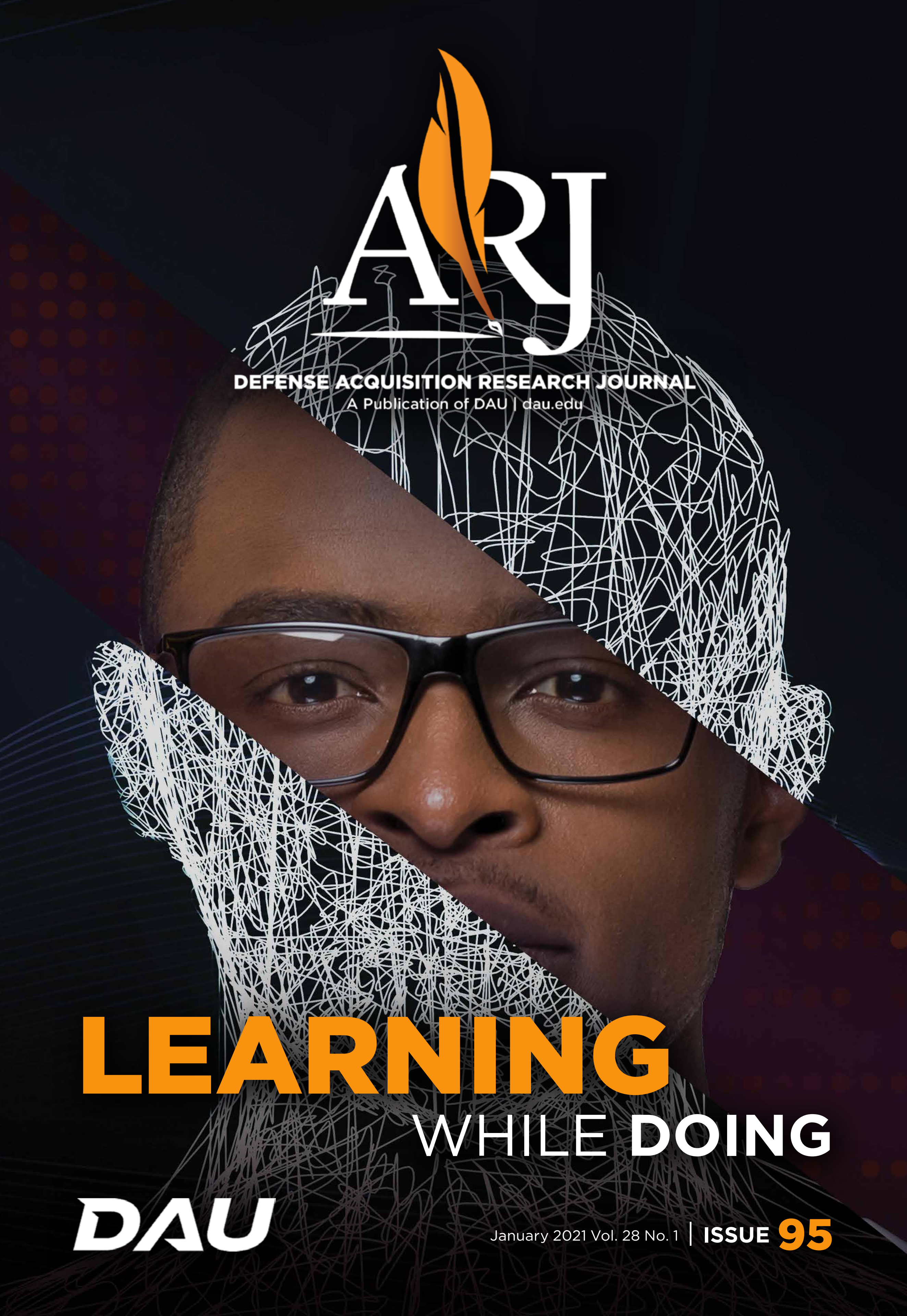


ARJ

DEFENSE ACQUISITION RESEARCH JOURNAL
A Publication of DAU | dau.edu

LEARNING WHILE DOING

DAU

January 2021 Vol. 28 No. 1 | ISSUE 95

ARTICLE LIST

ARJ EXTRA

DEFENSE ACQUISITION RESEARCH JOURNAL

A Publication of DAU | dau.edu

We're on the Web at:
http://www.dau.edu/library/arj

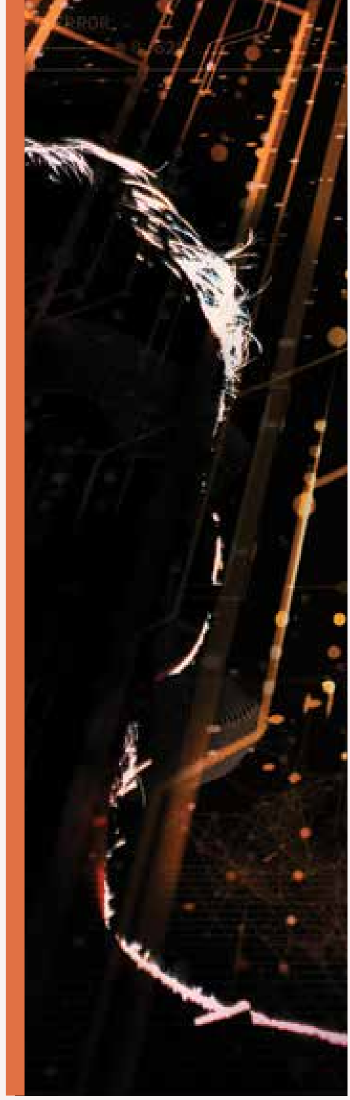
DAU

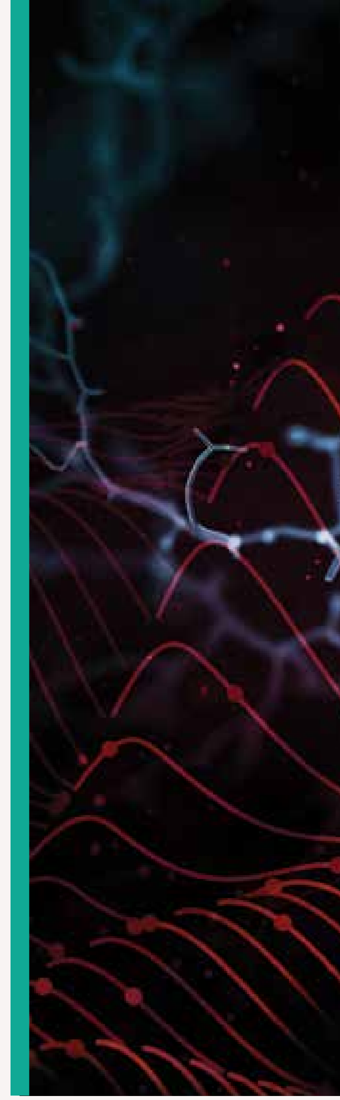
ISSN 2156-8391 (print) ISSN 2156-8405 (online)
DOI: https://doi.org/10.22594/dau.012021-95.28.01

The *Defense Acquisition Research Journal*, formerly the *Defense Acquisition Review Journal*, is published quarterly by the DAU Press and is an official publication of the Department of Defense. Postage is paid at the U.S. Postal facility, Fort Belvoir, VA, and at additional U.S. Postal facilities. Postmaster, send address changes to: Editor, *Defense Acquisition Research Journal*, DAU Press, 9820 Belvoir Road, Suite 3, Fort Belvoir, VA 22060-5565. The journal-level DOI is: https://doi.org/10.22594/dauARJ.issn.2156-8391. Some photos appearing in this publication may be digitally enhanced.

DEFENSE ACQUISITION RESEARCH JOURNAL
A Publication of DAU | dau.edu

CONTENTS | Featured Research





72

A Learning Curve Model Accounting for the Flattening Effect in Production Cycles

Capt Evan R. Boone, USAF, John J. Elshaw, Lt Col Clay M. Koschnick, USAF, Jonathan D. Ritschel, and Adedeji B. Badiru

This research created a new learning curve for production processes that incorporates a new model parameter. The new parameter allows for a steeper learning curve at the beginning of production and a flattening effect near the end of production.

CONTENTS | Featured Research

FROM THE CHAIRMAN AND EXECUTIVE EDITOR

Dr. Larrie D. Ferreiro

The theme for this issue is "Learning While Doing," an appropriate premise given that now in the era of COVID-19, many of us are getting on-the-job training in how to effectively work remotely from our teammates and organizations.

The first article, "Technology Trust: System Information Impact on Autonomous Systems Adoption in High-Risk Applications" by Michael G. Anderson and Johnathan C. Mun, addresses one of the more important issues in adopting autonomous systems in the military: how and when to deploy such technology, even as the systems become more capable. The use and adoption of an autonomous technology to replace people depends on both the system capability to perform the task, and the trust (based on experience) that it will do so. The development of experience-based trust in autonomous systems is costly and carries a high risk of harm to operators. This article examines a methodology for technology discovery that reduces the need for experience-based trust and contributes to increased adoption of autonomous systems.

The second article by Matthew R. Markman, Jonathan D. Ritschel, and Edward D. White, titled "Use of Factors in Development Estimates: Improving the Cost Analyst Toolkit," reports on research that expands the currently available toolkit for cost analysts, through the development of cost factors in the Engineering and Manufacturing Development (EMD) phase of the life cycle. The authors provide guidance on where cost analysts' efforts should be allocated, using factor creation and statistical testing in areas such as program management, systems engineering, data, and training.

The third article is "A Learning Curve Model Accounting for the Flattening Effect in Production Cycles" by Evan R. Boone, John J. Elshaw, Clay M. Koschnick, Jonathan D. Ritschel, and Adedeji B. Badiru. It describes the creation of a new learning curve for production processes that incorporates a new model parameter, that of the "flattening effect" later in the production process, i.e., a decreasing learning rate function over time, as opposed to a constant learning rate that is frequently used. The new parameter allows for a steeper learning curve at the beginning of production, and a flattening effect near the end of production. This model showed a statistically significant reduction in error when compared to Wright's learning curve, which is a popular method used by many organizations today.

The Research Agenda has been expanded to include cybersecurity and cyberanalytics.

This issue's Current Research Resources in Defense Acquisition focuses on Mid-Tier Acquisition.

The featured work in the Defense Acquisition Reading List book review is *The Story of Technology: How We Got Here and What the Future Holds* by Daniel Gerstein, reviewed by Janel C. Wallace.

Dr. Craig Arndt has left the Editorial Board. We thank him for his service.

We welcome Mr. David Lewis to the Editorial Board.

Dr. Larrie D. Ferreiro

Chairman and Executive Editor

Defense ARJ

DAU CENTER FOR DEFENSE ACQUISITION

RESEARCH AGENDA 2021

This Research Agenda is intended to make researchers aware of the topics that are, or should be, of particular concern to the broad defense acquisition community in the government, academic, and industrial sectors. It is compiled using inputs from Subject Matter Experts (SMEs) across those sectors. These topics are periodically vetted and updated as needed to ensure they address current areas of strategic interest.

The purpose of conducting research in these areas is to provide solid, empirically based findings to create a broad body of knowledge that can inform the development of policies, procedures, and processes in defense acquisition, and to help shape the thought leadership for the acquisition community. These research topics should be considered guidelines to help investigators form their own research questions. Some questions may cross topics and thus appear in multiple research areas.

Potential researchers are encouraged to contact the DAU Director of Research (research@dau.edu) to suggest additional research questions and topics, or with any questions on the topics.

Affordability and Cost Growth

- Define or bound "affordability" in the defense portfolio. What is it? How will we know if something is affordable or unaffordable?

- What means are there (or can be developed) to measure, manage, and control "affordability" at the Program Office level? At the industry level? How do we determine their effectiveness?
- What means are there (or can be developed) to measure, manage, and control "Should Cost" estimates at the Service, Component, Program Executive, Program Office, and industry levels? How do we determine their effectiveness?
- What means are there (or can be developed) to evaluate and compare incentives for achieving "Should Cost" at the Service, Component, Program Executive, Program Office, and industry levels?
- Recent acquisition studies have noted the vast number of programs and projects that don't make it through the acquisition system and are subsequently cancelled. What would systematic root cause analyses reveal about the underlying reasons, whether and how these cancellations are detrimental, and how acquisition leaders might rectify problems?
- Do joint programs—at the inter-Service and international levels—result in cost growth or cost savings compared with single-Service (or single-nation) acquisition? What are the specific mechanisms for cost savings or growth at each stage of acquisition? Do the data lend support to "jointness" across the board, or only at specific stages of a program, e.g., only at Research and Development (R&D), or only with specific aspects, such as critical systems or logistics?
- Can we compare systems with significantly increased capability developed in the commercial market to Department of Defense (DoD)-developed systems of similar characteristics?
- Is there a misalignment between industry and government priorities that causes the cost of such systems to grow significantly faster than inflation?
- If so, can we identify why this misalignment arises? What relationship (if any) does it have to industry's required focus on shareholder value and/or profit, versus the government's charter to deliver specific capabilities for the least total ownership costs?

Industrial Productivity and Innovation

Industry insight and oversight

- What means are there (or can be developed) to measure the level of insight and/or control that government has over subcontractors?
- What means are there (or can be developed) to measure costs of enforcement (e.g., auditors) versus actual savings from enforcement?
- What means are there (or can be developed) to evaluate and compare incentives for subcontractor/supply chain competition and efficiencies?
- What means are there (or can be developed) to evaluate and compare market-based incentives with regulatory incentives?
- How can we perform institutional analyses of the behaviors of acquisition organizations that incentivize productivity?
- What means are there (or can be developed) to evaluate and compare the barriers of entry for SMEs in defense acquisition versus other industrial sectors?
- Is there a way to measure how and where market incentives are more effective than regulation, and vice versa?
- Do we have (or can we develop) methods to measure the effect of government requirements on increased overhead costs, at both government and industrial levels?

- Examine the possibilities to rationalize and balance the portfolio of capabilities through buying larger quantities of common systems/subsystems/components across Defense Agencies and Services. Are there examples from commercial procurement and international defense acquisition that have produced positive outcomes?
- Can principal-agent theory be used to analyze defense procurement realities? How?
- What means are there (or can be developed) to measure the effect on defense acquisition costs of maintaining the industrial base in various sectors?
- What means are there (or can be developed) of measuring the effect of utilizing defense industrial infrastructure for commercial manufacture, particularly in growth industries? In other words, can we measure the effect of using defense manufacturing to expand the buyer base?
- What means are there (or can be developed) to measure the breadth and depth of the industrial base in various sectors that go beyond a simple head count of providers?
- Has change in the industrial base resulted in actual change in output? How is that measured?

Independent Research and Development

- What means do we require to measure the cost-effectiveness or Return on Investment (ROI) for DoD-reimbursed Independent Research and Development (IR&D)?
- Can we properly account for sales and revenues that are products of IR&D?
- Can we properly account for the barriers to entry for SMEs in terms of IR&D?
- Examine industry trends in IR&D, for example, percentage of revenue devoted to IR&D, collaboration with academia. How do they vary by industry sector—in particular, those associated with defense acquisition?
- What means are there (or can be developed) to measure the ROI for DoD-reimbursed IR&D versus directly funded defense R&D?
- What incentive structures will motivate industry to focus on and fund disruptive technologies?
- What has been the impact of IR&D on developing disruptive technologies?

Competition

Measuring the effects of competition

- What means are there (or can be developed) to measure the effect on defense acquisition costs of maintaining an industrial base in various sectors?
- What means are there (or can be developed) for measuring the effect of utilizing defense industrial infrastructure for commercial manufacture, particularly in growth industries? In other words, can we measure the effect of using defense manufacturing to expand the buyer base?
- What means are there (or can be developed) to determine the degree of openness that exists in competitive awards?
- What are the different effects of the two best value source selection processes (tradeoff versus lowest price technically acceptable) on program cost, schedule, and performance?

Strategic competition

- Is there evidence that competition between system portfolios is an effective means of controlling price and costs?
- Does lack of competition automatically mean higher prices? For example, is there evidence that sole source can result in lower overall administrative costs at both the government and industry levels, to the effect of lowering total costs?
- What are long-term historical trends for competition guidance and practice in defense acquisition policies and practices?
- To what extent are contracts awarded noncompetitively by congressional mandate, for policy interest reasons? What is the effect on contract price and performance?
- What means are there (or can be developed) to determine the degree to which competitive program costs are negatively affected by laws and regulations such as the Berry Amendment, Buy American Act, etc.?
- The DoD should have enormous buying power and the ability to influence supplier prices. Is this the case? Examine the potential change in cost performance due to greater centralization of buying organizations or strategies.

Effects of industrial base

- What are the effects on program cost, schedule, and performance of having more or fewer competitors? What measures are there to determine these effects?
- What means are there (or can be developed) to measure the breadth and depth of the industrial base in various sectors, that go beyond a simple head count of providers?
- Has the change in industrial base resulted in actual change in output? How is that measured?

Competitive contracting

- Commercial industry often cultivates long-term, exclusive (noncompetitive) supply chain relationships. Does this model have any application to defense acquisition? Under what conditions/circumstances?
- What is the effect on program cost performance of awards based on varying levels of competition: (a) “Effective Competition” (two or more offers; (b) “Ineffective Competition” (only one offer received in response to competitive solicitation; (c) “Split Awards” versus winner take all; and (d) “Sole Source.”

Improve DoD outreach for technology and products from global markets

- How have militaries in the past benefitted from global technology development?
- How/why have militaries missed the largest technological advances?
- What are the key areas that require DoD focus and attention in the coming years to maintain or enhance the technological advantage of its weapons systems and equipment?
- What types of efforts should DoD consider pursuing to increase the breadth and depth of technology push efforts in DoD acquisition programs?
- How effectively are DoD's global Science and Technology (S&T) investments transitioned into DoD acquisition programs?

- Are managers of DoD's applied R&D (i.e., acquisition program) investments effectively pursuing and using sources of global technology to affordably meet current and future DoD acquisition program requirements? If not, what steps could DoD take to improve its performance in these two areas?
- What are the strengths and weaknesses of DoD's global defense technology investment approach as compared to the approaches used by other nations?
- What are the strengths and weaknesses of DoD's global defense technology investment approach as compared to the approaches used by the private sector—both domestic and foreign entities (companies, universities, private-public partnerships, think tanks, etc.)?
- How does DoD currently assess the relative benefits and risks associated with global versus U.S. sourcing of key technologies used in DoD acquisition programs? How could DoD improve its policies and procedures in this area to enhance the benefits of global technology sourcing while minimizing potential risks?
- How could current DoD/U.S. Government Technology Security and Foreign Disclosure (TSFD) decision-making policies and processes be improved to help DoD better balance the benefits and risks associated with potential global sourcing of key technologies used in current and future DoD acquisition programs?
- How do DoD primes and key subcontractors currently assess the relative benefits and risks associated with global versus U.S. sourcing of key technologies used in DoD acquisition programs? How could they improve their contractor policies and procedures in this area to enhance the benefits of global technology sourcing while minimizing potential risks?
- How could current U.S. Government Export Control system decision-making policies and processes be improved to help DoD better balance the benefits and risks associated with potential global sourcing of key technologies used in current and future DoD acquisition programs?

Comparative studies

- Compare the industrial policies of military acquisition in different nations and the policy impacts on acquisition outcomes.
- Compare the cost and contract performance of highly regulated public utilities with nonregulated "natural monopolies" (e.g., military satellites, warship building).
- Compare contracting/competition practices of DoD with the commercial sector in regard to complex, custom-built products (e.g., offshore oil platforms).
- Compare program cost performance in various market sectors: highly competitive (multiple offerors), limited (two of three offerors), or monopoly?
- Compare the cost and contract performance of military acquisition programs in nations having single "purple" acquisition organizations with those having Service-level acquisition agencies.

Cybersecurity

General questions

- How can we perform analyses of the investment savings associated with institution of robust cybersecurity measures?
- How can we measure the cybersecurity benefits associated with using continuous integration and continuous deployment methodologies?

- How can we cost the discrete elements of cybersecurity that ensure system operational effectiveness within the categories of system functions, mission execution, system performance, and system resilience?
- How can we assess the most effective methodologies for identifying threats quickly, assessing system risk, and developing countermeasures?
- How can we establish a repeatable process for incorporating a continuous Authorization to Operate (ATO) construct for all software-centric acquisition programs?
- How can we articulate cyber risk versus operational risk so Combatant Commands (COCOMs) can be better informed when accepting new software?

Costs associated with cybersecurity

- What are the cost implications of (adding) cybersecurity to a program?
- What are reasonable benchmarks for cybersecurity cost as a percentage of Prime Mission Product (PMP)?
- What are the key cost drivers associated with cybersecurity?
- Is cybersecurity best estimated as a below-the-line common element (similar to Systems Engineering/Program Management or Training) or a PMP element?
- How are risks associated with not incorporating cybersecurity appropriately best quantified/monetized?

Acquisition of Services

Metrics

- What metrics are currently collected and available on services acquisition:
  - Within the Department of Defense?
  - Within the U.S. Government?
  - Outside of the U.S. Government?
- What and how much do these metrics tell us about services acquisition in general and about the specific programs for which the metrics are collected?
- What are the possible metrics that could be used in evaluating services acquisition programs?
  - How many metrics should be used?
  - What is the efficacy of each metric?
  - What is the predictive power of each metric?
  - What is the interdependence (overlap) between metrics?
- How do we collect data for services acquisition metrics?
  - What is being done with the data currently being collected?
  - Are the data being collected on services acquisition reliable?
  - Is the collection process affecting the data collected for services acquisition?
- How do we measure the impact of different government requirements on overhead costs and rates on services contracts?

Industrial base

- What is the right amount of contracted services for government organizations?
    - What are the parameters that affect Make/Buy decisions in government services?
    - How do the different parameters interact and affect government force management and industry research availability?
- What are the advantages, disadvantages, and impacts of capping pass-through costs, and how do they change with the value of the pass-through costs?
- For Base Operations and Support (BOS) contracts, is there a best size? Should large BOS contracts be broken up? What are the parameters that should be considered?
- In the management of large services contracts, what is the best organization? Is the System Program Office a good model? What parameters should be used in evaluating the advantages and disadvantages of an organization to manage large services contracts?
- What effect does strategic sourcing and category management have on small business if the small business is a strategic source or whether the small business is not a strategic source?
- Do the on-ramping and off-ramping requirements of some service contracts have an effect on the industrial base? If so, what are the impacts?

Industry practices

- What private sector business practices, other than maximizing profit, can the government effectively use to incentivize performance and otherwise improve business relationships with vendors?
- What are the best methods for evaluating different incentives to encourage small businesses to participate in government services contracts?
- What potential benefits can the government achieve from long-term supply chain relationships? What are the disadvantages?
- What benefits does industry get from the use of category managers and functional domain experts, and can the government achieve the same benefits?
- How can the government best capture, validate, and use demand management strategies?
- Are current services acquisition taxonomies comprehensive, or can they be improved?

Make/Buy

- What methods can best be used to define the cost value relationship in different classes of service contracts?
- Can we develop a method for determining the “should cost” of different services?
- Can we define and bound affordability of specific services?
- What are the characteristics of “inherently governmental” activities, and how can we evaluate the value of these services based on comparable characteristics in a competitive labor market?

- In services contracts, what are the inherent life-cycle costs, and how do we capture the life-cycle costs in make/buy decision making?
- In the case of government services contracting, what are the factors that contribute to less-than-optimum make/buy decision making?

Category management/strategic sourcing

- What effect does strategic sourcing/category management have on competition?
    - Effects on short term versus long term.
    - Effects on competition outside of the strategic sourcing/category management area of consideration.
- What metrics do different industries use for measuring the effectiveness of their supply chain management?
- Would the centralization of services acquisition contracts have measurable impacts on cost performance? Why or why not?
- What are the fundamental differences between the services taxonomy and the category management taxonomy, and are there means and good reasons to align the two taxonomies?

Contract management/efficacy

- What are the best ways to address the service parts of contracts that include both services and products (goods)?
- In the management of services contracts, what are the non-value-added tasks, and are there realistic ways to reduce the impact of these tasks on our process?
- When funds for services are provided via pass-throughs (i.e., from another organization), how are the requirements tracked, validated, and reviewed?
- Do Undefinitized Contract Actions have an effect on contractor pricing and willingness, or lack of willingness to provide support during proposal analysis?
- For multiaward, Indefinite-Delivery, Indefinite-Quantity (IDIQ)-type contracts, is there a method for optimizing the different characteristics (number of vendors, timelines, on-ramping, off-ramping, etc.) of these contracts?

Policy

- What current government policies inhibit alignment of contractors' approaches with the government's services acquisition programs?

Administrative Processes

- What means are there (or can be developed) to measure the efficiency and effectiveness of DoD oversight, at the Component, Service, and Office of the Secretary of Defense levels?
- What measures are there (or can be developed) to evaluate and compare the costs of oversight versus the cost savings from improved processes?
- What means are there (or can be developed) to empirically establish oversight process metrics as a basis for comparison? Can these be used to establish the relationship of oversight to cost/schedule/performance outcomes?
- What means are there (or can be developed) to study the organizational and governance frameworks, resulting in successful change management?

- To what extent (investment and performance) can scenario/simulation testing improve the delivery of complex projects?
- Is there a comparative statistical divergence between organizational honesty (reality) and contractual relationships (intent) in tendering?
- How does one formulate relational contracting frameworks to better account for and manage risk and liability in a collaborative environment?

Human Capital of Acquisition Workforce

- What means are there (or can be developed) to measure ROI for acquisition workforce training?
- What elements of the Professional Military Education framework can be applied to improve the professionalism of the civilian defense acquisition workforce?
- What factors contribute to the management and successful delivery of modern complex project management, including performance over the project life cycle?
- What behavioral leadership characteristics can be commonly observed in successful complex projects, contrasted against unsuccessful complex projects?
- What is the functional role of talent management in building organizational sustainability, performance, and leadership?
- How do we create incentives in the acquisition workforce (management, career, social, organizational) that provide real cost reductions?

Defense Business Systems

Organizational structure and culture in support of Agile software development methodologies

- At the beginning of the Business Capability Acquisition Cycle (BCAC) process, various steps are used to ensure accurate requirements are thoroughly documented and supported throughout the software development life cycle. How can these documentation requirements and processes be streamlined to support more direct-line communication between the end-user and software engineers? What are the hurdles to implementing these changes and how are they overcome? What are the effects of these changes on the organization or agency?
- Regarding new starts, how can the BCAC be modified specifically to support Agile development? How are these changes advantageous or disadvantageous to the customer and organization? Would these changes be helpful or detrimental to R&D versus a concurrent design and engineering software project?
- Generally, readiness review briefings within the BCAC are used to determine if a project is at an acceptable state to go to the next step in the process. If software is developed and released to production within a single Sprint (potentially every 2 weeks), how are Test Readiness Reviews, Systems Requirements Reviews, and Production Readiness Reviews handled? How have the changes to these events made them more or less relevant?

- How are organizations and agencies structured to support concurrent software design and development? What organizational structure would support R&D and non-R&D information technology (IT) capabilities?
- What steps are used to choose Agile as the default software development process versus any other software development methodology (e.g., Waterfall, Spiral, or Incremental) for your organization? What are the effects on project cost, schedule, and performance?
- Within DoD agencies and military branches, has the adoption of Agile resulted in faster deployment of new IT capabilities to the customer? How is this determined and measured?
- Industry often produces software using Agile. The DoD's BCAC process can produce an abundance of bureaucracy counter to Agile principles. How does hiring a contractor to implement or maintain IT capabilities and introducing Agile software development methods within a BCAC non-Agile process create conflict? How are these conflicts resolved or reconciled?
- How is IT engineering investment and innovation supported throughout DoD? What organizational or cultural aspects of an agency are specific to that support?

Defense Acquisition and Society

- To what extent should the DoD use the defense acquisition process to effectuate various social policies? The existing procurement regime favors a dizzying array of private interests ranging from organized labor; domestic manufacturers and firms located in areas of high unemployment; small businesses, including disadvantaged and women-owned firms; blind, severely handicapped, and prison industries; and, most recently, environmentally friendly vendors. Affirmatively steering the government's business from the open marketplace to preferred providers adds complexity, thus increasing transaction costs throughout the procurement process, which absorbs scarce resources. (Source: IBM Center for the Business of Government, http://www.businessofgovernment.org)
- How significant are the transaction costs resulting from the administration's commitment to transparency (generally, and specifically in the context of stimulus or recovery spending)? In a representative democracy, transparency is critical. But transparency is expensive and time-consuming, and the additional resources required to comply with the recently enhanced disclosure standards remain an unfunded mandate. Thus, the existing acquisition workforce must devote scarce resources to an (admittedly legitimate) end other than the pursuit of value for money or customer satisfaction. Is there an optimal balance or a point of diminishing returns? In other words, at what point does the cost of developing transparent systems and measures exceed the benefits of that transparency? (Source: IBM Center for the Business of Government, http://www.businessofgovernment.org)

Potential authors are encouraged to peruse the DAU Research website (***https://www.dau.edu/library/research/p/Research-Areas***) for information.

We’re on the Web at:
http://www.dau.edu/library/arj

ISSUE 95
JANUARY 2021
VOL. 28 NO. 1

TECHNOLOGY TRUST: SYSTEM INFORMATION IMPACT ON AUTONOMOUS SYSTEMS ADOPTION IN HIGH-RISK APPLICATIONS

Michael G. Anderson and Johnathan C. Mun

As autonomous systems become more capable, end users must make decisions about how and when to deploy such technology. The use and adoption of a technology to replace a human actor depends on its ability to perform a desired task and on the user's experience-based trust that it will do so. The development of experience-based trust in autonomous systems is costly, and it carries a high risk of physical harm to operators. This work focuses on identifying a methodology for technology discovery that reduces the need for experience-based trust and contributes to increased adoption of autonomous systems. The main research hypothesis is that manipulating the presentation of technical information can influence the initial formation of trust by functioning as a surrogate for experience-based trust, and that trust in technology can be captured through an anthropomorphic hierarchy of system attributes.

