailed analysis by providing information on the top 50 material cost drivers, as well as an explanation of negotiations for those parts. A benefit of having this information was sharing material cost information for common components across programs for more accurate estimating. In one instance, the Driver's Vision Enhancer (DVE) component's latest negotiated cost was provided to another program for an improved estimate.

The PNM contains negotiated labor hours by Contract WBS and cost center, and this supported a cross reference to the actual hours from the EVM system for leverage in future negotiations on labor hours. In another example of collaboration, based on the FPRP, the price analyst provided the contractor's additional costs for material acquisition, material handling, and general and administrative, which were then used to analyze and validate historical markups and understand the trends. These trends were then applied to direct labor rates from the PNM to estimate future costs. Cost analysts relied extensively on price analysts' data to support their program cost estimates.

Similar to the M88 program, price analysts did not rely extensively on cost analysis data but instead used the standard available information from the DCAA and DCMA. Additionally, for material costs, contractors were asked to provide information on high dollar items which was compared to past proposals with adjustments for quantities and inflation. Price analysts also looked at other contracts being executed and identified common parts across systems to achieve volume discounts based on total purchases by the contractor. Price analysts also estimated additional usage of parts based on Foreign Military Sales (FMS) to negotiate discounts on current contracts. Labor hours and costs are one area where active input was sought from cost analysts to review proposed manufacturing labor for inconsistencies based on current and historical data. The program management office was also engaged to provide a technical analysis of the non-manufacturing hours to ensure that the contractor is providing the support the Government requires.

In summary, as in other programs, cost and price analysts relied on traditional sources of information but also demonstrated best practices such as seeking volume discounts by combining common parts from several systems and by including additional unit volume from FMS. Similar to the M88 program, the estimates on percentage improvements for the PIM program by the cost analysts suggest that as the program cost fidelity increases from the EMD phase to LRIP and FRP phases, estimates of program costs show significant improvement. Thus, better access to negotiated costs from concluded contracts is supporting improved program cost estimates. Cost analysts are suggesting that availability of and access to price information is resulting in improved estimates and thus making the case for improved collaboration. For price analysis, the percent improvement suggests that, similar to the M88, current available information was sufficient and therefore price analysts predicted a moderate improvement in performance. However, price analysts are only now being made aware of the CSDRs and other analytical approaches used by cost analysts. Thus, a systemic access to actual costs and the corresponding detail from the CSDRs can perhaps lead to greater insights and better government positions.

HTV Program Discussion

The HTV program has been in production for many years, and therefore data from prior contracts has been a significant source of information for the price analysts. Cost analysts have relied on PNM data, along with the supporting detail, extensively for their analysis. In addition to the CSDRs, cost analysts have used EVM data to support analysis and program estimates. Program Management Office technical input on configurations and labor was also important for cost analysts. EVM data was a valuable source of actual data, and learning curve data was significant in estimating manufacturing costs.

Price analysts are only now getting access to CSDR information, and their analysis has relied on other traditional sources for price analysis. Their review of the CSDR suggests that available data can be segregated by truck variant and is detailed enough for direct comparisons on several large cost items such as engines and transmission and will prove to be a useful source of information. Material negotiations rely on current standard costs and quotes, but the recent history and actuals from the CSDR will support an analysis of the reasonableness of contractor proposals. The price analyst review of the CSDRs also suggested that actuals to date and the estimate at completion date would be useful in assessing prior negotiations and provide leverage for future negotiations. Cost analysis by bill of material would be beneficial, but the contractor report provides history of labor costs by category and assembly station, which is more representative of the manufacturing operation. This report, when combined with actual data of production units and their progress through assembly stations, can be a more accurate representation of manufacturing costs. The price analyst also stated that when the CDRLs are not available due to timing, the CSDRs can be an important source of indirect cost information. The traditional sources of information for the price analyst include prior contract prices, PNMs, spreadsheets showing agreed-to cost buildups from prior contracts or contract modifications, overhead rates with pools and bases, both for current rates proposals and prior year actuals, along with the IGCEs and market research.

The price analyst assessed that the HTV production benefits due to current cost information may be limited since the purchased material by cost is commercial or competitively procured and actual data may not show a significant improvement. The information could have helped for labor since the contractor runs all production vehicles down the same assembly line, thus supporting a comparison of actual labor costs to negotiated labor costs. However, labor is only about 20% of vehicle cost, and the realized improvement may not be significant.

In summary, the HTV has been in production for many years, and the contractor generally includes historical data in follow-on production proposals for labor hours by department and truck variant, and also provides recent purchase costs for individual material items. The DCAA provides indirect projected pools and bases to the ACC along with actual versus proposed historical information. Thus, the availability of historical data closely matches the benefits of current data and shows only a slight improvement in the production phase.

Similar to other programs, the estimates on percentage improvements for the HTV program by the cost analysts suggest that as the program cost fidelity increases from the EMD phase to the LRIP and FRP, forecasted estimates of program costs show significant improvement. Thus, better access to the PNM, EVM data, and actual costs are supporting improved program cost estimates. This implies that availability of information is resulting in improved estimates, which suggests the potential to realize significant value through improved collaboration. For price analysis, the limited percent improvement suggests that current information sources are sufficient to deliver effective results. However, price analysts are only now being made aware of the CSDRs and other analytical approaches used by cost analysts. Thus, a systemic access to actual costs and the corresponding detail from the CSDRs can perhaps lead to greater insights and better government positions.

Information Sources Used Price and Cost Analysts Other Than PNMs and CSDRs

In this section, the sources of information other than the PNMs and the CSDRs are documented.

- DCMA:
    - Forward Pricing Rate Proposals (FPRP) and Agreements (FPRA)
    - Hours per Vehicle Reports
- DCAA:
    - DCAA audit reports on labor and overhead rates
    - Actual Incurred Cost Reports
    - Purchase Orders for selected parts
- EVM:
    - Earned Value Management System Reports on actual costs by work breakdown structure
- IGCE (Initial Government Cost Estimates)
- BOE (Basis of Estimates)
- POP (Contract Period of Performance)
- CDR:
    - A007 for Stryker program
    - 0005 Parts Receipt Report
    - Systems Technical Services Monthly Cost Reports

Observations

The analyst estimates of percentage improvements in their process suggests that cost analysts anticipate or have experienced significant benefits with access to information from price analysts and other sources used by the pricing group.

The cost analysis function fundamentally requires projections of costs in the future. Hence, the emphasis on using all sources of information to improve the fidelity of future forecasts is paramount. Any information that can be used to improve the accuracy of projections is validated and included in the analysis.

Price analysts in general have developed fairly complete sources of information to support a near-term contracting action, hence the reliance on DCMA and DCAA reports, prior contracts, prior proposals, and market research. The availability of the CSDRs has been a recent development, and the CSDR data is not available for all programs. Awareness of the availability of the CSDRs, where available, also was not widespread. Thus, there is an opportunity to share lessons learned from programs that have leveraged the use of this data and encourage its use across other programs.

It was also clear that the PNMs and CSDRs are necessary for collaboration but may not be sufficient. Several additional and important sources of information from the DCMA, DCAA, EVM, and Contract CDRLs are used extensively by cost and price analysts to deliver effective results.

All programs have used collaboration, but were primarily driven by individual initiative and relationships. This suggests that developing means to capture the institutional experience in the current collaborative efforts and an environment that supports systemic collaboration are critical for continued improvement in price and cost analysis.

Several best practices were observed across many programs, and sharing these benefits broadly would be beneficial. In one program, the historical overhead rates were analyzed by identifying the costs used to determine those rates. Applying a regression analysis to this data created a predictive model of future overhead rates. Other programs identified common parts across many programs and used the total volume across all programs to negotiate an improved Government position. This approach was further enhanced by including anticipated FMS volumes in determining total volumes and realized a better government position.

Conclusions

Benefits of Collaboration

Collaboration has benefited both cost and price analysts. The benefits are along two dimensions: (1) information exchange of reports and data between the analysts and (2) analyst-to-analyst interactions. The questionnaire data and the follow up discussions suggest that the magnitude of the benefit varies between price analysts and cost analysts. It appears that there is a significant flow of information from the pricing group to the cost analysis group, but the information flow from the cost analysis group to the pricing group is limited. However, price analysts have taken advantage of other similar sources of information and interactions with program offices to make up for the limited access to the cost groups. The grid in Figure 1 graphically displays the positioning of the cost and price analysts relative to the two dimensions.

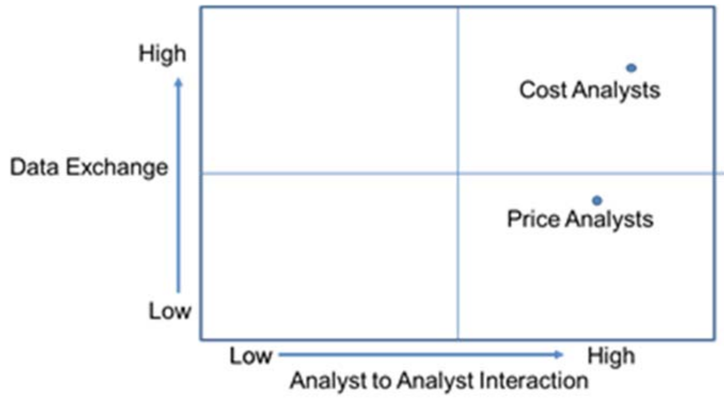


Figure 1. Positioning of Cost and Price Analysts Relative to the Two Dimensions

PNMs and CSDRs Are Necessary but Not Sufficient

The discussions with the analysts also revealed that there are several additional data sources that are significant in supporting analysts. Because the CSDRs and PNMs are the standard reports and should be available for all negotiated contracts, they were considered in our analysis as the primary sources of information. That said, the increased availability of the CSDRs is only a recent development, even though those reports are an increasingly common source of information for price analysts. Contract CDRLs, Earned Value Management System reports, and DCMA and DCAA reports on both forward and realized rates play an important role in supporting analysts.

Silos of Information

It is also clear that information exists in silos across the DoD enterprise, as evidenced by the information from the DCMA, DCAA, Program Offices, Acquisition Contracting Command, and Cost and Systems Analysis groups accessed by the analysts. Generally, only *experienced* analysts are able to obtain the required information based on relationships and knowledge of data. However, when new data sources are available, active use of these data sources demands proactive engagement. Given resource and time constraints, this may not always be possible. There is also not a readily available organizational mechanism to share best practices across the teams of analysts that support different programs.

Recommendations

- **Business processes and supporting information systems for rapid collection of and access to key program cost and pricing data would have several benefits.**
  - All analysts would have access to information on demand. (For contractor proprietary information, appropriate security and non-disclosure rules would need to be a part of the business processes.)
- **A notional information model is shown in Figure 2 where information is organized by Program ID.**
  - The information associated with every contracting action on the PNM is linked to the Program ID, which would include the PNM, EVM, DCMA, DCAA, and Contract CDRLs. As additional contracts are executed for the Program ID, information from each contract would be linked to the Program ID. As the program progresses through the acquisition framework, the CSDR, Bills of Material, LCCs, and POM

inputs would also be linked by Program ID. This information would be made available to analysts from across the enterprise for active use.

- Such an information organization would also lend itself to comparing the PNMs and Bills of Material, thus potentially automating the identification of changes and cost drivers.
- Bills of Materials comparisons could also be made across programs for tracking costs of common parts with similar form, fit, and function.
- Over time, the accumulated data could support large-scale data mining to understand configuration and cost trends.

