UMD-AM-13-102

Using Leading Indicators to Improve DoD Acquisitions

By:
Jacques S. Gansler and William Lucyshyn

Center for Public Policy and Private Enterprise
School of Public Policy

Revised October 2013

This research was partially sponsored by a grant from
The Naval Postgraduate School

University of Maryland 1856

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **OCT 2013**	2. REPORT TYPE	3. DATES COVERED **00-00-2013 to 00-00-2013**
4. TITLE AND SUBTITLE **Using Leading Indicators to Improve DoD Acquisitions**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Maryland,Center for Public Policy and Private Enterprise,2101 Van Munching Hall ,College Park,MD,20742**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**74**	

The Center for Public Policy and Private Enterprise at the University of Maryland's School of Public Policy provides the strategic linkage between the public and private sector to develop and improve solutions to increasingly complex problems associated with the delivery of public services—a responsibility increasingly shared by both sectors. Operating at the nexus of public and private interests, the Center researches, develops, and promotes best practices; develops policy recommendations; and strives to influence senior decision-makers toward improved government and industry results.

Table of Contents

Executive Summary

National budgetary challenges will continue to exert downward pressure on the Department of Defense's (DoD's) budgets. In the past, the DoD relied on personnel reductions in order to constrain costs. Today, however, the active military force structure is already near an all-time low, meaning that significant reductions are unlikely. At the same time, the challenges to the nation's security continue to grow. Consequently, the DoD must strive to develop an acquisition strategy that is not only affordable, but also provides the quality and quantity of forces required.

The current metrics that are used to evaluate DoD programs do not provide decision-makers with timely, consistent, reliable, or useful data in that they rely on lagging indicators (i.e., they provide information about past performance). In an effort to better control costs, schedule, and product quality, the DoD should develop and adopt effective "leading indicators"—reliable and predictive metrics that provide earlier warnings of programmatic problems and challenges. The successful use of leading indicators could provide program managers, the DoD, and Congress with earlier warnings of program difficulty.

While previous initiatives implemented by the DoD rely on lagging indicators, there are others that purport to make use of leading indicators. These initiatives are described below.

Prior to 1968, the DoD had no system for monitoring the progress of major systems (Hough, 1992). In order to facilitate internal cost control, the DoD introduced Selected Acquisition Reports (SARs), which summarized the latest estimates of cost, schedule, and performance. Soon thereafter, SARs were submitted to Congress on a regular basis. SARs have served as the primary source of research into cost growth for decades. Unfortunately, however, SAR data cannot be used to identify a program's cost drivers. This is because the reports classify cost growth using variance categories that show only the effects of secondary factors (Hough, 1992).

Another initiative, the Nunn-McCurdy Amendment (NM), enacted by Congress in 1982 and modified in 2006 and 2009, requires the DoD to notify Congress when the unit cost growth of any major defense acquisition program is expected to exceed certain cost growth thresholds.

Unfortunately, defense acquisition projects continue to experience high unit cost growth in spite of NM cost breeches. Acquisition problems are uncovered too late in the development process to allow program reforms to be effective. A 2011 RAND report identified common root causes of Nunn-McCurdy breaches, including the use of immature technologies, unanticipated integration issues, unstable funding, ambitious scheduling, and ill-conceived manufacturing processes and insufficient research, development, testing, and engineering (RDT&E). Needless to say, problems of this nature—after they have been observed on a program—cannot simply be corrected in order to bring costs back within initial expectations.

A third initiative, cost as an independent variable (CAIV), which was developed in the 1990s, strives to elevate the importance of cost within the trade space. All acquisitions are assessed based on their cost, schedule, and performance. Collectively, these three parameters make up the trade space. CAIV attempted to create a cost-saving environment by emphasizing the importance of cost over performance and schedule. However, it does not appear that the use of the CAIV approach has achieved the desired results. In the early 1990s, the DoD selected eight programs to serve as CAIV flagships. These programs, it was believed, would demonstrate how this initiative could contain costs. In 1999, the Government Accountability Office (GAO) identified program offices that were "leaders" in the application of various acquisition best practices, one of which was the CAIV approach (GAO, 1999a, p. 22). Yet none of the programs stayed below the initial estimated cost. Minimizing restrictions within the trade space by treating cost as an independent variable, is a good first step. However, in practice, it appears that this approach did not go far enough.

In the 1960s, the DoD developed earned value management (EVM) to serve as a leading indicator of program performance. In June 2002, the Office of Management and Budget mandated the use of EVM systems for all major information technology (IT) service and acquisition contracts. EVM measures the value of work accomplished in a given period and compares it with the planned value of work scheduled for that period and with the actual cost of work accomplished. As work is performed and measured against the baseline, the corresponding budget value is "earned." From the basic variance measurements, the program manager can identify significant drivers, forecast future cost and schedule performance, and construct corrective action plans to get the program back on track. But because the aggregated data is not necessarily predictive of a program's future performance, some government and commercial sector reports (e.g., Card, 2008) suggest that earned value should not be considered a leading indicator. It is also rather easy to "game the system" so that a program appears to be progressing, when there are actually significant, albeit hidden, problems.

Implemented by then-Under Secretary of Defense for Acquisition, Technology, and Logistics (USD[AT&L]) Ashton Carter in 2011, the should-cost/will-cost approach is similar to EVM but relies on two separate cost estimates: a non-advocate, historically based *will-cost* estimate, which provides the official basis for budgeting and programming; and a *should-cost* estimate for program management execution (Davies & Woods, 2011). The official budget baseline for the program is based on the non-advocate will-cost estimate. In contrast, the should-cost estimate is based on what the program manager believes is possible within "the context of creative, innovative, and disciplined measures to increase productivity" (Sledge, 2012, p. 1). Unfortunately, it does not appear that should-cost/will-cost provides managers with much incentive to build cost savings into their programs. On the one hand, program managers are required to budget to the historically based, and higher will-cost figure; on the other hand, they must drive their suppliers, and their own team, to the lower should-cost estimate.

These five initiatives have not significantly improved the DoD's acquisition outcomes, in part because monitoring and enforcement measures tend to be reactive (e.g., alerting the DoD and Congress of difficulties once they have occurred) as opposed to proactive (e.g., providing advanced warning of programs that are likely to encounter difficulty). Indeed, the longer a program is extended beyond its scheduled completion, the longer management should expect the program will take to complete. Or, in the words of economist Nassim Nicholas Taleb (2007), "the longer you wait, the longer you will be expected to wait" (p. 159). Of course, costs increase as the schedule slips.

This is why it is critical to identify issues and take corrective action as soon as future problems can be detected. Chalking up delays or cost increases to bumps in the road while hoping that the program will eventually find its way back on course is unrealistic and irresponsible. Instead, the program's path must be altered, sometimes significantly. The one-time cost increase or minor delay is a rare event indeed. Rather, schedule delays precipitate more schedule delays, and higher costs lead to still higher costs.

It is clear that new projects can be more effectively planned, budgeted, and scheduled when historical data from previous similar programs are used (Rhodes, Valerdi, & Roedler, 2009). Indeed, the widespread use of EVM in both the public and commercial sectors attests to this commonsense reality. Should-cost/will-cost takes this approach a step further by not only determining how much a program will cost based on historical data, but also how much it should cost if, for example, sourcing were to be carried out more efficiently or if process redundancies were to be eliminated.

However, historical data are not always reliable and, in fact, may be misleading, depending on how they are interpreted. This is especially true within the context of projects that rely on cutting-edge technology for which there is little precedent. By making use of better leading indicators, the DoD can better ensure that resources are consistent with the program's development plan so that success can be more easily achieved.

When product development is undertaken by the commercial sector, cost overruns, schedule delays, and program failure—while they do occur—are generally of a lower magnitude. Commercial sector processes seem to naturally constrain schedules and costs. Analyzing these processes will help inform the development of new leading indicators for the DoD.

Generally, leading commercial firms have identified three critical junctures at which they must have sufficient knowledge to make large development decisions or to continue the development process. First, before the initial program development, customers' expectations should be matched with the firms' resources, including technology, engineering, time, and funding. Once a development decision has been made, program designers should ensure that they have sufficient resources to meet the performance requirements of that product, and the design of that product is stable enough for routine production. After this stage, program developers must show that the product can be fully developed and produced within budget, schedule, and performance targets.

Paralleling these three critical junctures, there are three knowledge points. When the first point (technology development) is reached, future resources and needs match. At knowledge point 2 (product development), the design is stable. And at knowledge point 3 (production), the production processes are mature.

Not only do commercial firms determine the level of knowledge needed to progress from one phase of a project to the next, but they also often determine the unit cost of the product early on in development in order to ensure that the projected market size will be realized.

Whereas cost traditionally has been considered an outcome of product development, target costing treats it as an input. The target cost for a product is determined using a simple formula: Target Cost = Estimated Selling Price – Desired Profit. However, target costing is not merely the imposition of a cost ceiling. As Zengin and Ada (2010) pointed out, “manufacturers cannot make a trade-off between cost, quality, and functionality of the product with only cost considerations in mind” (p. 5594). Rather, target costing, as a strategy, promotes creativity and new ways of thinking to increase performance while discouraging the inclusion of non-value-added functions.

Mihm (2010) observed that “target costing does *not* require perfect knowledge about the component” (p. 1334). Fairly accurate component cost estimates can be developed via systematic value analyses of comparable existing parts (Mihm, 2010). And whereas the target cost of the product remains firm throughout the development process, component cost estimates are permitted to fluctuate as product development evolves. Typically, each product feature is ranked in terms of its relative importance. Some firms may go as far as to assign specific numeric weights to each feature. These weights are then used to determine where the firm can adjust costs while maintaining, or even enhancing, the product’s value (Ellram, 2006). In order to ensure the inclusion of the most valuable features, the target cost of one component may be increased while that of another is reduced.

Today, virtually every successful commercial firm employs a cost-driven approach to product development. For a variety of reasons, the DoD has been reluctant to do the same. But given current and impending budgetary constraints, it may soon have little choice in the matter. Admittedly, there are significant challenges that must be overcome in order for a cost requirement to be viable. In the commercial sector, not only is product development cost-driven, but it is also market-driven. Firms spend considerable sums in order to better understand what the customer is willing to pay for; a firm that adds extraneous features of little added value to the customer is punished in the market.