DOI: *https:// doi.org/10.22594/10.22594/dau.19-841.28.01*
Keywords: *Technology Trust, Autonomous Systems, Technology Risk Metrics, Anthropomorphic Hierarchy, Technology System Attributes*

Image designed by Nicole Brate

The use of technology by the Department of Defense (DoD) depends on its ability to perform a desired task. Many issues associated with trust in technology are increasing in importance as the U.S. military begins to acquire and deploy autonomous systems. To ensure the effective adoption of new innovations in technology, researchers need to establish a system of metrics that justify a level of technology trust. This article has the explicit goal of investigating and recommending trust metrics by applying advanced analytical methodologies to increase the speed and effectiveness of the adoption of new technologies. This investigation proceeds by participating in an evaluation of technologies for use in evolving, high-risk military applications. The trust metrics are measured in terms of the technology acceptance versus system control.

Technology Trust

The 2016 Defense Science Board report on autonomy (David & Nielsen, 2016) identifies trust as central to DoD's success in the broader adoption of autonomy. This article studies the potential for introducing trust metrics on the evaluation and selection of technologies. The work participates in an ongoing assessment of autonomous systems for use in high-risk military applications throughout fiscal year 2019. A model is developed that optimizes the cognitive impacts of these trust metrics as they relate to the technology selection and adoption process. The approach will be extensible and can be adopted into private industry.

"This article has the explicit goal of investigating and recommending trust metrics by applying advanced analytical methodologies to increase the speed and effectiveness of the adoption of new technologies.

Research Problem

The recent increase in the use and deployment of sophisticated technologies by other countries is a disruptive threat to the United States' technological superiority. The rapidly changing technology landscape requires DoD laboratories to increase the speed at which they adopt new technologies (David & Nielsen, 2016). With declining budgets in research, it is imperative that the DoD establish new methods for rapidly adopting and effectively deploying new and emerging technologies whenever possible. The goal of this article is to establish and measure a comprehensive trust metric for individual components of technologies, such as autonomous systems used in high-risk military applications. The development of a trust metric serves two purposes: first, as a surrogate for experience-based trust by contributing to the formation of initial-trust and, second, as a collection tool for capturing experience-based trust data.

This work emerges from the general question, "Can humans develop trust in complex systems without direct experience and a complete understanding of the technology?" Theories in anthropomorphism (assigned human attributes to technology) and system hierarchy hold promise in their ability to reduce complexity and improve the acceptance of complex systems. Thus, the specific research question posited by this article is "How does system information affect the adoption of autonomous systems used in high-risk military applications?"

To that end, this study attempts to answer the following questions:

1. How does the anthropomorphic categorization and presentation of technology affect the development of trust in technologies used in high-risk military applications? The constructs researched include:
    - Hardware
    - Algorithms
    - Links

2. How do varying levels of system control affect the development of trust in technologies used in high-risk military applications? The constructs researched include:
    - Perceived ease of use
    - Perceived usefulness
    - Intent to use
3. Does a causal relationship exist between an anthropomorphic hierarchy of system information and the acceptance of autonomous systems?

Literature Review

This article was initiated through informal interviews that attempted to identify the factors that contribute to the use of technology in high-risk environments. The participants were a small group of military personnel who have deployed with technology that posed great risk of physical harm should it fail. A majority of this group experienced significant injury due to the failure of technology, and the potential for bias was noted. A series of open-ended questions were provided to discuss what the users did or did not like about using technology in high-risk scenarios. The initial coding of interviews revealed the following three exploratory research themes:

1. Hands-on experience with technology is critical for establishing trust, and a team-based reputation for a technology is as important as personal experience.
2. Personal investment in a mission is key to learning and accepting new and complex technology.
3. Users operating in high-risk environments favor simple technology containing only the features needed to accomplish a mission and may reject new and complex technology in favor of older and more trusted systems.

These themes all have implications for the adoption of autonomous systems within the DoD. Advanced robotic systems have the ability to improve performance in a number of military roles while reducing risk to humans, and it is important to understand how to improve the adoption of such systems within the DoD. This initial research focused on technology in dangerous environments and reveals that adoption is highly dependent on the ability of the user to obtain the knowledge necessary to develop trust. This theme led to our initial literature review on understanding trust and how it applies to technology adoption.

Trust

Castelfranchi and Falcone (2010) review 72 definitions of what it means to know something well enough to trust, and their work found a great deal of confusion and ambiguity surrounding the use of this term. As a result, a limited unity on a definition of trust is accepted across research disciplines. However, two themes emerged from the many definitions of trust: (a) the basic premise of trust involves two actors, and (b) trust is a relationship in which one entity relies on someone, or something, based on a given criterion.

> **Advanced robotic systems have the ability to improve performance in a number of military roles while reducing risk to humans, and it is important to understand how to improve the adoption of such systems within the DoD.**

Adams and Webb (2002) describe two broad processes of developing trust between two persons. The first is defined as "experience-based trust," which develops through repeated engagements, and the second is called "reason-based trust," which develops in the absence of direct experience.

Rempel et al. (1985) address three factors that influence the development of experience-based trust: competence, benevolence, and integrity. Their work also discusses the significance of the mental motivation behind the desire to establish a relationship and finds it strongly correlated to the factors that influence trust. Their work confirms the second exploratory research theme that emphasizes the importance of personal investment.

Technology

The past research on interpersonal trust applies in many ways to trust in technology. This study examined literature that contributes to the development of a methodology of technology discovery leading to trust in technology. The potential for integrating interpersonal trust research into technology trust was discussed by McKnight et al. (2011). This research found that interpersonal trust is based on a trustor's expectations and reliance on a trustee to perform as expected through benevolence, even though the trustee possesses the volition to choose to do what is right or what is wrong. Because technology does not possess volition (ability to choose), Knight observed, some researchers went as far as to dismiss the idea of trust in technology as irrelevant.

A theory relevant to measuring and characterizing trust is found in the technology acceptance model (TAM) developed by Fred Davis in the late 1980s. This model plays a significant role in the majority of research investigating the factors and attributes that influence the acceptance of a technology. Venkatesh and Bala (2008) present the TAM's ability to predict and measure individual adoption and use of technology. The TAM assesses the behavioral intention to use a technology through two constructs: perceived usefulness (PU), which is defined as the extent to which a person believes that using a technology will enhance his or her job performance; and perceived ease of use (PEOU), which is defined as the degree to which a person believes that using a technology will be free of effort. These two variables are used to establish a relationship between external influences and potential system usage (Gefen et al., 2003).

> **In some military scenarios, developing experience-based trust presents high levels of risk for physical injury and harm.**

Tétard and Collan (2009) address the challenges of adopting new technology for high-risk scenarios in their work on the lazy-user, also called efficient-user theory. This theory states that users select the technology that demands the least amount of effort to do the job. The application of this theory places technology users at a disadvantage, particularly in high-risk military applications where our exploratory research indicates that users are known to avoid more capable technology for systems that are easier to understand. If an experience-based proxy can improve the accuracy of developing trust through increased technology literacy, it may lead to increased acceptance of more complex and capable technologies, thereby reducing the influence of the efficient-user theory. This leads to our third theme identified in exploratory research, "Users operating in high-risk environments favor simple technology containing only the features needed to accomplish a mission and may reject new and complex technology in favor of older and more trusted systems."

Experimental Design

The previous section discussed how a "trust-discovery" methodology could contribute to improved understanding of how people develop trust in machines. This understanding could lead to the development of a technology-literate workforce capable of accurately assessing new technology for a given operational scenario. The literature review strongly suggests that the manipulation of system information may influence technology trust.

This experiment investigates the formation of trust in technology and how it influences the adoption of autonomous systems for use in high-risk military applications. The formation of trust in technology is governed by two constructs: reason-based trust and experience-based trust. Existing literature presents the case for increased accuracy in technology selection through the development of experience-based trust. However, the development of experience-based trust is financially burdensome and takes much longer to form than reason-based trust. In some military scenarios, developing experience-based trust presents high levels of risk for physical injury and harm.

Experiment Introduction

This experiment is designed to research the manipulation of system information and study any influence on the formation of reason-based trust in autonomous systems used in high-risk military applications. The desired outcome of this work is the identification of causal relationships between system attributes and technology acceptance that can replace some of the burden required to develop experience-based trust. In other words, can a reason-based trust method be used to replace experience-based methods?

The experiment is designed in two-phases. Phase one is a group-administered experimental survey that employs manipulations of multiple theories of system information and technology acceptance to collect data on reason-based trust in systems with varying levels of system control. Phase two consists of administering the same survey, following extensive field testing and experimentation of the phase one systems, to collect data on experience-based trust. Trust is measured as an "intent to use" and based on responses to the TAM.

Anthropomorphism

The complexity of modern technology makes it difficult to establish generalizable categories capable of capturing system information and functioning as a proxy for experience-based trust. One area of research relevant to the establishment of technology categories involves anthropomorphism—the attribution of human traits to nonhuman entities to increase a trustor's ability to understand and accept complex technology.

Schaefer et al. (2016) and Waytz et al. (2014) identified anthropomorphism as a system factor that contributes to the development of human trust for robots. Reported cases in Pak et al. (2012) examine where the tendency to anthropomorphize technology leads to situations in which humans give a higher degree of trust to a technology than is warranted. The inverse of this situation also exists in the development of a lack of trust in a human teammate caused by the introduction of technology with more capability and reliability. In this experiment, anthropomorphism is assessed for its ability to influence technology trust.

System Hierarchy

In this work, we hypothesize that statistically significant differences will result in technology trust by introducing system information through an anthropomorphic hierarchy. Over a period of 10 years, the authors of this article provided instruction to third-year university engineering students on the topics of digital design and computer architecture. The predominant challenge reported by students in end-of-year course evaluations was difficulty synthesizing the highly complex components of a computer into a usable system. Based on student feedback, an anthropomorphic hierarchy was developed to structure the components of computer architecture to a more familiar format. This hierarchy provided students with the context needed to understand how the pieces of a computer function together to create a whole system. The work resulted in improved student ability to describe a computer from the elemental circuits up to the most advanced concepts of computer engineering such as compilers and operating systems.

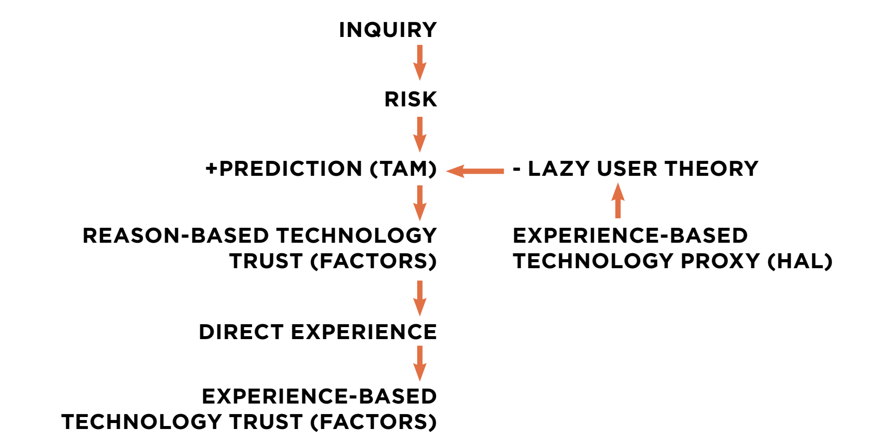
FIGURE 1. ANTHROPOMORPHIC TECHNOLOGY HIERARCHY

HARDWARE Processor Circuits	⟷	**BODY** Brain Nervous System
ALGORITHMS Data OS	⟷	**THOUGHT** Knowledge Wisdom
LINKS Inputs Outputs	⟷	**SENSES** Sight, Smell, Hear, Taste, Touch

To establish an invariant system hierarchy for use in measuring both reason-based and experience-based technology trust, we introduce the anthropomorphic categories of hardware, algorithms, and links (HAL) as illustrated in Figure 1.

To increase the value of this hierarchy, we further conceptualized a HAL score of trustworthiness. The values of each HAL metric are proposed to range from 0 to 100, and lead to an equally weighted maximum score (indicating most trustworthy) of 300. Future research is needed to identify the weights for the HAL score to accurately reflect the overall impact on trust. Since field experimentation has not been conducted, we introduce the HAL categories in the experiment without any associated "score." The HAL hierarchy is used to organize system information and provide a framework for future experience-based trust proxy research as shown in Figure 2.

FIGURE 2. HAL SCORE EXPERIMENTAL MODEL



The Experiment

The experiment was conducted using factorial design. Two independent variables are used: system information and level of system control. The level of system information is varied between two conditions: Less Information (Less Info) and More Information (More Info). The Less Info condition presented system information using vendor-provided datasheets. The More Info condition introduced the same system information but carefully organized under the HAL hierarchy. The system control is varied between three levels: Direct, Remote, and Autonomous. Two dependent variables are used: (a) level of risk associated with the loss of a system attribute, and (b) trust, as measured by the TAM.

Procedures

This article uses data provided by an ongoing external experiment (Appendix A). Phase one of that experiment is a group-administered survey that employs manipulations of system presentation. The phase one surveys were administered to a military unit responsible for the operational assessment of the three technologies (Direct-Control, Remote-Control, Autonomous). Each of the two phase-one surveys (Less Info, More Info) was completed by 20–25 subjects. Demographics such as age, military specialty, and exposure to similar technologies were captured to assess internal validity. Phase two consists of administering the same survey following extensive field testing by 15 subjects from the same military unit tasked in phase one. The phase two results are captured to provide external validity for the phase one results.

> **“One area of research relevant to the establishment of technology categories involves anthropomorphism—the attribution of human traits to nonhuman entities to increase a trustor's ability to understand and accept complex technology.**

Phase one of the experiment was conducted in a controlled and distraction-free classroom environment and involved the participation of two randomly selected groups of active-duty military tasked with a new high-risk mission. The two groups participated in separate morning sessions lasting 1 hour each. The start time for the second session was immediately following completion of the first session. Both groups were provided with identical overviews of a high-risk military scenario that would be completed by deploying three technology systems rather than human operators. The independent variable “system presentation” was manipulated between the first and second groups as Less Info and More Info. The second independent variable—“system control”—is provided to each participant in both groups in the form of the three different technologies. Appendix B lists the details of the survey questions as well as variable names and codes that are presented in the next section.

Results

Data Are Only Somewhat Normal

The research questions we are attempting to answer in this section follow.

- *Are the data considered sufficiently normal?*
- *Can we apply conventional parametric methods, or do we need more advanced nonparametric methods to analyze the data?*

Tables 1 and 2 show the results from randomly selected variables. The results are mixed. This means that although these are statistically significant in some areas, they may not be practically significant enough to justify normality.

TABLE 1. VAR1 NORMALITY TESTS

Best-Fitting Distributions: VAR1					
Rank	**Akaike**	**Anderson**	**Kolmogorov**	**Kuiper's**	**Schwartz**
1	Cosine	**Normal**	GenPareto	**Normal**	Cosine
2	Lognml3Arith	Logistic	Weibull	Logistic	Lognml3Arith
3	Weibull	TDist	GumbelMin	TDist	Weibull
4	**Normal**	Weibull	Triangular	Cosine	**Normal**
5	Gamma	GumbelMax	**Normal**	Weibull	Gamma
MAPE %					
1	19.0136%	19.0915%	N/A	19.4214%	19.0136%
2	19.3421%	19.2969%	19.5824%	19.4214%	19.3421%
3	19.3665%	19.4732%	24.8250%	19.4370%	19.3665%
4	19.4297%	20.0214%	21.2316%	19.4732%	19.4297%
5	19.4575%	21.8529%	19.6539%	19.6312%	19.4575%

Best Fit Rank : 5
Fit Name : Normal
Kolmogorov-Smirnov Statistic : 0.153350
Mean : 3.721371
Sigma : 1.250896
p value : **0.614791**
Actual to Theoretical Four Moments :
3.739130 1.053884 -0.190064 -1.168769;
3.721371 1.250896 0.000000 0.000000;

Nonparametric Shapiro-Wilk Test for Normality
(Royston Algorithm)
Shapiro-Wilks : 0.865946
SW *P*-value : **0.005368**
Null hypothesis: The data are normally distributed

Note. MAPE = Mean Absolute Percentage Error; VAR = Variable

TABLE 2. VAR105 NORMALITY TESTS

Best-Fitting Distributions: VAR105					
Rank	**Akaike**	**Anderson**	**Kolmogorov**	**Kuiper's**	**Schwartz**
1	Cosine	TDist	Weibull	TDist	Cosine
2	Uniform	Gamma	Uniform	GumbelMax	Uniform
3	Triangular	Normal	GumbelMax	Weibull	Triangular
4	Weibull	GumbelMin	LognmlArith	Laplace	Weibull
5	TDist	Logistic	**Normal**	GumbelMin	TDist
MAPE %					
1	20.2105%	20.4966%	25.4875%	20.4966%	20.2105%
2	20.3731%	21.5868%	19.8248%	21.8717%	20.3731%
3	20.4260%	22.5328%	25.6700%	22.2282%	20.4260%
4	20.4405%	22.6221%	23.5800%	23.3391%	20.4405%
5	20.4966%	22.9440%	20.7503%	24.0731%	20.4966%

Best Fit Rank : 5
Fit Name : Normal
Kolmogorov-Smirnov Statistic : 0.175000
Mean : 3.647780
Sigma : 1.105387
p value : **0.531299**
Actual to Theoretical Four Moments :
3.550000 1.050063 -0.146220 -1.072526;
3.647780 1.105387 0.000000 0.000000;

Nonparametric Shapiro-Wilk Test for Normality
(Royston Algorithm)
Shapiro-Wilks : 0.880332
SW *P*-value : **0.017937**
Null hypothesis: The data are normally distributed

We conclude that:

- The survey data are only somewhat normally distributed under certain circumstances, and we cannot state complete normality to fully justify standard modeling approaches.
- The data are ordinal and quasi-interval, with limited truncation between 1 and 5, and are limited to between 19 and 23 observations.
- Both parametric and nonparametric methods will be used, and this mixed approach is therefore justified.

Therefore, going forward, both parametric and nonparametric tests will be conducted whenever appropriate, and their results will be compared for corroboration.

Hotelling's T-Squared Distribution in Statistics

The research questions we are attempting to answer in this section are:

- *When all the survey responses for each subgroup are taken together as a whole, are there statistical differences in the responses?*
    - *Are the perceptions of the Direct-Control system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*
    - *Are the perceptions of the Remote-Control system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*
    - *Are the perceptions of the Autonomous system different when Less Info is provided, More Info is available, or a hands-on experiment is conducted?*

Tables 3 and 4 show a sampling of the results from the Hotelling T^2 test. The null hypothesis is that no statistical differences result from using a parametric Hotelling T^2 test, where all of the survey responses in each of the subcategories, when taken together, simultaneously do not indicate that any differences are discernible between the two groups tested.

TABLE 3. HOTELLING

Hotelling Test Groups	*P*-value	Variables Tested
Less Info vs. More Info	0.5863	VAR1:VAR14 vs. VAR101:VAR114
Less Info vs. More Info	0.7998	VAR15:VAR25 vs. VAR115:VAR125
Less Info vs. More Info	0.3515	VAR26:VAR36 vs. VAR126:VAR136
Less Info vs. More Info	0.2084	VAR37:VAR47 vs. VAR137:VAR147
Less Info vs. More Info	0.7095	VAR48:VAR51 vs. VAR148:VAR151
Less Info vs. More Info	0.4475	VAR52:VAR54 vs. VAR152:VAR154
Less Info vs. Experiment	0.0000	VAR15:VAR25 vs. VAR415:VAR425
Less Info vs. Experiment	0.0144	VAR26:VAR36 vs. VAR426:VAR436
Less Info vs. Experiment	0.0793	VAR37:VAR47 vs. VAR437:VAR447
More Info vs. Experiment	0.0000	VAR115:VAR125 vs. VAR415:VAR425
More Info vs. Experiment	0.1215	VAR126:VAR136 vs. VAR426:VAR436
More Info vs. Experiment	0.3232	VAR137:VAR147 vs. VAR437:VAR447

TABLE 4. HOTELLING FOR GROUP A6 VS. GROUP B6

VAR52; VAR53; VAR54 vs. VAR152; VAR153; VAR154
D1, D2, D3 vs. D1, D2, D3
Hotelling T-Square: Two Independent Variable
Equal Variance with Multiple Related Measures

Hotelling T2 2.85372
F Statistic 0.90484
P-value 0.44753
Null hypothesis tested is that there is zero difference between all the related variables compared across the two groups.

Covariance GROUP 1			
	VAR52	**VAR53**	**VAR54**
VAR52	0.00000	0.00000	0.00000
VAR53	0.00000	15.56621	13.61660
VAR54	0.00000	13.61660	17.73419
Covariance GROUP 2			
VAR152	0.23947	-0.26711	-1.00395
VAR153	-0.26711	9.35461	10.39408
VAR154	-1.00395	10.39408	15.11776
Covariance POOLED			
VAR152	0.11098	-0.12378	-0.46524
VAR153	-0.12378	12.68766	12.12324
VAR154	-0.46524	12.12324	16.52170

We conclude that:

- The results indicate that no perceivable differences exist between the Less Info and More Info groups in the Pre-experiment stage (comparing all subelements of group A to all subelements of group B).
- When comparing the Less Info Pre-experiment group against the Post-experiment group, we see a statistically significant difference among the responses. The trend seems to be that more difference is shown between group A (Less Info) and group C (Post-experiment) than between group B (More Info) and group C.
- In addition, the significance is higher for Direct-Control systems than Remote-Control systems, which in turn, is more significant than Autonomous systems.

Bonferroni Test

The research question we are attempting to answer in this section is:

- *When all the survey responses for each subgroup are taken individually, are there statistical differences in the responses?*

Table 5 shows a sampling of the results from the Bonferroni test. While the previous parametric Hotelling test looks at all subcategories in each group compared to all the subcategories in the second group, the parametric Bonferroni test compares one pair of the subgroups at a time, like the *t*-test. The difference is the Bonferroni test accounts for the added degrees of freedom with multiple simultaneous pairwise tests.

TABLE 5. BONFERRONI TEST

Simultaneous Confidence Intervals
Mean Difference of Null is 0
Model Inputs:

	VAR48; VAR 148; C1	**VAR49; VAR 149; C2**	**VAR50; VAR 150; C3**	**VAR51; VAR 151; C4**
Mean Difference	0.0522	-0.3283	-0.0152	-0.0457
Variance Group 1	1.6917	1.5336	0.8024	0.5850
Variance Group 2	1.4105	0.5553	0.3658	0.5553
Pooled Variance	1.2496	1.0393	0.7746	0.7558
F-Critical	2.6190	2.6190	2.6190	2.6190
T-Critical	3.3620	3.3620	3.3620	3.3620
Standard Error	0.3820	0.3178	0.2368	0.2311
Lower Confidence	-1.2323	-1.3966	-0.8115	-0.8225
Upper Confidence	1.3366	0.7401	0.7810	0.7312
Within Confidence?	Yes	Yes	Yes	Yes
Bonferroni Critical	2.6127	2.6127	2.6127	2.6127
Lower Confidence	-1.4760	-1.5993	-0.9626	-0.9700
Upper Confidence	1.5803	0.9428	0.9321	0.8786
Within Confidence?	Yes	Yes	Yes	Yes

Null hypothesis: The individual expected differences are equal to zero.

We conclude that:

- In all the tests, we did not detect any statistical significance, and find that all subgroups are statistically identical. This implies that additional testing is required.

The Three Systems Are Perceived Differently

The research question we are attempting to answer in this section is:

- *Are the three systems statistically different in their main characteristics?*

Forty-three separate Single Variable Multiple Treatment ANOVA models were run. Table 6 shows the statistically significant results from the ANOVA models. Out of the 43 models, 21 show statistical significance. ANOVA tests each of the survey questions in each of the three systems independently. For example, when testing VAR20, VAR31, VAR42, we see that at least one or more of these three variables are statistically different from one another.

TABLE 6. ANOVA I

ANOVA	*P*-value
VAR20; VAR31; VAR42	0.0008
VAR21; VAR32; VAR43	0.0903
VAR120; VAR131; VAR142	0.0264
VAR124; VAR135; VAR146	0.0362
VAR229; VAR240; VAR251	0.0000
VAR230; VAR241; VAR252	0.0000
AR231; VAR242; VAR253	0.0000
VAR232; VAR243; VAR254	0.0002
VAR233; VAR244; VAR255	0.0601
VAR237; VAR248; VAR259	0.0004
VAR238; VAR249; VAR260	0.0000
VAR239; VAR250; VAR261	0.0000
VAR266; VAR276; VAR286	0.0000
VAR267; VAR277; VAR287	0.0285
VAR268; VAR278; VAR288	0.0003
VAR269; VAR279; VAR289	0.0020
VAR270; VAR280; VAR290	0.0000
VAR271; VAR281; VAR291	0.0000
VAR272; VAR282; VAR292	0.0351
VAR273; VAR283; VAR293	0.0002
VAR274; VAR284; VAR294	0.0000

We conclude that:

- The ANOVA results support the results from the Hotelling T^2 tests, where we see that group A is statistically significantly different than group B and group C; and group B is statistically significantly different than group C.

The ANOVA test looks at the individual questions within each of these groups to identify which questions returned different responses for each of the systems in the three different testing environments (Pre-experiment less data, Pre-experiment more data, and Post-experiment).

The Three Surveys Provide New Significantly Valuable Information

The research question we are attempting to answer in this section is:

- *Do the added information and hands-on experimentation provide additional value-added insights?*

Thirty-three separate Single Variable Multiple Treatment ANOVA models were also run to test the individual questions among the three systems among the three groups (i.e., for each of the survey questions, if each of the three systems has similarities or differences among the Pre-experiment Less Info, Pre-experiment More Info, and Post-experiment groups). Table 7 shows the statistically significant results from the ANOVA models. Out of the 33 models, 16 show statistical significance ($\alpha = 0.05$).

TABLE 7. ANOVA II

Model	*P*-value
ANOVA on VAR15; VAR115; VAR415	0.0000
ANOVA on VAR16; VAR116; VAR416	0.0000
ANOVA on VAR17; VAR117; VAR417	0.0000
ANOVA on VAR18; VAR118; VAR418	0.0000
ANOVA on VAR18; VAR118; VAR418	0.0008
ANOVA on VAR20; VAR120; VAR420	0.0003
ANOVA on VAR21; VAR121; VAR421	0.0001
ANOVA on VAR22; VAR122; VAR422	0.0000
ANOVA on VAR23; VAR123; VAR423	0.0001
ANOVA on VAR24; VAR124; VAR424	0.0000
ANOVA on VAR28; VAR128; VAR428	0.0232
ANOVA on VAR29; VAR129; VAR429	0.0157
ANOVA on VAR31; VAR131; VAR431	0.0114
ANOVA on VAR35; VAR135; VAR435	0.0089
ANOVA on VAR38; VAR138; VAR438	0.0472
ANOVA on VAR43; VAR143; VAR443	0.0324

We conclude that:

- Direct-Control systems tend to benefit the most from the knowledge gained from additional information and hands-on experimentation.
- Remote-Control systems tend to benefit somewhat from the knowledge gained from additional information and hands-on experimentation.
- Autonomous systems tend to benefit the least from the knowledge gained from additional information and hands-on experimentation, and in fact, the additional work performed contributes added insights to only 18% of the cases.

> **“ The formation of trust in technology is governed by two constructs: reason-based trust and experience-based trust.**

The Three Systems Are Statistically Different with No Intervening Variables

The research questions we are attempting to answer in this section are:

- *Will different users of the technology with different backgrounds affect the results? That is, are there any controllable or blocking variables that need additional attention?*

Using the ANOVA with Blocking Variables model, we see the results in Table 8. In the experiment, the active-duty military either had experience with similar technology or they did not. The ANOVA test is run with blocking or controlling the user background.

TABLE 8. ANOVA WITH RANDOMIZED BLOCKS

Model Inputs:
VAR296; VAR297; VAR298
SUS(A), SUS(B), SUS(C)

ANOVA Randomized Blocks Multiple Treatments

	DF	SS	MS	F Stat	*P*-value
Block Factor (Row)	18	4384.65	243.59	1.5282	0.1367
Treatment Factor (Column)	2	11369.96	5684.98	5.6650	**0.0000**
Error	36	5738.38	159.40		
Total	56	21492.98			
F Critical (Treatment) @ 0.01	5.247893				
F Critical (Blocking) @ 0.01	2.479730				

Note. SUS = System Usability Score (for systems A, B, and C).

We conclude that:

- The treatment factor indicates that statistically significantly different results are shown among the three systems, but whether a soldier has experience with similar technology does not affect the results.

Nonparametric Kruskal–Wallis

The research question we are attempting to answer in this section is:

- *Does a nonparametric approach yield different results than a parametric model?*

Table 9 shows the results from the nonparametric Kruskal–Wallis test. As discussed, this test is the nonparametric equivalence of the ANOVA. Researchers use it to confirm the results of the ANOVA.

TABLE 9. ANOVA AND KRUSKAL-WALLIS I

VARIABLES TESTED	ANOVA	K-W
VAR20; VAR31; VAR42	0.0008	0.0008
VAR21; VAR32; VAR43	0.0903	0.0116
VAR120; VAR131; VAR142	0.0264	0.0057
VAR124; VAR135; VAR146	0.0362	0.0317
VAR229; VAR240; VAR251	0.0000	0.0000
VAR230; VAR241; VAR252	0.0000	0.0000
VAR231; VAR242; VAR253	0.0000	0.0000
VAR232; VAR243; VAR254	0.0002	0.0000
VAR233; VAR244; VAR255	0.0601	0.0851
VAR237; VAR248; VAR259	0.0004	0.0000
VAR238; VAR249; VAR260	0.0000	0.0000
VAR239; VAR250; VAR261	0.0000	0.0248
VAR266; VAR276; VAR286	0.0000	0.0000
VAR267; VAR277; VAR287	0.0285	0.0239
VAR268; VAR278; VAR288	0.0003	0.0022
VAR269; VAR279; VAR289	0.0020	0.0162
VAR270; VAR280; VAR290	0.0000	0.0000
VAR271; VAR281; VAR291	0.0000	0.0000
VAR272; VAR282; VAR292	0.0351	0.0208
VAR273; VAR283; VAR293	0.0002	0.0007
VAR274; VAR284; VAR294	0.0000	0.0000

We conclude that:

- Comparable to the ANOVA (from Table 6), the Kruskal–Wallis results show that out of the 43 models, the same 21 combinations have statistical significance.