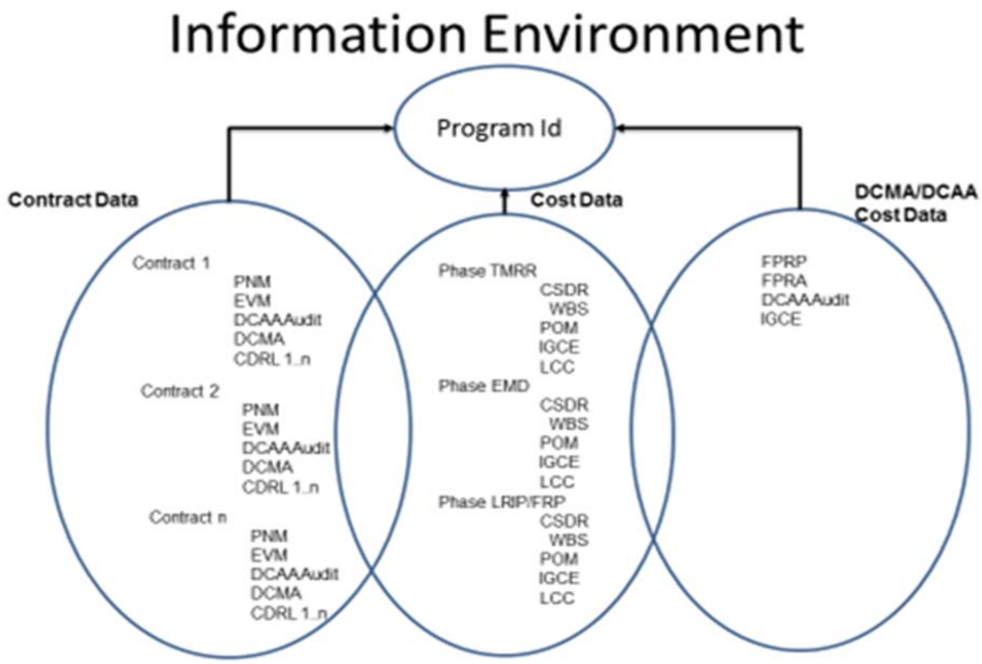


Figure 2. Notional Information Model Organized by Program ID

- **Collaborative Environment**
  - The benefits of co-location and collaboration were realized by the Stryker program, and while co-location may be constrained by availability of space, a simulated collaborative environment for price and cost analysts for all programs could be established. This environment could also include analysts from the Program Office, DCMA, and DCAA.
  - A technology environment that includes modern collaborative tools such as messaging, desktop video conferencing, and screen sharing applications to facilitate rapid communications should be considered.
- **Community of Practice (COP)**
  - The establishment of a Community of Practice (COP) to share best practices across the DoD enterprise where analysts could share insights, experiences, analysis, and successes should be considered.

Next Steps

This paper provided a window into the state of collaboration between price and cost analysts. It documented how collaboration is being practiced today and also identified several benefits of collaboration. It also recommended business process improvements and an information model to enhance the current state of collaboration. The next steps would involve describing in greater detail potential business process modifications and validating the expected improvements with the process changes. Other recommendations, such as the establishment of a community of practice and a clearinghouse for best practices, could be implemented in the short term.

Acknowledgments

The authors would like to thank the following staff and leadership of the Cost and Systems Analysis Group, the Acquisition Contracting Command, and Program Offices for their participation, insights, and support of this project:

Constance Tucker

Richard K. Kulczycki

Mary Durham

Paul Klott

John Lehman

Dale Lingaur

Lennie Schwerdtfeger

Carolin Stephan

Trevor VanAtta

Kelly Ziegenfelder

Panel 22. Improving Project Management of Complex Systems

Thursday, May 5, 2016	
3:30 p.m. – 5:00 p.m.	**Chair: William Taylor, Col, USMC (Ret.),** Program Executive Officer Land Systems, Marine Corps ***A Conceptual Framework for Adaptive Project Management in the Department of Defense*** Martin Brown, Jr., Project Manager, Program Executive Office for Enterprise Information Systems ***Program Affordability Tradeoffs*** Brian Schmidt, Economic Analyst, The MITRE Corporation Josie Sterling, Economic/Business Analyst, The MITRE Corporation Patricia Salamone, Business and Investment Analyst, The MITRE Corporation Ginny Wydler, Principal Analyst, The MITRE Corporation ***Squaring the Project Management Circle: Updating the Cost, Schedule, and Performance Methodology*** Charles Pickar, Senior Lecturer, NPS

William Taylor, Col, USMC (Ret.)—currently serves as Program Executive Officer Land Systems, Marine Corps (PEO LS), where he has been assigned since December 2008. He is the principal advisor to the Assistant Secretary of the Navy (Research, Development, & Acquisition) for the PEO portfolio of assigned major Marine Corps programs (ACAT I and II). Taylor was appointed to the Senior Executive Service (SES) in December 2008. Commissioned a second lieutenant in the U.S. Marine Corps in May 1979, he retired from active duty in September 2008 with the rank of colonel after capping his 29-year career by establishing and serving as the first Program Executive Officer Land Systems.

While in uniform, Taylor's extensive experience in acquisition management included assignments at every level within the Department of the Navy, from Program Office IPT Leader to the staff of the Secretariat. His acquisition career is highlighted by distinguished service as

NAVAIRSYSCOM's V-22 Joint Program Manager, leading the MV-22 Osprey Program from full-rate production to operational fielding. Prior to that, as H-46 Program Manager, he successfully forged a critical industry and government partnership, leading to the highly successful Engine Reliability Improvement Program. He also served in various capacities on the Assistant Secretary of the Navy for Research, Development, & Acquisition staff, including a tour as the Marine Military Assistant.

A veteran Marine helicopter pilot with nearly 5,000 flight hours, Taylor's operational experiences include combat operations in Beirut, Lebanon; missions in Cambodia in support of Joint Task Force Full Accounting; and presidential support as a Marine One Pilot assigned to Marine Helicopter Squadron One (HMX-1).

He holds a bachelor's degree from Rutgers University and a master's degree in defense systems acquisition management from the Naval Postgraduate School in Monterey, CA. Taylor's military decorations include the Defense Superior Service Medal in 2007; Legion of Merit with two gold stars in 2002, 2003, and 2008; Meritorious Service Medal in 1999; two Strike Flight Air Medals in 1983; Navy-Marine Corps Commendation Medal with a gold star in 1987 and 1991; and Combat Action Ribbon in 1983. Taylor is a member of the Senior Executive Association, the Marine Corps Association, and the Marine Corps Aviation Association.

A Conceptual Framework for Adaptive Project Management in the Department of Defense

Martin Brown, Jr.—is an experienced Project Manager with over 30 years of delivering successful results. During his Air Force career, he managed the transition to Air Mobility Command and was part of the multi-service group who defined the Joint Operations Planning and Execution System (JOPES). After leaving the Air Force, Brown spent 25 years successfully managing advanced research and production system projects. He was often called upon to rescue troubled projects. Brown became a Navy civilian in 2009 as the IIPT Lead for the Maritime Domain Awareness initiative. He was the Acquisition Lead for the Distributed Common Ground System—Navy, Increment 2 prior to joining the Navy Enterprise Networks team and the Mobility Lead for the Navy Marine Corps Intranet (NMCI). Brown holds both the Project Management Professional (PMP) and Agile Certified Practitioner (ACP) certifications from the Project Management Institute as well as a Level 3 DIAWA certification in Program Management. He has published several research papers on topics such as command and control in an information rich environment and adopting agile concepts in the DoD.

Abstract

Over the past 60 years, the conceptual framework defining project management has remained relatively unchanged despite a consistently poor success rate. The prescriptive, plan-based process has withstood several challenges because logically, it should work. In the past 10 years, the subject of complexity has received considerable attention from researchers. At the same time, project management is receiving attention from a fresh perspective. In the past, research focused on attempting to understand the underlying reasons for poor results. That has turned around with recent research focusing on project management success. Research has uncovered a set of traits found in consistently successful project managers indicating that successful managers approach project planning and execution from a different perspective than is taught in traditional project management curriculums. These successful project managers are able to adapt and adjust during execution to keep the effort progressing. This adaptive style of project management consistently performs well for highly complex environments, but it requires a perspective accompanied by skills that are not usually taught in traditional project management training curriculums. The purpose of this paper is to identify the characteristics of an adaptive project management framework and outline how those skills can be taught in the DoD acquisition environment.

Introduction

This research examines complexity as it impacts project management within the Department of Defense (DoD). The objective is to identify and explore the applicability of concepts emerging from recent research on project complexity and project management under conditions of complexity. The paper begins with an overview of traditional, prescriptive, plan focused project management concepts. The paper then explores a number of challenges to the traditional approach with an expanded discussion on agile principles applicable to project management.

The third section of the paper explores the concept of project complexity and research into the concept and its impact on project management. Following the discussion of project complexity, the paper explores project management and recent research identifying attributes of project managers with consistent records of success under conditions of complexity. Then, based on these attributes and borrowing from research, it describes an adaptive project management approach designed to address complex project environments. Finally, this paper makes initial recommendations for future project management training and education focusing on the specific success factors.

Traditional Project Management

The project management profession traces its roots back to the 1950s and the post-war environment. Project management evolved as a plan centric, prescriptive process. The focus is on creating the project plan and then executing to the plan. Project management theory grew from the "Scientific Management" approach set forth by Frederick Taylor. Scientific Management proponents believed that any process could be decomposed into its fundamental tasks. The ability to study, model, plan and implement improved task performance was key for improving efficiency and increasing profits. The transformation model of production served as the basis for an evolving concept of project management focused on managing work. In the transformation model, raw materials are converted into valuable, finished products through the efficient application of resources (primarily work). Early project management thought leaders like Fayol (work breakdown structure) and Gantt (schedule work) established a core set of principles that went unchallenged until the start of the 21st century.

This belief that managing projects is about managing work has become institutionalized, with a number of project management products designed to assist in the work management process. Work is planned and then managed to that plan for maximum efficiency. This plan prescribes the amount of work required and, through a few simple calculations, the budget and time required to complete the project. This plan centric approach to project management is reflected in the Project Management Institute's *Project Management Body of Knowledge* (PMBOK). PMBOK identifies 47 project management processes (Project Management Institute, 2013), and the majority (24) are directly linked to the planning phase of the project. Additionally, execution and controlling processes compare execution to the plan and seek to return to that plan, adjusting only as a last resort and only under carefully planned and documented processes. This plan centric thinking also serves as the foundation for a number of project related products, training courses, and certifications, resulting in the concepts becoming institutionalized.

The problem is that application of the processes, techniques, tools and methodologies have not resulted in a consistent pattern of success. Research by the Standish Group indicates that overall success rates for traditional project management methods (32%) are no different than those for an ad-hoc approach (44%; Standish Group, 2010). The 2010 IT Project Success survey by Dr. Dobb's Journal found that ad-hoc projects were 49% successful, while traditional approaches were 47% successful. The Dr. Dobb's IT Project Success Survey looked at Incremental and Agile project methodologies as well. The survey indicates that these two methods do improve project success rates to approximately 60%, which means that 40% of all projects will continue to fail or face significant challenges. Within the DoD, the results have been comparable. A November 2015 report by the Government Accountability Office (GAO) found that "the Federal Government invests more than $80 billion annually in IT. However, these investments frequently fail, incur cost overruns and schedule slippages, or contribute little to mission related outcomes" (Power, 2015).

A challenge facing advocates of revised thinking about project management is that these concepts have become institutionalized. Both the PMI and the DoD offer professional certificates based on demonstrated knowledge and experience in traditional project management. A number of products exist to assist in traditional project planning by delivering greater precision, sometimes at the cost of accuracy and predictability. Agencies such as the Government Accountability Office, faced with evidence of project management problems, focus on recommendations to improve the rigor and discipline used to apply traditional methodologies rather than exploring alternatives. Alternative concepts such as agile are re-cast are variations of the traditional approach. Because the plan centric concept has become institutionalized, many refuse to accept what the evidence tells us. Hill and Geras (2016), in their paper "System of Denial, Strategic Resistance to Military Innovation," found that "Dominant organizations have systems that focus organizational energy and attention on exploitation—that is, sustaining the status quo and continuing to improve what we already do." They go on to point out that this behavior inhibits continued learning and can generate "dysfunctional organizational responses to inconvenient information" (Hill & Geras, 2016). There have been improvements in DoD project performance in recent years, but innovative thought is constrained by conventional wisdom and organizational inertia.

Challenges to Conventional Wisdom

There have been challenges to this traditional view of project management, and there is no shortage of reasons proposed for the poor performance. Variations on the basic transformation have included the "Theory of Constraints" (Goldratt, 1997), a resurgence of the Flow model, originally proposed by Henry Ford and reintroduced as the Toyota (Ohno, 1988) way in the 1980s, and more recently a focus on value maximization as the basis of project management. In a briefing for International Project Management Day 2008, Harold Kerzner traces the evolution of views on Project Management. He found that traditional views of success being measured by the triple constraints (cost, schedule, scope) are giving way to a focus on delivering value within imposed constraints. This shift in thinking is significant in that it acknowledges that there is flexibility in the triple constraints that a knowledgeable project manager can use for business success.