The domestic defense market, however, is characterized by very few firms in each sector and only one customer (i.e., a monopsony). Because weapon systems are contracted for in advance of their production, the contractor generally is not incentivized to translate the diffuse desires of the customer—in this case, the DoD—into an effective and efficient product. Rather, the DoD specifies requirements upfront, and in great detail, for fear that they may never be developed. In fact, there is frequently an incentive to "gold-plate" products by adding every desired feature, including some of little marginal value. This is especially true within the context of complex product development, where neither the DoD nor the contractor fully understands the attributes and capabilities of the end product.

In the commercial sector, large retailers such as Wal-Mart have significant control over their supply bases because they have considerable buying power. Ellram (2006) noted that even in the manufacturing sector, large firms like Dell might be able to dictate pricing to companies like Intel. In the same way, the DoD must use all available strategies (e.g., competitive dual-sourcing) to leverage its size and buying power and exert downward pressure on the cost of weapons systems.

Leading indicators can help the DoD achieve product development success by drawing attention to essential elements that are automatically controlled within the commercial sector—and are thus often overlooked by the DoD. We note that the use of any cost control approach requires a trained and experienced acquisition workforce. The workforce must have sufficient understanding of industry behavior and incentives in order to achieve the desired results.

We have derived several features of product development that we believe can inform the creation of meaningful leading indicators. We contend that these indicators can be used in two distinct ways: first, the use of indicators will ensure that fewer programs will begin development on a weak case, thus avoiding a costly, though all too common, mistake—initiating a program that should have not been started; second, the use of leading indicators will provide program managers with earlier warnings of impending difficulties as a program progresses, which programs managers can take into account to correct minor difficulties before they become costly revisions. We describe several indicators below.

Initial Program Requirements

A primary reason for cost and schedule problems is the encouragement, within the acquisition environment, of overly ambitious product developments—sometimes referred to as "revolutionary" or "big bang" acquisition programs—that embody too many technical unknowns and insufficient knowledge about performance and production risks. Often, capabilities are not assigned to specific increments; rather, they are frontloaded onto the initial requirements document. By adopting an evolutionary approach, however, essential technologies can be fielded in the near term, delaying the instantiation of more time-intensive, costly, or technically challenging capabilities. An evolutionary approach also ensures that operational experience inform future versions of a product's requirements. Once agreed upon by the relevant stakeholders, the impulse to add requirements must be avoided. If requirements are added, the program should be assigned a higher level of risk.

Technology Readiness

Failure to properly mature new technologies almost invariably leads to cost and schedule over-runs in weapon system programs (GAO, 1999a). Recognizing this, the DoD already relies on technology readiness levels (TRLs) to assess the maturity of technologies, and the risks associated, before incorporating that technology into program development. TRLs provide a common understanding of technology status and thus help management in making decisions on the development and transition of technology. However, the DoD should also regard TRLs as leading indicators of potential problems, to be measured and monitored throughout early development and used to assign risk to programs accordingly.

In addition, Gove (2007) and Gove, Sauser, and Ramirez-Marquez (2010) have developed and suggested the use of integration readiness levels (IRLs) to access the interfacing of compatible interactions for various technologies and the consistent comparison of the maturity between integration points (see Table 3). On a scale similar to TRLs, IRLs could be used in conjunction with TRLs to help optimize the process of complex system integration.

Senior Leadership

DoD programs with strong senior leadership support were generally more stable and, as a result, were more successful. Those programs with strong senior leadership support, sometimes because they had a more immediate requirement, were viewed as a higher priority by senior leaders in the services and DoD. Further, this consistent leadership support from the DoD and the services promoted the development of better business plans and helped program managers adapt to the inevitable program perturbations (GAO, 2010c). Frequent changes in senior leadership can also lead to significant changes in an organization's priorities, goals, and strategies. These changes also can significantly impact relationships with partnering organizations.

Program Managers

Program managers of successful programs tend to share key attributes, such as experience, leadership continuity, and communication skills that facilitated open and honest decision-making. Their programs established sound, knowledge-based business plans before starting development and then executed those plans using disciplined approaches (GAO, 2010c). They also pursued incremental acquisition strategies and leveraged mature technologies, both of which are important leading indicators of program performance in and of themselves.

Supporting Organization

DoD programs are often large and complicated in nature, which calls for a team of professionals with diverse backgrounds and skills. Good staffing, of particular importance with today's systems, ensures the right people are in place to help meet organizational goals. Unfortunately, the number of experienced military acquisition personnel has been reduced significantly, which we believe has contributed directly to problems with effective management of DoD acquisition programs. A critical assessment of the education, experience, and quality of the supporting staff can provide a good indication of a program's likely performance.

Requirements Volatility

During program implementation, ineffective control of requirements changes leads to cost growth and program instabilities. One indicator that should be monitored is requirements volatility (i.e., adding, deleting, and modifying a system's requirements during the development process). These requirements changes generally will have an impact on several elements of the system. More problematic still is that the precise nature of the impact often cannot be anticipated (from a technical, schedule, or cost point-of-view). Developing an indicator that measures the rate of change of requirements over time can help forecast future program challenges (e.g., a surge in requirement changes could indicate potential risk to design, which in turn, could impact cost and schedule), as well as identify problems with the requirements generation process.

Contract Changes

Because the DoD's needs change from time to time, government acquisition contractors generally have to unilaterally make change orders with regard to specifications and other contract terms. These changes may be related to contract cost, delivery schedule, fee, terms and conditions, and personnel. In addition, changing technologies, funding, and mission requirements may necessitate changes to a contract. The complexity of contracts—which can involve a variety of people with diverse backgrounds from different functional areas on both the government and contractor sides—can lead to misinterpretations and miscommunications of requirements and administrative issues that do not become evident until the contract is under way.

Budget Stability

In order to prevent a vicious cycle wherein reductions in quantity lead to increases in program costs, programs should ensure that there is an adequate management reserve (MR) in the budget. The MR budget is typically used by contractor program managers to cover unknown problems that arise during development and that fall under the scope of work. Technical complexity should inform the amount of the MR budget.

Funding Flexibility

In general, there is no single funding source for large, complex systems. Rather, funding is programmed through the individual services or through individual program offices for the individual system. As a result, there is no advocate for joint or, as the case may be, enterprise-wide capabilities. Contracts are generally written that do not adequately specify how the individual products are going to be integrated and tested with other elements. Program managers must treat funding stability as a leading indicator. In the event that a reliable funding source is unavailable, programs should be assigned a higher level of risk.

Manufacturing Readiness

The success of acquisition programs also requires that manufacturing capability be managed effectively. Manufacturing readiness levels (MRLs) were designed to assess manufacturing readiness and manage manufacturing risk during acquisition. They were developed by a joint DoD/industry working group under the sponsorship of the Joint Defense Manufacturing Technology Panel (JDMTP) and were introduced to the defense community in 2005. Generally, MRLs serve three purposes: (1) to define the current level of manufacturing maturity, (2) to identify maturity shortfalls and associated costs and risks, and (3) to provide the basis for manufacturing maturation and risk management. The GAO (2009) recommended that weapon programs make use of MRLs. One possibility, in this regard, is incorporating the MRL scale into leading indicators that are measured and monitored across time in order to mitigate program risk.

The leading indicators that we have proposed are less prone to manipulation in that they are discrete values that can be objectively measured. Unlike “work performed” (the typical EVM metric), the number of requirements, technical readiness levels, years of management experience, contract changes, and so forth cannot be as easily manipulated.

Of course, there is always the danger that the leading indicators might be measured poorly or reported too late to be of use, but we contend that various safeguards can be put in place to prevent problems of this sort. For example, program managers might simply mandate that certain indicators be monitored and reported on a near-continuous basis. Further, under the leading indicators framework, there is explicit recognition that the indicators are designed to highlight deviations from planned values before these deviations become serious problems. The pressure to keep programs in the green—a common objective under EVM—is less of an issue. There is no motivation to hide problematic values because the values are indicative of potential problems, not actual ones.

It is also possible to aggregate leading indicators to determine a given program's "score," which might be helpful to gauge overall programmatic risk. Because the relative importance of a given indicator will likely vary across programs, each could be weighted prior to the launch of the program. The indicators could then be assessed and summed for an overall score at various points in time.

The Analytic Hierarchy Process (AHP) is but one technique that could be used. AHP allows users to compare incommensurable elements (in this case, requirements volatility, leadership, technical maturity, and so forth to the success of the program) in a rational and consistent way. By using AHP, program managers would have a better understanding of the relative importance of the chosen indicators.

The success of large projects is often assessed in terms of schedule, cost, and quality—the so-called iron triangle. However, a project that "fails" in any (or all) of these categories may go on to deliver large benefits to users, contractors, program personnel, and/or other stakeholders, especially as time passes. Conversely, projects that meet established requirements and that are completed on time and under budget may fail to meet the expectations of stakeholders.

The challenge, then, revolves around how the DoD should define and measure program success. Even if leading indicators are used narrowly to help predict program "performance" (i.e., how the program rates, in terms of quality, cost, and schedule) with the understanding that success is more difficult to define, there is the possibility that needed programs will be canceled. The underlying point is that success is often achieved in an environment that permits some degree of failure. And failures, in turn, occur in an environment that encourages moderate risk-taking.

Accordingly, programs that undertake the use of leading indicators must also consider the strategic importance of the program. In some instances, programs should take on higher levels of risk and be willing to accept moderate increases in schedule and cost. Unfortunately, such a suggestion rings hollow in an environment where most programs regularly exceed their budgets and schedules. We believe that implementing a system of leading indicators will help facilitate the needed change.

I. Introduction

> *"We've been living with unconstrained resources for 10 years, and, frankly, we've developed some bad habits. … In our acquisition programs … there is certainly room to become more efficient."*
>
> –General Martin Dempsey, Chairman, Joint Chiefs of Staff (Cook, 2013)

The United States is entering a critical period. National budgetary challenges will continue to exert downward pressure on the Department of Defense's (DoD's) budgets, which are projected to decrease, and then flatten (see Figure 1). But these projections are deceiving because even if the top-line numbers remain stable (which, with Social Security, Medicare, and debt payments rising, seems unlikely) the DoD's operation and support costs will continue to increase steadily. Within the DoD's budgets, there are two cost drivers that are, in effect, entitlements. The first is military health care (Tricare), which has almost tripled from $19 billion in 2001 to $53 billion in 2012. Healthcare for military personnel now consumes 10 percent of the entire defense budget. If left unchecked, healthcare will climb to $95 billion by 2030 (Cassata, 2013). Compensation for the DoD's military and civilian employees is the second major driver. As compensation continues to increase, a growing portion of "defense discretionary" spending must be diverted to fund personnel costs, limiting the resources available for recapitalization, modernization, and transformation of the military. In the past, the DoD could rely on personnel reductions in order to constrain costs. Today, however, the active military force structure is already near an all-time low, meaning that significant reductions are unlikely.