Table 10 shows the additional results from the nonparametric Kruskal–Wallis test. Similar to the ANOVA, the Kruskal–Wallis shows that out of the 33 models, the same 16 combinations show statistical significance.

TABLE 10. ANOVA AND KRUSKAL-WALLIS II

	ANOVA	KW
ANOVA & KW on VAR15; VAR115; VAR415	0.0000	0.0000
ANOVA & KW on VAR16; VAR116; VAR416	0.0000	0.0000
ANOVA & KW on VAR17; VAR117; VAR417	0.0000	0.0000
ANOVA & KW on VAR18; VAR118; VAR418	0.0000	0.0000
ANOVA & KW on VAR19; VAR119; VAR419	0.0008	0.0015
ANOVA & KW on VAR20; VAR120; VAR420	0.0003	0.0000
ANOVA & KW on VAR21; VAR121; VAR421	0.0001	0.0003
ANOVA & KW on VAR22; VAR122; VAR422	0.0000	0.0000
ANOVA & KW on VAR23; VAR123; VAR423	0.0001	0.0000
ANOVA & KW on VAR24; VAR124; VAR424	0.0000	0.0000
ANOVA & KW on VAR28; VAR128; VAR428	0.0232	0.0128
ANOVA & KW on VAR29; VAR129; VAR429	0.0157	0.0127
ANOVA & KW on VAR31; VAR131; VAR431	0.0114	0.0085
ANOVA & KW on VAR35; VAR135; VAR435	0.0089	0.0008
ANOVA & KW on VAR38; VAR138; VAR438	0.0472	0.0631
ANOVA & KW on VAR43; VAR143; VAR443	0.0324	0.0614

The Data Are Reliable and Valid

The research question we are attempting to answer in this section is:

- *Are the collected data reliable and valid for the research?*

The Interrater Reliability Test with Interclass Correlation (ICC) tests were run to determine if the data received were statistically reliable (Table 11). As mentioned, the ICC tests the reliability of the users' ratings by comparing the variability of all the ratings of the same subject to the total variation across all ratings and all users simultaneously. A high ICC indicates a high level of reliability (Mun, 2018).

TABLE 11. ICC AND RELIABILITY ANALYSIS

Intercorrelation ICC Reliability Measures (ICC)		
Pre-Experiment Less Info	**ICC**	***P*-value**
A1:: VAR1:VAR14	0.3544	0.0000
A2:: VAR15:VAR25	0.2886	0.0000
A3:: VAR26:VAR36	0.2302	0.0000
A4:: VAR37:VAR47	0.2692	0.0000
A5:: VAR48:VAR51		
Pre-Experiment More Info	**ICC**	***P*-value**
B1:: VAR101:VAR114	0.3207	0.0000
B2:: VAR115:VAR125	0.2568	0.0000
B3:: VAR126:VAR136	0.2528	0.0000
B4:: VAR137:VAR147	0.2975	0.0000
B5:: VAR148:VAR151	0.1581	0.0016
Post Experiment	**ICC**	***P*-value**
VAR201:VAR214	0.5067	0.0000
VAR215:VAR228	0.4584	0.0000
VAR229:VAR239	0.6709	0.0000
VAR240:VAR250	0.2593	0.0000
VAR251:VAR261	0.2200	0.0000
VAR262:VAR265	0.3146	0.0000
VAR266:VAR275	0.6925	0.0000
VAR276:VAR285	0.2264	0.0000
VAR286:VAR295	0.2328	0.0000
VAR296:VAR298	0.6081	0.0000

We conclude that:

- The data show statistical significance, and we conclude that the collected data are reliable and valid for the research.
- The ICC ranges from 0.1581 to 0.3544 for the Pre-experiment stage for both Less Info and More Info, compared to a range from 0.2200 to 0.6925 for the Post-experiment results. In other words, the more hands-on experimentation, the higher the validity of the collected data.

The Systems Are Independent and Uncorrelated

The research question we are attempting to answer in this section is:

- *Are the three systems somehow correlated in terms of their value to the warfighter?*

Table 12 shows a sampling of the results from the linear and nonlinear correlation matrices.

TABLE 12. LINEAR AND NONLINEAR CORRELATION MATRIX

Linear Correlation			
	VAR296	**VAR297**	**VAR298**
VAR296	1.000000	0.234553	0.279342
VAR297	0.234553	1.000000	0.065035
VAR298	0.279342	0.065035	1.000000
Linear Correlation *p*-Value			
VAR296	0.000000	0.333765	0.246782
VAR297	0.333765	0.000000	0.791381
VAR298	0.246782	0.791381	0.000000
Nonlinear Correlation			
VAR296	1.000000	0.206909	0.265491
VAR297	0.206909	1.000000	0.090518
VAR298	0.265491	0.090518	1.000000

We conclude that:

- It seems that very little correlation exists among the three final scores of the systems.

The results and conclusion make sense, as there should be very little relationship among the Direct-Control, Remote-Control, and Autonomous systems, especially when they are tested independently and at different times.

Each Level of Experimentation Yields Valuable Actionable Intelligence

The research questions we are attempting to answer in this section are:

- *Within each experimentation stage, are the three systems perceived to be different (Direct-Control vs. Remote-Control, Direct-Control vs. Autonomous, and Remote-Control vs. Autonomous systems)?*

- *Between the three levels of experimentation (Less Info, More Info, Live Experiments), are each of the subsections of the technology considered similar or different?*

Tables 13 and 14 show a summary of the results from the relevant *T*-tests and Mann–Whitney (MW) tests. Table 13 shows the results that answer the first question above whereas Table 14 answers the second research question above.

TABLE 13. PARAMETRIC T-TEST AND NONPARAMETRIC MANN–WHITNEY TEST I

Direct vs. Remote	T-Test *P*-value	MW *P*-value	Direct vs. Autonomous	T-Test *P*-value	MW *P*-value	Remote vs. Autonomous	T-Test *P*-value	MW *P*-value
VAR20; VAR31	0.4306	0.4388	VAR20; VAR42	**0.0008**	**0.0013**	VAR31; VAR42	**0.0011**	**0.0016**
VAR21; VAR32	0.1783	0.2914	VAR21; VAR43	**0.0146**	**0.0199**	VAR32; VAR43	0.1100	0.0782
VAR120; VAR131	0.1728	0.1584	VAR120; VAR142	**0.0025**	**0.0043**	VAR131; VAR142	**0.0442**	0.0684
VAR120; VAR131	**0.0069**	**0.0180**	VAR124; VAR146	**0.0496**	0.0902	VAR135; VAR146	0.1984	0.2030
VAR229; VAR240	**0.0008**	**0.0011**	VAR229; VAR251	**0.0000**	**0.0000**	VAR240; VAR251	**0.0469**	**0.0362**
VAR230; VAR241	**0.0000**	**0.0000**	VAR230; VAR252	**0.0000**	**0.0000**	VAR241; VAR252	0.3202	0.3413
VAR231; VAR242	**0.0000**	**0.0000**	VAR231; VAR253	**0.0000**	**0.0000**	VAR242; VAR253	0.0577	0.0626
VAR232; VAR243	**0.0000**	**0.0002**	VAR232; VAR254	**0.0000**	**0.0001**	VAR243; VAR254	0.1160	0.1336
VAR233; VAR244	0.1970	0.3795	VAR233; VAR255	**0.0064**	**0.0148**	VAR244; VAR255	0.0847	0.0722
VAR237; VAR248	**0.0211**	**0.0068**	VAR237; VAR259	**0.0000**	**0.0001**	VAR248; VAR259	**0.0207**	**0.0178**
VAR238; VAR249	**0.0010**	**0.0006**	VAR238; VAR260	**0.0000**	**0.0000**	VAR249; VAR260	**0.0126**	**0.0212**
VAR239; VAR250	0.3377	0.2651	VAR239; VAR261	0.4549	0.3521	VAR250; VAR261	0.3706	0.3851
VAR266; VAR276	**0.0012**	**0.0027**	VAR266; VAR286	**0.0000**	**0.0000**	VAR276; VAR286	**0.0007**	**0.0016**
VAR267; VAR277	**0.0461**	0.0994	VAR267; VAR287	**0.0015**	**0.0044**	VAR277; VAR287	0.1865	0.2195
VAR268; VAR278	0.1402	0.2110	VAR268; VAR288	**0.0001**	**0.0010**	VAR278; VAR288	**0.0037**	**0.0090**
VAR269; VAR279	0.3237	0.2919	VAR269; VAR289	**0.0006**	**0.0015**	VAR279; VAR289	**0.0036**	**0.0043**
VAR270; VAR280	**0.0007**	**0.0026**	VAR270; VAR290	**0.0000**	**0.0000**	VAR280; VAR290	0.1646	0.2060
VAR271; VAR281	**0.0000**	**0.0002**	VAR271; VAR291	**0.0000**	**0.0000**	VAR281; VAR291	**0.0008**	**0.0019**
VAR272; VAR282	0.3549	0.4883	VAR272; VAR292	**0.0070**	**0.0128**	VAR282; VAR292	**0.0305**	**0.0191**
VAR273; VAR283	0.4451	0.4362	VAR273; VAR293	**0.0001**	**0.0004**	VAR283; VAR293	**0.0001**	**0.0005**
VAR274; VAR284	**0.0048**	**0.0080**	VAR274; VAR294	**0.0000**	**0.0000**	VAR284; VAR294	**0.0034**	**0.0040**

TABLE 14. PARAMETRIC T-TEST AND NONPARAMETRIC MANN-WHITNEY TEST II								
Less Info vs. More Info	**T-Test *P*-value**	**MW *P*-value**	**Less Info vs. Live Experiment**	**T-Test *P*-value**	**MW *P*-value**	**More Info vs. Live Experiment**	**T-Test *P*-value**	**MW *P*-value**
VAR15; VAR115	0.2858	0.2593	VAR15; VAR415	**0.0000**	**0.0000**	VAR115; VAR415	**0.0000**	**0.0001**
VAR16; VAR116	0.3564	0.4038	VAR16; VAR416	**0.0000**	**0.0000**	VAR116; VAR416	**0.0000**	**0.0000**
VAR17; VAR117	0.2419	0.2438	VAR17; VAR417	**0.0000**	**0.0000**	VAR117; VAR417	**0.0000**	**0.0000**
VAR18; VAR118	0.2298	0.4038	VAR18; VAR418	**0.0000**	**0.0000**	VAR118; VAR418	**0.0000**	**0.0000**
VAR19; VAR119	0.0968	0.1094	VAR19; VAR419	**0.0037**	**0.0043**	VAR119; VAR419	**0.0005**	**0.0010**
VAR20; VAR120	0.4280	0.1704	VAR20; VAR420	**0.0000**	**0.0004**	VAR120; VAR420	**0.0000**	**0.0010**
VAR21; VAR121	0.3632	0.3711	VAR21; VAR421	**0.0000**	**0.0002**	VAR121; VAR421	**0.0000**	**0.0005**
VAR22; VAR122	0.4853	0.4227	VAR22; VAR422	**0.0000**	**0.0000**	VAR122; VAR422	**0.0000**	**0.0000**
VAR23; VAR123	0.1156	**0.0489**	VAR23; VAR423	**0.0012**	**0.0003**	VAR123; VAR423	**0.0000**	**0.0000**
VAR24; VAR124	0.0518	0.0610	VAR24; VAR424	**0.0000**	**0.0000**	VAR124; VAR424	**0.0000**	**0.0000**
VAR28; VAR128	**0.0148**	**0.0388**	VAR28; VAR428	**0.0078**	**0.0055**	VAR128; VAR428	**0.0271**	0.1005
VAR29; VAR129	**0.0102**	**0.0192**	VAR29; VAR429	**0.0059**	**0.0022**	VAR129; VAR429	0.1867	0.0515
VAR31; VAR131	0.1438	0.1420	VAR31; VAR431	**0.0016**	**0.0017**	VAR131; VAR431	**0.0362**	**0.0254**
VAR35; VAR135	**0.0383**	0.0771	VAR35; VAR435	**0.0032**	**0.0016**	VAR135; VAR435	0.0617	**0.0142**
VAR38; VAR138	0.0986	0.0883	VAR38; VAR438	**0.0130**	**0.0395**	VAR138; VAR438	0.0948	0.1769
VAR43; VAR143	0.1612	0.1650	VAR43; VAR443	**0.0085**	**0.0123**	VAR143; VAR443	**0.0484**	**0.0486**

We conclude that:

- Within each experimentation stage, the three systems are indeed perceived to be different.
    - Direct-Control vs. Autonomous shows the most amount of difference, regardless of the experimental stage.
    - A majority of the Direct-Control vs. Remote Control and Remote-Control vs. Autonomous systems also showed differences, although less than the Direct-Control vs. Autonomous systems.
- Between the three levels of experimentation (Less Info, More Info, live experiments), each subsection of the technology is considered statistically different.

- ○ Live experimentation shows a significant difference in the information and knowledge gathered.
- ○ Live experimentation can be concluded to have significant value and insight.
- ○ The difference between Less Info and More Info without any hands-on experimentation is only limited. In other words, having additional information on paper, without the ability to perform hands-on experimentation, yields little difference and only minor benefits.

Predictability Without Experimentation Is Very Limited

The research question we are attempting to answer in this section is:

- *Can the final outcome of a detailed experiment be predicted by performing some basic Pre-experimental survey?*

If the research question above is found to be predictable, this would save the DoD considerable time and expense. Results from detailed experimentation can be predicted from basic preliminary review of the technology.

Table 15 shows a sampling of the results from a multivariate regression model. Little to no statistical significance is discernible when using Pre-experimental data to predict the outcomes of the Post-experiment scores.

Multiple linear and nonlinear interacting multivariate regressions were also run, and none seems to exhibit coefficients of determination greater than 50% and adjusted coefficients of determination greater than 25%.

TABLE 15. LIMITED PREDICTABILITY WITH LINEAR AND NONLINEAR MULTIVARIATE REGRESSION

Model Inputs:
VAR296 vs. VAR15; VAR16; VAR17; VAR18; VAR19; VAR20; VAR21; VAR22; VAR23; VAR24; VAR25
SUSA vs. PU1, PU2, PU3, PU4, PEOU1, PEOU2, PEOU3, PEOU4, IU1, IU2, IU3

Multiple R	0.85341	Maximum Log Likelihood	-52.79311
R-Square	0.72830	Akaike Info Criterion (AIC)	6.82033
Adjusted R-Square	0.30135	Bayes Schwarz Criterion (BSC)	7.41682
Standard Error	7.28268	Hannan-Quinn Criterion (HQC)	6.92128

	Coeff	Std. Error	T-stat	*P*-value	Lower 5%	Upper 95%
Intercept	135.82272	26.21155	5.18179	0.00128	73.84226	197.80319
VAR X1	-1.78874	5.52795	-0.32358	0.75571	-14.86027	11.28279
VAR X2	0.02206	4.87473	0.00452	0.99652	-11.50484	11.54895
VAR X3	-13.67128	6.12796	-2.23097	0.06088	-28.16161	0.81904
VAR X4	-9.34621	6.28587	-1.48686	0.18065	-24.20993	5.51752
VAR X5	-1.40361	5.80732	-0.24170	0.81594	-15.13574	12.32853
VAR X6	-5.81092	3.63238	-1.59976	0.15369	-14.40012	2.77829
VAR X7	-2.34249	4.29174	-0.54581	0.60215	-12.49084	7.80587
VAR X8	1.71980	3.64092	0.47235	0.65105	-6.88960	10.32921
VAR X9	17.00398	5.38884	3.15541	0.01603	4.26140	29.74656
VAR X10	2.09003	3.88437	0.53806	0.60721	-7.09505	11.27512
VAR X11	-5.90165	2.22358	-2.65412	0.03275	-11.15957	-0.64372

ANOVA	DF	SS	MS	F	*p*-Value
Regression	11	995.19	90.47	1.70580	0.24525
Residual	7	371.26	53.04		
Total	18	1366.45			

Hypothesis Test
Critical F-statistic (99% confidence with DFR1 and DFR2) : 6.538166
Critical F-statistic (95% confidence with DFR1 and DFR2) : 3.603037
Critical F-statistic (90% confidence with DFR1 and DFR2) : 2.683924

Table 16 shows a principal component analysis and factor analysis result where the multiple variables were reduced further to see if there would be any improvements in the multivariate regression, but the results similarly indicate very low predictive power in the Pre-experiment results.

TABLE 16. PRINCIPAL COMPONENT ANALYSIS

Model Inputs:
VAR23:VAR33
PU1, PU2, PU3, PU4, PEOU1, PEOU2, PEOU3, PEOU4, IU1, IU2, IU3
* indicates negative values

Cum Proportions:										
55.05%	75.51%	85.26%	90.59%	94.76%	96.74%	97.92%	98.97%	99.57%	99.87%	100.00%
Eigenvectors:										
0.3537	*0.2475	*0.1379	*0.0953	0.1383	*0.3637	*0.0508	*0.4055	0.5314	*0.0935	*0.4195
0.3592	*0.2186	0.0260	0.1763	0.1098	*0.1431	*0.7667	0.1402	*0.1323	0.2344	0.2811
Eigenvalues (Arranged and Ranked):										
6.0552	2.2509	1.0725	0.5861	0.4586	0.2184	0.1292	0.1157	0.0666	0.0320	0.0148

A traditional ordinary least squares multivariate regression also does not make too much sense in that no one-to-one correspondence is detected among the data rows. That is, different active-duty military from the same unit participated in the three experimental stages. This means that the responses of one soldier will not correspond to the same perception of another soldier testing another system during a different stage. This explains partly the low predictability of Post-experiment results using Pre-experiment data.

> **"Having additional information on paper, without the ability to perform hands-on experimentation, yields little difference and only minor benefits.**

Additional sophisticated methods were performed, such as bootstrapping the regression, where an empirical bootstrap of the data was nonparametrically simulated and bootstrapped, then regression models were run. The process was repeated thousands of times. Figures 3, 4, and Table 17 illustrate the results. Only 9% to 12% of the time will a single variable be considered statistically significant, and the goodness-of-fit predictability levels vary widely, from 18% to 95%, depending on the specific issue under study. No consistent and valid predictive power is apparent in the Pre-experiment data. This concurs with the two-variable *T*-tests and MW tests shown previously where we do see significant and valuable insights exist when hands-on experimentation is performed, which means without these experiments, paper-based cursory system knowledge is insufficient to identify the true value and risks of a system.

TABLE 17. BOOTSTRAP REGRESSION III

Variable	IU1	IU2	IU3	PEOU1	PEOU2	PEOU3	PEOU4	PU1	PU2	PU3	PU4	R Square
Number of Datapoints	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Mean	0.5035	0.4958	0.4816	0.5066	0.4952	0.5217	0.5016	0.5064	0.5018	0.4944	0.5035	**0.6150**
Median	0.5132	0.4926	0.4652	0.5103	0.5026	0.5277	0.4935	0.5071	0.4971	0.4974	0.5057	**0.6226**
Standard Deviation	0.2886	0.2922	0.2931	0.2853	0.2889	0.2893	0.2867	0.2801	0.2862	0.2846	0.2917	0.1553
Variance	0.0833	0.0854	0.0859	0.0814	0.0840	0.0837	0.0822	0.0785	0.0819	0.0810	0.0851	2.41%
Coefficient of Variation	57.32%	58.94%	60.86%	56.32%	58.54%	55.46%	57.17%	55.32%	57.04%	57.57%	57.93%	0.2525
Maximum	1.0000	0.9974	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9712
Minimum	0.0021	0.0004	0.0004	0.0009	0.0005	0.0051	0.0003	0.0020	0.0009	0.0009	0.0005	0.1263
Range	0.9979	0.9970	0.9996	0.9991	0.9995	0.9949	0.9997	0.9980	0.9991	0.9991	0.9995	0.8449
Skewness	-0.0422	0.0362	0.0921	-0.0761	-0.0067	-0.0855	0.0071	0.0015	0.0005	0.0199	-0.0284	-0.2432
Kurtosis	-1.2087	-1.2455	-1.2177	-1.1780	-1.2168	-1.2010	-1.1436	-1.1199	-1.1846	-1.1535	-1.2117	-0.2463
25% Percentile	0.2408	0.2426	0.2211	0.2619	0.2352	0.2718	0.2533	0.2696	0.2630	0.2519	0.2550	0.5091
75% Percentile	0.7496	0.7566	0.7372	0.7596	0.7483	0.7652	0.7395	0.7330	0.7448	0.7325	0.7575	72.50%
Error Precision at 95%	3.56%	3.66%	3.78%	3.50%	3.63%	3.44%	3.55%	3.43%	3.54%	3.57%	3.60%	0.0157
5% Percentile	0.0463	0.0547	0.0409	0.0463	0.0443	0.0653	0.0459	0.0568	0.0501	0.0465	0.0430	0.3515
10% Percentile	0.1017	0.0985	0.0854	0.0986	0.0931	0.1071	0.1045	0.1205	0.1030	0.0985	0.0906	0.4041
20% Percentile	0.1984	0.1979	0.1840	0.2091	0.1917	0.2191	0.2021	0.2212	0.2137	0.2060	0.1933	0.4786
30% Percentile	0.3000	0.2837	0.2786	0.3223	0.2888	0.3305	0.3096	0.3198	0.3107	0.2929	0.2997	0.5361
40% Percentile	0.4165	0.3776	0.3701	0.4166	0.3913	0.4321	0.4187	0.4094	0.3993	0.3930	0.4010	0.5781
50% Percentile	0.5129	0.4914	0.4645	0.5098	0.5024	0.5260	0.4929	0.5062	0.4946	0.4962	0.5043	0.6224
60% Percentile	0.6109	0.5961	0.5733	0.6227	0.5969	0.6353	0.5875	0.6018	0.6012	0.5886	0.6142	0.6661
70% Percentile	0.7032	0.6981	0.6784	0.7089	0.6980	0.7279	0.6947	0.6931	0.7002	0.6800	0.7015	0.7013
80% Percentile	0.7940	0.8051	0.7959	0.7999	0.7913	0.8209	0.7956	0.7899	0.7930	0.7898	0.8053	0.7545
90% Percentile	0.8957	0.9014	0.8895	0.8876	0.8964	0.9119	0.9032	0.8988	0.9020	0.8868	0.9035	0.8156
95% Percentile	0.9418	0.9455	0.9498	0.9368	0.9448	0.9543	0.9536	0.9498	0.9419	0.9447	0.9570	0.8594
99% Percentile	0.9915	0.9894	0.9920	0.9912	0.9875	0.9914	0.9905	0.9923	0.9955	0.9914	0.9904	0.9344
Certainty Value 0.01	0.80%	1.40%	0.80%	1.10%	1.10%	0.40%	1.40%	1.20%	0.80%	1.30%	1.00%	
Certainty Value 0.05	5.52%	4.71%	6.32%	5.02%	5.52%	3.81%	5.22%	4.21%	4.81%	5.02%	5.52%	
Certainty Value 0.1	9.83%	10.13%	11.94%	10.03%	10.73%	9.33%	9.53%	7.92%	9.83%	10.23%	10.93%	

The main conclusion from the analysis is:

The final detailed experimental results cannot be sufficiently predicted by using Pre-experiment survey data, regardless of how much nonexperimental, paper-based information is provided to the user.

FIGURE 3. BOOTSTRAP REGRESSION I

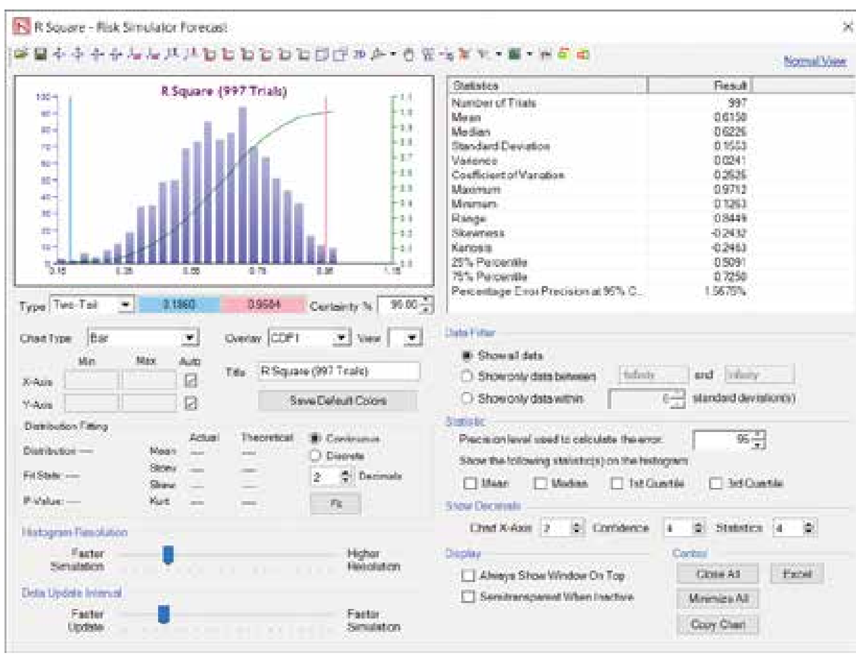
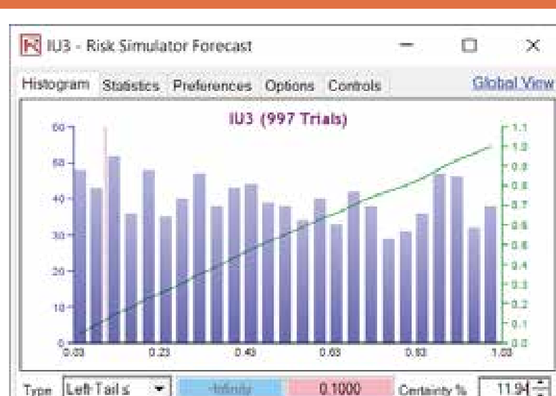


FIGURE 4. BOOTSTRAP REGRESSION II



Limitations of Research

The investigators used secondary data collected by the U.S. Department of Defense. The sample size was not within the control of the investigators and represented a smaller than desired number of participants.

Conclusions

The topic of trust in technology is increasingly important to the DoD as outlined in the Defense Science Board Study on Autonomy (David & Nielsen, 2016), which states, "There is a need to build trust in autonomous systems while also improving the trustworthiness of autonomous capabilities. These are enablers that align RDT&E [research, development, test & evaluation] processes to more rapidly deliver autonomous capabilities to DoD missions."

This work involves the introduction of novel ideas to existing theories that relate to the formation of trust. This research focuses on the impact of trust towards the adoption of autonomous systems. We have established that trust involves a user assuming some level of risk. The only literature available on technology trust involves situations that expose users to insignificant levels of risk. We posit that our research conducted on technology used in high-risk military application will reveal causality not identified in previous trust research.

This research tests theories of anthropomorphism and system hierarchy by manipulating the amount of information to observe the impact on the formation of initial, reason-based, technology trust. The article begins to answer the question of whether or not it is possible to predict and potentially capture trust in technology used for high-risk military applications. If a causal relationship exists between technology features and acceptance, it could greatly reduce the time and expense of adopting new technologies. The initial findings of this research indicate that manipulating familiarization with technology through the use of anthropomorphic categories, without the use of experience-based data or the ability to perform hands-on experimentation, yields little difference and only minor benefits. This article warrants further research to identify the influence of experience-based trust on the formation of reason-based trust.

References

Adams, B. D., & Webb, R. D. (2002, September 16–20). Trust in small military teams. *In Proceedings of the 7th International Command and Control Research and Technology Symposium* (pp. 1-20), Québec City, Canada. http://dodccrp.org/events/7th_ICCRTS/Tracks/pdf/006.PDF

Castelfranchi, C., & Falcone, R. (2010). *Trust theory: A socio-cognitive and computational model.* Wiley Online Library. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470519851

Cho, J.-H., Chan, K., & Adali, S. (2015). *A survey on trust modeling. ACM Computing Surveys (CSUR), 48*(2), 28. https://doi.org/10.1145/2815595

David, R. A., & Nielsen, P. (2016). *Defense science board summer study on autonomy.* Defense Science Board. https://dsb.cto.mil/reports/2010s/DSBSS15.pdf

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51–90. https://www.jstor.org/stable/30036519?seq=1

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS), 2*(2), 12. https://dl.acm.org/doi/10.1145/1985347.1985353

Mun, J. (2018). *Quantitative research methods.* Thompson-Shore & IIPER Press. https://www.amazon.com/gp/product/B083Q2MVBJ/

Pak, R., Fink, N., Price, M., Bass, B., & Sturre, L. (2012). Decision support aids with anthropomorphic characteristics influence trust and performance in younger and older adults. *Ergonomics, 55*(9), 1059–1072. https://psycnet.apa.org/record/2012-22389-007P

Rempel, J. K., Holmes, J. G., & Zanna, M. P. (1985). Trust in close relationships. *Journal of Personality and Social Psychology, 49*(1), 95–112. https://psycnet.apa.org/record/1985-30794-001

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 20*(10), 1–24. https://www.researchgate.net/publication/299371857_A_Meta-Analysis_of_Factors_Influencing_the_Development_of_Trust_in_Automation_Implications_for_Understanding_Autonomy_in_Future_Systems

Tétard, F., & Collan, M. (2009, January 5–8). Lazy user theory: A dynamic model to understand user selection of products and services. In Ralph H. Sprague, Jr. (Chair), *Proceedings of the 42nd Hawaii International Conference on System Sciences* (pp. 1–9), Waikoloa, Big Island, HI. https://ieeexplore.ieee.org/xpl/conhome/4755313/proceeding

Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences, 39*(2), 273–315. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2008.00192.x

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52,* 113–117. https://www.scholars.northwestern.edu/en/publications/the-mind-in-the-machine-anthropomorphism-increases-trust-in-an-au

APPENDIX A
Research Instrument

The investigators used secondary data collected by the U.S. Department of Defense. Data collection occurred in two phases. Phase one of the data collection was conducted in a controlled and distraction-free classroom environment and involved completion of a user survey by two randomly selected groups of active-duty military from within a single unit tasked with a high-risk mission. Both groups participated in separate morning sessions lasting 1 hour each. The second session started immediately following completion of the first session. Each group was provided with identical overviews of a high-risk military scenario that would be completed by deploying three technology systems rather than human operators. The independent variable "system presentation" was manipulated between the first and second groups. The second independent variable, "system control" was provided to all participants in the form of three separate technologies.