Recently, the move toward agile methods further threatens traditional views of project management. Although many organizations focus on specific agile methodologies and rituals (e.g., short iterations, daily stand up, retrospectives) the heart of agile implementation is the fundamental changes to project management called for in the Agile Manifesto and the Agile Principles.

Agile practitioners see the detailed planning, task decomposition and assignment of hours at the start of a project as unnecessary, often wasted effort that sacrifices accuracy with the illusion of precision. Work, at the task level, is best assigned by the team performing the work as close as possible to the actual start of that work when the most information about the tasks is available. Scrum, the most popular agile method in the United States, eliminates the project management role, instead assigning typical project management responsibilities to various participants in the process.

The Project Management Institute (PMI) has had a difficult time adjusting to agile. Agile challenges several key tenets of the project management conceptual framework. First, agile welcomes change. The plan centric methodology of traditional project management maintains alignment with the plan until there is a compelling reason to change. Principle two states that change is welcome, stressing the need to be flexible unless there is a compelling reason to stay with the plan. Principles five and 11 stress the concept that quality and best value will emerge from the agile process during execution.

The PMI does offer a certification as an Agile Certified Practitioner (ACP), but this is based on agile work experience and a review of agile principles and survey of various agile methodologies. Michele Sigler, in the "Software Project Manager's Bridge to Agility," sees the logical transition from project manager to scrum master who serves as a facilitator to the software development process. In many ways, this is a return to the concept that managers manage work. But, is this the correct role for the project manager? Agile provides a separation of project and production responsibilities. The scrum master and the development team(s) are responsible for the production elements. They follow the rituals and develop a predictable throughput, but this is independent of any specific project. The product owner in scrum is responsible for defining, prioritizing and accepting the individual features of the project being developed which is more in line with

the project manager responsibilities. The issue is that thought and concept development for agile methodologies has focused on the development process or production side of the equation.

Agile Principles

1. Our highest priority is to satisfy the customer with early and continuous delivery of valuable software
2. Welcome changing requirements, even late in development. Agile harnesses change for the customer's competitive advantage.
3. Deliver working software frequently, from a couple of weeks to a couple of months, with a preference for the shorter timescale.
4. Business people and developers must work together daily throughout the project.
5. Build projects around motivated people. Give them the environment and the support they need, and trust them to get the job done.
6. The most efficient and effective method of conveying information to or within the development team is face-to-face conversations.
7. Working software is the primary measure of progress.
8. Agile processes support sustainable development. The sponsors, developers and users should be able to maintain a constant pace indefinitely.
9. Continuous attention to technical excellence and good design enhances agility
10. Simplicity—the art of maximizing the amount of work not done—is essential.
11. The best architectures, requirements, and designs emerge from self-organized teams.
12. At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its

Figure 1. The Agile Principles
(Agile Alliance, n.d.)

The true power of agile is in something often referred to as the "agile mindset." Project managers can embrace agile principles within any development framework. The focus on value, adaptability to change, and frequent interactions with the customer and stakeholders to ensure that the project remains aligned with enterprise needs are equally applicable in waterfall as they are in scrum. Perhaps the most important principle for project managers to understand is “Simplicity—the art of maximizing the amount of work not done—is essential” (Agile Alliance, n.d.). Often misunderstood, this principle emphasizes a minimalist philosophy of agile. Agile is about focusing on the most important and valuable elements of the product and working very hard to identify and eliminate often costly “bells and whistles.” The concept of challenging early requirements and demonstrating meaningful, if only partial, implementations helps the customer eliminate the extras and focus on the core capabilities required. For example, if the customer’s specification calls for a “fully automated” analytical capability, and an increment delivers a semi-automated feature, the customer may find the semi-automated capability acceptable, thus eliminating significant cost and effort that can now be focused on other “high priority” items. The project manager needs to understand how to define and manage to a minimum acceptable feature set as the top priority for initial efforts. This helps to ensure that the project efforts are not wasted, even if funding is reduced or the project is terminated early. In many cases, the minimum acceptable feature set will allow the enterprise to suspend or terminate projects early, saving funds to be invested in other efforts with a higher value payoff.

Simplicity also applies to project initiation and planning efforts. Rather than expend the effort to create detailed task decompositions and budget estimates that hide inaccuracy behind the illusion of precision, an agile estimate will use expected productivity and estimated size/complexity to provide a range of features to be included given a fixed time or budget constraint. If neither of these is set, it is simple to estimate the time and budget ranges needed to deliver all functionality. As Figure 2 shows, the degree of uncertainty for a project decreases over the life of the project. Estimates performed at the start of the project can be underestimated by 1.6 times or over-estimated by 1.4 times so both the point estimate and the range estimates fall within this margin of error. Risk factors are often applied to the initial estimate to increase the confidence; however, recent research indicates that risk factors applied to elements of the WBS may significantly over and under state the total risk to the project.

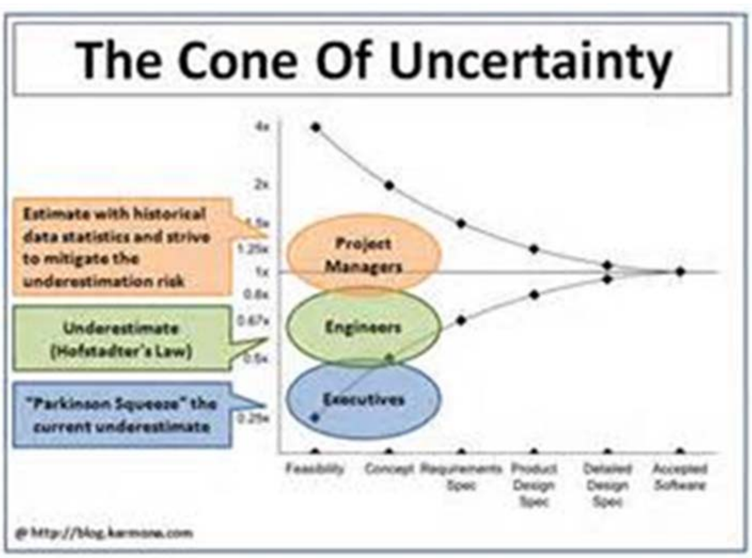


Figure 2. Project Uncertainty and Estimates

Project Complexity

What is project complexity? This question has been asked many times by researchers across the globe, and the answers have varied from paper to paper. Pich, Loch, and Meyer (2001) saw complexity as a result of the degree of project uncertainty. Rather than defining project objectives in terms of cost, schedule and cost, Pich, Loch, and Meyer proposed that complex project success is best defined as a payoff function that is dependent on the world state and decisions made by the project team. Their work proposed a shift in how projects are viewed, from a set of sequential tasks to a decision tree where information was revealed gradually over the course of the project. They stated that traditional project management methodologies, tools and techniques could deal with the known unknowns (referred to as risks), but failed repeatedly to address the unknown unknowns, those unforeseen events that are fairly common in project execution.

Other researchers have defined and categorized complexity causes. Hass (2008) identifies that "there is no widely accepted definition of project complexity that is research based and therefore defensible." Hass (2008) does identify several causes of complexity, such as

- Details—number of variables and interfaces
- Ambiguity—lack of awareness of events and causality
- Uncertainty—inability to pre-evaluate actions
- Unpredictability—the inability to know what will happen
- Dynamics—rapid rate of change
- Social Structure—numbers and types of interactions
- Interrelationships—many interdependencies and interconnections exist

Many of these same causes appear in a 2009 research study focused on defense acquisition in Australia. Members of the Commonwealth Department of Defense (including the Defense Materiel Organization [DMO]), the International Centre of Complex Project Managers (CCPM), and defense contractors such as Lockheed Martin, BAE, Boeing, and Raytheon, identified several themes related to project complexity, ranging from goals and stakeholders, to technology, management processes, and work practices and time (Remington, Zolin, & Turner, 2009).

Williamson (2012) sought to correlate the relationship between project complexity and project success. Working with the Project Management Institute, he conducted a survey in 2012 which established that increased complexity corresponded to lower success rates. An underlying message that emerges from the research on complexity is that our notion of a project as a sequential set of tasks is false. Several researchers (Benbya & McKelvey, 2006, pp. 12–34; Kautz & Madsen, 2010; Kautz, 2012) have explored the similarities of information systems development projects to complex adaptive systems (CAS). From a project management perspective, understanding the nature and structure of CAS is a critical element to successfully manage projects that exist in that domain.

Figure 3 shows both the traditional and complex views of the same project. The traditional view on the left has used reductionism (decomposition) to isolate the component tasks of the project and presents them in a sequential manner. This is the typical Gantt view used to sequence and manage work. When project tasks are reduced to this level, estimating the time and resources required for each task is straightforward. Unfortunately, the act of decomposition obscures the rich set of interrelationships, and the resulting cost and schedule estimates do not add up to the total value expected. The CAS view on the right shows how the elements of the project, the sequential tasks, can interact with other Comparative Project Views elements. The project manager and project team's understanding of these interrelationships is critical for success.

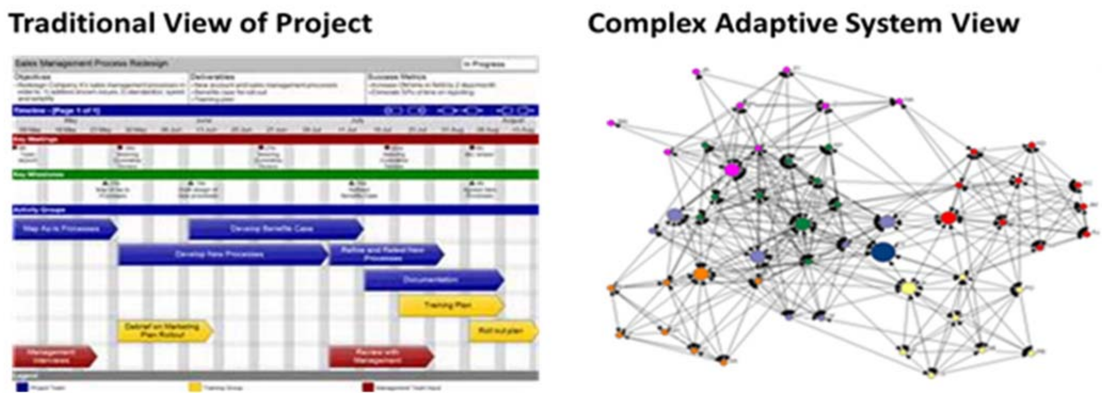


Figure 3. Projects From the Traditional and Complex Perspectives

Kautz (2012) identifies a set of characteristics found in CAS type projects. These characteristics establish the basis for the adaptive management concept.

- Interactions—The rich, dynamic, nonlinear and feedback behaviors of the development process as a whole cannot be known or predicted from an inspection of the components.
- Emergence—The emergent behavior and response to internal and external stimuli cannot be predicted or measured from an analysis of individual components.
- Interconnected autonomous agents (project team) have the ability to independently intervene and determine an action based on perception of the environment as well as sense and respond to change.
- Self-Organization—capacity of interconnected agents to evolve into an optimal organized form without external force to create disciplined interactions
- Co-Evolution—The entire project and its components alter structure and behavior in response to interactions both internal and external.
- Poise at the Edge of Chaos—The project exhibits both stability and instability at the same time. The project never locks into a predictable rhythm but never falls apart. Execution at the edge supports innovation and exploration.
- Time Pacing—The project settles into an internal rhythm that drives the momentum of change. Changes are time as well as event based.
- Poise at the Edge of Time—The project is rooted in the present but aware of the future.