At the same time, the challenges to the nation's security continue to proliferate. These challenges include regional instability, the continued threat of proliferation of weapons of mass destruction, transnational terrorism, cyber-attacks, and the emergence of China as a potential peer competitor. Consequently, the DoD must strive to develop an acquisition strategy that is not only affordable, but also provides the quality and quantity of forces required.

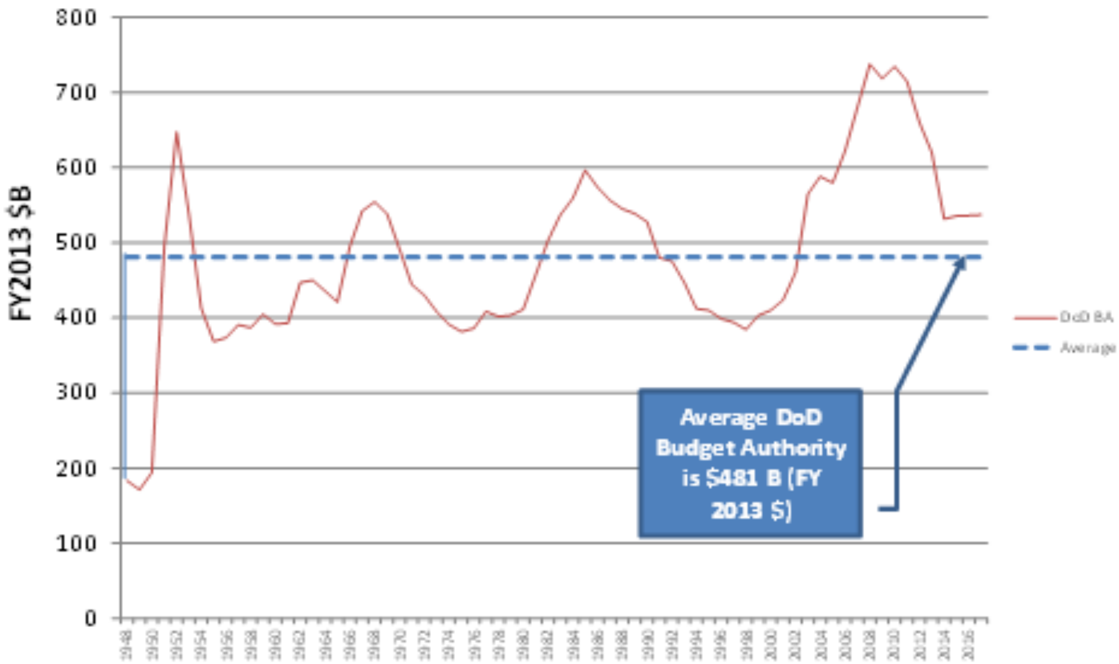

Figure 1. DoD's Budget Authority in Constant FY 2013 Dollars

The DoD's attempts to improve the efficient and effective allocation of resources are regularly thwarted by misguided political forces. For example, as the DoD reduced its force size, Congress refused to approve the recommended base closures; and in spite of the 30 percent cost savings that derive as a result of public/private competitions for non-inherently governmental work, Congress decided to stop such competitions from taking place.

The DoD also continues to face numerous difficulties in its major acquisition efforts. When one examines program cost growth—the positive difference between actual or projected costs and budgeted or initial estimated costs—one finds that most programs experience significant cost overruns, as well as lengthy schedule delays and reduced operational performance.

Numerous reports have revealed that the DoD's major weapon system programs have experienced high program cost-growth over an extended period of time. Figure 2 summarizes the findings of seven of these reports, which, collectively, examine programs ranging from 1946 to 2003. All of the studies adjusted program cost-growth for inflation and quantity change relative to the Milestone (MS) II baseline, although the studies did not necessarily make such adjustments in the same way (Arena, Leonard, Murray, & Younossi, 2006). A program's cost-growth was recorded as a cost-growth factor (CGF) relative to the total program's original estimate. From reports that record this information, the development program cost-growth factor ranged from 1.25 to 1.58, while procurement CGFs ranged from 1.18 to 1.65. A program's total CGF revealed a greater range of values—from a low of 1.14 to a high of 3.23. The most recent analysis (Arena et al., 2006) concluded that from 1968 to 2003, the average adjusted total cost growth for a set of completed weapon programs was 46 percent from MS II and 16 percent from MS III. In short, a variety of analyses from different time periods all recorded high program cost-growth.

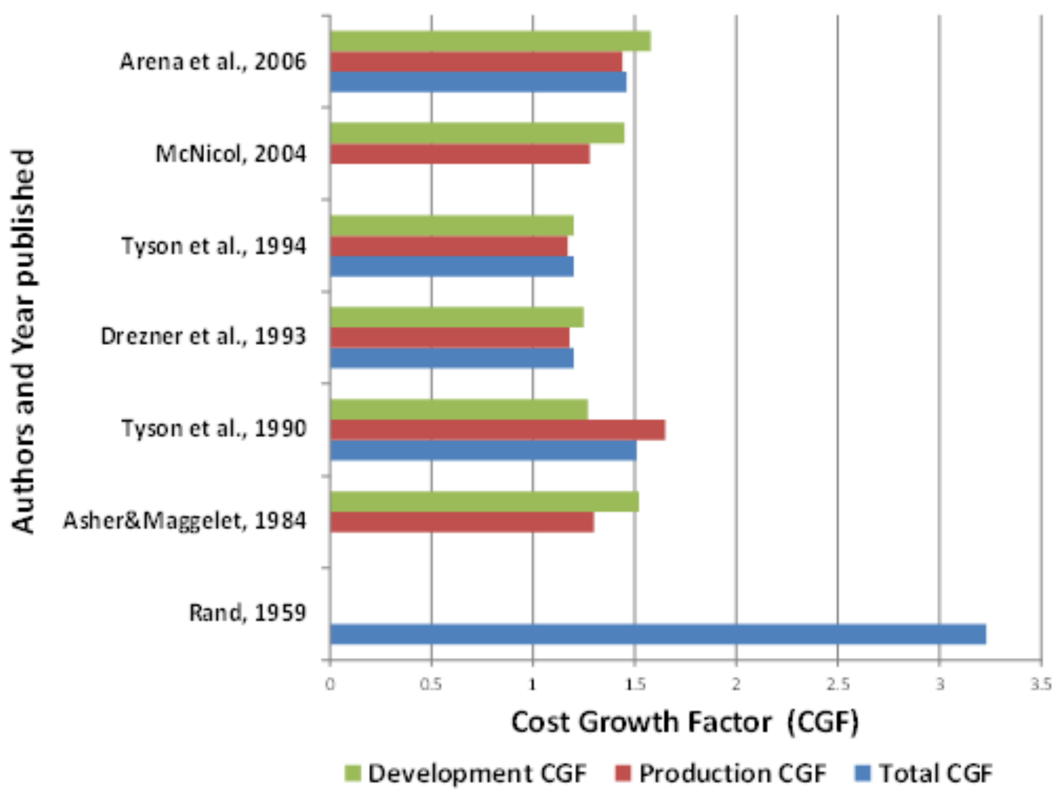

Figure 2. Former Studies Chronicling Cost Growth
Note. This information in this figure is from Arena et al. (2006).

A recent GAO report echoes this familiar theme. The report notes that "DoD's major weapon system programs continue to take longer, cost more, and deliver fewer quantities and capabilities than originally planned" (GAO, 2008). Furthermore, the latest GAO analysis reported that when assessed against its first full estimates, the total cost of the DoD's 86-program portfolio has increased by over $400 billion. This figure included more than $90 billion in development cost growth and almost $290 billion in procurement cost growth. Additionally, there was an average delay of 27 months in the delivery of initial operating capability. For instance, the F-22 program (begun over 20 years ago) experienced development cost growth of more than 60 percent, a quantity reduction of more than 70 percent, and an increase in the unit cost of almost 200 percent (GAO, 2013). Needless to say, cost growth of this magnitude will be particularly problematic in the anticipated fiscal environment.

Any number of problems might explain a program's poor performance—inappropriate contract type or poor design and mismanagement are often cited. However, consistent poor performance across multiple programs reveals a larger issue: the failure to match acquisition needs with developers' resources (e.g., technical knowledge, timing, and funding) when starting product development. In many cases, development is launched before knowing whether technologies and other capabilities will work as intended. The GAO (2010d) notes that despite the continued improvements in technology, design, and production, most DoD programs are still initiated with limited knowledge.

Defense programs are begun with the best of intentions, but in order to ensure that programs are able to deliver their planned capabilities, the DoD should develop and adopt effective leading indicators—reliable and predictive metrics that provide earlier warnings of programmatic problems and challenges. The successful use of leading indicators could provide program managers, the DoD, and Congress with earlier warnings of program difficulty. Indeed, the longer a problem lingers, the more difficult and costly it is to correct. Currently, limited metrics are used to gauge the impact of investments or the effectiveness of processes to develop, demonstrate, integrate, and transition technologies (GAO, 1999a). However, most of the potential indicators of programmatic challenges are not monitored in real time; or, if they are, they merely present a "snapshot" of a program's progress. Often, assessments are carried out after a program reaches a particular point in time (or development), or after the program concludes. In theory, these assessments, which rely on "trailing" or "lagging" indicators, might help program managers avoid similar problems on future programs—and yet such efforts have not been successful in solving the problem of escalating program costs. The bottom line is that current metrics used to evaluate programs do not provide decision-makers with timely, consistent, reliable, or useful data, in that they rely on lagging indicators (i.e., they provide information about past performance).

In an effort to better control costs, schedule, and product quality, leading indicators should be devised and implemented so that potential problems can be detected and corrective action taken before the problems fully develop. Identifying and mitigating problems early-on will obviate the need for costly "workarounds" (i.e., designing around a problem, technical or otherwise, that could have been avoided had the program altered course earlier).