Phase two of the experiment was conducted in the field and involved the hands-on testing of the three technologies introduced during the phase one survey. Phase two of the experiment was conducted 6 months after the classroom survey of phase one. A total of 15 participants were selected from the same military unit as in the phase one survey. This experiment was conducted over a 12-day period. The first 3 days were reserved for training, and the subsequent 9 days were used to test the operational capabilities of the systems in the high-risk scenario presented in phase one. The day after the field experimentation concluded, all participants gathered in a controlled classroom environment to respond to the same user survey provided in phase one.

The investigators used secondary data collected by the U.S. Department of Defense. That data collection activity was ruled not human subjects research by the governing Institutional Review Board (IRB) in accordance with Secretary of the Navy Instruction (SECNAVINST 3900.39E), December 19, 2017. The data to which the Naval Postgraduate School (NPS) investigators have access do not contain data that are personally identifiable. Therefore, the presented activity was deemed not human subjects research by NPS IRB.

APPENDIX B

Survey Questions

Survey			Pre-Experiment		Post-Experiment		
Question	**Category**	**Code**	**Less Info**	**More Info**	**Most Info**	**Usability Questions**	**Usability Scores**
Loss of system endurance (decreased operating time)	Risk of Failure	HA1	VAR1	VAR101	VAR401	Q1	VAR266
Loss of power (unable to overcome large obstacles)	Risk of Failure	HA2	VAR2	VAR102	VAR402	Q2	VAR266
Loss of agility (limited range of motion)	Risk of Failure	HA3	VAR3	VAR103	VAR403	Q3	VAR266
Loss of speed (operates slowly)	Risk of Failure	HA4	VAR4	VAR104	VAR404	Q4	VAR266
Only have direct control (radio/ autonomous have failed)	Risk of Failure	AL1	VAR5	VAR105	VAR405	Q5	VAR270
Only have radio control (direct/ autonomous have failed)	Risk of Failure	AL2	VAR6	VAR106	VAR406	Q6	VAR271
Only have autonomous operation (direct/radio have failed)	Risk of Failure	AL3	VAR7	VAR107	VAR407	Q7	VAR272
Loss of ability to store data (bad memory)	Risk of Failure	AL4	VAR8	VAR108	VAR408	Q8	VAR273
Slow response to commands (bad processor)	Risk of Failure	AL5	VAR9	VAR109	VAR409	Q9	VAR274
Loss of ability to obtain imagery (video)	Risk of Failure	LN1	VAR10	VAR110	VAR410	Q10	VAR275
Loss of ability to obtain environmental data	Risk of Failure	LN2	VAR11	VAR111	VAR411	Q1	VAR276
Loss of ability to geolocate/ navigate (GPS)	Risk of Failure	LN3	VAR12	VAR112	VAR412	Q2	VAR277
Loss of comms needed to send sensor data (no system transmit)	Risk of Failure	LN4	VAR13	VAR113	VAR413	Q3	VAR278
Loss of comms needed to control sensors (no system receive)	Risk of Failure	LN5	VAR14	VAR114	VAR414	Q4	VAR279
This system would improve my performance	Direct	PU1	VAR15	VAR115	VAR415	Q5	VAR280
The system would increase my accuracy	Direct	PU2	VAR16	VAR116	VAR416	Q6	VAR281
The system would enhance my effectiveness	Direct	PU13	VAR17	VAR117	VAR417	Q7	VAR282
Overall, this system would be useful	Direct	PU4	VAR18	VAR118	VAR418	Q8	VAR283
The operational use of this system is clear and understandable	Direct	PEOU1	VAR19	VAR119	VAR419	Q9	VAR284
Using this system should not require a lot of my mental effort	Direct	PEOU2	VAR20	VAR120	VAR420	Q10	VAR285
It should be easy to get this system to do what I want it to do	Direct	PEOU3	VAR21	VAR121	VAR421	Q1	VAR286
Overall, this system would be easy to use	Direct	PEOU4	VAR22	VAR122	VAR422	Q2	VAR287
Given the chance, I would use this system	Direct	IU1	VAR23	VAR123	VAR423	Q3	VAR288
It is likely that I would recommend this system	Direct	IU2	VAR24	VAR124	VAR424	Q4	VAR289
I have been exposed to this technology in the past	Direct	IU3	VAR25	VAR125	VAR425	Q5	VAR290
This system would improve my performance	Remote	PU1	VAR26	VAR126	VAR426	Q6	VAR291

Survey Questions (continued)

Survey			Pre-Experiment		Post-Experiment		
Question	**Category**	**Code**	**Less Info**	**More Info**	**Most Info**	**Usability Questions**	**Usability Scores**
The system would increase my accuracy	Remote	PU2	VAR27	VAR127	VAR427	Q7	VAR292
The system would enhance my effectiveness	Remote	PU3	VAR28	VAR128	VAR428	Q8	VAR293
Overall, this system would be useful	Remote	PU4	VAR29	VAR129	VAR429	Q9	VAR294
The operational use of this system is clear and understandable	Remote	PEOU1	VAR30	VAR130	VAR430	Q10	VAR295
Using this system should not require a lot of my mental effort	Remote	PEOU2	VAR31	VAR131	VAR431	SUSA	VAR296
It should be easy to get this system to do what I want it to do	Remote	PEOU3	VAR32	VAR132	VAR432	SUSB	VAR297
Overall, this system would be easy to use	Remote	PEOU4	VAR33	VAR133	VAR433	SUSC	VAR298
Given the chance, I would use this system	Remote	IU1	VAR34	VAR134	VAR434		
It is likely that I would recommend this system	Remote	IU2	VAR35	VAR135	VAR435		
I have been exposed to this technology in the past	Remote	IU3	VAR36	VAR136	VAR436		
This system would improve my performance	Autonomous	PU1	VAR37	VAR137	VAR437		
The system would increase my accuracy	Autonomous	PU2	VAR38	VAR138	VAR438		
The system would enhance my effectiveness	Autonomous	PU3	VAR39	VAR139	VAR439		
Overall, this system would be useful	Autonomous	PU4	VAR40	VAR140	VAR440		
The operational use of this system is clear and understandable	Autonomous	PEOU1	VAR41	VAR141	VAR441		
Using this system should not require a lot of my mental effort	Autonomous	PEOU2	VAR42	VAR142	VAR442		
It should be easy to get this system to do what I want it to do	Autonomous	PEOU3	VAR43	VAR143	VAR443		
Overall, this system would be easy to use	Autonomous	PEOU4	VAR44	VAR144	VAR444		
Given the chance, I would use this system	Autonomous	IU1	VAR45	VAR145	VAR445		
It is likely that I would recommend this system	Autonomous	IU2	VAR46	VAR146	VAR446		
I have been exposed to this technology in the past	Autonomous	IU3	VAR47	VAR147	VAR447		
Tasking is directly relevant to my job function	Control	C1	VAR48	VAR148	VAR262		
I am personally invested in learning how to conduct this mission	Control	C2	VAR49	VAR149	VAR263		
In general, I am comfortable learning how to use new technology	Control	C3	VAR50	VAR150	VAR264		
These technologies are critical for accomplishing this mission	Control	C4	VAR51	VAR151	VAR265		
What is your current job?	Demographic	D1	VAR52	VAR152			
How long in your current job?	Demographic	D2	VAR53	VAR153			
How long in the military?	Demographic	D3	VAR54	VAR154			

Author Biographies

Mr. Michael G. Anderson

is a PhD Candidate at the Naval Postgraduate School Information Sciences Department. His research interests include Technology Trust. Mr. Anderson has an extensive background in the design of wireless communication systems and the use and deployment of military technologies. His past experience includes over 10 years of work as adjunct faculty teaching university students and professional engineers and scientists in the areas of digital design and computer architecture. Mr. Anderson holds a BS and MS in Electrical Engineering from the University of Michigan, Ann Arbor.

(E-mail address: mganders@nps.edu)

Dr. Johnathan C. Mun

is a research professor at the Naval Postgraduate School and is a specialist in advanced decision analytics, quantitative risk modeling, strategic flexibility real options, predictive modeling, and portfolio optimization. He has authored 24 books; holds 22 patents and patents pending; created over a dozen software applications in advanced decision analytics; and has written over a hundred technical notes, journal articles, and white papers. He is currently the CEO of Real Options Valuation, Inc., and his prior positions include vice president of Analytics at Oracle/Crystal Ball and a senior manager at KPMG Consulting. Dr. Mun holds a PhD in Finance and Economics from Lehigh University, an MBA and MS from Southeastern University, and a BS in Physics and Biology from the University of Miami.

(E-mail address: jcmun@nps.edu)

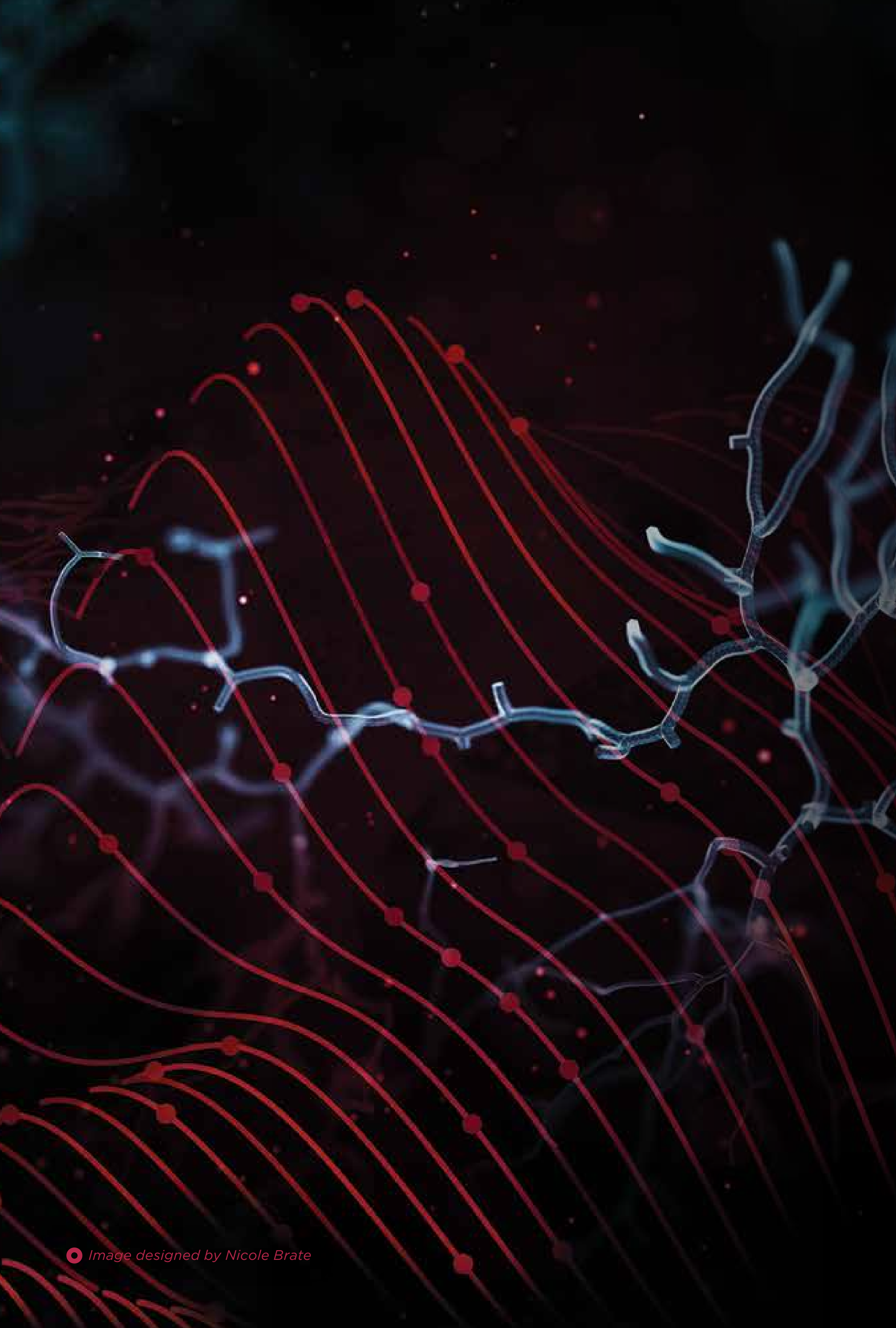


Image designed by Nicole Brate

USE OF FACTORS IN DEVELOPMENT ESTIMATES: IMPROVING THE COST ANALYST TOOLKIT

Capt Matthew R. Markman, USAF, Jonathan D. Ritschel, and Edward D. White

Factor Estimating is a technique commonly used by defense acquisition analysts to develop cost estimations. However, previous studies developing factors for the Engineering and Manufacturing Development (EMD) phase of the life cycle are limited. This research expands the current toolkit for cost analysts by developing cost factors in previously unexplored areas. More specifically, over 400 cost reports are utilized to create new standard cost factors that are delineated by five categories: commodity type, contract type, contractor type, development type, and Service. The factors are developed for those elements that are common in a wide array of projects such as program management, systems engineering, data, or training. This new factor dataset provides cost analysts with the information necessary to appropriately identify and select the most relevant factors to use when developing future cost estimates. Through statistical analysis, the research also helps identify those elements in which more analysts' time and energy should be allocated when developing their estimates.

DOI: *https:// doi.org/10.22594/10.22594/dau.19-848.28.01*
Keywords: *Errors in Estimating, Cost Factors, Engineering and Manufacturing Development, Statistical Testing, Cost Growth*

Cost growth in major defense acquisition programs is a well-documented concern (Ritschel et al., 2019; Younossi et al., 2007). This growth is problematic because it crowds out additional programs and leads to an inability to satisfy demands. As a result, cost analysts have conducted numerous studies to determine the causes behind defense program cost growth. Examples of identified causes include decisions by managers to change requirements, externally imposed funding changes, schedule perturbations, and *errors in estimating* or planning (Bolten et al., 2008). This article focuses on refining the cost analyst toolkit in an effort to reduce errors in estimating and thereby improve defense cost estimates and mitigate cost growth. More specifically, this article refines and expands the available set of *cost factors* for estimators to employ in Engineering and Manufacturing Development (EMD) cost estimates.

Defense cost analysts have a range of models and techniques they utilize to estimate program resources. One of these tools is the application of standard cost factors. Factors are traditionally used as primary and/or as cross-check methodologies when estimating major defense acquisition program (MDAP) "common" cost elements such as program management, systems engineering, training, site activation, and spare part costs.

Given that factors are just one of several cost-estimating techniques and that cost-estimating errors account for only a portion of program cost growth, one may question the magnitude of the impact that improvements in cost factors can provide. Research by Miller (2020) illuminates the potential impacts. To ascertain the estimating techniques used by cost analysts, Miller examines a sample of 60 defense development programs from 2003–2018. He finds the factor technique mean value across the 60 programs to be 16.9%. In other words, factor estimating is utilized to determine 16.9% of total cost in EMD cost estimates. The total dollar value of the 60 programs in Miller's (2020) sample is $48.8 billion. Therefore, even small (for example 1–2%) improvements in the accuracy of cost factors employed can result in millions of dollars of estimating error reductions.

> **This article focuses on refining the cost analyst toolkit in an effort to reduce errors in estimating and thereby improve defense cost estimates and mitigate cost growth.**

In what ways can extant factors be improved? What gaps exist? Currently, the research division of the Air Force Life Cycle Management Center (AFLCMC) periodically publishes standard cost factor tables for aircraft EMD that capture prime contractor data for a selection of clean-sheet design aircraft programs. Despite the utility of the AFLCMC-published tables, additional data exist that can assist in refining these factors, as well as developing new factors to include Army, Navy, and Joint programs. Other identified gaps in currently published EMD factors include neglected commodity categories (e.g., electronic/automated software, missiles, ordnance, space, and Unmanned Aerial Vehicles [UAVs]), development types (e.g., modification programs), and subcontractor data. Each additional category of data enables estimators to accomplish more in-depth analysis based on the type of program in question. Thus, expanding and refining factors for EMD programs gives estimators a more robust tool set to draw upon, ultimately leading to more precise estimates.

Literature Review

Several key documents designate and define the cost estimating methodologies utilized within the Department of Defense (DoD), including the *Air Force Cost Analysis Handbook* (AFCAH) and the *Government Accountability Office Cost Estimating and Assessment Guide*. These publications assist in setting a baseline for program offices and cost analysts to craft credible and consistent cost estimates. They also satisfy an overarching requirement for the DoD to have policies in place to safeguard the billions of taxpayer dollars allocated to MDAPs each year (U.S. Government Accountability Office, 2009). While the documents define the acceptable estimating methodologies, they do not represent an all-encompassing guidebook, as every MDAP presents unique challenges. The four primary techniques outlined in the AFCAH are analogy and factor, parametric, build-up (engineering), and expert opinion (subject matter expert) (Department of the Air Force, 2007). While each technique represents a different approach to cost estimating, with associated benefits and drawbacks, the merit of using multiple strategies to achieve greater confidence in an estimate cannot be overstated. The introduction of more than one estimating technique provides cost analysts with the ability to triangulate a point estimate that considers levels of detail not fully captured by individual techniques or estimates. Furthermore, this approach serves as a cross-check to ensure that estimates fall within percentage bounds set by the analyst.

Cost factor creation necessitates an understanding of Work Breakdown Structures (WBS). The WBS concept in MDAPs has remained relatively constant over the past several decades (DoD, 2005). It is a decomposition of a project into smaller, more manageable components, sometimes referred to as the management blueprint for the project (Mislick & Nussbaum, 2015). The WBS is mandated and governed by MIL-STD-881D, ultimately fulfilling broader requirements set forth in DoD Instruction 5000.2; this DoD publication aims to maintain uniformity in definition and consistency of approach for programs developing a WBS (DoD, 2018). For the sake of consistency, the DoD has revised and updated guidance regarding the WBS only when major technological advances or changes in the acquisition process warranted such action (DoD, 2005).

“While each technique represents a different approach to cost estimating, with associated benefits and drawbacks, the merit of using multiple strategies to achieve greater confidence in an estimate cannot be overstated.

The WBS consists of three primary hierarchical levels, with a fourth and fifth sometimes included in expanded forms; for this article, only the second level is addressed. Level two of the WBS captures major elements subordinate to the system identified by level one and consists of prime mission products, including all hardware and software elements. Level two also includes combinations of system-level services applicable to the program, including the following elements common to most programs: integration and assembly, system test and evaluation (ST&E), systems engineering/program management (SE/PM), common support equipment (CSE), peculiar support equipment (PSE), training, data, operational/site activation, and initial spares and repair parts (DoD, 2018). These common elements at level two of the WBS are the focus for developing factors in this article. Benefits

of the WBS mandated by MIL-STD-881D include ease of normalization of data and information across a variety of commodity types and DoD agencies, and the ability to reference past and current MDAPs to better understand and forecast their own costs, schedules, and overall program.

Research on MDAP cost factors in cost estimating is insufficient to fully and efficiently utilize the technique. The Air Force acquisition cost analyst community has conducted unpublished cost factor studies by Wren (1998) and Otte (2015) specific to MDAPs in the EMD phase. These studies, however, are very narrow in scope and apply solely to a limited subset of aircraft programs. Wren (1998) focused solely on developing factors relevant to common factors in 20 aircraft aviation programs. Otte (2015) updated the work of Wren, but his analysis remained narrowly focused on clean-sheet design aircraft programs. The efforts of Wren and Otte represent a sizable stepping stone towards an exhaustive reference table of factors for DoD analysts, but lack the breadth required to make the studies applicable to more than a specific set of programs based at Wright Patterson Air Force Base. Large gaps in cost factor creation exist for additional (e.g., nonaircraft) commodity types, modification programs, subcontractor data, and contract type.

Database

In an effort to reduce defense program cost growth, Congress enacted Pub. L. 111-23, Weapon Systems Acquisition Reform Act of 2009. This act created a Pentagon office—Office of Cost Assessment and Program Evaluation (CAPE). CAPE is chartered to provide independent analysis of resource allocation to deliver the optimal portfolio of defense capability through efficient and effective use of public funds (Office of the Secretary of

Defense [OSD], n.d.). CAPE initiated the development of the Cost Assessment Data Enterprise (CADE) system to help achieve its mission. CADE serves as an integrated web-based application for defense acquisition program cost, schedule, and technical data (OSD, n.d.). Within CADE are Cost Data Summary Reports (CDSR), which contain the data used in this analysis. EMD data were chosen as the only life-cycle phase to be analyzed based on the identified literature gap.

Contractor submittal of CDSRs is mandatory for all major contracts and subcontracts (regardless of contract type) valued at $50 million or more in programs designated as Acquisition Category I (DoD, 2011). The threshold for Acquisition Category I designation is *total* expenditures of $480 million in Research, Development, Test and Evaluation (RDT&E) fiscal year 2014 constant dollars or $2.79 billion in procurement (DoD, 2015). Due to these thresholds, no contracts under $50 million are used in the analysis.

Cost information in CDSRs is reported through a standardized WBS as governed by MIL-STD 881D. The level two WBS elements include system-level services applicable to the program, including elements common to most programs as shown in Table 1. These eight "common" WBS elements in Table 1 are the focus for analyzing factors in this article.

TABLE 1. WBS ELEMENTS

Level 2 Common WBS Elements
Systems Engineering/Program Management (SE/PM)
System Test and Evaluation (ST&E)
Training
Data
Peculiar Support Equipment (PSE)
Common Support Equipment (CSE)
Site Activation
Spares

The final dataset consists of programs spanning from 1961 to 2017, representing a broad range of programs across numerous commodity types and military services. The common WBS mandated by MIL-STD-881D enables consistency in data collection and normalization. The complete dataset within CADE contained 189 programs; however, only 102 of those programs fit the criteria for inclusion in the final dataset (see Appendix A for final program list). Table 2 depicts the exclusion criteria and remaining programs utilized for factor development.

TABLE 2. DATABASE EXCLUSIONS

Category	Number Removed	Remaining Programs
Available Programs in CADE		189
Excluded Commodity Types	35	154
No EMD Data	25	129
CCDR File Format Not .XLS	27	102
Final Dataset for Analysis		102

Several commodity types, such as system of systems, are excluded because they lie outside the scope of this analysis. Additionally, 25 programs lacked associated EMD phase costs and are excluded. Twenty-seven programs contained EMD data but have no accessible files within CADE, resulting in the entire program's exclusion from the dataset. These are primarily older programs with manually transcribed data from the 1980s or earlier, and in many instances the data are illegible.

Methods

The methodological approach has two stages. The first stage is creation of individual factors. The factors are calculated as a ratio of individual level two WBS elements from Table 1 to a base cost. The base cost is the program's Prime Mission Equipment (PME) value, which does not include the contractor's fee or miscellaneous expenses (general and administrative, undistributed budget, management reserve, or facilities capital cost of money). The general form of the calculation is shown in Equation 1.

$$\frac{WBS\ Level\ 2\ Element_{ij}}{PME_j} = Cost\ Factor_{ij}$$

where i = SE/PM, ST&E, Training, Data, PSE, CSE, Spares, and Site Activation

j = individual programs

After establishing cost factors for the level two WBS elements, it is possible to develop composite factors for a myriad of unique categories. Specific level two WBS elements can be examined in groupings to establish aggregate values that represent an average or percentage that can be used in formulating estimates. These groupings allow for analysis at several levels, such as fixed wing aircraft, rotary wing aircraft, a specified contractor for radar modifications, a specified contractor's role in a program (prime versus sub), a specified period for a certain commodity type, and many more.

Once the factors are established for each program, the mean, median, and standard deviation values for the various program groupings are calculated. In addition, interquartile ranges are calculated to examine variability among factors. This allows for descriptive analysis and provides a basis from which the programs are grouped and analyzed to compare differences.

The second stage of analysis subdivides the cost factors into categories for statistical testing to aid the cost analyst in determining appropriate levels of aggregation for practical use. The categories were determined through discussions with cost analyst practitioners in the field. These categories represent the way cost analysts may consider grouping or filtering their data when developing an estimate. The categories are Commodity Type, Service, Contractor Designation, Development Type, and Contract Type, with associated subcategories shown in Table 3.

TABLE 3. CATEGORIES FOR COMPARISON ANALYSIS

Categories				
Service	**Commodity Type**	**Contractor Designation**	**Contract Type**	**Development Type**
Army	Aircraft	Prime	CPAF (Cost Plus Award Fee)	Modification
Navy (includes Marine Corps)	Electronic/ Automated Software	Sub	CPFF (Cost Plus Fixed Fee)	New Design
Air Force	Missile		CPIF (Cost Plus Incentive Fee)	Prototype
Multiple	Ordnance		Cost-Other (Other than CPAF, CPFF, CPIF)	Subsystem
	Space		FFP (Firm Fixed Price)	New Mission Design Series (MDS) Designator
	Unmanned Aerial Vehicle (UAV)		FPI (Fixed Price Incentive)	Commercial Derivative
			FPIF (Fixed Price Incentive Firm Target)	
			Fixed - Other	
			Unknown	

For each of the categorical comparisons, hypothesis tests are used to identify differences in the elements detailed in Table 3. For example, differences in cost factors are tested based on whether the work was completed by a

> **The Kruskal-Wallis test is a rank-based nonparametric test to determine whether statistically significant differences exist between two or more groups of an independent variable on a continuous dependent variable.**

prime contractor or subcontractor (shown in the Contractor Designation column of Table 3). One of the most widely used hypothesis test techniques is a parametric test, such as the *t*-test. However, an underlying assumption of parametric tests is that the data are normally distributed. Therefore, a Shapiro-Wilk test was conducted to determine whether or not the data were normally distributed. The results of the test showed that the data were not normally distributed, thereby indicating parametric techniques should not be used.

As a result, nonparametric tests (which do not require the assumption of normality) are utilized throughout the remainder of the analysis. Specific nonparametric tests used include the Kruskal-Wallis and Steel-Dwass tests, which are similar to ANOVA and *t*-tests. The Kruskal-Wallis test is a rank-based nonparametric test to determine whether statistically significant differences exist between two or more groups of an independent variable on a continuous dependent variable. The dependent variable is the numerical cost factor value, while the independent variables are the various groups. For example, contractor type (prime versus subcontractor) is the independent variable, while the cost factor values are the dependent variable. Because the Kruskal-Wallis test does not identify where within the subcategory comparison differences occur, the Steel-Dwass test is employed. The Steel-Dwass multiple comparison test identifies which rank orders of the tested groups are statistically different for each instance of

subcategory comparison. The definition of statistical significance used throughout the analysis will be in reference to an $\alpha = 0.05$ level. This means that in order for the results to be deemed statistically significant, there is less than a 5% chance of concluding that a difference exists where there is no actual difference.

Results

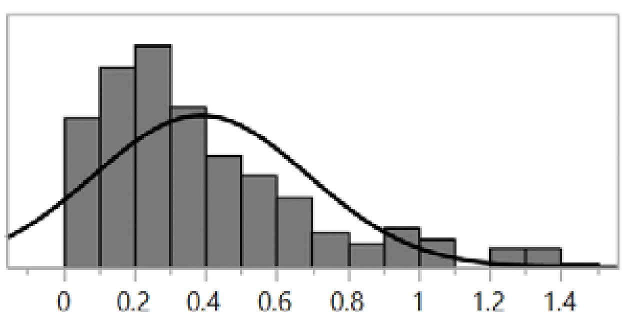
Factor development in stage one of the analysis applies Equation 1 to the dataset. More specifically, the eight level two WBS elements identified in Table 1 are combined with the final 102 program dataset. For example, a factor for ST&E (one of the WBS elements identified in Table 1) is developed for the C-17 program (one of the 102 programs identified in Appendix A) utilizing Equation 1. It is important to note that within an individual program, there may be multiple Cost Data Summary Reports (CDSR) reported in the CADE database. These reports serve as the primary means within the DoD to collect actual data reported by contractors in performing acquisition contracts. Therefore, the 102 programs used for analysis expands to 443 individual cost reports from which new, unique cost factors are created across the eight common WBS elements.

TABLE 4. FACTORS BY CATEGORY TYPE

Category	Total	Category	Total	Category	Total
Unique Factors Created	443	**Development Type**		**Contract Type**	
Commodity Type		Commercial Derivative	4	CPAF	74
Aircraft	245	Modification	135	CPFF	39
Electronic/ Automated Software	118	New Design	150	CPIF	66
Missile	22	Prototype	9	Cost-Other	135
Ordnance	12	Subsystem	105	FFP	27
Space	36	New Mission Design Series (MDS) Designator	40	FPI	20
Unmanned Aerial Vehicle (UAV)	10	**Service**		FPIF	19
Contractor Type		Air Force	196	Fixed-Other	6
Prime	308	Army	94	Unknown	57
Subcontractor	135	Multiple	24		
		Navy (includes Marine Corps)	129		

Individual factors from a CDSR, relevant only to the peculiarities of a specific program, are of limited utility to cost analysts. For example, the ST&E factor from the C-17 is undoubtedly useful to the C-17 program office; but relying on this single factor as the basis for analysis on a different program inserts additional uncertainty into that estimate. The credibility of a cost estimate is only as good as the data from which it is developed. Basing an estimate off a single data point goes against cost-estimating best practices. Therefore, the individual factors developed from the 443 CDSRs are mapped into composite factors. These composite factors are created according to the subcategories in Table 4, and descriptive statistics including mean, median, and standard deviation are calculated. (See Appendix B for the descriptive statistics for each of the eight common WBS elements.)