Complexity also limits the utility of traditional project management tools. Analysis and estimation tools such as the Program Evaluation and Review Technique (PERT) and Critical Path Method (CPM) are valid estimating and analysis tools when the specific tasks are known but the expected durations can vary. The Graphical Evaluation and Review Technique (GERT) added Monte Carlo simulation allowed project management professionals to generate distributions of probable project durations, accounting for path convergence and generalized task distributions. These methods moved from identifying the critical path to predicting if a given task would find itself on the critical path. Carracosa, Eppinger, and Whitney (1998) use the Design Structure Matrix framework to add overlapping tasks and rework into schedule simulations. Ludwig, Mohring, and Stork (1998) added “dynamic policies” (p. 609) for project scheduling that simulated a state where activities times became known gradually over time.

This research develops solid approaches to deal with anticipated risk; however, it fails to address unanticipated events and risk, the unknown unknowns. Additionally, these tools do not provide a set of rules or policies describing how the presence of these risk factors influence project management.

Complex Project Management

Understanding the characteristics and sources of complexity and developing the knowledge and skills to execute in this space with the proper tools is critical for project managers. Recent research based on the complexity framework has provided insight into the nature of projects and how knowledgeable project managers consistently deliver successful results. Terry Cooke-Davies et al. (2011) report the findings of a yearlong series of workshops sponsored by the Project Management Institute (PMI) in 2010–2011. One of the key findings highlights the difference between traditional project management and complex project management. “Traditional project management training emphasizes how to do many things that have been done many times before and for which a lot of standards and road signs are in place” (Cooke-Davies et al., 2011). Those managers who demonstrate consistent success in complex environments have “a different perspective and clear realization that much of what is required involves exploration and ‘living off the land,’ that is creating what is needed from what the local environment provides at that moment” (Cooke-Davies et al., 2011).

Cooke-Davies et al. (2011) and other recent studies have begun to identify a set of characteristics possessed by project managers with consistently successful results. There are several variations on the list of project manager traits for success. The CIO, in an article titled “Six Attributes of Successful Project Management” (Levinson, 2008), provided the following list:

1. They possess the gift of foresight. They are able to anticipate and head off problems.
2. They are organized, focused on the “Big Picture” and able to prioritize competing priorities.
3. They know how to lead.
4. They are good communicators.
5. They are pragmatic and do not try to overanalyze.
6. They are empathetic. Most importantly, they understand stakeholder concerns and work to address them.

The CIO is not the only organization to publish project management success factors. The Standish Group, in their annual Chaos Report, list project success factors. Table 1 shows how this list has evolved over time by comparing the 1995 version to the 2009 version. It is interesting to note that “Clear Business Objectives” replaced “Clear Statement of Requirements” at number three and that “Proper Planning and “Small Project Milestones” gave way to “Project Management Expertise” and “Execution” in the more recent list.

Table 1. Evolution of Project Management Success Factors

1995	2009
1. User Involvement	1. User Involvement
2. Executive Management Support	2. Executive Support
3. Clear Statement of Requirements	3. Clear Business Objectives
4. Proper Planning	4. Emotional Maturity
5. Realistic Expectations	5. Optimization
6. Small Project Milestones	6. Agile Processes
7. Competent Staff	7. Project Management Expertise
8. Ownership	8. Skilled Resources
9. Clear Vision and Objectives	9. Execution
10. Hardworking, Focused Staff	10. Tools and Infrastructure

The other noteworthy element is the shift to leadership skills in the 2009 list. This is directly related to the evolving thought that the project manager is a leader who motivates and guides the execution of project activities. The shift toward leadership correlates to the inclusion of agile processes as a success factor. The agile movement, starting in 2001, has challenged traditional concepts of project management in ways that are often overlooked as organizations rush to be agile. Too often, agile rituals such as shorter iterations are adopted without thinking through the fundamental changes required for these rituals to be effective.

What emerges from these recent studies is a new profile for project (and program) managers which is supported by research. The traits of a project manager likely to succeed in complex environments include

- Business Focus: Project management decisions are business decisions that flow from the organization strategy and recognize the business value of the effort. This shift reflects a growing perception that the project manager's responsibility is to deliver value within defined constraints and not manage to a pre-defined cost schedule and performance.
- Focus on the Big Picture: Successful project managers constantly focus on value delivery, looking at execution tasks from the perspective, "How does successful completion of this task contribute to creating value?" The project manager understands that there are alternative paths toward value and that his/her primary mission is to move in directions that maximize overall payoff.
- Perceptive, Seems to Anticipate Need for Change: Successful project managers are quick to assess the impact of events, both internal and external, and are ready to adjust. This ability is a result of careful planning to identify essential elements of information and then recognizing them early.
- Leadership: Project managers are leaders. The good ones display empathy, conviction, a positive attitude, and an adaptable style that is appropriate for the situation and the team.
- Communications: Successful project managers know how and when to communicate and, more importantly, how to listen.
- Pragmatic: Successful project managers are not afraid of decisions. They don't over analyze or wait for others. They also empower the team to make tactical decisions because, as leaders, they have communicated the "manager's intent."

How are these traits put into practice? Pich, Lock, and De Meyer's (2001) "model of project uncertainty and complexity" compared several project management approaches under differing conditions of uncertainty (pp. 5–11). Their results support the concept of a pre-defined project plan and executing to that plan when there is adequate knowledge of the project terrain to create a plan that maximizes the payoff function. They caution that these circumstances rarely exist, especially for information technology projects. A second approach is to have a project plan with specified contingency actions. Again, this approach is most useful when project uncertainty can be anticipated with a degree of certainty. When the uncertainty and associated complexity of a project includes a significant number of unforeseeable events/influences, predetermined plans prove not to be the best project management approach. Under these circumstances, the best results are obtained when the project manager and team integrate a "learning" approach to their execution. Information gathering through either "scanning the horizon" or specific focused knowledge acquisition activities allow the team to learn and the project path to evolve. This "exploration of uncharted terrain" approach consistently achieved the best results under conditions of uncertainty.

A Conceptual Framework for Adaptive Project Management

Through these various studies on project complexity and project management success factors, a well-defined set of project manager skills and knowledge emerges. These skills and knowledge provide the basis for an adaptive approach to project management.

Strong Understanding of Business Value

Perhaps the most important question for a project manager is "Why?" Traditional project management focuses on "What," as in "What is the scope?," "What is the Budget?," and "What is the deadline?" This is adequate under conditions of certainty, where execution simply means following the plan. Unfortunately, to paraphrase a common belief for contingency operations, "plans rarely survive contact with the project." Understanding the underlying business reason and the desired value of the project allows the project manager in a complex environment to adjust and adapt within the value construct and to identify when key stakeholders need to be brought into the discussion because the available options result in the need to modify the value expectations. Harold Kerzner, speaking at the International Project Management Day conference in 2008, noted that project managers

- are involved in strategy and project selection processes and are expected to provide execution perspectives
- have expertise in business with some technical knowledge. Project managers are first and foremost expected to make sound business decisions.

As the Chaos report on success qualities in project managers indicates, there has been a shift toward project managers having in depth business skills with some technical knowledge. In part, this is due to a growing realization that project management decisions are business decisions related to the defining and prioritizing of activities and not the management of work.

Plan Is a Verb, Not a Noun

Traditional, prescriptive project management centers on developing and executing the project plan. Successful project managers in complex environments "Focus on the end goal and manage all elements to that end rather than trying to manage the individual components" (Cooke-Davies et al., 2011). Project managers who demonstrate consistent success in complex environments tend to plan and think in terms of the big picture. Planning focuses on understanding the intended flow of the project as well as how internal and external events can influence that flow. The natural tendency to address complexity through reductionism, which means to decompose complex elements into simple subsets, tends to restrict vision to a prescribed path, as shown in Figure 4.

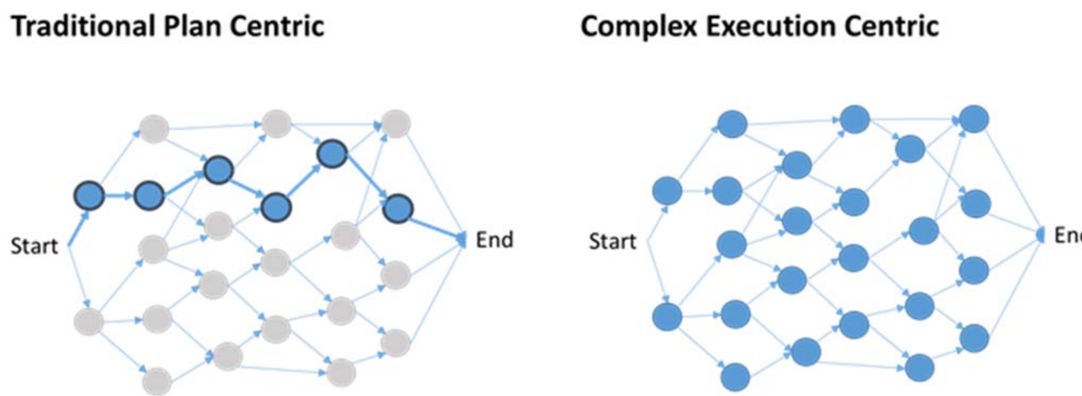


Figure 4. A Plan Centric View Obscures Options and Alternative Paths; Execution Centric Helps Identify Decision Points

Developing the traditional project plan requires a number of assumptions that establish the preference for one alternative over other available alternatives. As the plan is refined through increasingly detailed analysis and estimates, the project team becomes blind to the assumptions and how much error those assumptions have introduced. As execution progresses and assumptions prove to be in error, the project manager will often resort to expensive (in terms of time and resources) efforts to return to the plan because an alternative and less costly route is not readily apparent. Again, this was reiterated in research by Pich et al. (2001) and others.

In contrast, studies have found that successful project managers in complex environments plan at the macro level, focusing on identifying potential alternatives and the assessments required to decide. During this initial planning effort, unknown elements are identified and analyzed to see where they fit in the process flow, as are the project decisions that the unknown factors impact. Rather than make assumptions that support a specific project path, unknown elements are mapped to decision points based on how they impact the project, and external and internal factors that could influence the project end state are identified. Alistair Cockburn (2006) talks about the three elements of any project being the product, product knowledge, and process knowledge. Each is important for project success and therefore needs to be incorporated into project planning. A knowledge acquisition plan begins to unfold based on a policy stressing the value of knowledge and the cost of acquisition. Knowledge acquisition is not free. There are significant differences in the cost of knowledge based on the acquisition method used. Scanning the horizon or general observation is relatively inexpensive. Dedicated knowledge acquisition activities are significantly more expensive and therefore need to be used judiciously. There is also a decreasing value of additional knowledge. Once the project manager or team recognizes an event and identifies a potential impact, the law of diminishing returns applies to additional attempts to refine the information.

Nothing in this discussion of planning in complex environments is meant to imply that detailed planning does not occur. The CIO, in its discussion of success characteristics, noted that complex project managers are capable of producing detailed decompositions of project tasks quickly and accurately. The article goes on to point out that these project managers understand that these detailed decompositions reflect only one of a number of potential paths to completion (Levinson, 2008).

There are a number of similarities between leading a force in a contingency operation and executing a complex project. Army Field Manual 100-7*, Decisive Force: The Army in Theater Operations,* provided a model for complex project planning. It characterizes "Operational Art" as tactical and operational engagements designed to achieve strategic objectives (Department of the Army, 2005). The concept of "branches and sequels" is just as valid in project management as it is in contingency operations. Simply stated, branches are contingency plans for changing disposition, orientation, or direction of movement based on specific indicators and warnings. Sequels are actions taken after an event based on possible outcomes—victory, defeat, or stalemate (Department of the Army, 2005). Project managers who successfully navigate complex projects include branches and sequels in their plans. A change in direction or branch may be indicated by external events, while sequels are planned following key decision points in the project.

An added benefit of complex planning methodologies is that the analysis of decision points identifies a set of logical project review points often calling for stakeholder decisions on project direction. These natural governance points will normally be event based rather than calendar driven.

Empathy and a Pragmatic Approach

There is general agreement that success in complex project environments is often the result of the creativity, imagination, openness and flexibility of the project manager and the team. A common finding was that successful project managers displayed both passive and active empathy. Passive empathy is the level of consciousness in anticipating and predicting situations and taking control before they become problems. In reality, the experienced project manager has perfected his or her OODA Loop (Figure 5). Originated by Colonel John Boyd (USAF), the OODA (Observe, Orient, Decide and Act) loop represented the decision making and action cycle of a fighter pilot. Colonel Boyd stated that he could win any air engagement, starting from a position of disadvantage, simply because he operated on a faster decision cycle. Since originally proposed, the OODA loop concept has been used in a number of professions, including project management.