Report Roadmap

The DoD continues to struggle to contain the costs of its weapons programs. We believe that leading indicators—measures that are predictive of future system performance before that performance is realized—can help the DoD better meet cost and schedule objectives and minimize the risk of program failure. In Part II, we describe past DoD strategies to control costs and outline what we believe to be a better approach. Looking to the commercial sector to derive some of the essential leading indicators is part of this approach, which we explore in Part III. In Part IV, we define a number of leading indicators, recognizing that these will vary across programs. We also provide a framework for how these indicators will be monitored and how they can inform the decision-making process. Finally, in Part V, we provide our concluding remarks.

II. Previous Initiatives

In its effort to control the cost of weapon systems, the DoD has implemented a number of strategies, some of which rely on lagging indicators and others that purport to make use of leading indicators. For instance, Selected Acquisition Reports (SARs) and the Nunn-McCurdy (NM) amendment use information on past performance in order to evaluate programs. Another strategy, cost as an independent variable (CAIV) is a strategy that sets a cost objective for an acquisition program, then manages the requirements and performance to meet that objective. Other strategies, such as earned value management (EVM) and, more recently, should-cost/will-cost, are both designed to provide leading indicators of program performance. In the following section, we discuss these strategies, their advantages, their shortcomings, and some of their results to date.

Selected Acquisition Reports

SARs summarize the latest estimates of cost, schedule, and performance status. Prior to 1968, the DoD had no system for monitoring the progress of major systems (Hough, 1992). In order to facilitate internal cost control, the DoD introduced SARs in 1968. Soon thereafter, SARs were submitted to Congress on a regular basis. The GAO praised the SAR system, describing it as a meaningful management tool for measuring and tracking the progress of major acquisitions (Hough, 1992), and then-Deputy Secretary of Defense David Packard asserted that the new system could be used to "clearly identify and explain the causes for increased costs that occur in the future" (*Acquisitions Weapons Systems*, 1969, p. 72). In 1975, the military departments were legally required to submit SARs. These reports generally are prepared annually; however, quarterly exception reports are required for those programs that have incurred unit cost increases of at least 15 percent, or schedule delays of at least six months. Quarterly, SARs are also submitted for programs that are re-baselined at major milestone decision points.

The total program cost estimates provided in the SARs include research and development, procurement, military construction, and acquisition-related operation and maintenance. Total program costs reflect not only actual costs to date, but future anticipated costs as well. SARs have served as the primary source of research into cost growth for decades. Unfortunately, however, SAR data cannot be used to identify a program's cost drivers. This is because the reports classify cost growth using variance categories that only show the effects of secondary factors (Hough, 1992). These categories include economic escalation, quantity change, schedule slippage, engineering modification, and estimating changes. In pointing out the irony of David Packard's initial hopes for the system, Hough (1992) asserted that "the failure to identify root causes of cost growth in the SAR … limits its utility to macroanalysis and identification of persistent problems" (p. 23).

Nunn-McCurdy Amendment

The NM, implemented in 1982 and modified in 2006 and 2009, requires the DoD to notify Congress when the unit cost growth of any major defense acquisition program is expected to exceed certain cost growth thresholds. Specifically, NM stipulates two levels of unit cost growth breach, referred to as the "significant level" and the "critical level." A "significant" unit cost breach occurs if a program experiences cost growth over 15 percent of the current baseline estimate, whereas a "critical" unit cost breach occurs if a program experiences cost growth of 25 percent over the current baseline estimate. This unit cost breach occurs if a program experiences unit cost growth above specified thresholds either as measured by total program acquisition unit cost (PAUC) or average procurement unit cost (APUC).

The NM law requires a program manager to fulfill specific criteria when a program breaches. From 1982 to 2006, implementation of NM did not seem to have significant impact on acquisition outcomes. The most consistent criticism of NM was that the measure was ineffective because programs would avoid incurring an NM breach by establishing a new "current" baseline. The NM statute was amended in 2006 by the DoD Authorization Act of 2005 (Public Law 109-163) to introduce new criteria. The new provision specified a second condition for incurring an NM breach: unit cost growth over the original baseline estimate. The revision did not change the reporting requirements for either the "significant" or "critical" unit-cost breach.

Congress amended NM again with the Major Weapons Systems Acquisition Reform Act of 2009, adding two new requirements to the process of recertifying programs that incur an NM breach. A program with an NM unit-cost breach now must (a) rescind the most recent milestone approval and (b) receive a new milestone approval before any actions regarding the contract may continue. The new milestone approval requires a certification that the costs of the program are reasonable, and the certification must be supported by an independent cost estimate that includes a confidence level for the estimate.

However, despite these efforts, defense programs continue to experience high unit-cost growth. Unit-cost growth has remained high since NM was implemented in 1982. At the same time, the true impact of NM is unclear. The DoD's data collection has been inconsistent because the DoD does not track acquisition information accurately or consistently across the entire department, nor is such information always provided in a timely manner. Definitions and baselines typically change multiple times over a program's development cycle. The data that is reported tends to be of only marginal value. Moreover, most reported information is input oriented, and, as a result, no consistent linkages exist between the data and the actual performance of a program.

In addition, NM generally identifies acquisition problems too late in the development process to allow program reforms to be effective. Although NM specifically states that Congress should be notified if a program manager believes that acquisition difficulties may occur, Congress is often not informed of a program's unit cost growth until an NM unit cost breach is imminent, or has actually taken place.

Ultimately, by the time a program manager reports that a program will likely trigger an NM breach, the program is likely to be too far along in its development to significantly alter its course. In 2011, a RAND report identified common root causes of NM breaches, including the use of immature technologies, unanticipated integration issues, unstable funding, ambitious scheduling, and ill-conceived manufacturing processes and insufficient research, development, testing, and engineering (RDT&E). Needless to say, problems of this nature simply cannot be corrected in order to bring costs back within initial expectations.

Cost as an Independent Variable

Developed in the 1990s, CAIV strives to elevate the importance of cost within the trade space. All acquisitions are assessed based on their cost, schedule, and performance. Collectively, these three parameters make up the trade space. Historically, performance has received the most emphasis and is often considered the independent variable. The other two parameters (i.e., the dependent variables) were varied as the program progressed, in order to maintain the desired performance. The goal of CAIV was to shift this emphasis from performance to cost, allowing variance in performance and schedule so that cost can be better maintained (Kaye, Sobota, Graham, & Gotwald, 2000). In short, this strategy attempted to create a cost-saving environment by emphasizing the importance of cost as well as flexibility with regard to performance and schedule.

Under design-to-cost (DTC), the predecessor to CAIV, the primary focus centered on meeting the projected average unit procurement costs. It has been argued that DTC led managers to focus on reducing near-term production costs, to the exclusion of system life-cycle costs. However, under CAIV, program managers take into account the estimated complete life-cycle cost of the program and adjust cost and performance accordingly. Moreover, there is specific recognition that the best time to reduce life-cycle costs is early in the acquisition process (Land, 1997, p. 27). In fact, according to Newnes et al. (2008), "50–70% of the avoidable costs of a product are in-built within the concept design stage" (p. 100). Similarly, research by Kluge (1997) suggested that most of the complexity in a product (and thus its cost) is generated by its design and not by customer demand. According to Kluge (1997), "complexity can, therefore, often be reduced without customers noticing much difference in the finished item … for instance, by standardizing parts and subassemblies" (p. 214). CAIV encourages program managers, when appropriate, to spend more money upfront in an effort to reduce production or operations and support costs.

The key tenets of CAIV, according to the DoD's *Defense Acquisition Deskbook* (1999), are as follows:

- Requirements are stated in terms of capabilities and may be exchanged, substituted, or adjusted for the sake of another. Capabilities should be established at the system level and not at lower levels.
- Early and continuous customer/warfighter participation in setting and adjusting program goals throughout the program is imperative.

- Trade space (i.e., cost gradient with respect to performance) around the cost objective is encouraged.
- Realistic but aggressive cost objectives are set early and updated for each phase of an acquisition program. (p. 37)

In 2002, then-Under Secretary of Defense for Acquisition, Technology, and Logistics (USD [AT&L]) Edward "Pete" C. Aldridge Jr. required that all Acquisition Category (ACAT) 1 programs use CAIV in order to control costs. In retrospect, however, it does not appear that use of the CAIV approach has achieved the desired results. To the contrary, a number of programs that relied on CAIV experienced, and continue to experience, major cost overruns. In the early 1990s, the DoD selected eight programs to serve as CAIV flagships. These programs, it was believed, would demonstrate how this initiative could contain costs. In 1999, the GAO identified additional program offices that were "leaders" in the application of various acquisition best practices, one of which was the CAIV approach (GAO, 1999b, p. 22).

SARs from 2010 featured five of the original flagship programs: the AIM-9X Sidewinder missile, the MIDS communications terminal, the JASSM cruise missile, the F-35 Joint Strike Fighter, and the SBIRS satellite program. The SARs also featured two of the programs that the GAO identified in 1999: the Advanced Amphibious Assault Vehicle (which is now known as the Expeditionary Fighting Vehicle, or EFV), and the Advanced Medium-Range Air-to-Air Missile (AMRAAM). The changes in quantity and the percentage change in cost (adjusted for quantity) for each of these programs are provided in Table 1.

Program	Change in Quantity	Percent Change in Program Cost (adjusted for quantity)
AIM-9X	+93	+13%
MIDS	+1,666	+12%
SBIRS	+1	+151%
JASSM	-429	+64%
JSF	-409	+58%
EFV	-432	+169%
AMRAAM	+2390	+41%

Table 1. CAIV Programs Cost Growth
(*Note*: The information in this table is from GAO [1999b] and Selected Acquisition Reports from 2010.)

None of the programs are below the initial estimated cost. In fact, among the seven programs, costs have grown by 47.6 percent on average—just slightly over the historic statistic (46 percent) cited in the 2007 RAND report referenced previously. Based on the wide range of percentages in Figure 3, one is led to conclude that the CAIV initiative is having little discernible impact, positive or negative, on program cost growth. Moreover, it is clear that CAIV did not enable the acquisition of planned quantities. In fact, if the JASSM, JSF, and EFV programs were to revert to their initial planned quantities, the percent change in cost would be significantly higher. Minimizing restrictions within the trade space by treating cost as an independent variable is a good first step. However, in practice, it appears that this approach, alone, does not go far enough.