FIGURE. SE/PM SHAPIRO-WILK TEST



Goodness-of-Fit Test

	W	Prob<W
Shapiro-Wilk W Test	0.8855	<.0001*

The subcategories in Table 4 primarily represent subcategories established within the data hierarchy of the CADE database. These subcategories can be statistically tested to identify where differences exist. If differences are not found between the subcategories, then analysts can use composite factors comprising a wider dataset. However, if differences exist, then analysts should only use factors comprising programs within that unique subcategory. First, normality of the eight common WBS elements is tested with the Shapiro-Wilk test at an $\alpha = 0.05$ threshold. Results for the first element tested, SE/PM, are shown in the Figure.

As shown in the Figure, the null hypothesis is rejected with a p-value of less than 0.0001. (Note: The null hypothesis states that there is no significant difference between a normal distribution and the data; a rejection of the null therefore indicates that differences are present and the data are not normally distributed.) Similar Shapiro-Wilk test results for the subsequent seven WBS elements (not shown) rejected the null hypothesis, necessitating nonparametric testing throughout the remainder of the analysis. Nonparametric testing identifies similarities of locations in the data elements analyzed. Histograms of the data in this analysis reveal a consistent right-skewed profile. Due to the similarities in the shape of the histograms, the nonparametric tests can be considered to be testing medians (Hollander et al., 2014). Therefore, subsequent discussion of nonparametric results will focus on differences in the medians of the data.

Commodity Type

The first category analyzed is commodity type. The Kruskal-Wallis test reveals statistical differences between WBS element median values (Table 5). Specific differences are identified within the SE/PM, ST&E, and Site Activation WBS elements.

TABLE 5. KRUSKAL-WALLIS RESULTS FOR COMMODITY TYPE

WBS Element	Alpha	N	Chi-Square	*P*-value	Null Hypothesis Test Result
SE/PM	0.05	406	49.2441	<0.0001	Reject
ST&E	0.05	374	32.3203	<0.0001	Reject
Training	0.05	192	6.9636	0.2234	Do Not Reject
Data	0.05	267	6.1052	0.2961	Do Not Reject
PSE	0.05	149	2.2603	0.8121	Do Not Reject
CSE	0.05	50	1.0203	0.9609	Do Not Reject
Site Activation	0.05	47	14.4899	0.0059	Reject
Spares	0.05	84	3.7434	0.2905	Do Not Reject

TABLE 6. COMMODITY DIFFERENCES SUMMARY

	Aircraft	Electronic/ Automated Software	Missile	Ordnance	Space	UAV
SE/PM	2	1	1	0	0	0
ST&E	2	1	1	0	3	1
Site Activation	1	1	0	0	0	0

After determining that statistical differences exist, the Steel-Dwass multiple comparison test is employed to identify which commodity types exhibited differences. The identification of differences through the statistical tests tells the analyst that utilizing the more readily available aggregated factors is ill-advised. Rather, it indicates that the analysts should take more time to refine and narrow the dataset to account for the differences and isolate the relevant data. Table 6 summarizes the findings for each WBS element with the number of differences annotated by commodity type. The aircraft commodity type contains the most statistical differences, with five instances where the WBS median values were statistically different from the other subcategories (for example, the median SE/PM cost factor for aircraft is different than both the SE/PM cost factor for electronic/automated software systems and missiles). The space and electronic/automated software contain the second most statistical differences with three each. For the WBS elements, SE/PM and ST&E contain 85.7% of all differences. The implications for practical usage are that standard factors for SE/PM and ST&E should be careful to ensure delineation by commodity type and not modeled at aggregated levels. This is especially important for these two WBS elements, as they have the highest factor values with respect to PME among all the elements. In other words, these two elements have the largest cost impacts of all the WBS elements. Thus, taking the extra time and effort to refine the cost factor by commodity type is suggested in these areas.

Contract Type

The second category analyzed is contract type. Contract types are designated on the Contractor Cost Data Reporting (CCDR) system. There are two broad categories of contract type: cost reimbursable contracts and fixed price contracts. Further subdivision of these categories ranges from firm-fixed-price, in which the contractor has full responsibility for the performance costs and resulting profit (or loss), to cost plus-fixed-fee, in which the contractor has minimal responsibility for the performance costs and the negotiated profit is fixed. In between are the various incentive contracts where the contractor's responsibility for the performance costs and the profit or fee incentives offered are tailored to the uncertainties involved in contract performance. Examples include cost plus award fee or cost plus incentive fee.

> **"The identification of differences through the statistical tests tells the analyst that utilizing the more readily available aggregated factors is ill-advised. Rather, it indicates that the analysts should take more time to refine and narrow the dataset to account for the differences and isolate the relevant data.**

The Kruskal-Wallis test results in rejection of the null hypothesis in four areas. Differences in median values are found for SE/PM, ST&E, Data, and Peculiar Support Equipment (PSE) (Table 7).

Conducting the Steel-Dwass multiple comparison test across all contract types reveals statistically significant differences across all but one contract type (Table 8). Fixed Price Incentive (FPI) contracts display the most statistical differences with eight. Any project expecting an FPI contract should place increased scrutiny on the programs that contribute to the composite factor calculation and the specific contract type utilized. Additionally, SE/PM and ST&E find 10 differences each. The concentration of differences in the SE/PM and ST&E WBS elements suggests estimators should afford extra time and research for estimates in those areas. [Note that the PSE WBS element displays statistical differences according to the Kruskal-Wallis test in Table 7, but no individual pair differences are found with the Steel-Dwass test. This is due to the extremely low *n* values for several subcategories.]

TABLE 7. KRUSKAL-WALLIS RESULTS FOR CONTRACT TYPE

WBS Element	Alpha	N	Chi-Square	*P*-value	Null Hypothesis Test Result
SE/PM	0.05	406	32.8151	<0.0001	Reject
ST&E	0.05	374	34.4853	<0.0001	Reject
Training	0.05	192	5.6801	0.683	Do Not Reject
Data	0.05	267	19.4757	0.0125	Reject
PSE	0.05	149	18.7037	0.0165	Reject
CSE	0.05	50	6.8419	0.4455	Do Not Reject
Site Activation	0.05	47	9.8514	0.1972	Do Not Reject
Spares	0.05	84	9.4857	0.2196	Do Not Reject

TABLE 8. CONTRACT TYPE DIFFERENCES SUMMARY

	CFAP	CPFF	CPIF	Cost-Other	FFP	EPI	FPIF	Unknown
SE/PM	2	2	0	1	2	3	0	0
ST&E	1	1	0	1	1	5	0	1
Data	0	0	1	0	0	0	0	1

Development Type

The third category analyzed is development type consisting of six subcategories. New Design programs are those with capabilities new to the DoD, while Modifications are defined as programs undergoing a major change to core capabilities or performance. Prototypes are programs intended to test an emerging capability for future utilization. The New Mission Design Series (MDS) Designator subcategory captures existing major programs undergoing minor changes, such as the F-16B, which accommodates two pilots, instead of one, for training purposes. Commercial Derivatives are defined as programs initiated in the commercial market that are adapted for subsequent military use. Lastly, the Subsystem designation is assigned to those programs whose efforts are accomplished independent of the primary project, such as an engine upgrade. The Kruskal-Wallis test reveals differences in five WBS areas: SE/PM, ST&E, Data, PSE, and Spares (Table 9).

TABLE 9. KRUSKAL-WALLIS RESULTS FOR DEVELOPMENT TYPE

WBS Element	Alpha	N	Chi-Square	*P*-value	Null Hypothesis Test Result
SE/PM	0.05	406	18.3391	0.0026	Reject
ST&E	0.05	374	15.3905	0.0088	Reject
Training	0.05	192	6.7041	0.2436	Do Not Reject
Data	0.05	267	13.8759	0.0164	Reject
PSE	0.05	149	11.4644	0.0429	Reject
CSE	0.05	50	6.3575	0.273	Do Not Reject
Site Activation	0.05	47	8.5601	0.128	Do Not Reject
Spares	0.05	84	13.0157	0.0232	Reject

TABLE 10. DEVELOPMENT TYPE DIFFERENCES SUMMARY

	Modification	New Design	Prototype	Subsystem	New MDS Designator	Commercial Derivative
SE/PM	1	2	0	0	1	0
ST&E	0	0	0	1	1	0
Data	0	0	1	0	1	0
PSE	1	0	0	0	1	0
Spares	1	1	0	0	0	0

The Steel-Dwass test identifies median value statistical differences for each development category (Table 10). All development categories have at least one statistically significant difference except for commercial derivatives, which is the smallest category comprising less than 1% of the dataset. The new MDS designator and new design subcategories have the most differences at four and three, respectively. Projects in these two subcategories should ensure factor development does not have other development types in its composite factors.

Contractor Type

The fourth category analyzed is contractor type. The CCDR dataset consisted of prime contractor data and subcontractor data. The majority of the data—69.5%—is prime data. Because the fourth category had only two subcategories, the Steel-Dwass test is not needed. The identification of differences through the Kruskal-Wallis test is sufficient. Results are shown in Table 11.

Differences in the contractor type category are found for only two WBS elements: ST&E and PSE. The small number of differences suggests that composite factor development does not require large amounts of time and effort dedicated to determining whether the data are from the prime or a sub. Rather, aggregated factor models consisting of both contractor types may be sufficient.

TABLE 11. KRUSKAL-WALLIS RESULTS FOR CONTRACTOR TYPE

WBS Element	Alpha	N	Chi-Square	*P*-value	Null Hypothesis Test Result
SE/PM	0.05	406	0.7777	0.3778	Do Not Reject
ST&E	0.05	374	12.064	0.0005	Reject
Training	0.05	192	0.0811	0.7759	Do Not Reject
Data	0.05	267	2.66	0.1029	Do Not Reject
PSE	0.05	149	5.3186	0.0211	Reject
CSE	0.05	50	1.6912	0.1934	Do Not Reject
Site Activation	0.05	47	0.0571	0.8111	Do Not Reject
Spares	0.05	84	0.087	0.768	Do Not Reject

Service

The last category analyzed is military service. The data are subcategorized by Air Force, Army, Navy, and Multiple as designated on the CCDRs. The Kruskal-Wallis test for the Service category identifies statistically different median values in two areas: SE/PM and ST&E (Table 12).

Despite only two WBS elements containing statistical differences in median values, the Steel-Dwass multiple comparison test is able to identify a total of 12 statistically significant instances (Table 13). The Army SE/PM factor is found to be different from all other Services, while the ST&E factor for multiple Services is also different from all others. For these two WBS elements, practitioners should ensure delineation by Service in composite factor development.

TABLE 12. KRUSKAL-WALLIS RESULTS FOR SERVICE

WBS Element	Alpha	N	Chi-Square	*P*-value	Null Hypothesis Test Result
SE/PM	0.05	406	20.1146	0.0002	Reject
ST&E	0.05	374	9.1187	0.0278	Reject
Training	0.05	192	3.7819	0.286	Do Not Reject
Data	0.05	267	1.6337	0.6518	Do Not Reject
PSE	0.05	149	2.666	0.446	Do Not Reject
CSE	0.05	50	2.1053	0.5508	Do Not Reject
Site Activation	0.05	47	1.222	0.7477	Do Not Reject
Spares	0.05	84	1.0621	0.588	Do Not Reject

TABLE 13. SERVICE DIFFERENCES SUMMARY

	Air Force	Army	Navy	Multiple
SE/PM	1	3	1	1
Spares	1	1	1	3

Timeframe-Specific Analysis

The initial dataset exclusion criteria (Table 2) removed 27 programs due to inaccessible files or illegible data entries. These excluded programs are primarily from the 1980s or earlier. Exclusion of these programs may raise concerns of bias in the analysis. To determine whether the exclusion of these older programs has an effect on the factors developed, a timeframe-specific analysis on a subset of the data spanning the past two decades is accomplished using 1998 as the cut-off date. Table 14 displays the descriptive statistics for the SE/PM WBS element for the original dataset, as well as the revised dataset spanning from 1998 to 2017.

TABLE 14. SE/PM DESCRIPTIVE STATISTICS COMPARISON

Commodity	Original Mean	1998–Present Mean	Original Median	1998–Present Median
Aircraft	0.3025	0.3433	0.2292	0.2727
Electronic/Automated Software	0.5463	0.5479	0.4875	0.4875
Missile	0.5014	0.5014	0.3897	0.3897
Ordnance	0.3426	0.3484	0.285	0.3409
Space	0.3825	0.4059	0.3109	0.3109
UAV	0.4913	0.5154	0.3655	0.3887

The descriptive statistics of the subset of data for SE/PM are similar in most cases, and identical in some, to the original dataset. Analysis of other WBS elements (not shown) yields similar results. The consistency displayed between the subset and original dataset leads to the conclusion that the 27 programs excluded due to inaccessible files or illegible entries are unlikely to affect the descriptive statistics or statistical analysis results.

Discussion and Conclusions

This article sought to improve the current state of cost estimating with a focus on furthering EMD cost factors. These improvements are achieved through several avenues. First, new standard cost factors were created from a diverse set of program types comprising over 400 CDSRs. The development and publication of these new factors are useful on their own merit. But additional gains in cost estimation accuracy are possible by determining which factors should be used in various circumstances. This second benefit is determined through statistical testing of relevant categorical grouping (commodity, contract type, development type, contractor, and

Service). When statistical testing *does not* reveal differences in categories, then aggregated composite factors are sufficient. However, when differences *are detected*, then analysts should allocate more time and effort to ensure properly refined composite factors are utilized, rather than relying on the readily available aggregated factors.

The following example illustrates the potential gains to be achieved. In this hypothetical scenario, the analyst is estimating SE/PM for an aircraft. The mean SE/PM cost factor value for the aggregated dataset is 0.3802. While this is a good starting point, the analyst knows through the statistical testing results in this article that SE/PM is frequently found to be unique in a multitude of categories. If only the commodity type of aircraft is known, then the mean SE/PM aircraft cost factor value of 0.3025 would be the value chosen. But perhaps the analyst also knows the type of contract is CPAF. The results in this article indicate that the SE/PM cost factor has statistically different values based on contract type. The analysts, therefore, would be advised to allocate further effort to refining the dataset to include only those programs composed of aircraft with CPAF contracts. In this hypothetical example, the final cost factor value would be 0.2945. The refining of criteria in this example led to a 22.5% difference in mean values of included data points, which if examined in the context of a $30 million program, reflects a $2.57 million difference in the estimate for SE/PM.

Each MDAP presents unique characteristics that must be explored and understood to make the inclusion of its data truly meaningful in the context of constructing a cost estimate.

As shown in the example, each MDAP presents unique characteristics that must be explored and understood to make the inclusion of its data truly meaningful in the context of constructing a cost estimate. Generic composite factors represent a starting point for analysts in instances where MDAP characteristics may be unrefined. Once a program's requirements have been solidified and the manner in which they will be accomplished is well-defined, analysts can refine their dataset to MDAPs with direct application to their program.

As reviewed at the beginning of this article, Miller (2020) found the cost factor technique is commonly used for EMD programs. Thus, even small improvements in the accuracy of cost factors employed can have positive impacts. These better estimates should lead to better program outcomes. As a result, the cost growth due to estimating inaccuracies, as identified by Bolten et al. (2008), should be reduced.

While the discussion thus far has focused on an illustrative example and potential program-level impacts, some specific findings deserve increased attention and can impact where cost analysts allocate effort in refining cost factors germane to their specific estimate. First, knowledge of contract type is highly desirable, as the contract type category contained the highest number of statistical differences between the subcategories. While it would be most advantageous to develop composite factors based on the *precise* contract type (e.g., cost plus award fee), even broader classifications into the *two general categories* of cost reimbursable or fixed price contracts are useful. Second, the commodity type category was found to have the second most differences in median values after contract type. Commodity information should be readily available for any project, allowing for ease of analyst calibration. The results also indicate those areas where analysts should economize their time. Specifically, the results showed fewer differences in the contractor type category. The implication is that deriving the factor from prime or subcontract data has little effect.

> **“Future research should focus not only on factor development in other phases of the life cycle, but also on those elements of cost growth that are not attributable to estimator toolkit deficiencies.**

Lastly, the statistical testing also illuminates which of the eight individual WBS elements deserve the most attention from cost analysts. Interestingly, the SE/PM and ST&E elements were flagged in virtually every categorical test. Making the distinction more compelling is the fact that these two elements typically have the highest in raw dollar value of the WBS elements analyzed. Coupling the high dollar value with the statistical testing results suggests that analysts should spend their time and energy on these areas.

In contrast, elements such as data and training were rarely flagged with statistically significant differences. Aggregated factors are therefore likely to be sufficient in these areas.

Several limitations to this study are noted. First, the analysis applies only to development projects. Projects in the production stage are likely to have different factors. Future research is recommended in this area. Second, the CCDR database that was utilized contained only contract values greater than $50 million. Smaller projects were not considered. Third, 27 older programs could not be analyzed due to inaccessible files or illegible data. Timeframe testing was conducted to analyze the effect with results showing little potential for bias. Lastly, an anonymous reviewer suggested exploring the effects of dollar-weighted factors, rather than equal weighting of individual contracts. This is an area to explore in future research.

The cost factor development and analysis presented here is one step toward improving public procurement in the DoD. Future research should focus not only on factor development in other phases of the life cycle, but also on those elements of cost growth that are not attributable to estimator toolkit deficiencies. Ultimately, it will be the combination of improvements in all these areas that is necessary to achieving efficiency gains in public procurement.

Acknowledgement

This material is based on work supported by the Acquisition Research Program at the Naval Postgraduate School.

References

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* (Report No. MG-670). RAND. https://www.rand.org/pubs/monographs/MG670.html

Department of the Air Force. (2007). *Air Force cost analysis handbook* (AFCAH). Washington, DC: U.S. Department of the Air Force.

Department of Defense. (2005). *Department of Defense standard practice: Work breakdown structures for defense materiel items* (MIL-HDBK-881A). Office of the Assistant Secretary of Defense for Acquisition, Performance Assessments & Root Cause Analysis. https://www.secnav.navy.mil/rda/OneSource/Documents/New%20MIL-STD-881C%20Work%20Breakdown%20Structures%20for%20Defense%20Materiel%20Items.pdf

Department of Defense. (2011). *Cost and software data reporting (CSDR) manual* (DoDM 5000.04-M-1). Office of the Director, Cost Assessment and Program Evaluation. https://cade.osd.mil/content/cade/files/csdr/guidance/DoDM%205000.04-M-1%20CSDR%20Manual.pdf

Department of Defense. (2015). *Operation of the defense acquisition system* (DoDI 5000-02). Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. https://www.acq.osd.mil/fo/docs/500002p.pdf

Department of Defense. (2018). *Department of Defense standard practice: Work breakdown structures for defense materiel items* (MIL-STD-881D). U.S. Government Printing Office. http://everyspec.com/MIL-STD/MIL-STD-0800-0899/MIL-STD-881D_55780/

Hollander, H., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric statistical methods* (3rd Ed.). Wiley. https://www.wiley.com/en-us/Nonparametric+Statistical+Methods%2C+3rd+Edition-p-9780470387375

Miller, C. W. (2020). *An investigation into subject matter expert elicitation in cost risk analysis* [Master's thesis]. Air Force Institute of Technology. https://apps.dtic.mil/sti/pdfs/AD1101597.pdf

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools.* Wiley. https://www.wiley.com/en-us/Cost+Estimation%3A+Methods+and+Tools-p-9781118536131

Office of the Secretary of Defense. (n.d.). *Cost assessment and program evaluation.* Department of Defense. https://www.cape.osd.mil/

Otte, J. (2015). *Factor study September 2015* [Unpublished manuscript]. Research Group, Air Force Life Cycle Management Center.

Ritschel, J. D., Lucas, B. M., White, E. D., & Mrla, D. (2019). The impact of WSARA on the cost of Air Force weapon systems. *Journal of Public Procurement, 19*(1), 2–14. https://doi.org/10.1108/JOPP-03-2019-022

U.S. Government Accountability Office. (2009). *Cost estimating and assessment guide* (Report No. GAO-09-3SP). https://www.gao.gov/products/GAO-09-3SP

Wren, D. (1998). *Avionics support cost element factors.* [Unpublished manuscript]. Research Group, Aeronautical Systems Center.

Younossi, O., Arena, M. V., Leonard, R. S., Roll, C. R., Jr., Jain, A., & Sollinger, J. M. (2007). *Is weapon system cost growth increasing? A quantitative assessment of completed and ongoing programs* (Report No. MG-588-AF). RAND. https://www.rand.org/pubs/monographs/MG588.html

APPENDIX A

List of Programs

AIRCRAFT
A-6A Full Scale Development
A-6E Full Scale Development
AH-64E Apache (Formerly AB3)
ARH – Armed Reconnaissance Helicopter
B-1 CMUP – B-1 LANCER Penetrating Bomber Conventional Mission Upgrade Program
B-1B Integrated Battle Station (IBS)
B-2 DMS: Defensive Management System
B-2 EHF SATCOM AND COMPUTER INCREMENT I – B-2 Advanced Extremely High Frequency SatCom Capability
B-2 RMP – B-2 Radar Modernization Program
B-52 Combat Network Communications Technology (CONECT)
B-58A Full Scale Development
BLACK HAWK UPGRADE (UH-60M) – Utility Helicopter Upgrade Program
C-130 AMP – C-130 Aircraft Avionics Modernization Program
C-130J – HERCULES Cargo Aircraft Program
C-17A – GLOBEMASTER III Advanced Cargo Aircraft Program
C-5 AMP – C-5 Aircraft Avionics Modernization Program
C-5 RERP – C-5 Aircraft Reliability Enhancement and Re-engineering Program
CH-47F – Cargo Helicopter. CH-47D Helicopter Upgrade Program
CH-53K – Heavy Lift Replacement Program
Comanche – Reconnaissance Attack Helicopter (RAH-66)
CRH – Combat Rescue Helicopter
E-10 – Multi-Sensor Command and Control Aircraft Program
E-2D AHE – E-2D Advanced Hawkeye
F/A-18E/F – SUPER HORNET Naval Strike Fighter
F-22 – RAPTOR Advanced Tactical Fighter
F-22A Increment 3.2B
F-117A Full Scale Development
F-35 – Lightning II Joint Strike Fighter (JSF) Program
H-1 UPGRADES (4BW/4BN) – United States Marine Corps Mid-life Upgrade to AH-1W Attack Helicopter and UH-1N Utility Helicopter
JSTARS – Joint Surveillance Target Attack Radar System
KC-135A Full Scale Development
MH-60R – Multi-Mission Helicopter Upgrade
MH-60S – Multi-Mission Combat Support Helicopter
P-8A – Poseidon Program
RQ-4A/B Full Scale Development
V-22 – OSPREY Joint Advanced Vertical Lift Aircraft
VH 71 – Presidential Helicopter Fleet Replacement Program
VH-92A Presidential Helicopter
YA-10 Development
ELECTRONIC/AUTOMATED SOFTWARE
3DELRR – Three-Dimensional Expeditionary Long-Range Radar
ADS (AN/WQR-3) – Advanced Deployable System
AMDR – Air & Missile Defense Radar
AMF JTRS – Joint Tactical Radio System Airborne & Maritime/Fixed Station
AOC-WS – Air and Space Operations Center-Weapon System
CAC2S – Common Aviation Command and Control System
CANES – Consolidated Afloat Network Enterprise Services
CEC – Cooperative Engagement Capability
CIRCM – Common Infrared Countermeasures
DCGS ARMY – Distributed Common Ground System Army
F-15 EPAWSS – Eagle Passive Active Warning Survivability System

FAB-T – Family of Beyond Line-of-Sight Terminals	
FBCB2 – Force XXI Battle Command Brigade and Below Program	
G/ATOR – Ground/Air Task Oriented Radar	
GCSS ARMY – Global Combat Support System Army	
GSE – Ground Soldier Ensemble	
IAMD – Integrated Air & Missile Defense	
ITEP – Improved Turbine Engine Program	
JATAS – Joint and Allied Threat Awareness System	
JLENS – Joint Land Attack Cruise Missile Defense Elevated Netted Sensor System	
JPALS – Joint Precision Approach and Landing System	
JTRS GMR – Joint Tactical Radio System Ground Mobile Radio	
JTRS NED – Joint Tactical Radio System Network Enterprise Domain	
Land Warrior – Integrated Soldier Fighting System for the Infantryman	
LMP – Logistics Modernization Program	
MIDS – Multi-Functional Information Distribution System (Includes Low Volume Terminal and JTRS)	
MP RTIP – Multi-Platform Radar Technology Insertion Program	
MPS – Mission Planning System	
NGJ – Next Generation Jammer	
NMT – Navy Multiband Terminal	
Space Fence Inc. 1 – Space Fence Ground-Based Radar System Increment 1	
WIN-T – Warfighter Information Network-Tactical	
MISSILE	
APKWS – Advanced Precision Kill Weapon System	
AGM-88E AARGM – AGM-88E Advanced Anti-Radiation Guided Missile (AARGM) Program	
AIM-9X – Air-to-Air Missile Upgrade	
GMLRS/GMLRS AW – Guided Multiple Launch Rocket System/Guided Multiple Launch Rocket System Alternative Warhead	
ICBM – Fuse Modernization Program	
JAGM – Joint Air-to-Ground Missile	
JASSM (JASSM/JASSM-ER) – Joint Air-to-Surface Standoff Missile	
JCM – AGM-169 Joint Common Missile	
Offensive Anti-Surface Warfare Increment 1 (Long Range Anti-Ship Missile)	
Patriot PAC-3 – Patriot Advanced Capability 3	
SM-6 – Standard Missile-6	
ORDNANCE	
B61 Mod 12 Life Extension Program Toolkit Assembly	
ERM – Extended Range Munition	
EXCALIBUR – Family of Precision, 155 mm Projectiles	
SDB I – Small Diameter Bomb Increment I	
SDB II – Small Diameter Bomb, Increment II	
SPACE	
AEHF – Advanced Extremely High Frequency (AEHF) Satellite Program	
EPS – Enhanced Polar System	
AIM-9X – Air-to-Air Missile Upgrade	
GPS OCX – Global Positioning Satellite Next Generation Control Segment	
GPS-IIIA – Global Positioning Satellite III	
MUOS – Mobile User Objective System	
NAVSTAR GPS – Global Positioning System	
NPOESS – National Polar-Orbiting Operational Environmental Satellite System	
SBIRS HIGH – Space-Based Infrared System Program, High Component	
TSAT – Transformational Satellite Communications System	
UAV	
GLOBAL HAWK (RQ-4A/B) – High Altitude Endurance Unmanned Aircraft System	
MQ-1C Gray Eagle	
MQ-4C Triton (formerly Broad Area Maritime Surveillance - BAMS)	
NAVY UCAS – Navy Unmanned Combat Air System	
REAPER (MQ-9 UAS) – Unmanned Aircraft System	
VTUAV – Vertical Takeoff and Land Tactical Unmanned Air Vehicle (Fire Scout)	

APPENDIX B

Summary Tables

	Training Summary Table				Data Summary Table			
	Mean	Median	Std Dev	N	Mean	Median	Std Dev	N
Service								
Air Force	0.0319	0.0093	0.0643	95	0.0385	0.0217	0.0608	126
Army	0.0398	0.0148	0.0673	45	0.0405	0.0180	0.0646	50
Navy	0.0329	0.0071	0.0653	50	0.0319	0.0148	0.0473	85
Multiple	0.0482	0.0482	0.0647	2	0.0194	0.0189	0.0103	6
Development Type								
Modification	0.0245	0.0051	0.0406	64	0.0448	0.0243	0.0664	84
New Design	0.0395	0.0166	0.0772	76	0.0297	0.0134	0.0457	85
Prototype	0.0029	0.0029	0.0019	2	0.0060	0.0042	0.0065	6
Subsystem	0.0277	0.0063	0.0475	23	0.0333	0.0180	0.0616	54
Variant	0.0543	0.0166	0.0886	24	0.0441	0.0269	0.0543	34
Commercial Derivative	0.0134	0.0133	0.0118	3	0.0240	0.0152	0.0187	4
Contractor Type								
Prime	0.0344	0.0100	0.0406	163	0.0384	0.0205	0.0572	206
Subcontractor	0.0329	0.0109	0.0772	29	0.0296	0.0175	0.0555	61
Commodity Type								
Aircraft	0.0307	0.0055	0.0544	111	0.0355	0.0206	0.0498	174
Electronic/Automated Software	0.0527	0.0254	0.0922	53	0.0407	0.0164	0.0736	59
Missile	0.0117	0.0079	0.0122	7	0.0418	0.0107	0.0861	12
Ordnance	0.0081	0.0062	0.0039	6	0.0100	0.0071	0.0109	4
Space	0.0142	0.0146	0.0119	9	0.0240	0.0076	0.0291	10
UAV	0.0176	0.0123	0.0180	6	0.0449	0.0280	0.0534	8
Contract Type								
CPAF	0.0468	0.0275	0.0785	30	0.0376	0.0217	0.0635	39
CPFF	0.0491	0.0167	0.0981	18	0.0362	0.0246	0.0401	19
CPIF	0.0371	0.0079	0.0736	27	0.0243	0.0092	0.0409	43
Cost-Other	0.0313	0.0065	0.0608	59	0.0351	0.0206	0.0571	74
FFP	0.0526	0.0178	0.0640	8	0.0262	0.0133	0.0396	18
FPI	0.0142	0.0159	0.0124	15	0.0358	0.0333	0.0251	19
FPIF	0.0266	0.0102	0.0554	13	0.0691	0.0167	0.1041	16
Fixed-Other	0.0016	0.0016	-	1	0.0060	0.0049	0.0040	4
Unknown	0.0210	0.0047	0.0271	21	0.0468	0.0294	0.0631	35