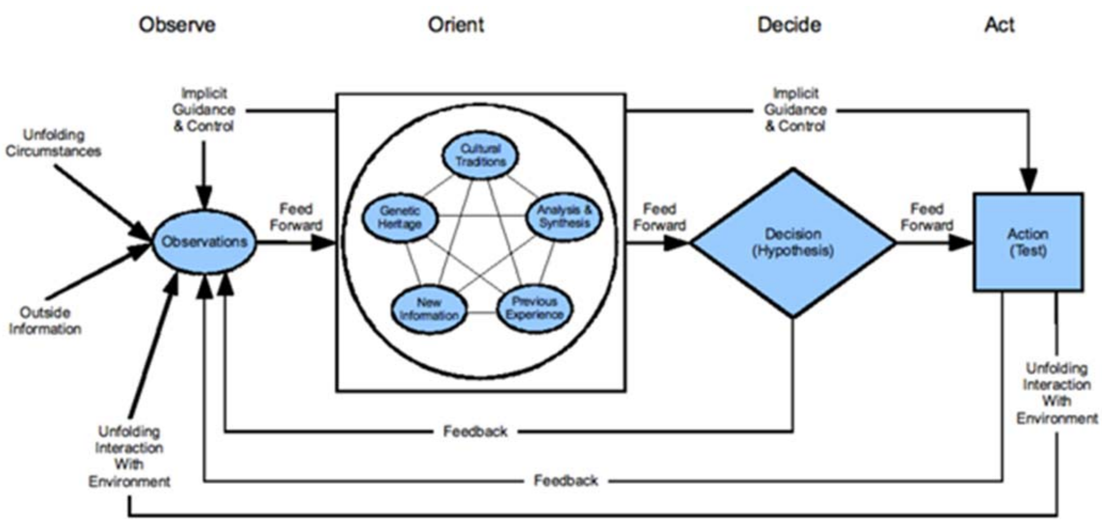


Figure 5. The OODA Loop
(Kallokain, 2008)

For the project manager, the challenge is to understand what to observe and how to operate in tune with the project flow as an autonomous agent, steering project execution consistently toward the goal. Additionally, as the leader of the effort, the project manager has to influence how members of the project team observe and respond to unfolding events and emerging information.

One of the key functions of the decision process is in knowledge acquisition throughout the planning and execution phases of the project. As stated previously, unknowns and uncertainty are not cloaked by assumptions. Instead, yet to be revealed information is mapped to the decisions it impacts, and plans for discovery are integrated into the project. In some cases, these may be implemented as what agile proponents call technical spikes; however, many times the information can be discovered through inquiry and expanding the project horizon. The goal is to gain sufficient knowledge to support decisions at the last responsible moment. The last responsible moment is an agile term used in lean development to reflect the requirement to decide at the point where further delay results in the loss of a valuable alternative. The concept is that decisions made too soon in the process do not take advantage of potentially valuable information, while procrastination leads to the loss of choices. Project managers need to guide this process, especially when key stakeholders are involved in the decision, to be sure that all decisions are made with the best available knowledge.

Management of the decision process to align with the concept of the last responsible moment is, in many ways, a corollary to normal decision models which focus on the right side of the decision by assessing the consequences of various choices. Here we are focusing on the left side to ensure that decision analysis benefits from the most knowledge possible.

Communications Is Key

Project management and successful execution in complex environments relies on communications. The project manager sets the tone and leads by example, but all members of the project team have the responsibility to communicate frequent and meaningful content. The project manager, just as the commander in a contingency operation, must clearly communicate his/her intent and continue to communicate intent throughout the planning and execution. Stephen Covey (1992) stated that “much of true leadership is exercised by communicating a vision and plan that appeals to the values of people through principles” (p. 24).

Effective, efficient communications is the unifying force that helps bring the self-organizing team of autonomous agents together in a synchronized group. Shared observations help all members expand the observe phase of the OODA loop, while communications regarding decisions and actions at the tactical level of execution aids the entire team in assessing the impact and reinforcing progress through synchronized actions.

Communications and transparency are also key elements to keep key stakeholders involved and engaged throughout the project lifecycle. Observables or “information radiators” enable team members and stakeholders to quickly come up to speed on progress and to identify elements needing additional attention. Software development teams use burn down charts to show progress on delivering required features. Kanban charts help visualize where specific elements are in the development lifecycle. At the project level, Gantt and milestone charts, with their sequential representation, are poor representations of project activities. PERT charts, with activities on nodes, help display the interconnections among the elements and show progress across the many project engagements. The problem is current project management software tools are good at creating and displaying Gantt charts, but do a poor job in providing clear, easily understood PERT depictions.

Appropriate Project Management Tools

A key training issue facing organizations today is how to prepare project managers to succeed on complex projects. Certifications such as offered by DIAWA and the PMI provide a solid baseline but are generally focused on traditional, prescriptive project methodologies. Training available on agile methodologies tends to focus on the software development side and does not provide insight for project managers.

A project manager in a complex environment needs to understand the various methodologies, their limitations and benefits, and how they can be adapted for a specific implementation. He/she also needs to understand how tools used in the traditional project environment can be adapted to function in a complex environment. For example, earned value is often used to assess the feasibility of the plan. This same set of calculations, applied to the range of efforts in a complex environment, can help identify where attention is needed because progress is lagging. Unlike the traditional use, earned value provides insight into where adjustments are needed to the execution, either by increasing the effort in a specific set of activities or to pull back, regroup, and try an alternative path.

Governance in an Adaptive Environment

Organizational oversight and governance is always an issue as one moves away from the prescriptive project management model. Wysocki (2014) states that the “current business climate is one of unbridled complexity, change, and speed. … This situation has placed a significant challenge on organizations and their project managers in that traditional project management tools, templates, and processes are no longer effective” (pp. 3–4).

Organizational complexity is a factor that sound governance policies can minimize. Project Value Delivery, in a 2013 white paper, stressed the need to minimize internal organizational complexity to help reduce overall project complexity. They cite multiple reviews, multiple overlapping review panels, and hierarchical review process as examples of organizational complexity impacting project execution that can be streamlined or eliminated. They recommend a “sound governance structure” potentially tailored for the project.

Wysocki (2014) believes that regular stakeholder reviews are critical in ensuring that the project remains aligned with the enterprise vision. Reviews are needed at Project Initiation to assess the affordability of the project, when the project plan (similar to a campaign plan) is reviewed, and then when needed for key decisions. The final review, after project close out, serves as a retrospective where the project manager and key stakeholders review what went well and where there is need for improvement.

Training Project Managers for Complexity

Project managers who can successfully navigate and deliver results in complex environments are not born. Today, most are accidents of experience, thrown into complex projects without a net and surviving. This doesn’t have to be the case. Within the DoD, we train leaders and help them develop and apply these same skills, usually in combat command positions. The parallels between project leadership and troop leadership are clear.

Traditional project management training emphasizes how to do many things that have been done before and for which a lot of standards and road signs exist. Managers come out of these training environments believing that every problem has a solution for which there is a paved road or high-speed rail line that will get them to their destination. As

Wysocki (2014) characterizes the situation, we are training cooks when we need chefs: "A cook is trained and experienced to follow recipes developed by someone else. A chef is that someone else" (p. 31).

Training needs to emphasize leadership, critical thinking, observation and situational awareness. Much of this are the same things we teach combat commanders, but with a project focus. Additionally, project managers need to understand the various tools available to them and when and how they are used. Project managers need to understand how various development and project methodologies function and how to tailor for a specific project.

Finally, training is needed on tools and techniques for stakeholder interactions and how to drive to key functionality. For example, in agile development, it is often useful to gather user representatives in a room, hand them a stack of play money (representing the budget), and have the various functions of the development effort arrayed on the table and priced. The users are asked to prioritize the functions, deciding what is above and below the line. The value for the project manager and the team is not the final prioritization, but the discussions that take place describing what could be cut from a high priority item and what elements from lower features would be elevated. This insight is invaluable during execution when tough decisions are required.

Beyond formal training, the key to developing project managers able to succeed in complexity is on the job training and mentoring. The Project Management Institute recommends establishing a mentoring program. In fact, *The Project Manager Competency Development Framework* cites an effective mentoring program as a leadership performance criterion (Project Management Institute, 2007). Mentors need to be trained to be effective, and they must have the correct temperament to be effective.

Job assignments need to be managed to provide project managers with the opportunity to learn by doing.

Conclusion

Researchers have made significant progress in understanding the nature of project complexity and the skills and characteristics project managers need to succeed. Project managers are leaders, and additional research is needed to understand how military commanders at all levels perform in complex contingency environments. Formal project manager training programs need to address the skills and competencies needed in complex environments. Assessing skills and knowledge needs to move away from multiple choice tests to practical exercises where there is no "school solution." In the DoD, this level of training is available to senior program managers at the 400 level of DAU classes. That type of training needs to flow downward to intermediate level classes.

More research is needed to refine project management education and training. Specifically, research is needed to refine the success traits of successful project managers. Specific tools and techniques used by these managers need to be catalogued, along with the concepts that led to the selection of specific tools and how the use was adapted. DoD project managers need to understand how to identify project specific indicators and warnings and how to apply the OODA loop in a project context. Finally, research is needed to identify and recommend solutions to eliminate controllable complexity in defense acquisition.

References

2010 IT Project Success Survey. (2010, August 2). *Dr. Dobb's Journal*.

Agile Alliance. (n.d.). Principles behind the Agile Manifestor. Retrieved from http://www.agilemanifesto.org/principles.html

Benbya, H., & McKelvey, B. (2006). Toward a complexity theory of information systems development. *Information Technology and People, 19*(1), 12–34.

Carracosa, M., Eppinger, S. D., & Whitney, D. E. (1998). Using the design structure matrix to estimate product development time. In *Proceedings of DETC 98, AMSE Design Engineering Technical Conference*. Atlanta, GA.

Cockburn, A. (2006). *Agile software development, the cooperative game* (2nd ed.). Upper Saddle River, NJ: Addison-Wesley Professional.

Cooke-Davies, T., et al. (2011). *Aspects of complexity: Managing projects in a complex world.* Newtown Square, PA: Project Management Institute.

Covey, S. (1992). *Principle centered leadership*. New York, NY: Simon and Schuster.

Department of the Army. (2005). *Decisive force: The Army in theater operations* (Field manual 100-7). Washington, DC: Author.

Goldratt, E. M. (1997). *Critical chain*. Great Barrington, MA: North River Press.

Hass, K. (2008). Introducing the new project complexity model, part 1. Retrieved from http://www.projecttimes.com/articles/introducing-the-new-project-complexity-model.-part-i.html

Hill, A., & Geras, S. (2016). System of denial, strategic resistance to military innovation. *Naval War College Review, 69*(1).

Kallokain. (2008, April 15). One revolution through the OODA loop [Web log post]. Retrieved from http://kallokain.blogspot.com/2008/04/one-revolution-through-ooda-loop.html

Kautz, K. (2012). Information systems development projects as complex adaptive systems. 23rd Australasian Conference on Information Systems. Geelong.

Kautz, K., & Madsen, S. (2010). Understanding agile software development in practice. In *Proceedings of the 2010 International Conference on Information Resource Management*. Rose Bay, Jamaica.

Kerzner, H. (2008). *Value-driven project management*. Retrieved from International Institute for Learning Inc. website: http://www.iil.com/IPMDay2008/pdf/2008_IPMD_Kerzner_Final.pdf

Levinson, M. (2008, September 3). Six attributes of successful project managers. Retrieved from http://www.cio.com/article/2433916/project-management/six-attributes-of-successful-project-managers.html

Ludwig, A., Mohring, R., & Stork, F. (1998). *A computational study on bounding the makespan distributions in stochastic project networks* [Technical report]. Berlin, Germany.

Ohno, T. (1988). *Toyota production system, beyond large-scale production*. Boca Raton, FL: Productivity Press.

Pich, M., Loch, C., & De Meyer, A. (2001). *On uncertainty and complexity in project management.* Fountainebleau, France: The Center for Integrated Manufacturing and Service Operations.