Earned Value Management

EVM is a managerial tool that provides a systematic approach to the integration and measurement of cost, schedule, and performance on a project or task and uses those to estimate the completion time and cost. In June 2002, the Office of Management and Budget mandated the use of EVM systems for all major IT service and acquisition contracts. The DoD requires EVM on contracts worth more than $50 million and the application of at least some EVM principles on contracts worth more than $20 million. Over a decade ago, the secretary of defense decided to cancel the Navy A-12 Avenger II Program because of performance problems detected by EVM.

In the 1960s, the DoD developed the EVM process. EVM measures the value of work accomplished in a given period and compares it with the planned value of work scheduled for that period and with the actual cost of work accomplished. When using EVM, an integrated baseline is developed by time-phasing budget resources for the defined work, during the planning phase. As work is performed and measured against the baseline, the corresponding budget value is "earned." From this earned value metric, EVM allows (1) relating time-phased budgets to specific tasks and to requirements contained in a statement of work; (2) providing accurate, reliable, and timely data; and (3) measuring project progress and performance with related costs, schedule, and technical accomplishments. From the basic variance measurements, the program manager can identify significant drivers, forecast future cost and schedule performance, and construct corrective action plans to get the program back on track. Therefore, it equips both the government and contractors with the ability to examine detailed schedule information, critical program and technical milestones, and cost-to-date data.

EVM is superior to independent schedule and cost control for evaluating work progress and identifying potential schedule slippage and budget overruns. At the same time, however, program status reports derived through EVM are based on aggregated past performance, as opposed to discrete measures. Thus, it is not always possible to determine the immediate or root cause of cost overruns. Moreover, because the aggregated data is not necessarily predictive of a program's future performance, or cost, some government and commercial sector reports (e.g., Card, 2008) suggest that earned value should not be considered a leading indicator.

In addition, EVM, though widely advocated, has been applied very inconsistently over time, undermining its effectiveness. Several unfavorable findings from recent audits further indicate that EVM is not serving its intended function in the internal control process. GAO (2010d) identifies 11 key requirements for effective implementation of EVM in acquisition programs, grouped into three categories: establishing a sound EVM system, ensuring reliable data, and using earned value data to make decisions.

The GAO's (2010d) evaluation of EVM implementation in federal agencies finds that most programs do not fully implement the key practices needed to establish comprehensive EVM systems to help reduce acquisition risk. For example, some programs do not adequately determine an objective measure of earned value and develop the performance baseline. Without having such baseline review, programs have not evaluated the validity of their baseline plan sufficiently to determine whether all significant risks contained in the plan have been identified and mitigated. Moreover, some programs do not define the scope of effort using a work breakdown structure. As such, programs lack a basis for planning the performance baseline and assigning responsibility for that work, both of which are necessary to accomplish a program's objectives.

The data problem obstructs the effectiveness of EVM substantially. The fidelity of the information produced by EVM is critical to providing an objective assessment of a program's performance from which well-informed management decisions can be made. Thus, a critical precondition for effective use of EVM is that EVM data must be reliable before being used for decision-making. However, many acquisition programs did not fully ensure that their EVM data was reliable. Generally, most programs have established standard procedures to review earned value data, identify and record cost and schedule variances, and forecast estimates at completion. However, analysis of the EVM performance data and the variances from the performance baseline is still not adequate.

In short, until EVM systems are fully implemented, acquisition programs face an increased risk that program managers cannot effectively use EVM as a management tool to mitigate and reverse poor cost and schedule performance trends. The GAO's (2009) study of earned value data trends of the 16 federal programs indicates that most are currently experiencing cost overruns and schedule slippages.

In addition to these two problems associated with EVM implementation, EVM as a managerial tool may suffer from other limitations. EVM may not be able to tell the whole story of program development. First, EVM has no provision to track project quality. Thus, it fails to show program managers qualitatively whether a program is in good shape or bad shape. For example, when EVM reports that a requirement in a program is 60 percent completed, it cannot show whether the completed part of the requirement represents close to what is acceptable in terms of quality. It could be that program performance (in terms of cost and schedule) is achieved at the expense of quality. Second, EVM metrics cannot reveal the reasons why a program might be experiencing schedule or cost variances. And it may not accurately represent the cost and schedules that are most likely necessary for a project to achieve a particular functionality. By and large, it seems that heavy reliance on EVM has not improved program acquisition efforts.

It is also relatively easy to "game the system" so that a program appears to be progressing, when there are actually significant, albeit hidden, problems. Programs are often coded as green, yellow, or red so that upper-level management can keep tabs on the status of multiple programs with relative ease. According to CIOinsight (2005), there are several "tricks" that can be used to keep a program in the green, which are as follows:

- "Pad the schedule" by telling management that a three-month project will take four, which allows the program to keep up appearances and beat expectations, even if things are going wrong.
- "Push problem tasks forward" so that a project can remain in the green for a longer period of time.
- "Bump the task completion percentages" by changing the completion percentage of tasks that cannot be easily or objectively measured.
- "Re-baseline the project," after a small change request, by significantly elongating the schedule, which turns a red program green.
- "Integrate late in the game," which allows interoperability problems to remain hidden for the life of the program.

While it might be unfair to suggest that program personnel consciously engage in devious practices, it is nevertheless the case that they are incentivized to keep their program in the green, rather than report problems to management early on. Indeed, even senior management must report to their superiors as well as to members of Congress, who no doubt prefer to see green programs. After years of development, a "green" program may yield a product of less than the desired value, with everyone involved feigning ignorance as to what went wrong. If EVM is to be resurrected, it must be viewed as a tool, and not a shield to hide behind. It should also be emphasized that EVM, like the other approaches, is complex. The successful use of EVM requires knowledgeable and experienced acquisition personnel to make the progress judgments.

Should-Cost/Will-Cost

Implemented by then-USD(AT&L) Ashton Carter in 2011, the should-cost/will-cost approach was devised in response to anticipated national budgetary constraints identified by Congress. As of 2011, all ACAT I, II, and III programs use this approach. Simply put, should-cost/will-cost identifies low-value, high-cost elements of a program and seeks to increase value or decrease costs.

Under this approach, two separate cost estimates are developed: a non-advocate *will-cost* estimate, which provides the official basis for budgeting and programming, and a *should-cost* estimate for program management execution (Davies & Woods, 2011). The official budget baseline for the program is based on the non-advocate, historically based, will-cost estimate, which is usually developed by the Office of the Secretary of Defense's (OSD's) Cost Assessment and Program Evaluation (CAPE) office. CAPE estimates are typically derived by taking into account the costs of analogous programs. In contrast, the should-cost estimate is based on what the program manager believes is possible within "the context of creative, innovative, and disciplined measures to increase productivity" (Sledge, 2012, p. 1). In preparing their should-cost estimate, managers are encouraged to identify cost savings without relying on previous templates; rather, a should-cost review "attempts to break the cycle of historical-based cost estimation by challenging existing cost structures" (Sledge, 2012, p. 2). Accordingly, a should-cost estimate can include alternative material solutions, the trading of subcomponents, or reductions in performance expectations (Carter, 2011). Under should-cost/will-cost, program managers pay close attention to the difference between the should-cost and will-cost estimate. At every milestone decision, the difference is calculated and used as a criterion by which to evaluate the program.

According to a 1972 report by the Army Safeguard Office, "cost growth, the positive difference between ultimate cost and initial cost, is a function of the prevailing incentive systems, and incentive systems can be changed" (p. 3). Unfortunately, it does not appear that should-cost/will-cost provides managers with much incentive to build cost savings into their programs. On the one hand, program managers are required to budget to the historically based (and higher) will-cost figure; on the other hand, they must drive their suppliers to the lower, should-cost estimate.

Retired Army Colonel Nathanial Sledge (2012) writes that the new approach “reduces their management trade space, making it more challenging to demonstrate year-over-year progress” (p. 2). In other words, a program manager who works “to achieve a baseline of should-cost initiatives is shooting himself or herself in the foot” (Sledge, 2012, p. 3).

The should-cost/will-cost approach has other disadvantages. For instance, the will-cost estimate is created early in the program and therefore is prone to inaccuracy for a multitude of reasons, including unstable requirements and unknown sourcing. Because program “savings” under should-cost/will-cost are expressed as the difference between the two estimates, an inaccurate will-cost estimate can make achieving cost savings impossible, or even too easy. Either way, one cannot help but think that the outcome is somewhat artificial.

Finally, because system requirements are fixed but cost is not, it is virtually impossible to trade higher performance for lower costs. Just as it has in the past, this limitation will lead to the initiative’s eventual demise.

A New Approach

These four reforms have not improved the DoD’s acquisition outcomes significantly, in part because monitoring and enforcement measures tend to be reactive (e.g., alerting the DoD and Congress of difficulties once they have occurred) as opposed to proactive (e.g., providing advanced warning of programs that are likely to encounter difficulty). Indeed, the longer a program is extended beyond its scheduled completion, the longer management should expect the program will take to complete. Or, in the words of economist Nassim Nicholas Taleb (2007), “the longer you wait, the longer you will be expected to wait.” This is somewhat counterintuitive and perhaps explains why program management persists in the sort of wishful thinking described in the previous paragraphs. In other words, it appears that some managers tend to think, perhaps unconsciously, that a program’s duration (or cost, for that matter) is constrained by an upper limit. That is, the more days that pass, the closer the program must be to its end. Unfortunately, this is not the case.

This is why it is critical to identify issues and take corrective action as soon as future problems can be detected. Chalking up delays or cost increases to bumps in the road while hoping that the program will eventually find its way back on course is unrealistic and irresponsible. Instead, the program's path must be altered, sometimes significantly. The one-time cost increase or minor delay is a rare event indeed. Rather, schedule delays precipitate more schedule delays, and higher costs lead to still higher costs.

So far, the discussion has revolved around the nature of prediction errors, but in the abstract. Note, however, that within the DoD, vicious cycles of ever-increasing costs can be attributed to very real characteristics of DoD programs. For instance, the majority of programs seek to develop, manufacture, and field a certain quantity of discrete platforms (e.g., airplanes, tanks, radios, etc.). Yet, because program costs are not adequately controlled during the design and development process, DoD programs often must reduce planned quantities in order to stay within their planned overall program budgets (GAO, 2001). The Air Force's air superiority fighter, the F-22 Raptor, suffered this fate. As costs increased, quantities were reduced, causing program costs (adjusted for quantity) to increase, which, in turn, triggered further reductions in quantity. Originally, the Air Force planned to order 750 F-22s at a cost of $26.2 billion (Williams, 2002). Beginning in 1991, the Air Force reduced its order to 650 aircraft, then to 438 in 1994, and finally down to 183 in 2011. As late as 2006, the costs continued to climb from $361 million per aircraft, to $412 million per aircraft in 2012 (GAO, 2011). In the end, the F-22 was not procured in the numbers required to replace the F-15s. Moreover, the F-22, while praised by DoD officials and pilots alike, included far fewer capabilities than originally planned.