	SE/PM Summary Table				ST&E Summary Table			
	Mean	**Median**	**Std Dev**	**N**	**Mean**	**Median**	**Std Dev**	**N**
Service								
Air Force	0.3685	0.2972	0.2755	177	0.2251	0.1672	0.2074	166
Army	0.5080	0.4426	0.3372	91	0.2157	0.1992	0.1915	80
Navy	0.3393	0.2551	0.3039	115	0.2201	0.1582	0.2150	105
Multiple	0.3142	0.2699	0.2053	23	0.1059	0.0642	0.1027	23
Development Type								
Modification	0.3484	0.2845	0.2555	124	0.2155	0.1396	0.2193	119
New Design	0.4738	0.3759	0.3472	131	0.2143	0.1817	0.1880	114
Prototype	0.1906	0.1783	0.1472	8	0.2673	0.2820	0.1028	9
Subsystem	0.3730	0.2793	0.2816	101	0.1744	0.1038	0.1883	89
Variant	0.3249	0.2517	0.2924	39	0.2934	0.2456	0.2281	39
Commercial Derivative	0.1840	0.2128	0.1011	3	0.1804	0.1585	0.1432	4
Contractor Type								
Prime	0.3849	0.2947	0.3068	284	0.2294	0.1838	0.2019	274
Subcontractor	0.3966	0.3336	0.2898	122	0.1733	0.0999	0.2001	100
Commodity Type								
Aircraft	0.3025	0.2292	0.2385	227	0.2498	0.2036	0.2139	225
Electronic/Automated Software	0.5463	0.4875	0.3511	107	0.1702	0.1038	0.1924	88
Missile	0.5014	0.3897	0.3297	20	0.2041	0.1842	0.1772	18
Ordnance	0.3426	0.2850	0.1737	11	0.1513	0.0961	0.0998	11
Space	0.3825	0.3109	0.3093	31	0.0778	0.0448	0.0879	23
UAV	0.4913	0.3655	0.3217	10	0.2068	0.1893	0.01273	9
Contract Type								
CPAF	0.4128	0.3649	0.2641	66	0.1802	0.1072	0.1964	63
CPFF	0.5189	0.4233	0.3896	37	0.1671	0.0791	0.2095	31
CPIF	0.3905	0.2729	0.2987	61	0.2586	0.1997	0.2200	55
Cost-Other	0.4082	0.3175	0.3103	126	0.1824	0.1277	0.1748	113
FFP	0.2457	0.1560	0.2531	25	0.1777	0.1300	0.1503	20
FPI	0.2118	0.1694	0.2232	17	0.3907	0.3267	0.1991	20
FPIF	0.4203	0.3931	0.2811	19	0.2876	0.2167	0.2168	17
Fixed-Other	0.5720	0.5427	0.2327	2	0.2714	0.2227	0.2483	4
Unknown	0.3131	0.2430	0.2573	51	0.2248	0.1608	0.2163	51
Averages	**0.3802**	**0.3121**	**0.2732**	**75.1852**	**0.2117**	**0.1621**	**0.1822**	**69.2593**

	Training Summary Table				Data Summary Table			
	Mean	Median	Std Dev	N	Mean	Median	Std Dev	N
Service								
Air Force	0.0646	0.0282	0.0922	79	0.0136	0.0014	0.0313	22
Army	0.0399	0.0115	0.0626	28	0.0211	0.0088	0.0331	14
Navy	0.0592	0.0177	0.0917	40	0.01186	0.0011	0.0224	13
Multiple	0.0593	0.0593	0.0565	2	0.0063	0.0063	-	1
Development Type								
Modification	0.0477	0.0177	0.0880	60	0.0129	0.0013	0.0319	19
New Design	0.0573	0.0286	0.0770	46	0.0148	0.0067	0.0206	18
Prototype	0.0118	0.0090	0.0049	3	0.0001	0.0001	0.0001	2
Subsystem	0.0485	0.0194	0.0609	13	0.0378	0.0063	0.0537	5
Variant	0.0978	0.0481	0.1070	26	0.0108	0.0038	0.0171	5
Commercial Derivative	0.0039	0.0039	-	1	0.0018	0.0018	-	1
Contractor Type								
Prime	0.0497	0.0186	0.0778	120	0.0133	0.0015	0.0268	41
Subcontractor	0.0945	0.0545	0.1110	29	0.0235	0.0095	0.0390	9
Commodity Type								
Aircraft	0.0549	0.216	0.0789	98	0.0125	0.0018	0.0309	31
Electronic/Automated Software	0.0468	0.0094	0.0565	12	0.0149	0.0015	0.0280	7
Missile	0.0716	0.0085	0.0993	11	0.0218	0.0202	0.0212	6
Ordnance	0.0235	0.0182	0.0193	9	0.0353	0.0353	0.0493	2
Space	0.01247	0.0477	0.1673	11	0.0013	0.0013	-	1
UAV	0.0496	0.0213	0.0632	8	0.0209	0.021	0.0327	3
Contract Type								
CPAF	0.0540	0.0347	0.0637	14	0.0069	0.0024	0.0103	10
CPFF	0.0203	0.0092	0.0279	13	0.0365	0.0365	0.0301	2
CPIF	0.0398	0.0214	0.0542	28	0.0215	0.0081	0.0404	9
Cost-Other	0.0699	0.0186	0.1099	44	0.0103	0.0017	0.0193	14
FFP	0.0238	0.0175	0.0249	11	0.0004	0.0006	0.0002	3
FPI	0.1098	0.0619	0.1167	14	0.0028	0.0028	-	1
FPIF	0.0338	0.0042	0.0686	9	0.0290	0.0018	0.0459	9
Fixed-Other	0.0041	0.0041	-	1	-	-	-	-
Unknown	0.0929	0.0798	0.0925	15	0.0057	0.0057	0.0064	2

	Site Activation Summary Table				Spares Summary Table			
	Mean	**Median**	**Std Dev**	**N**	**Mean**	**Median**	**Std Dev**	**N**
Service								
Air Force	0.0490	0.0235	0.0798	23	0.0430	0.0113	0.0558	33
Army	0.0299	0.0250	0.0319	4	0.0221	0.0107	0.0259	10
Navy	0.0309	0.0020	0.0686	18	0.0341	0.0225	0.0347	41
Multiple	0.0065	0.0065	0.0049	2	-	-	-	-
Development Type								
Modification	0.0495	0.0141	0.0968	12	0.0222	0.0046	0.0479	25
New Design	0.0500	0.0241	0.0590	19	0.0438	0.0332	0.0394	34
Prototype	0.0040	0.0040	-	1	0.0279	0.0279	-	1
Subsystem	0.0046	0.0410	0.0040	4	0.0283	0.0225	0.0288	7
Variant	0.0276	0.0013	0.07878	9	0.0504	0.0303	0.0493	15
Commercial Derivative	0.0001	0.0001	<0.0001	2	0.0054	0.0054	0.0069	2
Contractor Type								
Prime	0.0405	0.0042	0.0737	40	0.0372	0.0174	0.0468	62
Subcontractor	0.0277	0.0030	0.0519	7	0.0331	0.0195	0.0336	22
Commodity Type								
Aircraft	0.0186	0.0015	0.0476	26	0.0397	0.0168	0.0498	52
Electronic/Automated Software	0.0917	0.0687	0.1018	11	0.0239	0.0152	0.0284	21
Missile	0.0009	0.0009	-	1	-	-	-	-
Ordnance	-	-	-	-	-	-	-	-
Space	0.0602	0.0494	0.0591	6	0.0356	0.0250	0.0304	6
UAV	0.0024	0.0028	0.0017	3	0.0519	0.0302	0.0353	5
Contract Type								
CPAF	0.0498	0.04260	0.0511	5	0.0255	0.0113	0.0298	17
CPFF	0.0277	0.0152	0.0316	6	0.0045	0.0012	0.0074	4
CPIF	0.0723	0.0649	0.0777	6	0.0255	0.0275	0.0192	11
Cost-Other	0.0355	0.0040	0.06750	15	0.0439	0.0226	0.0438	18
FFP	0.0008	0.0005	0.0009	3	0.0410	0.0047	0.0824	7
FPI	0.0023	0.0004	0.0040	4	0.0593	0.0432	0.0545	10
FPIF	0.0090	0.0002	0.0152	3	0.0152	0.0092	0.0195	4
Fixed-Other	-	-	-	-	-	-	-	-
Unknown	0.0790	0.0044	0.1505	5	0.0440	0.0236	0.0428	13

Author Biographies

Capt Matthew R. Markman, USAF

is a cost analyst at Wright Patterson Air Force Base, Ohio. Capt Markman graduated with a BS in Business Management from the Air Force Academy, and an MS in Cost Analysis from the Air Force Institute of Technology (AFIT).

(E-mail address: Matthew.Markman.1@us.af.mil)

Dr. Jonathan D. Ritschel

is an assistant professor of Cost Analysis in the AFIT Department of Systems Engineering and Management. He received a BBA in Accountancy from the University of Notre Dame, an MS in Cost Analysis from AFIT, and a PhD in Economics from George Mason University. Dr. Ritschel's research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, and economic institutional analysis.

(E-mail address: Jonathan.Ritschel@afit.edu)

Dr. Edward D. White

is a professor of Statistics in the AFIT Department of Mathematics and Statistics. He received a BS in Mathematics from the University of Tampa, an MAS from The Ohio State University, and a PhD in Statistics from Texas A&M University. Dr. White's primary research interests include statistical modeling, simulation, and data analytics.

(E-mail address: Edward.White@afit.edu)

MARCOM AWARD FOR PEER REVIEWED ACADEMIC JOURNAL

WINNER 2020

Norene Johnson, Emily Beliles, and Michael Krukowski

DAU Press

Fort Belvoir, VA

Image designed by Nicole Brate

A LEARNING CURVE MODEL ACCOUNTING FOR THE FLATTENING EFFECT IN PRODUCTION CYCLES

Capt Evan R. Boone, USAF, John J. Elshaw, Lt Col Clay M. Koschnick, USAF, Jonathan D. Ritschel, and Adedeji B. Badiru

The authors investigate production cost estimates to identify and model modifications to a prescribed learning curve. Their new model examines the learning rate as a decreasing function over time as opposed to a constant rate that is frequently used. The purpose of this research is to determine whether a new learning curve model could be implemented to reduce the error in cost estimates for production processes. A new model was created that mathematically allows for a "flattening effect," which typically occurs later in the production process. This model was then compared to Wright's learning curve, which is a popular method used by many organizations today. The results showed a statistically significant reduction in error through the measurement of the two error terms, Sum of Squared Errors and Mean Absolute Percentage Error.

DOI: *https:// doi.org/10.22594/10.22594/dau.20-850.28.01*
Keywords: *Learning Curve, Cost Estimation, Acquisition, Wright's Learning Curve, Boone's Learning Curve*

Many manufacturing firms today operate in a fiscally constrained and financially conscious environment. Managers throughout these organizations are expected to maximize the utility from every dollar as budgets and profit margins continue to shrink. Increased financial scrutiny adds greater emphasis on the accuracy of program and project management cost estimates to ensure acquisition programs are sufficiently funded. Cost estimating models and tools used by organizations must be evaluated for their relevance and accuracy to ensure reliable cost estimates. Many of the cost estimating procedures for learning curves were developed in the 1930s (Wright, 1936) and are still in use today as a primary method to model learning. As automation and robotics increasingly replace human touch-labor in the manufacturing process, the current 80-year-old learning curve model may no longer provide the most accurate approach for estimates. New learning curve methods that incorporate automated production and other factors that lead to reduced learning should be examined as an alternative for cost estimators in the acquisition process.

> **"Increased financial scrutiny adds greater emphasis on the accuracy of program and project management cost estimates to ensure acquisition programs are sufficiently funded.**

Since Wright's (1936) original learning curve model was developed, researchers have found other functions to model learning within the manufacturing process (Carr, 1946; Chalmers & DeCarteret, 1949; Crawford, 1944; DeJong, 1957; Towill, 1990; Towill & Cherrington, 1994). The purpose of this research is to address a gap in the literature that fails to account for the nonconstant rate of learning, which results in a flattening effect at the end of production cycles. We will investigate learning curve estimating methodology, develop learning curve theory, and pursue the development of a new estimation model that examines learning at a nonconstant rate.

This research identifies and models modifications to a learning curve model such that the estimated learning rate is modeled as a decreasing learning rate function over time, as opposed to a constant learning rate that is currently in use. Wright's (1936) learning curve model in use today mathematically states that for every doubling of units there will be a constant gain in efficiency. For example, if a manufacturer observes a 10% reduction in labor hours in the time to produce unit 10 from the time to produce unit 5, then it should expect to see the same 10% reduction in labor hours in the time to produce unit 20 from the time to produce unit 10. We propose that

more accurate cost estimates would result if a more flexible exponent were taken into consideration in developing the learning curve model. The proposed general modification would take the form:

$$Cost(x) = Ax^{f(x)} \tag{1}$$

Where:

$Cost(x)$ = cumulative average cost per unit

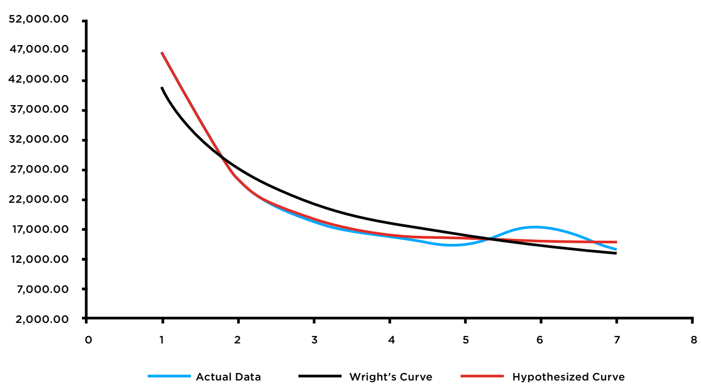
A = theoretical cost to produce the first unit

x = cumulative number of units produced

$f(x)$ = learning curve effect as a function of units produced

The exponent function in Equation 1 will be explored in this article. Figure 1 demonstrates the phenomena this research will examine. The black (flatter) line depicts the traditional curve where learning occurs at a constant rate; the red (steeper) line represents the hypothesized learning structure where the rate of learning changes as a function of the number of units produced; and the blue line represents notional data used to fit the two curves.

FIGURE 1. LEARNING CURVE DEPICTION



To address this research gap, our study aims to model a function that has the added precision of diminishing learning effects over time by introducing a learning curve decay factor that more closely models actual production cycle learning. We will accomplish this by developing a new learning curve model that minimizes the amount of error compared to current estimation models. Learning curves, specifically when estimating the expected cost per unit of complex manufactured items such as aircraft, are frequently modeled with a mathematical power function. The intent of these models is to capture the expected reduction in costs over time due to learning effects, particularly in areas with a high percentage of human touch labor. Typically, as production increases, manufacturers identify labor efficiencies and improve the process. If labor efficiencies are identified, it translates to unit cost savings over time. The general form of the learning curve model frequently used today is based on Wright's theory and is shown in Equation 2. Note that the structure of the exponent b ensures that as the number of units produced doubles, the cost will decrease by a given percentage defined as the learning curve slope (LCS). For example, when LCS is 0.8, then the cost per unit will decrease by 80% between units 2 and 4.

$$Cost(x) = Ax^{b} \tag{2}$$

Where:

$Cost(x)$ = cumulative average cost per unit

A = theoretical cost to produce the first unit

x = cumulative number of units produced

$b = \frac{\ln \textit{Learning Curve Slope}}{\ln 2}$

The cost of a particular production unit is modeled as a power function that decreases at a constant exponential rate. The problem is that the rate of decrease is not likely to be constant over time. We propose that the majority of cost improvements are to be found early in the production process, and fewer revelations are made later as the manufacturer becomes more familiar with the process. As time progresses, the production process should normalize to a steady state and additional cost reductions prove less likely.

> **Our study aims to model a function that has the added precision of diminishing learning effects over time by introducing a learning curve decay factor that more closely models actual production cycle learning.**

For relatively short production runs, the basic form of the learning curve may be sufficient because the hypothesized efficiencies will not have time to materialize. However, when estimating production runs over longer periods of time, the basic learning curve could underestimate the unit costs of those furthermost in the future. The underestimation occurs because the model assumes a constant learning rate, while actual learning would diminish, causing the actuals to be higher than the estimate. Current models may underestimate a significant amount when dealing with high unit cost items such as those in major acquisition programs; a small error in an estimate can be large in terms of dollars. Through the use of curve fitting techniques, a comparison can be made to determine which model best predicts learning within a particular production process. The remainder of this article is organized as follows. A literature review of the most common learning curve processes is presented in the next section, followed by methodology and model formulation. We then provide an in-depth analysis of the learning curve models, followed by future research directions, conclusions, and limitations of this research.

Literature Review

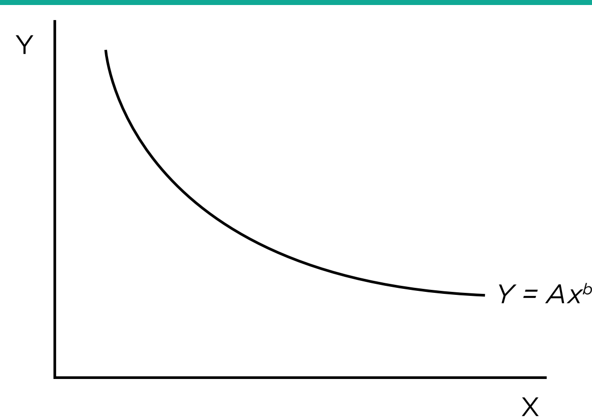
Learning curve research dates back to 1936, when Theodore Paul Wright published the original learning curve equation that predicted the production effects of learning. Wright recognized the mathematical relationship that exists between the time it takes for a worker to complete a single task and the number of times the worker had previously performed that task (Wright, 1936). The mathematical relationship developed from this hypothesis is that as workers complete the same process, they get better at it. Specifically, Wright realized that the rate at which they get better at that task is constant.

The relationship between these two variables is as follows: as the number of units produced doubles, the worker will do it faster by a constant rate. He proposed that this relationship takes the form of:

$$F = N^x \quad \text{or} \quad x = \frac{Log\,F}{Log\,N};$$

"where F = a factor of cost variation proportional to the quantity N. The reciprocal of F then represents a direct percent variation of cost vs. quantity" (Wright 1936). The relationship between these variables can be modified to predict the expected cost of a given unit number in production by multiplying the factor of cost variation by the theoretical cost of the first unit produced—this relationship was stated in Equation 2 and is shown in Figure 2. It is a log linear relationship through an algebraic manipulation. The logarithmic form of this equation (taking the natural log of both sides of the equation) allows practitioners to run linear regression analysis on the data to find what slope best fits the data using a straight line (Martin, n.d.).

FIGURE 2. WRIGHT'S LEARNING CURVE MODEL



Note. (Martin, n.d.)

The goal of using learning curves is to increase the accuracy of cost estimates. Having accurate cost estimates allows an organization to efficiently budget while providing as much operational capability as possible because it can allocate resources to higher priorities. While the use of learning curves focuses on creating accurate cost estimates, learning curves often use the number of labor hours it takes to perform a task. When Wright originated the theory, he proposed the output in terms of time to produce, not production cost. However, many organizations perform learning curve analysis on both production cost and time to produce, depending on the data available. Nevertheless, labor hour cost is relevant because it is based on factors such as labor rates and other associated values. The use of labor hours in learning curve development allows a common comparison over time without the effects of inflation convoluting the results. However, the same goal can be achieved by using inflation-adjusted cost values.

Wright's model has been compared to some of the more contemporary models that have surfaced in recent years since the original learning curve theory was established (Moore et al., 2015). Moore compared the Stanford-B, Dejong, and the S-Curve models to Wright's model to see if any of these functions could provide a more accurate estimate of the learning phenomenon. Both the Dejong and the S-curve models use an incompressibility factor in the calculation. Incompressibility is a factor used to account for the percentage of automation in the production process. Values of the incompressibility factor can range from zero to one where zero is all touch labor and one is complete automation. Moore found that when using an incompressibility factor between zero and 0.1, the Dejong and S-Curve models were more accurate (Moore et al., 2015). In other words, when a production process had very little automation and high amounts of touch labor, the newer learning curve models tended to be more accurate. For all other values of incompressibility, Wright's model was more accurate.

More recently, Johnson (2016) proposed that a flattening effect is evident at the end of the production process where learning does not continue to occur at a constant rate near the end of a production cycle. Using the same models as Moore, Johnson explored the difference in accuracy between Wright's model and contemporary models early in the production process versus later in the production process. He had similar findings to Moore in that Wright's model was most accurate except in cases where the incompressibility factors were extremely low. When the incompressibility factor is low, more touch labor is involved in the process allowing for the possibility of additional learning to occur. He also found that Wright's learning curve was more accurate early in the production process whereas the Dejong and S-Curve models were more accurate later in the production process (Johnson, 2016).

Another key concept in learning curve estimation and modeling is the idea of a forgetting curve (Honious et al., 2016). A forgetting curve explains how configuration changes in the production process can cause a break in learning, which leads to loss of efficiency that had previously been gained. When a configuration change occurs, the production process changes. Changes may include factors such as using different materials, different tooling, adding steps to a process, or might even be attributed to workforce turnover. The new process affects how workers complete their tasks and causes previously learned efficiencies to be lost. If manufacturers fail to take these breaks into account, they may underestimate the total effort needed to produce a product. Honious et al. (2016) found that configuration changes significantly changed the learning curve, and that the new learning curve slope was steeper than the previous steady slope prior to a configuration change. The distinction between pre- and post-configuration change is important to ensure both types of effects are taken into account.

> **"When a production process had very little automation and high amounts of touch labor, the newer learning curve models tended to be more accurate. For all other values of incompressibility, Wright's model was more accurate.**

The International Cost Estimating and Analysis Association (ICEAA) published learning curve training material in 2013. While presenting the basics of learning curve theory, it also presented some rules of thumb for learning. The first rule is that learning curves are steepest when the production process is touch-labor intensive. Conversely, learning curves are the flattest when the production process is highly automated (ICEAA, 2013). Another key piece of information is that adding new work to the process can affect the overall cost. ICEAA states that this essentially adds a new curve for the added work, which increases the original curve by the amount of the new curve (ICEAA, 2013). The equation is as follows:

$$Cost(x) = A_1 x^{b_1} + A_2 (x-L)^{b_2} \tag{3}$$

Where:

$Cost(x)$ = cumulative average cost per unit

A_1 = theoretical cost to produce the first unit prior to addition of new work

x = cumulative number of units produced

L = last unit produced before addition of new work

A_2 = theoretical cost to produce the first unit after addition of new work

$$b_1 = \frac{\ln \textit{Learning Curve Slope prior to additional work}}{\ln 2}$$

$$b_2 = \frac{\ln \textit{Learning Curve Slope prior to additional work}}{\ln 2}$$

(typically same as b_1)

Equation 3 is important to consider when generating an estimate after a major configuration change or engineering change proposal (ECP). For example, while producing the eighth unit of an aircraft, the customer realizes they need to drastically change the radar on the aircraft. Learning has already taken place on the first eight aircraft; the new radar has not yet been installed, and therefore no learning has taken place. To accurately take into account the new learning, the radar would be treated as a second part to the equation, ensuring we account for the learning on the eight aircraft while also accounting for no learning on the new radar.

Lastly, Anderlohr (1969) and Mislick and Nussbaum (2015) write about production breaks and the effects they have on a learning curve. These production breaks can cause a direct loss of learning, which can fully or partially reset the learning curve. For example, a 50% loss of learning would result in a loss of half of the cost reduction that has occurred (ICEAA, 2013). This information is important when analyzing past data to ensure that breaks in production are accounted for.

Thus far, we have laid out the fundamental building blocks for learning curve theory and how they might apply in a production environment. Wright's learning curve formula established the method by which many organizations account for learning during the procurement process. Following Wright's findings, other methods have emerged that account for breaks in production, natural loss of learning over time, incompressibility factors, and half-life analysis (Benkard, 2000). This article adds to the discussion by examining the flattening effect and how various models predict learning at different points in the production process.

When examining learning curve theory and the effects learning has on production, it is critical to understand the production process being estimated. Since Wright established learning curve theory in 1936, factory automation and technology have grown tremendously and continue to grow. Contemporary learning curve methods try to account for this automation. To get the best understanding, we will examine the aircraft industry, specifically how it behaves in relation to the rest of the manufacturing industry.

The aircraft industry has relatively low automation (Kronemer & Henneberer, 1993), especially compared to other industries. Kronemer and Henneberer (1993) state that the aircraft industry is a fairly labor-intensive process with relatively little reliance on automated production techniques, despite it being a high-tech industry. Specifically, they list three main reasons why manufacturing aircraft is so labor-intensive. First, aircraft manufacturers usually build multiple models of the same aircraft, typically for the commercial sector alone. These different aircraft models mean different tooling and configurations are needed to meet the demand of the customer. Second, aircraft manufacturers deal with a very low unit volume when compared to other industries in manufacturing. The final reason

> **Following Wright's findings, other methods have emerged that account for breaks in production, natural loss of learning over time, incompressibility factors, and half-life analysis (Benkard, 2000).**

for low levels of automation is the fact that aircraft are highly complex and have very tight tolerances. To attain these specifications, manufacturers must continue to use highly skilled touch laborers or spend extremely large amounts of money on machinery to replace them (Henneberer & Kronemer, 1993). For these reasons, we should typically see or use low incompressibility factors in the learning curve models when estimating within the aircraft industry.

Although the aircraft industry remains largely unaffected by the shift to machine production from human touch labor, many industries are seeing a rise in the percentage of manufacturing processes that are automated. In a *Wall Street Journal* article posted in 2012, the author showed how companies have been increasing the amount of money spent on machines and software while spending less on manpower. They proposed that part of the reason behind this shift was a temporary tax break "that allowed companies in 2011 to write off 100% of investments in the first year" (Aeppel, 2012). Tax breaks combined with extremely low interest rates provided industry with incentive to invest in future production. Investment in production technology increases the incompressibility factor that should be used when estimating the effects of learning. In a separate article for the *Wall Street Journal*, Kathleen Madigan also pointed out the increase in spending on capital investments in relation to labor. She stated that "businesses had increased their real spending on equipment and software by a strong 26%, while they have added almost nothing to their payrolls" (Madigan, 2011).

Methodology

Model Formulation

Before we can begin the process of developing a new learning curve equation, we need to examine the characteristics of the curve we expected to best fit the data. Our hypothesis is that a learning curve whose slope decreases over time would fit the data better than Wright's curve. To adjust the rate at which the curve flattens, the *b* value from Wright's learning curve, or the exponent in the power function, needs to be adjusted. Specifically, to make the curve move in a flatter direction, the exponent in the power curve must decrease as the number of units produced increases. Initially we modified Wright's existing formula by dividing the exponent by the unit number as shown in Equation 4.

$$Cost(x) = Ax^{b/x} \tag{4}$$

Where:

$Cost(x)$ = cumulative average cost per unit

A = theoretical cost of the first unit

x = cumulative number of units produced

b = Wright's learning curve constant as described in Equation 2

Using Wright's learning curve, b is a negative constant that has a larger magnitude for larger amounts of learning (i.e., as LCS decreases, b becomes more negative). Therefore, in Equation 4, when b is divided by x, the amount of learning is reduced. In fact, the flattening effect is fairly drastic. For example, when applying Equation 4, a standard 80% Wright's learning curve exhibits 90% learning by the second unit and flattens to 97% by the fourth unit. To implement a learning curve that has the flexibility to not flatten as quickly, we instead divide b by $1+x/c$ where c is a positive constant (see Equation 5). The term $1+x/c$ is always greater than 1 and is increasing as x increases; therefore, a flattening effect always occurs (i.e., learning decreases as the number of units produced increases). The choice of the constant c is critical in determining how quickly the learning decreases. For example, when $c = 4$, a standard 80% Wright's learning curve exhibits 86% learning by the second unit and approximately 89% learning by the fourth unit. For the same standard 80% curve when $c = 40$, the learning decreases to 80.9% by the second unit and to 81.6% by the fourth unit. The new equation (which we also refer to as Boone's learning curve hereafter) took the form:

$$Cost(x) = Ax^{b/(1+x/c)} \tag{5}$$

Where:

$Cost(x)$ = cumulative average cost per unit

A = theoretical cost of the first unit

x = cumulative number of units produced

b = Wright's learning curve constant as described in Equation 2

c = decay value (positive constant)

The function that modifies the traditional learning curve exponent in Equation 5—i.e., $1+x/c$ – has a key characteristic—ensures that the rate of learning associated with traditional learning curve theory decreases as each additional unit is produced. Specifically, $1+x/c$ is always greater than 1 since x/c is always positive. Note that c is an estimated parameter and x increases as more units are produced, so the term x/c is decreasing. When c is large, Boone's learning curve would effectively behave like Wright's learning curve. For example, if the fitted value of c is 5,000, then $1+x/c$ equals 1.0002 after the first unit has been produced and 1.004 after the twentieth unit has been produced. This equates to a decrease in the learning rate of the traditional theory (i.e., b) of less than 0.07%. More formally, as c goes to infinity, $b/(1 + x/c)$ goes to b.

Note that the previous discussion assumed that b was the same value for both Wright's and Boone's learning curve to help demonstrate the flattening effect. In practice, nothing precludes each of the learning curves from having different b values. For instance, if we desire a learning curve that possesses more learning early in production and less learning later in production (compared to Wright's curve), then the b parameters could be different—this was shown in Figure 1. In this case, Boone's curve would have a b value less than Wright's curve (i.e., a more negative value representing more learning). Then the flattening effect of dividing by $1+x/c$ as production increases would eventually result in a curve with less learning than Wright's curve. For example, consider an 80% Wright's learning curve and a Boone's learning curve that initially has 70% learning and a decay value of 8; by the eighth production unit, Boone's curve would be at 82% learning.