Power, D. (2015). (2015, November 4). *Information technology: Implementation of reform legislation needed to improve acquisitions and operations* (GAO 16-204T). Washington, DC: GAO.

Project Management Institute. (2007). *The project manager competency development framework*. Newtown Square, PA: Author.

Project Management Institute. (2013). *A guide to the project management body of knowledge* (PMBOK guide) (5th ed.). Newtown Square, PA: Author.

Project Value Delivery. (2013). *Minimizing complexity—The core of complex project preparation* (White paper 2013-13). Retrieved from http://projectvaluedelivery.com/

Remington, K., Zolin, R., & Turner, R. (2009). A model of project complexity: Distinguishing dimensions of complexity from severity. *In Proceedings of the 9th International Research Network of Project Management Conference*.

The Standish Group. *Standish Group 2010 chaos report*. Retrieved from http://www.standishgroup.com/

Williamson, D. J. (2012). *Assessing the relationships among information technology project complexity, complication, and success*. Newtown Square, PA: Project Management Institute.

Wysocki, R. (2014). *Effective complex project management, an adaptive framework for delivering business value.* Plantation, FL: J. Ross.

Program Affordability Tradeoffs

Brian Schmidt—Economic Analyst at the MITRE Corporation, has worked at MITRE for over 30 years. For the past 15 years, his work has centered around the development and application of the Portfolio Analysis Machine (PALMA™), a computer program whose purpose is to help government agencies select an optimal portfolio of investments based on cost and mission-level benefit. PALMA has been applied to projects for the Army, NOAA, and many other government agencies. He has a bachelor's degree in mathematics from the University of Chicago and a doctorate in mathematics from Princeton University. [bschmidt@mitre.org]

Josie Sterling—has been an Economic/Business Analyst at the MITRE Corporation since 1982, supporting cost analysis, analysis of alternatives, economic analysis, cost/benefit analysis, affordability engineering, trade-studies, and risk assessment and management for a variety of civilian and DoD sponsors. She holds a master's degree in mathematics from Tufts University. [sterling@mitre.org]

Patricia Salamone—Business and Investment Analyst at the MITRE Corporation, is a member of MITRE's Center for Acquisition and Management Sciences. She typically provides expertise and analytic support to multiple projects in areas including decision support, investment analysis, systems engineering, and acquisition support. She holds a bachelor's degree in physics from Worcester Polytechnic Institute and a master's degree in physics from Boston College. [salamone@mitre.org]

Ginny Wydler—Principal Analyst at the MITRE Corporation, supports Defense sponsors with their acquisition strategies and source selections. She has over 25 years of experience in acquisition and is a former Navy Contracting Officer. She holds a master's degree in National Security Strategy, Industrial College of the Armed Forces; a master's degree in Acquisition and Contracting, Naval Postgraduate School; and a bachelor's degree in Business Administration, University of Maryland. She has published in the Defense Acquisition Research Journal and National Contract Management Association (NCMA) Journal. She is a Certified Professional Contracts Manager (CPCM) and a Fellow with NCMA. She was DAWIA Level III Contracting. [vwydler@mitre.org]

Abstract

In today's fiscal environment, federal programs must be postured to conduct on-going tradeoff analyses to stay affordable as budgets are reduced and capabilities change or become more challenging to implement. This research focuses on recommended practices for conducting economic resource-constrained tradeoff analyses. The goal is to offer guidance to programs in making cost-effective affordability decisions that keep the program within its budget, or to find economic efficiencies if the program is currently affordable and within its budget.

Background

Over the years, there have been major efforts within the federal government to reduce the cost of acquiring systems. The Government Accountability Office has shown repeated problems in meeting program milestones and keeping programs within cost and schedule requirements. Since 2010, the Department of Defense (DoD) has issued three versions of Better Buying Power for the DoD acquisition community, emphasizing the need for "affordability" (Carter, 2010; Kendall, 2012, 2015).

In response to these problems, the MITRE Corporation conducted an internal research project resulting in the Affordability Engineering Framework (AEF). The AEF, shown at a high-level in Figure 1, provides a structured framework with approaches and tools to address program affordability challenges over the life cycle (MITRE Corporation, 2012).

Figure 1. AEF Framework

Step 1 of the AEF is to identify and assess potential risks to program affordability, and to initiate actions to mitigate those risks. Step 2 is to examine the sufficiency of the program baseline and corresponding cost and schedule estimates, then to compare this cost estimate with the budget profile to evaluate life cycle affordability. Step 3 is to conduct a tradeoff analysis of courses of action for the purpose of making the program affordable or, if the program is currently affordable, to explore opportunities to improve program efficiencies. Finally, the objective of Step 4 is to help the decision-maker select the appropriate course(s) of action.

The research described in this summary is intended to facilitate Step 3 of the AEF. Specifically, the purpose of our research is to (part 1) gain an understanding of how government program offices currently conduct tradeoff analyses and (part 2) develop a guidance document and a software tool to help them with this process. This summary describes shortfalls in current practice that came to our attention during part 1 and recommendations for correcting these shortfalls. These recommendations will influence part 2, which is now under way.

Approach and Findings

To better understand how program offices deal with affordability challenges and tradeoff analyses, we conducted interviews with MITRE staff supporting 19 government sponsors. The main topics of the interviews were to understand the following: what affordability tradeoff analyses programs typically conduct and what decisions are supported by the results of the analyses; what factors (inputs) are considered in conducting tradeoff analyses and how are they considered; and what resources (people, tools, time) are available to conduct these analyses. The information we gained from conducting these interviews will assist us in part 2 of our research (developing a guidance document and tool). Meanwhile, this summary paper deals with shortfalls in current practice that came to our attention in the course of the interviews. These shortfalls were in three areas: measuring benefit, combining metrics, and assessing risk in tradeoff analyses.

Measuring Benefit

In many studies, the only measure of benefit was a technical performance measure, such as the speed of an aircraft. No attempt was made to connect this to a measure of effectiveness, which expresses how well the system carries out its operational mission. This approach leads to a “more is better” outlook. It does not provide insight into how much operational value is diminished when a lower cost alternative is selected and whether this is acceptable. In studies that used metrics that were not easy to express in scientific units, such as “the ability to conduct close air support,” the system that was adopted for scoring was often not carefully constructed. For example, numerical scores were not given clear interpretations.

The authors recommend adherence to established decision analysis methods for rating value or utility. For further information on these methods, see Von Winterfeldt and Edwards’ (1986) *Decision Analysis and Behavioral Research*, Chapter 7.

Combining Metrics

In studies that combined the scores of several metrics, weighted averages were almost always used. Although the weighted average is known to be the correct function to use when certain independence conditions hold (see, for example, Kirkwood, 1997, p. 243), there are cases in which such conditions do not hold, as is illustrated by a classic study of the Mexico City airport (Keeney & Raiffa, 1993, Chapter 8). The inappropriate choice of a function can lead to misleading assessments of overall benefit. For example, an alternative that improves the overall benefit score may be one that is improving metrics that are already at an acceptable level, while leaving other metrics below acceptable levels.

To ensure that tradeoffs are represented realistically, the authors recommend that analysts be aware of the existence of functions other than averages that can be used to combine metrics. Some examples are the Multiplicative Utility Function (Keeney & Raiffa, 1993), the exponential average (Schmidt, 2015), and the max-average (Lamar, 2009). We are continuing research to investigate methods for making these concepts more understandable and useable for program tradeoffs.

Assessing Risk in Tradeoff Analyses

Risk was often not considered, or was considered improperly, in affordability tradeoff analyses. For example, in some studies risk was assessed for one candidate system but not for another. In other studies, only one type of risk (e.g., schedule risk) was considered, while other types (e.g., cost, technical maturity, interoperability, and statutory/regulatory) were ignored.

Our recommendation is to consider what we call the execution risk framework (Henry, 2011). This method evaluates each alternative across a number of risk sources or categories. For each alternative, the risk for each category is assessed using a utility-like scale. Once an assessment is made for each category, these scores can be combined using a variety of methods, including the max-average (Lamar, 2009) or exponential average (Schmidt, 2015). Risk scores can then be used to calculate risk-adjusted benefit. In addition, understanding where there is risk for a given alternative guides the formulation of new risk-reduction alternatives, which include risk mitigation activities and costs for those activities.

Next Steps

The next step in this research (part 2) will be to construct a guidebook on recommended practices and a software tool to help program offices make analytically-driven tradeoff decisions. This research will result in a simple-to-use tool enabling programs to conduct affordability tradeoff analyses on a regular basis. Although this study focused on DoD program offices, our intent is that all federal agencies will gain from the findings of this research and the products that will become available.

References

Carter, A. B. (2010). *Better buying power: Mandate for restoring affordability and productivity in defense spending* [Memorandum]. Washington, DC: DoD.

Henry, R. (2011). *Looking at the full picture: Evaluating alternatives with the execution risk assessment framework* (Public Release 11-4485). MITRE Corporation.

Keeney, R., & Raiffa, H. (1993). *Decisions with multiple objectives*. New York, NY: Cambridge University Press.

Kendall, F. (2012). *Better buying power 2.0: Continuing the pursuit for greater efficiency and productivity in defense spending* [Memorandum]. Washington, DC: DoD.

Kendall, F. (2015). *Better buying power 3.0: Achieving dominant capabilities through technical excellence and innovation* [Memorandum]. Washington, DC: DoD.

Kirkwood, C. W. (1997). *Strategic decision making: Multiobjective decision analysis with spreadsheets*. Belmont, CA: Duxbury Press.

Lamar, B. W. (2009). *Min-additive utility functions* (Public Release 09-0383). MITRE Corporation.

MITRE Corporation. (2012). *Affordability Engineering Framework (AEF) overview* (Public Release 12-1207).

Schmidt, B. K. (2015). *Exponential average* (Public Release 15-3793). MITRE Corporation.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York, NY: Cambridge University Press.

Disclaimer

Squaring the Project Management Circle: Updating the Cost, Schedule, and Performance Methodology

Charles K. Pickar—is a member of the Graduate School of Business & Public Policy faculty at the Naval Postgraduate School. He is a retired Army Officer with extensive experience in defense project management, systems engineering, international security affairs, and strategic planning. He holds a bachelor's degree from the University of Maryland, a master's from The Johns Hopkins University in systems engineering, as well as a master's in national security affairs from the Naval Postgraduate School and a doctorate from Nova Southeastern University. He is fluent in German and Italian. His professional experience includes executive management positions at Lockheed Martin and SAIC. His research interests focus on systems thinking and engineering approaches to management problems. [ckpickar@nps.edu]

Abstract

In the ever increasing complexity of defense acquisition, the traditional metrics for defense projects of cost, schedule, and performance are insufficient. This paper explores the concept of cost, schedule, and performance to determine if these three quintessential project management criteria are sufficient to serve as the guiding principles for defense project managers. The ever increasing complexity of the weapons system development environment, from the necessity for specialization to the intricacies of new technology, requires a broader view than that offered by cost, schedule, and performance. In squaring the project management circle, we must add a fourth variable, context, to provide focus. Because if we measure it, it will get done.

Introduction

This paper is an examination of the complexity of defense acquisition and its relationship to the measures of cost, schedule, and performance—the project management circle (see Figure 1). Rather than use the more traditional name of project management triangle (or triple constraint, or even the iron triangle), we refer to the concept as the project management circle. The circle recognizes the interrelationships, necessary equilibrium, and the influencing and balancing effects that these three variables provide. This paper seeks to demonstrate that these three variables are missing an important consideration that should be held in equal regard. Recognizing that fourth consideration will allow the circle to be squared.

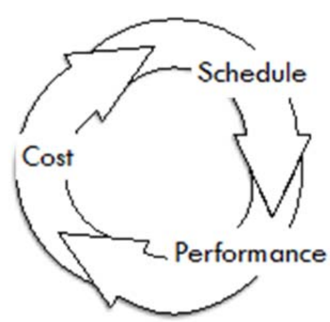


Figure 1. The Project Management Circle

Any discussion of cost, schedule, and performance must include the concept of project success as ultimately, cost, schedule, and performance are meant to ensure success. These three ideas form the heart of the concept of project management and are seen as not only management concepts, but as the definition of success in project management. In fact, cost, schedule, and performance are the prevailing criteria used to evaluate project success in the U.S. Department of Defense, and indeed throughout the U.S. Government. Congress uses cost, schedule, and performance; the media tend to focus on cost, schedule, and performance; and the GAO almost exclusively uses cost, schedule, and performance to measure execution. This paper suggests that the cost, schedule, and performance paradigm, while still effective as a measure of managing programs, needs to be expanded or changed.