It is clear that new projects can be more effectively planned, budgeted, and scheduled when historical data from previous similar programs are used (Rhodes, Valerdi, & Roedler, 2009). Indeed, the widespread use of EVM in both the public and commercial sectors attests to this commonsense reality. Should-cost/will-cost takes this approach a step further by not only determining how much a program will cost, based on historical data, but also how much it should cost if, for example, sourcing were to be carried out more efficiently, or if process redundancies were to be eliminated.

However, historical data are not always reliable and may in fact be misleading, depending on how they are interpreted. This is especially true within the context of projects that rely on cutting-edge technology for which there is little precedent. Those who devise the earned value or should-cost and will-cost estimates may, for instance, attribute a past program's success to a certain process or program manager, while perhaps failing to recognize that the technology in question was less complex than the one that is currently under development.

More broadly, it is important to remember that people select and recount events sequentially, thereby creating a narrative within which one assumes that one event was caused by a previous one. And even if one can accurately identify the proximate causes of particular program failures (or successes), it does not follow that the overall failure (or success) of the program can be attributed to the string of outcomes constructed by program personnel.

The DoD should strive to eliminate potential problems early on so as to remove the guesswork that takes place after the fact. In all likelihood, a program's failure can be attributed to a host of elements that interact in unpredictable ways. However, as mentioned previously, the root cause of program failure invariably comes down to a mismatch between the DoD's needs and desires and its available resources, technical or otherwise. By making use of better leading indicators, the DoD can better ensure that resources are consistent with the program's development plan so that success can be more easily achieved. Simply put, it is far easier to correct problems early on, and ensure that things go right, than it is to correct the problems, once they have had a chance to fully develop.

III. Product Development in the Commercial Sector

When product development is undertaken by the commercial sector, cost overruns, schedule delays, and program failure—although they do occur—are generally of a lower magnitude. The commercial sector processes seem to naturally constrain schedules and costs. In this section, we examine processes leading commercial firms use, and use this analysis to help inform the development of new leading indicators for the DoD.

Knowledge-Based Development

In order to survive and flourish under fierce market competition, commercial firms have to develop increasingly complex products in less time and under tight budgets. To achieve this goal, leading firms adopt a knowledge-based development process, ensuring that a sufficient level of knowledge exists at critical junctures throughout the acquisition process.

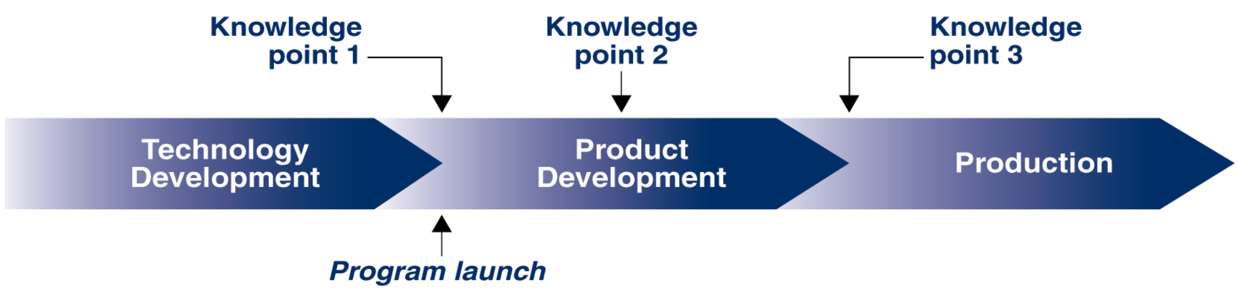

Figure 3. Commercial Sector Approach to Product Development (GAO, 2004)

Generally, leading commercial firms have identified three critical junctures at which they must have sufficient knowledge to make large development decisions or continue the development process (see Figure 5). First, before the initial program development, customers' expectations should be matched with the firms' resources, including technology, engineering, time, and funding. Once a development decision has been made, program designers should ensure that they have sufficient capacity to meet the performance requirements of that product, and the design of that product should be stable enough for routine production. After this stage, program developers must show that the product can be produced within budget, schedule, and performance targets. This process actually is an evolutionary and incremental one. Basic requirements in each phase should be achieved first, before the program is allowed to move forward. Together, these practices ensure a high level of knowledge and thus a low degree of risks introduced into the acquisition process that could further result in favorable cost, schedule, and performance outcomes.

Commercial firms find that this evolutionary development strategy is especially useful when the goal is to use new technology and reduce development duration.

Paralleling these three critical junctures, there are three knowledge points. The requirements at each knowledge point must be met prior to moving forward. The knowledge points are described as follows.

- ***Knowledge Point 1 (technology development): Resources and needs match.***

  This point appears when a match is made between the market needs and commercial firms' development capacities (knowledge, time, and budget). Leading companies use extensive communication mechanisms to reach customers in order to ensure that their needs could be aligned with the companies' resources and abilities. Launching a program before requirements and resources are matched may result in a product that fails to perform as expected (e.g., more costs and longer development duration).

- ***Knowledge Point 2 (product development): Product design is stable.***

  Knowledge point 2 occurs as a product's design is stable and reliable; the design meets customers' requirements (including unit cost and being within the companies' cost and time schedule as well). This is achieved at the product's critical design review. Failure to ensure the stability of the design would result in costly design changes in development and production stages.

- ***Knowledge Point 3 (production): Production processes are mature.***

  Knowledge point 3 happens when the product can be manufactured with manageable cost, schedule, and quality. This is achieved by ensuring the statistical control in the manufacturing processes.

Target Costing

Not only do commercial firms determine the level of knowledge needed to progress from one phase of a project to the next, but they often determine the unit cost of the product early on in development in order to ensure that the projected market size will be realized. In the commercial sector, this approach, known as target costing, was first introduced in Japan in the early 1960s. Today, it is widely used by commercial firms throughout the developed world, but is rarely implemented successfully within the DoD.

Whereas cost traditionally has been considered an outcome of product development, target costing treats it as an input. The target cost for a product is determined using a simple formula: Target Cost = Estimated Selling Price – Desired Profit. But the number made public (and especially to the customer) is the selling price. The target cost is an internal target for the developer.

However, target costing is not merely the imposition of a cost ceiling. As Zengin and Ada (2010) pointed out, “manufacturers cannot make a trade-off between cost, quality, and functionality of the product with only cost considerations in mind” (p. 5594). Indeed, in today’s competitive global markets, a business that pursues such a strategy would quickly fold. Rather, target costing, as a strategy, promotes creativity and new ways of thinking to increase performance while discouraging the inclusion of non-value-added functions. As a result, today’s customers are able to purchase lower-cost, higher-quality products that meet their needs. Cooper and Chew (1996) described the logic behind target costing as follows: “Looking at today’s marketplace, the organization maps customer segments and targets the most attractive ones … and then determine[s] what level of quality and functionality will succeed within each segment, given a fixed target price, volume, and launch date” (p. 1). Gordon (2000) noted that many firms use target costing “as a way to focus on managing costs, rather than recovering costs through some form of cost-plus pricing mechanism” (p. 169).

After the target cost is determined, it must be apportioned among the many internal cost centers, including marketing, manufacturing, general and administrative, logistics, distribution, and the price of purchased items (Ellram, 2006). Following this high-level allocation to features or functions, costs are apportioned further at the level of the individual component, material, or service.

Mihm (2010) observed that “target costing does *not* require perfect knowledge about the component” (p. 1334). Fairly accurate component cost estimates can be developed via systematic value analyses of comparable existing parts (Mihm, 2010). And whereas the target cost of the product remains firm throughout the development process, component cost estimates are permitted to fluctuate as product development evolves. Typically, each product feature is ranked in terms of its relative importance. Some firms may go as far as to assign specific numeric weights to each feature. These weights are then used to determine where the firm can adjust costs while maintaining, or even enhancing, the product’s value (Ellram, 2006). In order to ensure the inclusion of the most valuable features, the target cost of one component may be increased while that of another is reduced. The most successful firms continually rely on their sense of customer value as the basis for their cost-allocation decisions (Cooper & Chew, 1996). Indeed, even after the product is released, firms strive to increase the product value and incorporate any improvements into future iterations. Even within a single iteration, firms work to improve the manufacturing and other processes in order to reduce costs. The target costing process is summarized in Figure 4.