Population and Sample

To test the new learning curve in Equation 5, we looked at quantitative data from several DoD airframes to gain a comprehensive understanding of how learning affects the cost of lot production. The costs used in this analysis are the direct lot costs and exclude costs for items such as Research, Development, Test, & Evaluation (RDT&E), support items, and spares. These data specifically include Prime Mission Equipment (PME) only as these costs are directly related to labor, and can be influenced directly through learning. To ensure we are comparing properly across time, we used inflation and rate-adjusted PME cost data for each production lot of the selected aircraft systems. The PME cost data were adjusted using escalation rates for materials using Office of the Secretary of Defense (OSD) rate tables, when applicable. We used data from fighter, bomber, and cargo aircraft, as well as missiles and munitions. This diverse dataset allowed comparison among multiple systems in different production environments.

Data Collection

Data used were gathered from the Cost Assessment Data Enterprise (CADE). CADE is a resource available to DoD cost analysts that stores historical data on weapon systems. Some of the older data also came from a DoD research library in the form of cost summary reports. The data used can be broken out by Work Breakdown Structure (WBS) or Contract Line Item Number (CLIN). For this research, the PME cost data were broken out by WBS element, then rolled up into top line, finished product elements and used for the regression analysis. In total, 46 weapon system platforms were analyzed (see Table).

TABLE. RESULTS

PROGRAM	Wright's SSE	Wright's MAPE	Boone SSE	Boone MAPE	SSE Difference	MAPE Difference
Platform A	2.78E+08	5.3%	2.17E+08	4.8%	-22%	-10%
Platform B	4.88E+08	5.4%	4.90E+08	5.6%	0%	5%
Platform C	1.58E+07	10.8%	4.51E+05	2.1%	-97%	-80%
Platform D	6.56E+10	22.1%	6.02E+10	24.5%	-8%	11%
Platform E	1.14E+09	6.2%	1.10E+09	5.6%	-4%	-9%
Platform F	1.94E+06	4.6%	1.95E+06	4.6%	0%	1%
Platform G	7.14E+08	13.6%	6.28E+08	12.9%	-12%	-5%
Platform H	5.49E+06	4.6%	5.00E+06	4.0%	-9%	-13%
Platform I	1.30E+09	18.6%	1.21E+09	23.8%	-7%	28%
Platform J	7.90E+06	3.9%	6.12E+06	3.6%	-23%	-8%
Platform K	2.18E+07	6.0%	7.48E+06	3.2%	-66%	-47%
Platform L	1.06E+08	9.6%	1.05E+08	9.7%	-1%	0%
Platform M	1.49E+07	10.7%	1.48E+07	13.4%	0%	26%
Platform N	9.92E+08	16.3%	7.67E+07	10.0%	-92%	-39%
Platform O	1.81E+08	13.0%	1.78E+08	14.0%	-1%	7%

TABLE. RESULTS (CONTINUED)

PROGRAM	Wright's SSE	Wright's MAPE	Boone SSE	Boone MAPE	SSE Difference	MAPE Difference
Platform P	1.71E+07	6.3%	7.96E+06	4.7%	-53%	-26%
Platform Q	8.00E+06	10.1%	4.11E+06	7.6%	-49%	-25%
Platform R	1.48E+09	18.8%	1.31E+09	18.3%	-12%	-2%
Platform S	5.00E+07	6.2%	4.89E+07	6.1%	-2%	-2%
Platform T	4.01E+07	11.1%	5.45E+06	6.5%	-86%	-41%
Platform U	1.19E+06	8.8%	1.34E+06	7.8%	13%	-11%
Platform V	1.60E+09	10.6%	1.74E+02	0.0%	-100%	-100%
Platform W	1.39E+09	6.4%	1.38E+09	6.4%	-1%	0%
Platform X	7.61E+08	18.1%	3.18E-01	0.0%	-100%	-100%
Platform Y	6.81E+05	3.3%	1.10E+06	4.1%	4.1%	26%
Platform Z	2.12E+06	7.5%	1.57E+06	6.8%	6.8%	-9%
Platform AA	2.66E+07	5.0%	2.73E+07	5.5%	5.5%	10%
Platform AB	1.48E+09	18.8%	1.31E+09	18.3%	18.3%	-2%
Platform AC	3.81E+07	5.9%	2.45E+07	4.5%	4.5%	-24%
Platform AD	3.03E+11	21.9%	1.34E+11	16.7%	16.7%	-24%
Platform AE	1.04E+09	10.0%	1.03E+09	10.3%	10.3%	3%
Platform AF	9.01E+05	5.1%	6.94E+05	4.0%	4.0%	-23%
Platform AG	8.20E+06	5.9%	1.77E+06	3.7%	3.7%	-37%
Platform AH	6.40E+06	10.8%	6.11E+06	9.8%	9.8%	-9%
Platform AI	1.47E+07	8.2%	5.22E+06	5.4%	5.4%	-35%
Platform AJ	4.95E+07	10.0%	4.98E+07	10.7%	10.7%	6%
Platform AK	5.99E+07	19.8%	5.69E+07	20.4%	20.4%	3%
Platform AL	1.50E+10	12.9%	1.43E+10	14.8%	14.8%	15%
Platform AM	1.29E+07	5.5%	1.28E+07	5.4%	5.4%	-3%
Platform AN	4.99E+06	3.7%	3.02E+06	3.4%	3.4%	-9%
Platform AO	9.63E+07	21.9%	9.45E+07	21.5%	21.5%	-2%
Platform AP	1.18E+06	3.1%	1.22E+06	3.4%	3.4%	7%
Platform AQ	2.77E+03	3.4%	1.19E-05	0.0%	0.0%	-100%
Platform AR	1.84E+06	17.3%	1.82E+06	18.0%	18.0%	4%
Platform AS	3.27E+06	1.3%	1.09E+00	0.0%	0.0%	-100%
Platform AT	1.98E+03	2.8%	1.19E+03	1.7%	1.7%	-40%

Note. The actual names of each system and contractor have been removed and replaced with a designator of Platform A…Platform AT.

Analysis

Regression analysis was used to test which learning curve model was most accurate in estimating the data. The goal is to minimize the sum of squared errors (SSE) in the regression to examine how well a model estimates a given set of data. The SSE is calculated by taking the vertical distance between the actual data point (in this case lot midpoint PME cost) and the prediction line (or estimate) (Mislick & Nussbaum, 2015). This error term is then squared and the sum of these squared error terms is the value for comparing which model is a more accurate predictor. However, since an extra parameter is available in fitting the regression for the new model, it should be able to maintain or decrease the SSE in most cases. As previously mentioned, as the decay parameter in Equation 5 approaches infinity, Boone's learning curve approaches Wright's learning curve formula. With this in mind, we also examined the Mean Absolute Percentage Error (MAPE). MAPE takes the same error term as the SSE calculation but then divides it by the actual value; then the mean of the absolute value of these modified error terms is calculated. By examining the error in terms of a percentage, comparisons between different types and sizes of systems are more robust. If Boone's curve reduces both SSE and MAPE when compared to the SSE and MAPE of Wright's curve, it would indicate the new model may be better suited for modeling learning and the associated costs.

As stated previously, Wright's learning curve is suitable for a log-log model. A log-log model is used when a logarithmic transformation of both sides of an equation results in a model that is linear in the parameters. As Wright

proposed, this linear transformation occurs because learning happens at a constant rate throughout the production cycle. If learning happens at a non-constant rate (as in Boone's learning curve), then the curve in log-log space would no longer be linear. This constraint means typical linear regression methods would not be suitable for estimating Boone's learning curve; therefore, we had to use nonlinear methods to fit these curves.

Specifically, we used the Generalized Reduced Gradient (GRG) nonlinear solver package in Excel to minimize the SSE by fitting the A, b, and c parameters from Equation 5. To use this solver, bounds for the three parameters had to be established. These are values that are easy to obtain for any dataset, as they are provided by Microsoft Excel when fitting a power function or by using the "linest()" function in Excel. We used this as a starting point because Wright's curve is currently used throughout the DoD. For the A variable, the lower bound was one-half of Wright's A and the upper bound was 2 times Wright's A. These values were used to give the solver model a wide enough range to avoid limiting the value but small enough to ease the search for the optimal values. Neither of these limits was found to be binding. For the exponent parameter b, we chose values between 3 and -3 times Wright's exponent value. In theory, the value of the exponent should never go above 0 due to positive learning leading to a decrease in cost, but in practice some datasets go up over time and we wanted to be able to account for those scenarios, if necessary. Again, these values between 3 and -3 times Wright's exponent value were never found to be binding limits for the model. Finally, for the decay parameter c, fitted values were bounded between 0 and 5,000; the 5,000 upper bound was found to be a binding constraint in the solver on several occasions. In practice, analysts could bound the value as high as possible to reduce error, but in the case of this research, we used 5,000 as no significant change was evidenced from 5,000 to infinity—relaxing this bound would have only further reduced the SSE for Boone's learning curve.

Statistical Significance Testing

Once the SSE and MAPE values were calculated for each learning curve equation, we tested for significance to determine whether the difference between the error values for the two equations were statistically different.

Specifically, we conducted t-tests on the differences in error terms between Wright's and Boone's learning curve equations. This *t*-test was conducted for both SSE and MAPE values separately. A nonsignificant *t*-test indicates no statistically significant difference between the two learning curves.

Analysis and Results

The Table shows the SSE and MAPE values for both Wright's and Boone's learning curve for each system in the dataset. The last two columns are the percentage difference in SSE and MAPE between the two learning curve methods. This percentage was calculated by taking the difference of Boone's error term minus Wright's error term divided by Wright's error term. Negative values represent programs where Boone's learning curve had less error than Wright's learning curve, and positive values represent programs where Wright's curve had less error than Boone's curve.

Based on this analysis, we observed that Boone's learning curve reduced the SSE in approximately 84% of programs and reduced MAPE in 67% of programs. The mean reduction of SSE and MAPE was 27% and 17%, respectively. As previously mentioned, these values were based on using both learning curve equations to minimize the SSE for each system in the dataset. This is standard practice in the DoD as prescribed by the U.S. Government Accountability Office (GAO, 2009) *Cost Estimating and Assessment Guide* when predicting the cost of subsequent units or subsequent lots.

> **"The incompressibility factor represents the amount of automation in the production process, which limits how much learning can occur (Badiru et al., 2013).**

We conducted additional tests to determine if a statistical difference existed between the means of both curve estimation techniques. On average, programs estimated using Boone's learning curve had a lower error rate (M = 4.73, SD = 2.15) than those estimated using Wright's learning curve (M = 8.64, SD = 4.55). Additionally, the difference between these two error rates expressed as a percentage and compared to a hypothesized value of 0 (no difference) was significant, $t(46) = -4.87$, $p < .0001$, and represented an effect of $d = 1.10$. We then applied the same test to the difference in the MAPE values from Boone's learning curve and Wright's learning curve. On average, programs estimated using Boone's learning curve had a lower MAPE value (M = .08, SD = .07) than those estimated using Wright's MAPE value (M =

.10, SD = .06). The difference between these two estimates has a mean of -.17, which translates to Boone's curve reducing MAPE by 17% more on average. Additionally, the difference between these two error rates expressed as a percentage and compared to a hypothesized value of 0 (no difference) was significant, $t(46) = -3.48$, $p < .0005$, and represented an effect of $d = .22$. The results indicate that in both SSE and MAPE, Boone's learning curve reduced the error, and that each of those values was statistically significant when using an alpha value of 0.05.

Discussion

As stated previously, an average of a 27% reduction in the SSE resulted from among the 46 programs analyzed. These results were statistically significant. Also, a 17% reduction in the MAPE resulted from among the programs analyzed, which was also found to be statistically significant. Based on these results, we can conclude that Boone's learning curve equation was able to reduce the overall error in cost estimates for our sample. This information is critical to allow the DoD to calculate more accurate cost estimates and better allocate its resources. These conclusions help answer our three guiding research questions. Specifically, we were looking for the point where Wright's model became less accurate than other models. We found that adding a decay factor caused the learning curve to flatten out over time, which resulted in less error than Wright's model. Additionally, we found that Boone's learning curve was more accurate throughout the entire production process, not just during the tail end when production was winding down. Boone's learning curve was steeper during the early stages of production when it's hypothesized that the most learning occurs. Toward the end of the

> **Future research should identify decay values for different types of weapon systems—similar to the way learning curve rates are established for different categories of programs.**

production process, Boone's curve flattens out more than Wright's curve, supporting our contention that learning toward the end of the production cycle yields diminishing returns. While Wright's curve assumes constant learning throughout the entire process, Boone's curve treats learning in a nonlinear fashion that slows down over time. By reducing the error in the estimates and properly allocating resources, the DoD could potentially minimize risk for all parties involved. The benefit of Boone's learning curve is accuracy in the estimation process. If labor estimates aren't accurate in the production process, risks escalate, such as schedule slip, cost overruns, and increased costs for all involved. Accuracy in the cost estimate should be the goal of both the contractor and government, thereby facilitating the acquisition process with better data.

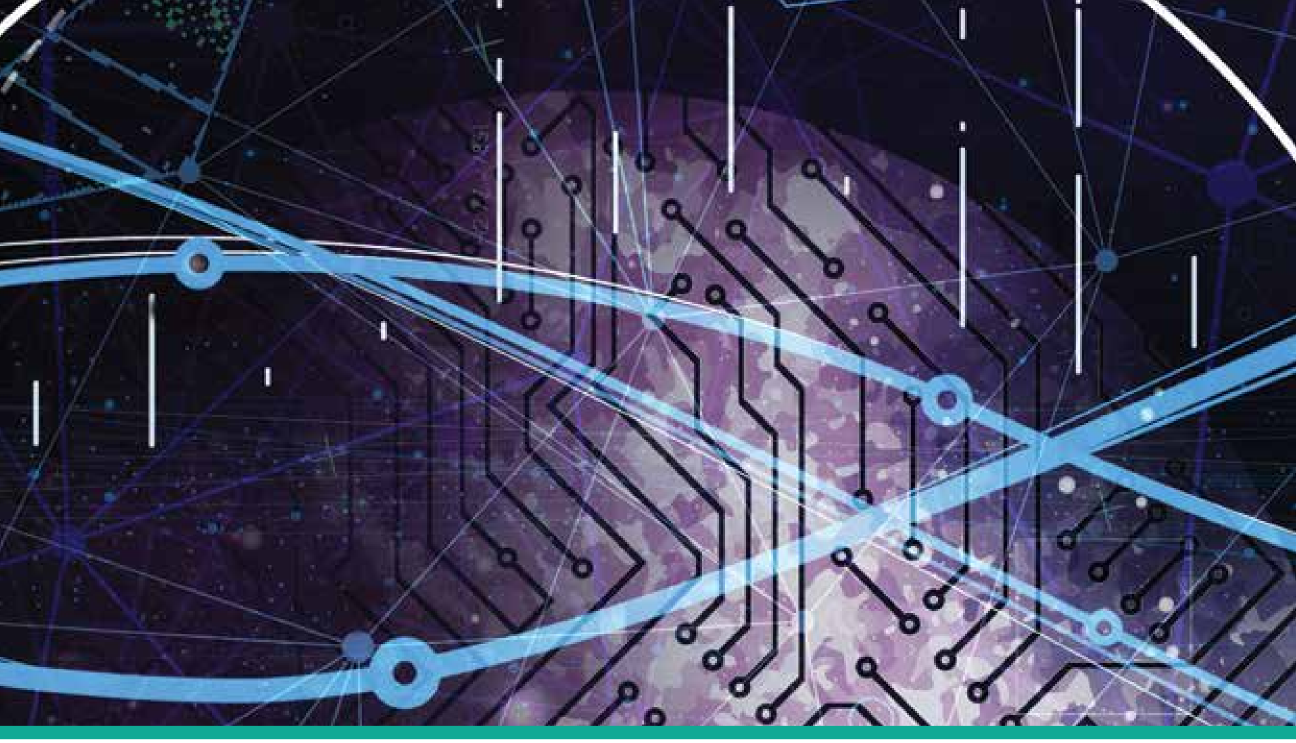
Limitations

One limitation of this study is that all 46 of the weapon systems analyzed were U.S. Air Force systems. While the list included many platforms spanning decades, we hesitate to draw conclusions outside of the U.S. Air Force without further research and analysis. That said, we see no reason our model wouldn't apply equally well in any aircraft production environment, both within and outside the DoD. Another limitation in this research is the use of PME cost as opposed to labor hours. Labor-hour data are not readily available across many platforms, which led to the use of PME cost. Contractor data provided to the government normally come in the form of lots, which is the lowest level tracked by cost estimators. To compare learning curves across multiple platforms, the same level of analysis is required to ensure a fair comparison. Future research should attempt to examine data at the individual level of analysis between systems and exclude those where only lot data are available. Because there are inherently less lots than units, this may affect how the equation behaves when applied at the unit level. For this research, we used the lot midpoint formula/method (Mislick & Nussbaum, 2015), but further research should be conducted to evaluate the performance of Boone's learning curve with unitary data. Finally, we only performed a comparison to Wright's learning curve since that is a primary method of estimation in the DoD. A comparison with other learning curve models may yield different results, although previous research found those curves were not statistically better than Wright's.

Recommendations for Future Research

Data outside of the U.S. Air Force should be examined to test whether this equation applies broadly to programs, and not just to Air Force programs. Also, conducting the analysis with unitary data could confirm that this works for predicting subsequent units as well as subsequent lots, while reducing error over Wright's method. We also made an attempt to select weapon systems that had minimal automation in the production process. However, DeJong's Learning Formula is another derivation from Wright's original in which an incompressibility factor is introduced. The incompressibility factor represents the amount of automation in the production process, which limits how much learning can occur (Badiru et al. 2013). Other models such as the S-Curve model (Carr, 1946) and a more recent version (Towill, 1990; Towill & Cherrington, 1994) also account for some form of incompressibility. Additional research could also include modifications to Boone's formula to try and further reduce the error types listed in this research. Furthermore, fitting Boone's curve in this analysis was based on past data whereas cost estimates are used to project future costs. Therefore, future research should identify decay values for different types of weapon systems—similar to the way learning curve rates are established for different categories of programs. Lastly, further research could examine whether the incorporation of multiple learning curve equations at different points in the production process would be beneficial to reducing additional error in the estimates.

We developed a new learning curve equation utilizing the concept of learning decay. This equation was tested against Wright's learning equation to see which equation provided the least amount of error when looking at both the SSE and MAPE. We found that Boone's learning curve reduced error in both cases and that this reduction in error was statistically significant. Follow-on research in this field could lead to further discoveries and allow for broader use of this equation in the cost community.



References

Aeppel, T. (2012, January 17). Man vs. machine: Behind the jobless recovery. *Wall Street Journal.* http://www.wsj.com/articles/SB10001424052970204468004577164710231081398

Anderlohr, G. (1969). What production breaks cost. *Industrial Engineering, 1*(9), 34–36.

Badiru, A., Elshaw, J., & Mack, E. (2013). Half-life learning curve computations for airframe life-cycle costing of composite manufacturing. *Journal of Aviation and Aerospace Perspective, 3*(2), 6–37.

Benkard, C. L. (2000). Learning and forgetting: The dynamics of aircraft production. *American Economic Review, 90*(4), 1034–1054. https://www.aeaweb.org/articles?id=10.1257/aer.90.4.1034

Carr, G. W. (1946). Peacetime cost estimating requires new learning curves. *Aviation, 45*(4), 220–228.

Chalmers, G., & DeCarteret, N. (1949). *Relationship for determining the optimum expansibility of the elements of a peacetime aircraft procurement program.* Wright-Patterson AFB, OH: USAF Air Materiel Command.

Crawford, J. R. (1944). *Estimating, budgeting and scheduling.* Burbank, CA: Lockheed Aircraft Co.

De Jong, J. R. (1957). The effects of increasing skill on cycle time and its consequences for time standards. *Ergonomics, 1*(1), 51–60. https://doi.org/10.1080/00140135708964571

Honious, C., Johnson, B., Elshaw, J., & Badiru, A. (2016, May 4–5). The impact of learning curve model selection and criteria for cost estimation accuracy in the DoD. In Maj Gen C. Blake, USAF (Chair), *Proceedings of the 13th Annual Acquisition Research Symposium,* Monterey, CA. https://apps.dtic.mil/dtic/tr/fulltext/u2/1016823.pdf

International Cost Estimating and Analysis Association (2013). *Cost estimating body of knowledge* [PowerPoint slides]. ICEAA. https://www.iceaaonline.com/cebok/

Johnson, B. J. (2016). *A comparative study of learning curve models and factors in defense cost estimating based on program integration, assembly, and checkout* [Master's thesis]. Air Force Institute of Technology. https://apps.dtic.mil/dtic/tr/fulltext/u2/1056447.pdf

Kronemer, A, & Henneberger, J. E. (1993). Productivity in aircraft manufacturing. *Monthly Labor Review, 16*(6), 24–33. http://www.bls.gov/mfp/mprkh93.pdf

Madigan, K. (2011, September 28). It's man vs. machine and man is losing. *Wall Street Journal.* https://blogs.wsj.com/economics/2011/09/28/its-man-vs-machine-and-man-is-losing/

Martin, J. R. (n.d). *What is a learning curve?* Management and Accounting Web. http://maaw.info/LearningCurveSummary.htm

Mislick, G. K., & Nussbaum, D. A. (2015). *Cost estimation: Methods and tools.* Wiley. https://www.academia.edu/24430098/Cost_Estimation_Methods_and_Tools_by_Gregory_K_Mislick_and_Daniel_A_Nussbaum_1st_Edition

Moore, J. R., Elshaw, J. Badiru, A. B., & Ritschel, J. D. (2015, October). Acquisition challenge: The importance of incompressibility in comparing learning curve models. *Defense Acquisition Research Journal, 22(*4), 416–449. https://www.dau.edu/library/arj/ARJ/ARJ75/ARJ75-ONLINE-FULL.pdf

Towill, D. R. (1990). Forecasting learning curves. *International Journal of Forecasting, 6*(1), 25–38. https://www.sciencedirect.com/science/article/abs/pii/016920709090095S

Towill, D. R., & Cherrington, J. E. (1994). Learning curve models for predicting the performance of AMT. *The International Journal of Advanced Manufacturing Technology, 9*, 195–203. https://doi.org/10.1007/BF01754598

U.S. Government Accountability Office. (2009). *GAO cost estimating and assessment guide* (GAO-09-3SP). Government Printing Office. https://www.gao.gov/new.items/d093sp.pdf

Wright, T. P. (1936). Factors affecting the cost of airplanes. *Journal of the Aeronautical Sciences, 3*(4), 122–128. https://arc.aiaa.org/doi/10.2514/8.155

Author Biographies

Capt Evan R. Boone, USAF

is a cost analyst at the Air Force Life Cycle Management Center, Air Force Materiel Command, Wright Patterson AFB, Ohio. His research interests include engineering economics and learning curves as applied in a production environment. Capt Boone holds an MS in Cost Analysis from the Air Force Institute of Technology (AFIT), and a BS from the United States Air Force Academy.

(E-mail address: evan.boone.1@us.af.mil)

Dr. John J. Elshaw

is an assistant professor of Systems Engineering in the Department of Systems Engineering and Management, AFIT. 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. Dr. Elshaw is a graduate of the United States Air Force Squadron Officer School and Air Command and Staff College.

(E-mail address: john.elshaw@afit.edu)

Lt Col Clay M. Koschnick, USAF

is an assistant professor, Systems Engineering, in the Department of Systems Engineering and Management, AFIT. His research interests include engineering economics, decision analysis, and econometrics. Lt Col Koschnick holds a PhD in Industrial and Systems Engineering from the University of Florida, a BS in Operations Research from the United States Air Force Academy, and an MS in Operations Research from the Georgia Institute of Technology.

(E-mail address: clay.koschnick@afit.edu)

Dr. Jonathan D. Ritschel

is an assistant professor, Cost Analysis Program, in the Department of Systems Engineering and Management, AFIT. His research interests include public choice, the effects of acquisition reforms on cost growth in DoD weapon systems, research and development cost estimation, and economic institutional analysis. Dr. Ritschel holds a PhD in Economics from George Mason University, a BBA in Accountancy from the University of Notre Dame, and an MS in Cost Analysis from AFIT.

(E-mail address: jonathan.ritschel@afit.edu)

Dr. Adedeji B. Badiru

is dean of the Graduate School of Engineering and Management, AFIT. He is a registered professional engineer, a Fellow of the Institute of Industrial Engineers, and a Fellow of the Nigerian Academy of Engineering. He holds a BS in Industrial Engineering; an MS in Mathematics and an MS in Industrial Engineering from Tennessee Technological University; and a PhD in Industrial Engineering from the University of Central Florida.

(E-mail address: adedeji.badiru@afit.edu)

PROFESSIONAL READING LIST

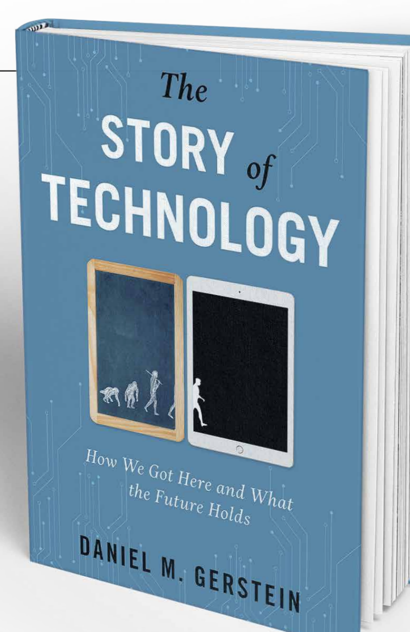
The Defense Acquisition Professional Reading List is intended to enrich the knowledge and understanding of the civilian, military, contractor, and industrial workforce who participate in the entire defense acquisition enterprise. These book recommendations are designed to complement the education and training vital to developing essential competencies and skills of the acquisition workforce. Each issue of the *Defense Acquisition Research Journal* will include one or more reviews of suggested books, with more available on our Website: http://dau.edu/library.

We encourage our readers to submit book reviews they believe should be required reading for the defense acquisition professional. The books themselves should be in print or generally available to a wide audience; address subjects and themes that have broad applicability to defense acquisition professionals; and provide context for the reader, not prescriptive practices. Book reviews should be 450 words or fewer, describe the book and its major ideas, and explain its relevancy to defense acquisition. Please send your reviews to the managing editor, *Defense Acquisition Research Journal* at DefenseARJ@dau.edu.

Featured Book

The Story of Technology: How We Got Here and What the Future Holds

Author: Dr. Daniel M. Gerstein
Publisher: Prometheus Books
Copyright Date: 2019
Hard/Softcover/Digital: Hardcover, 360 pages
ISBN-13: 978-1633885783
Reviewed by: Janel C. Wallace, Professor of Contract Management, DAU



Review:

In defense acquisition, the government must become a better consumer by taking advantage of technology initially developed for civilian applications, being opportunistic, and collaborating. Democratization of technology is not considered an unusual concept. Research and development (R&D) is becoming more specialized. The government will find itself investing in R&D at the later, more mature technology readiness levels. Our focus should align with the highest priorities of the DoD including the use of breakthrough technologies for national security. Consideration should be given to using other transaction authority, where appropriate, for added flexibility in government procurement and to the use of public-private partnerships to achieve innovation.

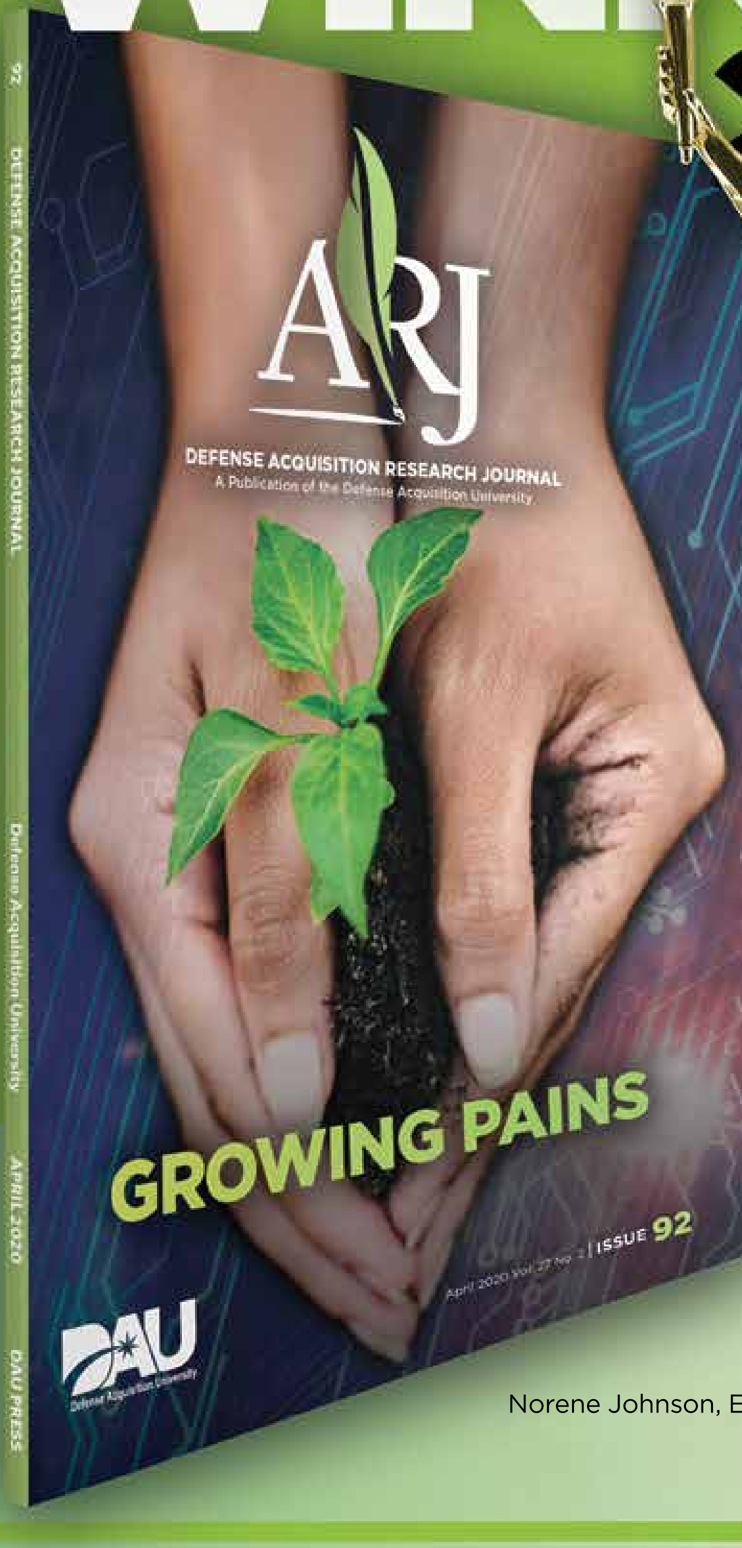
Gerstein discusses the evolution of technology and points of inflection, which is where our society fundamentally changes. This inflection can be good, bad, or a mix of both. Technologies today are combinations and recombinations of other technologies, put together for practical reasons to achieve operational needs. Gerstein points out that it's through the science and technology, R&D, innovation, and transformation—all of which are considered technology development—that integration, coordination, and synchronization of the system of systems actually occur. When technology development is connected, it relates directly to an operational need where developers work closely with operators to identify real-world problems and look for solutions. When technology development is disconnected, it is considered to have no practical purpose. Diverse talents, experience, and world views allow for more robust thinking. Different groupings of personalities can introduce nonlinear or orthogonal—also known as "outside the box"—thinking into the solution of a problem.