The Problem

The current practice of project management assumes a simple, structured, and stable environment where the basic ideas of cost, schedule, and performance are sufficient to capture the workings of weapons system development, as well as serving to define success. However, the DoD environment is complex, dynamic, and constantly changing, and defining success is problematic. In this 21st Century environment, the obligation of management in general and that of project managers specifically is to deal with complexity. Getting the system developed and fielded, regardless of complexity, is the focus of the project management effort.

Simultaneously, while cost and schedule measures remain important, they are insufficient as measures of project success. Nevertheless, the management concepts of cost, schedule, and performance remain the same. We continue to manage and define success in acquisition using cost, schedule, and performance. While important measures, cost schedule, and performance are insufficient criteria for both program management and program success. This is the dilemma facing project managers today.

The Approach

The research seeks to examine the complexity of defense project management and relate that complexity to the key variables of cost, schedule, and performance. The intent is to explore other variables that will better help to explain DoD project success (or failure) and provide the DoD weapons system project manager the ability to manage more effectively. This research attempts to “square” the project management circle by identifying a fourth critical variable that must be addressed by project managers. The research methodology consists of a system-focused approach based on an extensive review of the literature of cost, schedule, and performance, and project complexity.

The analysis consists of three parts. The first section will examine the concept of cost, schedule, and performance. The second section explores DoD project management complexity. Project success is also examined as it relates to complexity and defense acquisition.

Lastly, a fourth variable, the concept of project management context, is introduced. We explore the idea of project context, identifying and categorizing the context of the defense project using a systems framework. The results of the analysis will identify those context variables that would contribute to a project management model addressing complex weapons systems development. The expected result is the identification of variables, beyond cost, schedule, and performance, that contribute to project success and enable the complex systems project manager to better address the management challenges evident in most DoD systems development.

Cost, Schedule & Performance

Cost, schedule, and a third measure—performance, scope, or quality, among others—are as old as the practice of project management. Project managers apply management principles and knowledge to effect change. The Project Management Institute (PMI; n.d.) defines a project as "a temporary endeavor undertaken to create a unique product, service or result." The temporary aspect emphasizes the limits of project management, the application of finite resources—both time (schedule) and budget (cost). The unique property underlines the focus and purpose of the project, as well as the result—performance. No one would dispute the importance of these concepts, but are these criteria sufficient in today's complex environment? The reality is projects in general, and complex projects in particular (including defense projects), are often completed late, over budget or both (Morris & Hough, 1988).

While well known to the practice of project management, cost, schedule, and performance are elusive concepts in the context of the academic literature. Cost, schedule, and scope in the construction industry are prevalent, but cost, schedule, and performance as used in defense are not. Most research on cost, schedule, and performance generally focuses on the specifics of earned value and the mechanics of managing weapons systems projects (Atkinson, Crawford, & Ward, 2006; Aubry, Hobbs, & Thuillier, 2007; Gardiner & Stewart, 2000).

The business of defense project management is to create a product—a system comprised of advanced technology for the most part, a weapons system (Gaddis, 1959). Defense project managers manage work (or scope), technology, people, interfaces and the overall system to ensure a viable result. Cost, schedule, and performance are metaphors for trade-offs. The project manager must manage many obvious and sometimes not so obvious constraints and trade those constraints against each other (Caccamese & Bragantini, 2012). This decision function is the essence of project management.

Systems take inputs and transform those inputs into outputs. Mastery of project management requires recognition that the project is a system and that the "black box" (transformation) process of systems development sometimes produces unforeseen results (outputs). These unforeseen results include a continuum that ranges from success to failure in both capability and management. Figure 2 is an Integrated Definition for Function Modeling (IDEF0) representation of the system. Inputs are combined with resources to produce outputs. Controls are the constraints of the system; budget and time are most often the primary constraints, with performance both a constraint as well as an output of the system. The mechanisms include the actual work and application of skills to transform the system. Management of the diverse factors that form the process of weapons system development results in functional integration. Systems integration further highlights the systems nature of defense project management.

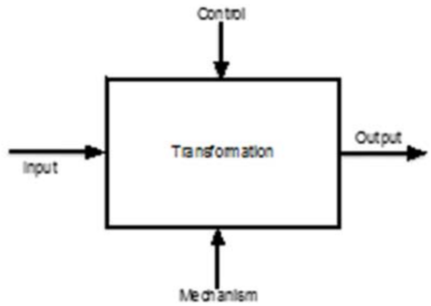


Figure 2. The Systems Nature of Project Management

We manage this system by monitoring, measuring and attempting to control cost, schedule, and performance. However, we often fail to meet these measures. Could it be that DoD projects continue to fail because our management tools consist of this limited set of measurement criteria? After all, the old phrase attributed to past systems thinkers and management experts, "what gets measured gets managed," is as actual today as when it was first voiced (Willcocks & Lester, 1996). Using the same thinking, a corresponding phrase, "what gets measured gets done," also rings weak in that we are certainly measuring, but in many cases, project success isn't getting done.

Further, even if the cost, schedule, and performance criteria are achieved, the only thing demonstrated is that we are meeting goals rather than accurately measuring success (Atkinson, 1999). Are meeting these criteria—cost, schedule, and performance—critical for success, without which the weapons system development is classified as a failure? Senior DoD officials routinely point out that notwithstanding cost and schedule overruns, the weapons systems the DoD produces are the best in the world.

Cost, schedule, and performance were sufficient for the management of simple programs and, in some cases, complicated programs. Complicated programs, while sometimes large and consisting of many moving parts, operate in predictable ways (Sargut & McGrath, 2011). The operation of a weapons system, while complicated (and difficult), is predictable. Complex programs are different. While complex systems may operate in predictable ways, the interactions of elements of the complex system are unpredictable, as they constantly change (Sargut & McGrath, 2011). By their very nature, cost, schedule, and performance are metrics, and as metrics, become predictors. However, it is almost impossible to predict with accuracy the end state of cost, schedule, and performance in complex systems.

Complexity

Complexity is the major dynamic of weapons system development in the 21st Century. Complexity is ever-present, but at the same time constantly changing. The continued growth of complexity has changed the process and the organizations of project management in important ways. Managerial and technical complexity, and the resultant recognition of the limits of human capability, has resulted in necessary changes in both human and organizational capacity. From the human perspective, complexity has spawned specialists—experts in a particular field—able to address those smaller aspects of a complex system that can be handled by a single person. The need to deal with technical complexity while ensuring system capability is the basis for the field of systems engineering, among other technical specializations.

Specialization has a limiting function, in that the specialists in a project organization are measured by, and capable of addressing, only those issues in their specific area. As a result, the project management offices have increased in size to meet the needs of specialization. This has resulted in a corresponding decrease in the visibility over the entire project, a “can’t see the forest for the trees” analogy from the individuals’ perspective. This has the potential of causing a potential decrease in efficiency in the execution of the project.

Complexity in project management refers to those organizational, informational and technical characteristics of the project and, by extension, the project management organization and the technical staff (Baccarini, 1996). Included in the organizational construct are the categories of stakeholders and other interested parties. Complexity has a direct effect on management and decisions as the more complex the system, the potentially more complex the management effort and decisions required. The mixture of human-socio-political complexity found in weapons systems development offices further adds to this complexity (Atkinson, 1999; Pinto, 2000). Finally, complexity reduces the predictability of the outcome of decisions made (Sargut & McGrath, 2011).

Definitions and explanations of complexity, managerial, engineering and technological abound, from Williams to Gell-Mann, to Holland, to Hughes (Gell-Mann, 1995; Holland, 1993; Hughes, 1998; Sargut & McGrath, 2011; Williams, 2002). From the project management perspective, Baccarini (1996) identifies two elements of complexity, organizational and technological complexity. He further subdivides these functions into differentiation and interdependency. Differentiation refers to the varied size and structure of projects and the organizations that manage them, while interdependency describes the activities between these varied elements (Baccarini, 1996).

Williams builds on the Baccarini topology and defines project complexity as categories in two key areas, structural complexity and uncertainty (Williams, 2005). Structural complexity is a result of the number of elements of a project, the pieces, including the people, the organizations, and the technology, coupled with the way these pieces interact, their interdependencies. This combination of interactions of the varied elements is structural complexity (Williams, 2002). Williams’ (2002) second aspect of complexity is uncertainty of the goals and the methods necessary to reach those goals.

Table 1. Project Management Complexity
(Baccarini, 1996; Sargut & McGrath, 2011; Sheard & Mostashari, 2009; Williams, 2002)

Type	Sub-type	Acquisition Management Example
Structural	Size	Organization (number of people) Budget Scope of work Contractor (size and number of people)
	Connectivity/ Actions/ Approvals	Acquisition organizations Requirements organizations Industry organization Review processes (both programmatic and technical)
	Organizational	Stakeholder Organizations Boundaries/ different commands/ different agencies Executive Branch Congress
Uncertainty	Budget	Funding
	Technical Complexity	Variety of tasks Interdependencies between tasks
	Objectives	System Requirements
Dynamic	Short-term	Daily problems Personnel changeover Engineer shortage Materials failures Short requirement dynamics Rework
	Long-Term	Changing budget Environment
Socio-Political	Social-Political	Personnel changeover "the new PEO/ PM" Change and change management Regulations/ Policy changes
System	Interdependency	Emergence Unanticipated actions and consequences a result of incomplete appreciation of system

Sargut and McGrath (2011) identify three properties, multiplicity, interdependence and diversity, as key. Multiplicity refers to the number of interacting elements or scale. This is similar to the Williams construct of structural complexity. Interdependence is the connectivity of different elements. And diversity is a measure of the difference in the elements (Sargut & McGrath, 2011).

From the systems side, Sheard and Mostashari (2009) explain project complexity from the systems engineering perspective. That view acknowledges structural complexity, but adds dynamic and socio-political complexity as factors influencing complex systems development. Dynamic complexity recognizes the active nature, the change-over-time, of systems development. Socio-political complexity reflects the human side of complexity, focusing on the importance and challenges of social interaction in systems development, including the cognitive challenge complexity causes, as well as the effect of everyday politics.

To allow for a more complete analysis, the complexity frameworks developed by Williams based on project management, by Sheard and Mostashari based on systems engineering, and discussed by Sargut and McGrath based on business considerations are combined to illustrate project management complexity in the Department of Defense (Baccarini, 1996; Williams, 2002; Sargut & McGrath, 2011; Sheard & Mostashari, 2009). While the Sheard and Mostashari (2009) framework is focused on systems engineering, it is valuable because it provides an important link from engineering to project management. The resulting framework (Table 1) includes a topology of different kinds of structural complexity, uncertainty, dynamic and socio-political complexity, and overall system complexity.

This grouping of complexity factors combines the management and engineering considerations in the development of weapons systems. In many cases, the resultant complexity is manifested in more than one type. For example, personnel issues including leadership changes have complexity effects in uncertainty, dynamic complexity, and socio-political complexity.

Structural complexity includes the scale, connectivity, organizational structure, and objectives of the development. Size is about magnitude of the acquisition system and its policies, bureaucracy and hierarchy to include the private sector side of defense acquisition. Connectivity acknowledges that the volume of staff actions between these organizations is significant and includes both issues relating to managing ongoing development. The connectivity aspect of structural complexity is influenced by the nature of defense acquisition systems. Since the technology development infrastructure (i.e., laboratories, R&D centers, and manufacturing) is for the most part privately owned, structural complexity also describes the network connectivity necessary for the system to function. Beyond the hierarchies, project organizations are major business entities directly controlling budgeting, spending and, in most cases, the award of fee to defense companies. Project organizations are spread throughout the United States and overseas, further adding to the complexity. Finally, the focus on defense project management by the DoD essentially means project organizations extend from the task level of the project to the highest levels of the bureaucracy. A recent GAO study recognized the challenges of structural complexity in finding the reviews for some programs include up to 56 organizations at eight levels. These structural requirements, reviews and responding to information requests can add up to two years to the development time (GAO, 2015).