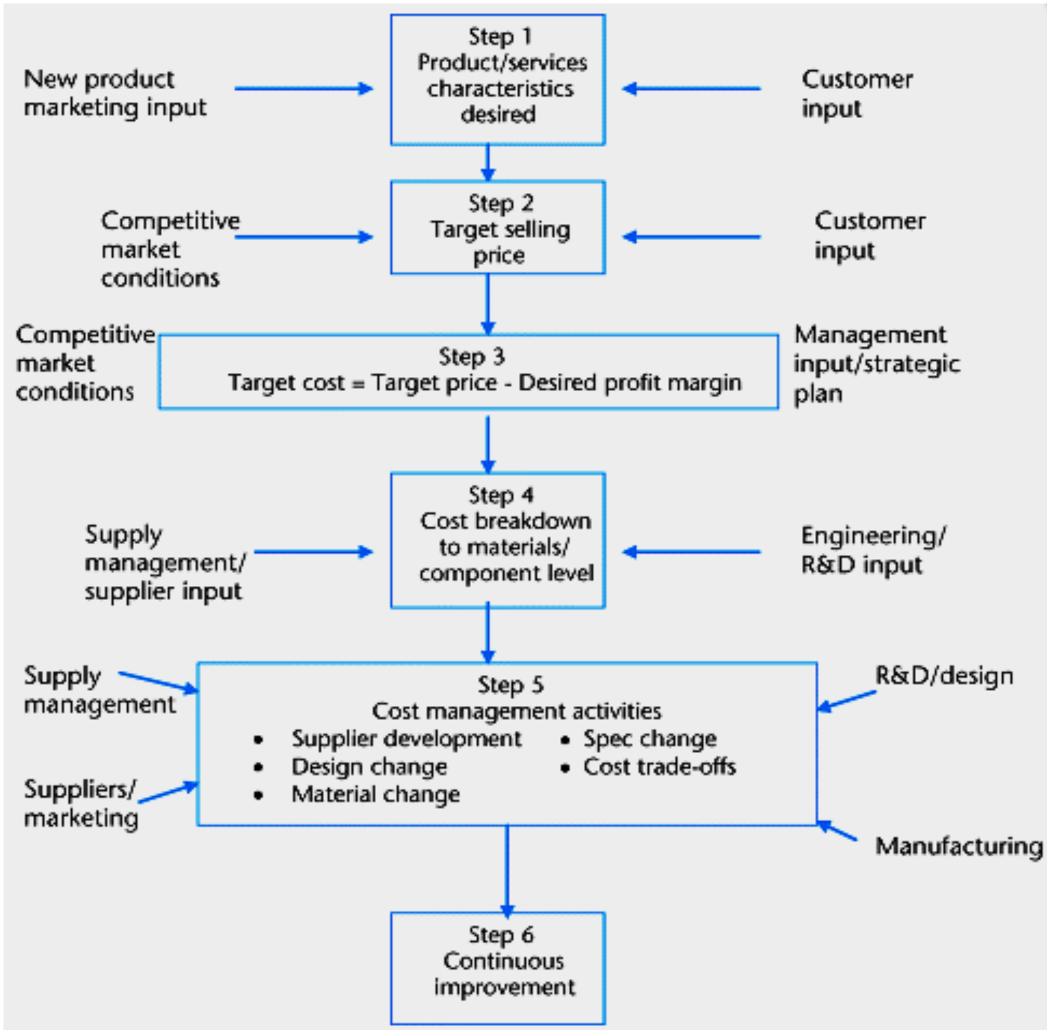

Figure 4. Target Costing Process (Ellram, 2006)

The automobile industry illustrates the benefits of target costing. With relatively few new manufacturers gaining ground in the market, reliability, cost, and performance are all major contributors to the quantity of vehicles a manufacturer sells. Accordingly, components and potential material solutions are stringently analyzed in terms of their cost and value to the customer. It is no mystery that the Japanese firm Toyota has had considerable, prolonged success, largely on account of its costing approach. Indeed, when asked why Toyota is a top-selling car company, everyday Americans readily respond that it offers customers higher quality at lower cost.

Today, virtually every successful commercial firm employs a cost-driven approach to product development. For a variety of reasons, the DoD has been reluctant to do the same. But given current and impending budgetary constraints, it may soon have little choice in the matter. Admittedly, there are significant challenges that must be overcome in order for a cost requirement to be viable. In the commercial sector, not only is product development cost driven, but it is also market driven. Firms spend considerable sums in order to better understand what the customer is willing to pay for; a firm that adds extraneous features of little added value to the customer is punished in the market.

The defense market, however, is characterized by very few firms in each sector and only one customer (i.e., a monopsony). Because weapon systems are contracted for in advance of their production, the contractor generally is not incentivized to translate the diffuse desires of the customer—in this case, the DoD—into an effective and efficient product. Rather, the DoD specifies requirements upfront, and in great detail, for fear that they never may be developed. In fact, there is frequently a perverse incentive to "gold-plate" products by adding every desired feature, including some of little marginal value. This is especially true within the context of complex, new product development where neither the DoD, nor the contractor, fully understands the attributes and capabilities of the end product.

Take, for example, the development of the C-5, which, to this day, has a number of unique features. For example, the nose swings open on hinges so that, in addition to an aft ramp, a front ramp can be extended for easy loading and unloading of equipment. Another innovation is an automated, built-in test capability that "electronically monitors 600 test points, locates any troubles, and prints out repair instructions" (Shults, 1976, p. 4). The initial aircraft specifications, however, also called for a number of innovative features that in retrospect were a clear case of over-specification by the Air Force. For example, included in the original requirements document was the requirement for an in-flight airdrop capability—the aircraft would have to be able to airdrop single loads of up to 50,000 pounds from the rear cargo bay. There was also a requirement for advanced avionics that would allow the C-5 crews to identify drop zones and conduct airdrop operations at night or in poor weather. Further, there was a requirement for a terrain-following radar so that the C-5 could fly at low altitudes to evade detection by the enemy (Shults, 1976). Additionally, there was a requirement for the C-5 to be capable of landing on short, unimproved runways. Early criticism surrounding the inclusion of these features—many believed that they would never actually be used—was, for the most part, overlooked initially. As it turned out, including these capabilities proved technically challenging and, ultimately, very costly to develop.

Even though product-marketing input is less of a determinant in the cost of DoD systems, the DoD's input can play a large role. Indeed, Figure 4 suggests that both types of input—product marketing and customer input—are essential. In the commercial sector, large retailers such as Wal-Mart have significant control over their supply bases because they have considerable buying power. Ellram (2006) noted that even in the manufacturing sector, large firms like Dell might be able to dictate pricing to companies like Intel. In the same way, the DoD must use all available strategies (e.g., competitive dual-sourcing) to leverage its size and buying power, and exert downward pressure on the cost of weapons systems.

The pace of technological innovation is another example of market forces at work within the commercial sector. The accelerating rate at which new personal computers, smartphones, and MP3 players appear on store shelves is as much a function of new technology (creating the demand for new capabilities) as it is the accumulation by industry of users' feedback and desires, the essential core of which is reflected in the design of the product. Once the two processes—user input and technological innovation—merge, an uninterrupted loop spurs ever increasing gains in efficiency and performance. Because development is incremental, commercial firms are typically well-positioned to estimate costs.

Firms can further refine the accuracy of these estimates by relying on standardized components that are manufactured by other firms (Rush, 1997). Automakers, for instance, often use standardized components because they are cheaper than built-to-order parts and generally come with a warranty, which reduces the manufacturer's cost of long-term operations and support while maintaining the reliability of their products.

For the DoD, it is often more challenging to pursue an incremental approach to development because the customer base for each product is relatively small and systems have relatively long life cycles. It can also be more challenging to incorporate commercial-off-the-shelf (COTS) components into defense systems; barriers to their use include proprietary interfaces, stringent military environmental requirements, specialized cost accounting requirements, export controls, and continued cultural resistance.

These challenges notwithstanding, the DoD must strive to approximate the approach to cost management techniques that are used by commercial firms. Too often, the perceived uniqueness of the defense market is used to justify relaxed policies with regard to cost control. However, the commercial sector's experience indicates that holding fixed the unit cost of a product is not only a possibility but also a preferable strategy in today's competitive market. Although competition within the defense market is less fierce in some respects, the cost constraints faced by the DoD are no less significant.

In a 2005 report, the GAO provided a telling glimpse into some of the dysfunction that plagues typical DoD programs:

> Leadership rarely separates long-term wants from needs based on credible, future threats. As a result, the DoD starts many more programs than it can afford—creating a competition for funds that pressures program managers to produce optimistic cost estimates and to overpromise capabilities. Moreover, our work has shown that DOD allows programs to begin without establishing a formal business case. And once they begin, requirements and funding change over time. (p. 2)

This description stands in stark contrast to the commercial sector's targeted, knowledge-based approach to product development described in the previous section. Because the DoD does not operate in a market-driven, customer-competitive environment, in which technology, requirements, innovation, and cost constraints are shaped by both internal and external forces, it must steer product development through the conscious implementation of certain policies, procedures, and safeguards. The specifications that are built into Toyota's next vehicle—in terms of cost and technology—are constrained, at least partially, by those that are offered by its competitors. In addition, Toyota can forecast the cost of these specifications with a high level of accuracy. Leading indicators can help the DoD achieve product development success by drawing attention to essential elements that are naturally regulated within the commercial sector—and thus often overlooked by the DoD.

IV. Leading Indicators

Most of the indicators that DoD programs currently use reflect current status and historical trends. The goal of developing leading indicators is to enable better forecasts of a program's performance in the future. In Section III, we described several features of product development that we believe can inform the creation of meaningful leading indicators. These indicators, which may rely on both qualitative and quantitative assessments, provide an indication of a program's future performance.

The DoD should develop indicators to monitor these features of product development. We also contend that these indicators can be used in two distinct ways:

- First, the use of indicators will ensure that fewer programs will begin development on a weak case, thus avoiding a costly, though all too common, mistake—initiating a program that should have not been started.
- Second, the use of leading indicators will provide program managers with earlier warnings of impending difficulties as a program progresses, which program managers can take into account to correct minor difficulties before they become costly revisions.

The leading indicators that are used, as well as their relative importance, will undoubtedly vary across programs. However, there are some common indicators that will likely play a role in most DoD programs. Below, we describe these common indicators. Then, we discuss how program leadership might go about implementing leading indicators, determining their relative importance, and tracking them over time.

Pre-Milestone B

As mentioned, certain leading indicators can be measured prior to program launch in order to help determine a program's long-term viability. These indicators include

- initial requirements,
- technology readiness,
- senior leadership support,

- skilled program management, and
- experienced supporting organization.

Initial Program Requirements

A primary reason for cost and schedule problems is the lack of clear tradeoffs among cost, schedule, risk, and requirements supported by rigorous system engineering, budget, and management processes during program initiation. This can be caused by either an incomplete set of requirements or a larger set of requirements (over-specified or "gold plated") with a corresponding growth in the number and scope of key performance parameters. There is a tendency within the acquisition environment to develop overly ambitious products—sometimes referred to as "revolutionary" or "big bang" acquisition programs—that embody too many technical unknowns and insufficient knowledge about performance and production risks. Identifying all requirements as early as possible during the design phase is a difficult but desirable goal. Incomplete requirements can be a significant source of program difficulties, resulting in requirements volatility as the program matures and sees cost and schedule over-runs. The knowledge gaps are largely the result of a lack of early and disciplined systems engineering analysis of a weapon system's requirements prior to beginning system development, which translates customer needs into a producible weapon system. If this early systems engineering is not performed, as has often been the case with the DoD's major acquisitions in the past, significant cost increases can occur as the system's requirements become better understood by the government and the contractor (GAO, 2010a).