Progress impacts our decisions, causing us to respond to technology advances before the potential impacts are fully considered. Leadership attention and additional resources, such as money, may not accelerate development to maturity, even when desired. In finding vaccines, there is a lengthy period for demonstrating safety and efficiency. Fielding not yet proven technologies under development introduces significant risk. It is suggested that nations assess technology risks to their national security, economic competitiveness, and societal well-being.

Gerstein posits that technological surprises will continue to occur, resulting in disruptive technologies, which change how we normally conduct business. This may result in intentional misuse and accidents but possibly other uses. Our increased knowledge and technology in robotics, autonomous systems, and artificial intelligence could change what it means to be human, displace the workforce, and impact our expectations of privacy. In the future, the workforce will be technologists having a broad mix of skills. They will need to be innovators with vision and entrepreneurial spirit yet responsible to society. Just because a technology can be developed or fielded does not mean it is beneficial or worth the costs. Generally, society reacts only when potential dangers of a technology have been demonstrated. Technologists should apply the proper tools and techniques; provide accurate, reliable, and reproducible research for building future knowledge; and be made to expose all good and bad uses for their technology.

MARCOM AWARD FOR PEER REVIEWED ACADEMIC JOURNAL

WINNER 2020



Norene Johnson, Emily Beliles, and Michael Krukowski

DAU Press

Fort Belvoir, VA

Current Research Resources in DEFENSE ACQUISITION

MIDDLE TIER ACQUISITION

Each issue of the *Defense Acquisition Research Journal* will bring to the attention of the defense acquisition community a topic of current research, which has been undertaken by the DAU Knowledge Repository (KR) librarian team in collaboration with DAU's Director of Research. Both government civilian and military Defense Acquisition Workforce (DAW) readers will be able to access papers publicly and from licensed resources on the DAU KR Website: https://identity.dau.edu/EmpowerIDWebIdPForms/Login/KRsite.

Nongovernment DAW readers should be able to use their local knowledge management centers/libraries to download, borrow, or obtain copies. We regret that DAU cannot furnish downloads or copies.

Defense Acquisition Research Journal readers are encouraged to submit proposed topics for future research by the DAU Knowledge Repository librarian team. Please send your suggestion with a short write-up (less than 100 words) explaining the topic's relevance to current defense acquisition to: Managing Editor, *Defense Acquisition Research Journal*, DefenseARJ@dau.edu.

Defense Acquisitions Annual Assessment: Drive to Deliver Capabilities Faster Increases Importance of Program Knowledge and Consistent Data for Oversight

Shelby S. Oakley

Summary:

In response to section 833 of the National Defense Authorization Act (NDAA) for Fiscal Year (FY) 2019, this report provides insight into 121 of the Department of Defense (DoD)'s most costly weapon and information technology (IT) acquisition programs. Specifically, this report covers the following four sets of programs:

- 85 major defense acquisition programs (MDAP),
- 8 future MDAPs,
- 13 middle-tier acquisition (MTA) programs, and
- 15 major IT programs.

APA Citation:

Oakley, S. S. (2020). *Defense acquisitions annual assessment: Drive to deliver capabilities faster increases importance of program knowledge and consistent data for oversight* (Report No. GAO-20-439). U.S. Government Accountability Office. https://www.gao.gov/a ssets/710/707359.pdf

Assessing the Industrial Base Implications of the Army's Future Vertical Lift Plans

Rhys McCormick and Andrew Philip Hunter

Summary:

This paper presents a detailed analysis of the industrial base implications of the Army's approach to vertical lift modernization. It begins by examining the Army's addressable market for the vertical lift aircraft industry. Second, it looks at the opportunities and challenges presented by restructuring and optimizing the rotocraft industrial base. Next, it looks at what the Army's use of new approaches for FVL—such as modular open systems architecture (MOSA) and Middle Tier of Acquisition (MTA)—might mean for key industry dynamics.

APA Citation:

McCormick, R., & Hunter, A. P. (2020). *Assessing the industrial base implications of the Army's future vertical lift plans*. Center for Strategic and International Studies. https://csis-website-prod.s3.amazonaws.com/s3fs-public/publication/200506_Industrial%20Base%20Army%20FVL_WEB_v3_%20FINAL.pdf

Application of Technology Demonstrations and Prototyping in Middle Tier Acquisitions

Douglas S. Miller

Summary:

The background for this research is in support of an effort to expand the body of acquisition knowledge within a specific region of program management dealing with the 2016–2018 National Defense Authorization Act Middle Tier Section 804 rapid prototyping and rapid fielding initiative. Specifically, the research aims to improve understanding of the nature and role of technology demonstrations and prototyping as acquisition tools supporting rapid prototyping and fielding.

APA Citation:

Miller, D. S. (2019). *Application of technology demonstrations and prototyping in middle tier acquisitions.* DAU Senior Service College Fellowship. https://search.dtic.mil/documents/rest/v1/download?caller=sni-user&id=/citation/TR/AD1074470.xml

Issues with Access to Acquisition Information in the Department of Defense: A Series on Considerations for Managing Program Data in the Emerging Acquisition Environment

Jeffrey A. Drezner, Megan McKernan, Jerry M. Sollinger, and Sydne Newberry

Summary:

This report outlines issues and opportunities in data requirements, governance, and management to strive for more efficient, effective, and informed acquisition while reducing burden and ad hoc data requests. We address general data governance and management challenges, as well as specific challenges associated with the Middle Tier of Acquisition for rapid prototyping and rapid fielding, the Selected Acquisition Report (SAR), and the Defense Acquisition Executive Summary (DAES) process and data.

APA Citation:

Drezner, J. A., McKernan, M., Sollinger, J. M., & Newberry, S. (2020). *Issues with access to acquisition information in the Department of Defense: A series on considerations for managing program data in the emerging acquisition environment.* RAND. https://www.rand.org/pubs/research_reports/RR3130.html

DoD Acquisition Reform: Leadership Attention Needed to Effectively Implement Changes to Acquisition Oversight

Shelby S. Oakley

Summary:

This report addresses (1) the progress DoD has made implementing selected oversight reforms related to major defense acquisition programs; (2) how DoD has used middle-tier acquisition pathways; and (3) challenges DoD faces related to reform implementation. GAO reviewed five reforms: milestone decision authority designation; cost, fielding, and performance goals; independent technical risk assessments; restructuring of acquisition oversight offices; and middle-tier acquisition.

APA Citation:

Oakley, S. S. (2019). *DoD acquisition reform: Leadership attention needed to effectively implement changes to acquisition oversight* (Report No. GAO-19-439). U.S. Government Accountability Office. https://www.gao.gov/assets/700/699527.pdf

A Happy Medium: Middle-tier Acquisition Authority Features Flexible Prototype and Fielding Options

Douglas W. Burbey, Mindy Gabbert, and Kathryn Bailey

Summary:

Situated between the acquisition pathways of "urgent" and "tailorable traditional DoDI 5000.02," middle-tier acquisition is for programs that house mature prototypes from government and industry that should not require much additional development to begin production. In May, the Army Acquisition Executive empowered the Program Executive Office for Command, Control and Communications – Tactical (PEO C3T) to use the process for two of its top efforts—the Integrated Tactical Network and Unified Network Operations—both of which support the Network Cross-Functional Team and Army network modernization initiatives.

APA Citation:

Burbey, D. W., Gabbert, M., & Bailey, K. (2019, October). A happy medium: Middle-tier acquisition authority features flexible prototype and fielding options. *Army AL&T Magazine*. U.S. Army Acquisition Support Center. http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip&db=mth&AN=138948885&site=ehost-live&scope=site

MARCOM AWARD FOR PEER REVIEWED ACADEMIC JOURNAL

WINNER 2020

Norene Johnson, Emily Beliles, and Michael Krukowski

DAU Press

Fort Belvoir, VA


ALUMNI
DAU
ASSOCIATION
NOW ACCEPTING SUBMISSIONS
UNTIL MAY 15, 2021

2021 EDWARD HIRSCH ACQUISITION AND WRITING COMPETITION

CALL FOR PAPERS

Suggested research topics may include, but are not limited to:

- Defense Acquisition Case Histories
- Other Transaction Authorities
- Mid-tier Acquisition
- Agile Software Development and Program Management
- System Cyber Hardness
- Cyber Training and Concepts
- Controlling Costs Throughout the Product Life Cycle
- Acquisition of Defense Business Systems
- Intellectual Property
- Acquisition Readiness
- Emerging Changes to Earned Value Management
- Career Path and Incentives
- Improving Professionalism of the Total Acquisition Workforce
- Incorporating Foreign Military Sales and Direct Contractor Sales Strategies into Programs
- Services Management
- Should Cost Management
- Artificial Intelligence
- Data Visualization
- Digital Twin Engineering

GROUND RULES

- The competition is open to anyone interested in the DoD acquisition system and is not limited to government or contractor personnel.
- Employees of the federal government (including military personnel) are encouraged to compete and are eligible for cash awards unless the paper was researched or written as part of the employee's official duties or was done on government time. If the research effort is performed as part of official duties or on government time, the employee is eligible for a non-cash prize, i.e., certificate and donation of cash prize to a Combined Federal Campaign registered charity of winner's choice.
- ***First place Jacques S. Gansler Award is $1,000***. Second and third prizes, if awarded, are each $500.
- Papers are to be submitted to the DAU Director of Research: ***research@dau.edu.***
- The format of the paper must be in accordance with guidelines for articles submitted for publication in the ***Defense Acquisition Research Journal***.
- Papers will be evaluated by a panel selected by the DAUAA Board of Directors and the DAU Director of Research.
- Award winners will present their papers at a DAU event in June/July 2021 at the DAU Fort Belvoir campus.
- Papers must be ***submitted by May 20, 2021,*** and awards will be announced by June 2021.

JACQUES S. GANSLER AWARD

FIRST PLACE

$1,000 & $500 2nd & 3rd Prize

DON'T MISS OUT ON GETTING PUBLISHED!

Defense ARJ Guidelines **FOR CONTRIBUTORS**

The Defense Acquisition Research Journal (ARJ) is a scholarly peer-reviewed journal published by DAU. All submissions receive a double-blind review to ensure impartial evaluation.

IN GENERAL

We welcome submissions describing original research or case histories from anyone involved in the defense acquisition process. Defense acquisition is broadly defined as any actions, processes, or techniques relevant to as the conceptualization, initiation, design, development, testing, contracting, production, deployment, logistics support, modification, and disposal of weapons and other systems, supplies, or services needed for a nation's defense and security, or intended for use to support military missions.

Research involves the creation of new knowledge. This generally requires either original analysis of material from primary sources, including program documents, policy papers, memoranda, surveys, interviews, etc.; or analysis of new data collected by the researcher. Articles are characterized by a systematic inquiry into a subject to establish facts or test theories that have implications for the development of acquisition policy and/or process.

The *Defense ARJ* also welcomes case history submissions from anyone involved in the defense acquisition process. Case histories differ from case studies, which are primarily intended for classroom and pedagogical use. Case histories must be based on defense acquisition programs or efforts. Cases from all acquisition career fields and/or phases of the acquisition life cycle will be considered. They may be decision-based, descriptive or explanatory in nature. Cases must be sufficiently focused and complete (i.e., not open-ended like classroom case studies) with relevant analysis and conclusions. All cases must be factual and authentic. Fictional cases will not be considered.

We encourage prospective writers to coauthor, adding depth to manuscripts. We recommend that junior researchers select a mentor who has been previously published or has expertise in the manuscript's subject. Authors should be familiar with the style and format of previous *Defense ARJs* and adhere to the use of endnotes versus footnotes, formatting of reference lists, and the use of designated style guides. It is also the responsibility of the corresponding author to furnish any required government agency/employer clearances with each submission.

MANUSCRIPTS

Manuscripts should reflect research of empirically supported experience in one or more of the areas of acquisition discussed above. The *Defense ARJ* is a scholarly research journal and as such does not publish position papers, essays, or other writings not supported by research firmly based in empirical data. Authors should clearly state in their submission whether they are submitting a research article or a case history. The requirements for each are outlined below.

Research Articles

Empirical research findings are based on acquired knowledge and experience versus results founded on theory and belief. Critical characteristics of empirical research articles:

- clearly state the question,
- define the research methodology,

- describe the research instruments (e.g., program documentation, surveys, interviews),
- describe the limitations of the research (e.g., access to data, sample size),
- summarize protocols to protect human subjects (e.g., in surveys and interviews), if applicable,
- ensure results are clearly described, both quantitatively and qualitatively,
- determine if results are generalizable to the defense acquisition community
- determine if the study can be replicated, and
- discuss suggestions for future research (if applicable).

Research articles may be published either in print and online, or as a Web-only version. Articles that are 5,000 words or fewer (excluding abstracts, references, and endnotes) will be considered for print as well as Web publication. Articles between 5,000 and 10,000 words will be considered for Web only publication, with a two sentence summary included in the print version of the *Defense ARJ*. In no case should article submissions exceed 10,000 words.

Case Histories

Care should be taken not to disclose any personally identifiable information regarding research participants or organizations involved unless written consent has been obtained. If names of the involved organization and participants are changed for confidentiality, this should be highlighted in an endnote. Authors are required to state in writing that they have complied with APA ethical standards. A copy of the APA Ethical Principles may be obtained at http://www.apa.org/ethics/.

All case histories, if accepted, will receive a double-blind review as do all manuscripts submitted to the *Defense ARJ*.

Each case history should contain the following components:

- Introduction
- Background
- Characters
- Situation/problem
- Analysis
- Conclusions
- References

Book Reviews

Defense ARJ readers are encouraged to submit book reviews they believe should be required reading for the defense acquisition professional. The reviews should be 500 words or fewer describing the book and its major ideas, and explaining why it is relevant to defense acquisition. In general, book reviews should reflect specific in-depth knowledge and understanding that is uniquely applicable to the acquisition and life cycle of large complex defense systems and services. Please include the title, ISBN number, and all necessary identifying information for the book that you are reviewing as well as your current title or position for the byline.

Audience and Writing Style

The readers of the *Defense ARJ* are primarily practitioners within the defense acquisition community. Authors should therefore strive to demonstrate, clearly and concisely, how their work affects this community. At the same time, do not take an overly scholarly approach in either content or language.

Format

Please submit your manuscript according to the submissions guidelines below, with references in APA format (author date-page number form of citation) as outlined in the latest edition of the *Publication Manual of the American Psychological Association*. References should include Digital Object Identifier (DOI) numbers when available. The author(s) should not use automatic reference/bibliography fields in text or references as they can be error-prone. Any fields should be converted to static text before submission, and the document should be stripped of any outline formatting. All headings should conform to APA style. For all other style questions, please refer to the latest edition of the *Chicago Manual of Style*.

Contributors are encouraged to seek the advice of a reference librarian in completing citation of government documents because standard formulas of citations may provide incomplete information in reference to government works. Helpful guidance is also available in *The Complete Guide to Citing Government Information Resources: A Manual for Writers and Librarians* (Garner & Smith, 1993), Bethesda, MD: Congressional Information Service.

The author (or corresponding author in cases of multiple authors) should attach a cover letter to the manuscript that provides all of the authors' names, mailing and e-mail addresses, as well as telephone numbers. The letter should verify that (1) the submission is an original product of the author(s); (2) all the named authors materially contributed to the research and writing of the paper; (3) the submission has not been previously published in another journal (monographs and conference proceedings serve as exceptions to this policy and are eligible for consideration for publication in the *Defense ARJ*); (4) it is not under consideration by another journal for publication. If the manuscript is a case history, the author must state that they have complied with APA ethical standards in conducting their work. A copy of the APA Ethical Principles may be obtained at http://www.apa.org/ethics/. Finally, the corresponding author as well as each coauthor is required to sign the copyright release form available at our website: www.dau.edu/library/arj.

COPYRIGHT

The *Defense ARJ* is a publication of the United States Government and as such is not copyrighted. We will not accept copyrighted manuscripts that require special posting requirements or restrictions. If we do publish your copyrighted article, we will print only the usual caveats. The work of federal employees undertaken as part of their official duties is not subject to copyright except in rare cases.

Web-only publications will be held to the same high standards and scrutiny as articles that appear in the printed version of the journal and will be posted to the DAU website at www.dau.edu.

In citing the work of others, please be precise when following the author date-page number format. It is the contributor's responsibility to obtain permission from a copyright holder if the proposed use exceeds the fair use provisions of the law (see the latest edition of *Circular 92: Copyright Law of the United States of America and Related Laws Contained in Title 17 of the United States Code,* Washington, DC: U.S. Government Printing Office). Contributors will be required to submit a copy of the writer's permission to the managing editor before publication.

We reserve the right to decline any article that fails to meet the following copyright requirements:

- The author cannot obtain permission to use previously copyrighted material (e.g., graphs or illustrations) in the article.
- The author will not allow DAU to post the article in our *Defense ARJ* issue on our Internet homepage.
- The author requires that the usual copyright notices be posted with the article.
- To publish the article requires copyright payment by the DAU Press.

SUBMISSION

All manuscript submissions should include the following:

- Completed submission checklist
- Completed copyright release form
- Cover letter containing the complete mailing address, e-mail address, and telephone number for each author
- Biographical sketch for each author (70 words or fewer)
- Headshot for each author saved as a 300 dpi (dots per inch) high resolution JPEG or Tiff file no smaller than 5x7 inches with a plain background in business dress for men (shirt, tie, and jacket) and business appropriate attire for women. All active duty military should submit headshots in Class A uniforms. Please note: low-resolution images from Web, PowerPoint, or Word will not be accepted due to low image quality.
- One copy of the typed manuscript, including:
    - Title (12 words or fewer)
    - Abstract (150 to 250 words)
    - Two sentence summary
    - Keywords (5 words or fewer—please include descriptive words that do not appear in the manuscript title, to make the article easier to find)
- Figures and tables saved as separate individual files and appropriately labeled

The manuscript should be submitted in Microsoft Word (please do not send PDFs), double-spaced Times New Roman, 12-point font size (5,000 words or fewer for the printed edition and 10,000 words or fewer for online-only content excluding abstracts, figures, tables, and references).

Figures or tables should not be inserted or embedded into the text, but submitted as separate files in the original software format in which they were created. For additional information on the preparation of figures or tables, refer to the Scientific Illustration Committee, 1988, Illustrating Science: Standards for Publication, Bethesda, MD: Council of Biology Editors, Inc. Restructure briefing charts and slides to look similar to those in previous issues of the *Defense ARJ*.

All forms are available at our website: www.dau.edu/library/arj. Submissions should be sent electronically, as appropriately labeled files, to the *Defense ARJ* managing editor at: DefenseARJ@dau.edu.

Defense ARJ
PRINT SCHEDULE

The *Defense ARJ* is published in quarterly theme editions.

All submissions are due by the first day of the month.
See print schedule below.

Author Deadline	**Issue**
July	**January**
October	**April**
January	**July**
April	**October**

In most cases, the author will be notified that the submission has been received within 48 hours of its arrival. Following an initial review, submissions will be referred to peer reviewers and for subsequent consideration by the Executive Editor, ***Defense ARJ***.

Contributors may direct their questions to the Managing Editor, *Defense ARJ*, at the address shown below, or by calling 703-805-3801 (fax: 703-805-2917), or via the Internet at norene.johnson@dau.edu.

The DAU Homepage can be accessed at:
https://www.dau.edu

DEPARTMENT OF DEFENSE
DAU
ATTN: DAU PRESS (***Defense ARJ***)
9820 BELVOIR RD STE 3
FORT BELVOIR, VA 22060-5565


CALL FOR
AUTHORS

We are currently soliciting articles and subject matter experts for the 2021-2022 *Defense Acquisition Research Journal (ARJ)* print year.

Please see our guidelines for contributors for submission deadlines.

Even if your agency does not require you to publish, consider these career-enhancing possibilities:

- Share your acquisition research results with the Acquisition and Sustainment (A&S) community.
- Change the way Department of Defense (DoD) does business.
- Help others avoid pitfalls with lessons learned or best practices from your project or program.
- Teach others with a step-by-step tutorial on a process or approach.
- Share new information that your program has uncovered or discovered through the implementation of new initiatives.
- Condense your graduate project into something beneficial to acquisition professionals.

ENJOY THESE BENEFITS:

- Earn 30 continuous learning points for publishing in a refereed (peer reviewed) journal.
- Earn a promotion or an award.
- Become part of a focus group sharing similar interests.
- Become a nationally recognized expert in your field or specialty.
- Be asked to speak at a conference or symposium.

We welcome submissions from anyone involved with or interested in the defense acquisition process—the conceptualization, initiation, design, testing, contracting, production, deployment, logistics support, modification, and disposal of weapons and other systems, supplies, or services (including construction) needed by the DoD, or intended for use to support military missions.

If you are interested, contact the *Defense ARJ* managing editor ***(DefenseARJ@dau.edu)*** and provide contact information and a brief description of your article. Please visit the *Defense ARJ* Submissions page at ***https://www.dau.edu/library/arj/p/Defense-ARJ-Submissions***.

Statement Required by the Act of August 12, 1970 Section 3685, Title 39, U.S.C. Showing Ownership, Management, and Circulation

The Defense Acquisition Research Journal (ARJ) is published at DAU, Fort Belvoir, VA. DAU publishes four issues annually. The Chief of the DAU Press and Managing Editor of the Defense ARJ is Norene Johnson, the Assistant Editor is Emily Beliles, and the publisher is the DAU Press. All are collocated at the following address:

DAU
Attn: DAU Press (Defense ARJ)
9820 Belvoir Road, Suite 3
Fort Belvoir, VA 22060-5565

Average Number of Copies Each Issue During the Preceding 12 Months:

A. Total number of copies printed (net press run) 2,991
B. Paid and/or requested circulation:
   1. Sales through dealers and carriers, street vendors, and counter sales0
   2. Mail subscriptions paid and/or requested. 2,923
C. Total paid and/or requested circulation 2,923
D. Free distribution by mail, carrier, or other means; samples, complimentary, and other free copies. 21
E. Total distribution 2,944
F. Copies not distributed:
   1. Office use, leftover, unaccounted, spoiled after printing. 47
   2. Returns from news agents 0
G. Total 2,991

Actual Number of Copies of Single Issue Published Nearest to Filing Date:

A. Total number of copies printed (net press run) 2,908
B. Paid and/or requested circulation:
   1. Sales through dealers and carriers, street vendors, and counter sales . 0
   2. Mail subscriptions paid and/or requested 2,837
C. Total paid and/or requested circulation 2,837
D. Free distribution by mail, carrier, or other means; samples, complimentary, and other free copies. 30
E. Total distribution 2,867
F. Copies not distributed:
   1. Office use, leftover, unaccounted, spoiled after printing. 41
   2. Returns from news agents 0
G. Total 2,908

RECOGNITION OF REVIEWERS
FOR PRINT YEAR 2020

We would like to express our appreciation to all of the subject matter experts who volunteered to participate in the Defense Acquisition Research Journal peer review process. The assistance of these individuals provided impartial evaluation of the articles published during the 2020 print year. We would also like to acknowledge those referees who wished to remain anonymous. Your continued support is greatly appreciated, and we look forward to working with many of you again in print year 2021.

Mr. Pat Armstrong
Economic/Business Analyst, Principal
The MITRE Corporation

Dr. Donald Birchler
Senior Research Scientist
The CNA Corporation

Mr. Irv Blickstein
Senior Engineer
The RAND Corporation

Mr. Alexander M. Brofos
Senior Principal Engineer
The MITRE Corporation

Professor Joel D. Brown
Contracting and Facilities Engineering
DAU

Mr. Kevin Cincotta
Principal Cost Analyst
The MITRE Corporation

Dr. Lisa Colabella
Senior Operations Researcher
The RAND Corporation

Dr. David Coleman
Chief Engineer, Naval Enterprise Sea Systems
The MITRE Corporation

Lt Col John M. Colombi, USAF (Ret.)
Professor and Program Chair, Systems Engineering
Air Force Institute of Technology

Dr. Sidhartha R. Das
Professor
George Mason University

Dr. Timothy L. J. Ferris
Senior Lecturer
Cranfield University

Dr. David R. Gallay
Assistant Professor of Finance
George Mason University

Dr. Gerald A. Hanweck, Sr.
Professor of Finance
School of Business, George Mason University

Dr. Jennifer Hitchcock
Professor of Program Management
DAU

Mr. Mark Holbrook
Instructor
DAU

Mr. Maynard Holliday
Senior Engineer
The RAND Corporation

Dr. Joseph L. Ilk
Professor of Contract Management
DAU

Mr. John L. Larson
Professor of Systems Engineering Management
DAU

Mr. Terrence Leary
Acquisition Strategy and Management Principal
The MITRE Corporation

Mr. Keith Lindsay
Professor of Contract Management
DAU

Mr. William Lucyshyn
Research Professor
University of Maryland

Mr. Gregg Maffett
Professor of Contracts Management
DAU

Mr. Tom Neff
Professor of Program Management
DAU

Dr. Alexander Odeh
Principal Economist/Business Analyst
The MITRE Corporation

Ms. Monica M. Pieper
Economics/Business Analyst, Principal
The MITRE Corporation

Dr. Edward H. Powley
Associate Professor of Management
Naval Postgraduate School

Ms. Emily Rayburn
Contracts Analyst
The MITRE Corporation

Mr. David M. Riel
Professor of Acquisition Management
DAU

Dr. Jeanette Robinson
Professor of Small Business
DAU

Mr. Michael Roche
Professor
DAU

Dr. Michael Santens
Professor of Contract Management
DAU

Mr. Jonathan Specht
Professor of Systems Engineering Management
DAU

Mrs. Dana Stewart
Professor of Financial Management
DAU

Professor Trevor Taylor
Professorial Fellow
Royal United Services Institute

Dr. Laura Werber
Senior Management Scientist
The RAND Corporation

ONLINE EDITIONS

Defense ARJ and Defense Acquisition Magazine

PLEASE SUBSCRIBE

or resubscribe so you will not miss out on accessing future publications.

Online presence for easier use on mobile and desktop devices:

https://www.dau.edu/library/arj

https://www.dau.edu/library/defense-atl

Send an e-mail to **defensearj@dau.edu** and/or **defacq@dau.edu**, giving the e-mail address with which you want us to notify you when a new issue is posted

Type "Add to **LISTSERV**" in the subject line

Also use this address to notify us if you change your e-mail address

Defense Acquisition Research Journal
SURVEY

Please rate this publication based on the following scores:

5 — Exceptional   4 — Great   3 — Good   2 — Fair   1 — Poor

Please circle the appropriate response.

1. How would you rate the overall publication?	5	4	3	2	1
2. How would you rate the design of the publication?	5	4	3	2	1

	True	False
a) This publication is easy to read		
b) This publication is useful to my career		
c) This publication contributes to my job effectiveness		
d) I read most of this publication		
e) I recommend this publication to others in the acquisition field		

If hand written, please write legibly.

3. What topics would you like to see get more coverage in future ***Defense ARJs***?

4. What topics would you like to see get less coverage in future ***Defense ARJs***?

5. Provide any constructive criticism to help us to improve this publication:

6. Please provide e-mail address for follow up (optional):

YOU CAN ALSO FIND THE SURVEY ONLINE AT

https://www.dau.edu/library/arj

FREE ONLINE SUBSCRIPTION

☐ ***Defense ARJ*** ***Defense Acquisition*** ☐

Thank you for your interest in ***Defense Acquisition Research Journal*** and ***Defense Acquisition*** magazine. To receive your complimentary online subscription, please write legibly if hand written and answer all questions below—incomplete forms cannot be processed.

***When registering, please do not include your rank, grade, service, or other personal identifiers.**

☐ *New Online Subscription* ☐ *Cancellation* ☐ *Change E-mail Address*

Date ____

Last Name: ____

First Name: ____

Day/Work Phone: ____

E-mail Address: ____

Signature: ____
(Required)

PLEASE FAX TO: 703-805-2917
OR EMAIL TO: defensearj@dau.edu

The Privacy Act and Freedom of Information Act

In accordance with the Privacy Act and Freedom of Information Act, we will only contact you regarding your ***Defense ARJ*** and ***Defense Acquisition*** subscriptions. If you provide us with your business e-mail address, you may become part of a mailing list we are required to provide to other agencies who request the lists as public information. If you prefer not to be part of these lists, please use your personal e-mail address.

SUBSCRIPTION

DEFENSE ACQUISITION RESEARCH JOURNAL

A Publication of DAU | dau.edu

We're on the Web at:
http://www.dau.edu/library/arj

ARJ

DEFENSE ACQUISITION RESEARCH JOURNAL

A Publication of DAU | dau.edu

Current. Connected. Innovative.

2020 MARCOM

PLATINUM

WINNER