Uncertainty focuses on three major areas: budget, technical complexity, and overall system objectives. In defense acquisition, budget is a major concern and source of uncertainty because of the year-to-year budget cycle, as well as political considerations. Uncertainty also stems from the military rotation policy, where senior leaders change jobs approximately every two to three years. Most new leaders are driven to make a mark on the organization and may be therefore unwittingly contributing to the uncertainty of the staff. Technical complexity is a fact of life in defense systems, and the reality is while we plan for technological development, it is in fact an estimate only. As we develop systems, we learn more about the technologies, and are then better able to plan for schedule, and cost.

Dynamic complexity is classified as short and long term and generally refers to change and time available. This concept is divided into the short and the long term because of the differences in perspective, as well as the universe of potential reactions to dynamic complexity (Sheard & Mostashari, 2009). Whether it is a tactical response to a development problem or an administrative response to directives, the project management system is in constant flux. This dynamic is a function of the diverse and always changing aspects of ongoing development. Further, each individual (the human element) will interpret and emphasize different aspects of the problem and how to address that problem. This has a potentially significant impact on the management system.

Socio-political complexity is the nexus between management, and the non-engineering human factors of policy, process and practice of the system is most critical (Maier, 1995). Socio-political complexity also recognizes the politics of project management, starting with the budget process, through Congress, and back into the development organizations. An oft overlooked, but critical aspect of context is politics. In fact, politics is by far the most powerful factor in the category of context. Most engineers and project managers dismiss politics as the realm of higher-level decision makers. In fact, many refuse to engage in politics as they find the practice distasteful (Pinto, 2000). However, dismissing those political activities can have consequences. Whenever people are put in an organization and asked to function as a team, there is an inevitable use of power and political behavior (Pinto, 2000). Notwithstanding a general distaste for political behavior in the workplace, the reality is the practice of politics is a prime force in any weapons development.

The last aspect of complexity in the context of program management is overall system complexity. System complexity is the result of the interaction of all the stated elements of project complexity. When different systems interact, or when different aspects of complexity act on each other, there are two results. The first is the cumulative effect of the interaction. For the project organization, the interdependencies between those managing the development and those executing the development should result in repeatable, consistent results—continued progress in system development (Rebovich, 2008). However, when the link between those managing and those executing is broken or, as can happen, ignored, the interdependency is broken.

In today's environment, the concentration of the defense project manager role on the program (work/scope), the technology to be developed, the interfaces of the system including those non-technical interfaces, and the project environment—the ecosystem of the development. Thus, complexity requires a different approach to project management, one that acknowledges the importance of managing resources (cost and schedule) to optimize system performance, but at the same recognizes the crucial known and unknown constraints and interdependencies of the environment—the context of the system development.

Project Success

Cost, schedule, and performance are both a management tool as well as a predictor of success. The management science discipline has sought to quantify the activities of the various management disciplines, including project management. Tishler et al. (1996) observed that in order to identify the managerial factors (and by extension the processes leading to those factors), success must be defined. They further cite research by Pinto and Slevin (1998) that definitions of success change during different phases of the lifecycle.

A major focus of the literature on project success has been on the idea of success criteria, or critical success factors (CSF; Jugdev & Müller, 2005). Identified success factors include cost, schedule, and performance, as well as project functionality and its management (Morris & Hough, 1988). Further studies added criteria such as customer satisfaction, efficiency of execution, and effectiveness of the project organization (Pinto & Slevin, 1998). Tishler et al. (1996) suggest groupings of four critical success factors for defense oriented projects, preparation, quality of the system development team and the user customer organization, management policy and project control. This clustering of success factors represents the amalgamation of the broader literature on project success criteria (Atkinson, 1999; Cooke-Davies, 2002; De Wit, 1988; Morris & Hough, 1988; Pinto & Mantel, 1990; Pinto & Slevin, 1998).

Preparation for project execution includes the necessary planning for initiating the project, as well as the necessary coordination. Included in the idea of preparation is an assessment of the urgency of need as urgency in defense projects overcomes many constraints. Team quality refers to both the management as well as technical capabilities of the development organization. Management policy is focused on quality, producibility, and design-to-cost considerations. Project control relates to the systematic use of control methods for cost, schedule, and performance. Together, these factors represent the criteria generally necessary for projects to be successful. This suggests that the exclusive and rigid adherence to cost, schedule, and performance as indicators of success (and the hallmark of defense project management) alone does not reflect the totality of success in defense project management.

Most importantly, the literature identifies two kinds of success, project success and project management success (Cooke-Davies, 2002; De Wit, 1988; Jugdev & Müller, 2005). Project success is measured as achieving technical performance and/or mission performance goals, coupled with customer (warfighter) satisfaction (De Wit, 1988). Project success is measured against the overall objectives of the development (Cooke-Davies, 2002). The nature of defense acquisition requires weapons systems that function as intended. This is measured in very real terms of life or death and battlefield success or failure. Rigid adherence to and sole focus on cost, schedule, and performance mean little if the system does not perform when needed (Cleland & King, 1983). In the greater scheme of things, what matters in defense acquisition is whether the system functions as the warfighter needs.

A major, complex project's principal success criteria will vary over time (Atkinson, 1999; De Wit, 1988; Morris & Hough, 1988). To paraphrase de Wit (1988), defense projects have at least three specific indicators of success: identifying the technology, developing the technology, and developing the weapons system using the technology. Delivering capable weapons systems is project success. But, identifying the technology and developing the technology are also measures of project success.

Cost, schedule, and performance measure project management success. At its heart, project management success is a measure of how efficiently the project has been managed (Baccarini, 1999). Project success is different from project management success. Project success will be determined by the warfighter community.

Project management success includes overcoming issues such as supply-chain challenges and effective coordination within the project management office. Project management success is focused on the development process. The DoD, the GAO, and Congress measure project management success, rather than project success. Project management success, however, is and must remain subordinate to project success. Cost, schedule, and performance are inadequate indicators even for project management success.

Context, the Fourth Criterion

The factors of complexity identified in Table 1 provide a starting point for identification of factors that are beyond the basics of cost, schedule, and performance, yet influence project and project management success. From a systems perspective, execution of weapons system development must be considered from the viewpoint of all stakeholders (Owens et al., 2011). This viewpoint includes an appreciation of the identified elements of complexity, including structure, uncertainty, dynamics, socio-political and system. These factors of complexity constantly shape the project environment and form the basis of the context of the system within which the project manager must operate. We group these complexity factors that are the result of the execution of a project together and define them as context (Owens et al., 2011).

Context includes those project organization activities that are essential to administer programs, but are not directly related to the cost, schedule, and performance of the project. Context ranges from tracking budget requests through the bureaucracy to responding to stakeholder inquiries on how resources are being used. In weapons system development, context includes those activities that, while not tied directly to cost, schedule, and performance, are essential for execution.

Each project is unique, a mix of many factors. More than cost, schedule, and performance, context reflects the ecosystem of the project organization and the project. If cost, schedule, and performance are measures and criteria for project management success, context is a criterion for project success.

The goal of this paper was to explore the concept of cost, schedule, and performance to determine if these three quintessential project management criteria were sufficient to serve as the guiding principles for defense project managers. The ever increasing complexity of the weapons system development environment, from the necessity for specialization to the intricacies of new technology, requires a broader view than that offered by cost, schedule, and performance. In squaring the project management circle, we must add as fourth variable, context, to provide focus because if we measure it, it will get done.

References

Atkinson, R. (1999). Project management: Cost, time and quality, two best guesses and a phenomenon, its time to accept other success criteria. *International Journal of Project Management, 17*(6), 337–342. http://doi.org/10.1016/s0263-7863(98)00069-6

Atkinson, R., Crawford, L., & Ward, S. (2006). Fundamental uncertainties in projects and the scope of project management*. International Journal of Project Management, 24*(8), 687–698. http://doi.org/10.1016/j.ijproman.2006.09.011

Aubry, M., Hobbs, B., & Thuillier, D. (2007). A new framework for understanding organisational project management through the PMO. *International Journal of Project Management, 25*(4), 328–336. http://doi.org/10.1016/j.ijproman.2007.01.004

Baccarini, D. (1996). The concept of project complexity–A review. *International Journal of Project Management, 14*(4), 201–204.

Baccarini, D. (1999). The logical framework method for defining project success. *Project Management Journal.*

Caccamese, A., & Bragantini, D. (2012). *Beyond the iron triangle: Year zero*. Project Management Institute.

Cleland, D. I., & King, W. R. (1983). *Project management handbook*. New York, NY: Van Nostrand Reinhold.

Cooke-Davies, T. (2002). The “real” success factors on projects. *International Journal of Project Management, 20*(3), 185–190.

De Wit, A. (1988). Measurement of project success. *International Journal of Project Management, 6*(3), 164–170. http://doi.org/10.1016/0263-7863(88)90043-9

Gaddis, P. O. (1959, May–June). The project manager. *Harvard Business Review*.

GAO. (2015). *Acquisition reform: DOD should streamline its decision-making process for weapon systems to reduce inefficiencies* (GAO-15-192). Washington, DC: Author.

Gardiner, P. D., & Stewart, K. (2000). Revisiting the golden triangle of cost, time and quality: The role of NPV in project control, success and failure. *International Journal of Project Management, 18*(4), 251–256. http://doi.org/10.1016/s0263-7863(99)00022-8

Gell-Mann, M. (1995). What is complexity. *Complexity, 1*(1), 16–19.

Holland, J. H. (1993). Complex adaptive systems. *A new era in computation*.

Hughes, T. P. (1998). *Rescuing Prometheus*. New York, NY: Pantheon.

Jugdev, K., & Müller, R. (2005). A retrospective look at our evolving understanding of project success.

Maier, M. (1995). Social systems engineering and architecture. In *Proceedings of the Fifth International Symposium of the National Council on System Engineering* (Vol. 1, pp. 139–146).

Morris, P., & Hough, G. H. (1988). *The anatomy of major projects*. John Wiley & Sons.

Owens, J., Ahn, J., Shane, J. S., Strong, K. C., & Gransberg, D. D. (2011). Defining complex project management of large U.S. transportation projects: A comparative case study analysis. *Public Works Management & Policy, 17*(2), 170–188. http://doi.org/10.1177/1087724x11419306

Pinto, J. K. (2000). Understanding the role of politics in successful project management. *International Journal of Project Management, 18*(2), 85–91. http://doi.org/10.1016/s0263-7863(98)00073-8

Pinto, J. K., & Mantel, S. J. J. (1990). The causes of project failure. *IEEE Transactions on Engineering Management, 37*(4), 269–276. http://doi.org/10.1109/17.62322

Pinto, J. K., & Slevin, J. K. (1998). Project success: Definitions and measurement techniques. *Project Management Journal*, 1–7.

Rebovich, G. (2008). *The evolution of systems engineering* (pp. 1–5). Paper presented at the 2008 Annual IEEE Systems Conference. http://doi.org/10.1109/SYSTEMS.2008.4518992

Sargut, G., & McGrath, R. G. (2011). Learning to live with complexity. *Harvard Business Review, 89*(9), 68–76, 136.

Sheard, S. A., & Mostashari, A. (2009). A complexity typology for systems engineering. *Systems Engineering*.

Tishler, A., Dvir, D., Shenhar, A., & Lipovetsky, S. (1996). Identifying critical success factors in defense development projects: A multivariate analysis. *Technological Forecasting and Social Change, 51*(2), 151–171. http://doi.org/10.1016/0040-1625(95)00197-2

Willcocks, L., & Lester, S. (1996). Beyond the IT productivity paradox. *European Management Journal, 14*(3), 279–290. http://doi.org/10.1016/0263-2373(96)00007-2

Williams, T. (2002). *Modelling complex projects*. New York, NY: John Wiley & Sons.

Williams, T. (2005). Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE Transactions on Engineering Management, 52*(4), 497–508. http://doi.org/10.1109/TEM.2005.856572

Project Management Institute (PMI) . (n.d.). What is project management? Retrieved from http://www.pmi.org/About-Us/About-Us-What-is-Project-Management.aspx

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net