Such was the case with the DoD's Joint Tactical Radio System (JTRS). With JTRS, functions that are traditionally built into a radio's hardware were, instead, implemented through software. An open systems framework known as the Software Communications Architecture (SCA) was key to the system's interoperability; it "told designers how elements of hardware and software are to operate in harmony" (Brown, Sticklan, & Babich, 2006, p. 1), thus enabling users of different JTRS variants (airborne, maritime, ground, fixed, etc.) to load and run the same software applications. However, the JTRS program's failure to define the specific limitations of the available technology, and, instead, rely heavily on the SCA—a "responsive" and "flexible" architecture—leading to the belief that difficult technical problems could be addressed further downstream (Gansler, Lucyshyn, & Rigilano, 2011). Unfortunately, this was not the case, and the program was eventually terminated.

Even though evolutionary development is the preferred process for the development of the DoD's systems, partial solutions are not always embraced. Often, all of the required capabilities are frontloaded onto the initial requirements document. By adopting an evolutionary approach, however, essential technologies can be fielded in the near term, delaying the instantiation of more time-intensive, costly, or technically challenging capabilities. Without an evolutionary approach, by the time the product is ready for production, the "next best thing" has already taken root in developers' minds, hastening its obsolescence. Adopting an evolutionary approach whereby a basic capability is fielded—and incremental capability improvements are periodically made in subsequent blocks—can actually mitigate risk in the long-run. By shortening development timetables and ensuring the use of mature technologies, such an approach would reduce the risk of program delay and cost overruns. An evolutionary approach also ensures that operational experience informs future versions of a product's requirements.

Program decisions to begin design and/or production are too often made without sufficient knowledge. As a result, requirements tend to be overly ambitious and thus unachievable. The DoD has two choices: improve knowledge or lower user expectations. Because a system should not be designed as a final solution but as an initial response to a problem (Keating et al., 2003), the latter is more appropriate. However, within the DoD, requirements are often added throughout the development process that increase system complexity. This added complexity translates to longer schedules and higher costs.

The number of a program's requirements, and perhaps, more importantly, the extent to which all of the requirements are planned for the initial release, provide an indication of future schedule and cost growth. Once agreed upon by the relevant stakeholders, the impulse to add requirements must be avoided. And, as requirements are added and changed, the program should be assigned a higher level of risk.

Technology Readiness

Failure to properly mature new technologies almost invariably leads to cost and schedule over-runs in weapons system programs (GAO, 1999a). Recognizing this, the DoD already relies on Technology Readiness Levels (TRLs) to assess the maturity of technologies, and the risks associated, before incorporating that technology into program development. TRLs provide a common understanding of technology status and thus help management in making decisions on the development and transition of technology. The TRL classification system was originated by the National Aeronautics and Space Administration (NASA) in the 1980s and was later adopted by the DoD (see Table 2). The DoD treats TRLs as benchmarks. In fact, DoD Instruction (DoDI) 5000.02 prevents the award of an engineering and manufacturing development (EMD) contract before the relevant technologies reach TRL 6. However, the DoD should also regard TRLs as leading indicators of potential problems, to be measured and monitored throughout early development, and assign risk to programs accordingly.

It is important to recognize that TRLs do not assess uncertainty involved in maturing and integrating a technology into a system and thus cannot assess maturity at the system level (Gove, Sauser, & Ramirez-Marquez, 2010). DoD systems have become more complex, generally supporting a variety of users and requiring large numbers of developers and maintainers. The intricacy of the many internal and external interfaces contributes significantly to this complexity. Most importantly, today’s systems are software-intensive, creating a greater integration challenge based on the innumerable, potential logic paths. Moreover, functionality that, in the past, was deeply embedded in the physical configuration of components “has begun to emerge as software, enabling synergies among components that would have been unimaginable only a few years ago” (National Research Council, 2008, p. 18). This creates an increasing integration challenge for the developers of the DoD’s systems.

TRL	Definition	Description
9	Actual system proven through successful mission operations.	Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions.
8	Actual system completed and qualified through test and demonstration.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
7	System prototype demonstration in an operational environment.	Prototype near, or at, planned operational system. Represents a major step up from TRL 6, requiring demonstration of an actual system prototype in an operational environment such as an aircraft, vehicle, or space. Examples include testing the prototype in a test bed aircraft.
6	System/subsystem model or prototype demonstration in a relevant environment.	Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in simulated operational environment.
5	Component and/or breadboard validation in relevant environment.	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so it can be tested in a simulated environment. Examples include "high fidelity" laboratory integration of components.
4	Component and/or breadboard validation in laboratory environment.	Basic technological components are integrated to establish that they will work together. This is relatively "low fidelity" compared with the eventual system. Examples include integration of "ad hoc" hardware in the laboratory.
3	Analytical and experimental critical function and/or characteristic proof of concept.	Active R&D is initiated. This includes analytical studies and laboratory studies to physically validate the analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
2	Technology concept and/or application formulated.	Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies.
1	Basic principles observed and reported	Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a technology's basic properties.

Table 2. Technology Readiness Levels in the DoD
(*Note*: The information in this table is from DoD, 2011.)

Indeed, Mosher (2000) highlighted system integration as the most difficult part of any development program. Henderson and Clark (1990) emphasized that systems often fail because attention is given to the technology while knowledge of the linkages/integrations is overlooked. It is clear that an additional indicator is needed for system-level assessment. Gove (2007) and Gove, Sauser, and Ramirez-Marquez (2010) have developed and suggested the use of Integration Readiness Levels (IRLs) to access the interfacing of compatible interactions for various technologies and the consistent comparison of the maturity between integration points (see Table 3).

	IRL	Definitition	Description
Pragmatic	9	Integration is Mission Proven through successful mission operations.	IRL 9 represents the integrated technologies being used in the system environment successfully. In order for a technology to move to TRL 9 it must first be integrated into the system, and then proven in the relevant environment, so attempting to move to IRL 9 also implies maturing the component technology to TRL 9.
	8	Actual integration completed and Mission Qualified through test and demonstration , in the system environment.	IRL 8 represents not only the integration meeting requirements, but also a system-level demonstration in the relevant environment. This will reveal any unknown bugs/defect that could not be discovered until the interaction of the two integrating technologies was observed in the system environment.
Syntactic	7	The integration of technologies has been Verified and Validated and an acquisition/insertion decision can be made.	IRL 7 represents a significant step beyond IRL 6; the integration has to work from a technical perspective, but also from a requirements perspective. IRL 7 represents the integration meeting requirements such as performance, throughput, and reliability.
	6	The integrating technologies can Accept, Translate, and Structure Information for its intended application.	IRL 6 is the highest technical level to be achieved, it includes the ability to not only control integration, but specify what information to exchange, unit labels to specify what the information is, and the ability to translate from a foreign data structure to a local one.
	5	There is sufficient Control between technologies necessary to establish, manage, and terminate the integration.	IRL 5 simply denotes the ability of one or more of the integrating technologies to control the integration itself; this includes establishing, maintaining, and terminating.
	4	There is sufficient detail in the Quality and Assurance of the integration between technologies	Many technology integration failures never progress past IRL 3, due to the assumption that if two technologies can exchange information successfully, then they are fully integrated. IRL 4 goes beyond simple data exchange and requires that the data sent is the data received and there exists a mechanism for checking it.
Semantic	3	There is **Compatibility** (i.e. common language) between technologies to orderly and efficiently integrate and interact.	IRL 3 represents the minimum required level to provide successful integration. This means that the two technologies are able to not only influence each other, but also communicate interpretable data. IRL 3 represents the first tangible step in the maturity process.
	2	There is some level of specificity to characterize the **Interaction** (i.e. ability to influence) between technologies through their interface	Once a medium has been defined, a "signaling" method must be selected such that two integrating technologies are able to influence each other over that medium. Since IRL 2 represents the ability of two technologies to influence each other over a given medium, this represents integration proof of- concept
	1	An **Interface** between technologies has been identified with sufficient detail to allow characterization of the relationship.	This is the lowest level of integration readiness and describes the selection of a medium for integration.

Table 3. Integration Readiness Levels
(*Note*: The information in this table came from Gove, Saucer, & Ramirez-Marquez, 2007)

On a scale similar to TRLs, IRLs could be used in conjunction with TRLs to help assess the integration of complex system. Specifically, IRLs could serve the following goals:

- Provide an integration-specific metric to determine the integration maturity between two or more configuration items, components, and/or subsystems.
- Provide a means to reduce the uncertainty involved in maturing and integrating a technology into a system.
- Provide the ability to consider the meeting of system requirements in the integration assessment so as to reduce the integration of obsolete technology over less mature technology.
- Provide a common platform for both new system development and technology insertion maturity assessment.

Senior Leadership

DoD programs with strong senior leadership support generally are more stable and as a result are more successful. Those programs with strong senior leadership support, sometimes because they had a more immediate requirement, were viewed as a higher priority by senior leaders in the services and DoD. Further, this consistent leadership support from DoD and the services promoted the development of better business plans and helped program managers to adapt to the inevitable program perturbations (GAO, 2010c).

This senior leadership support translates into greater attention throughout the acquisition hierarchy. This was vividly demonstrated by the Mine-Resistant, Ambush-Protected (MRAP) vehicle program, perhaps the largest and fastest industrial mobilization effort since World War II. Secretary of Defense Gates galvanized support for the MRAP program and became its most important champion. In the words of General Petraeus (2010), "Secretary Gates' direction was a key catalyst and a pretty key factor in production of the MRAPs." In May of 2007, his fifth month in office, Gates declared the MRAP the DoD's top acquisition priority, and called for "any and all options to accelerate the production and fielding of this capability" (Osborn, 2007).

Once the requirement was formally approved, the vehicles began to arrive in theater within six months (Gansler, Lucyshyn, & Varettoni, 2010). Although the Secretary of Defense cannot be expected to be personally involved to this degree, with every major DoD program, high-level support helps to ensure a program's success.

Within the DoD, very high-level acquisition positions require private sector industry experience. Indeed, the most qualified senior leaders have often spent a considerable amount of time working in the private sector. Unfortunately, political appointees in senior DoD leadership positions serve an average of only 18–24 months (Aerospace Industries Association, 2007) and, as a result, may have limited sustained impact on a specific program. In addition, DoD programs may not begin to field useful capabilities for several years. As a result, political appointees may be less inclined to launch large new programs in the first place.

In addition, career government executives sometimes adopt a wait-and-see attitude with regard to incoming appointees. Executives, who "personify the cultures of their departments" and have "intimate knowledge on how things really are accomplished in day-to-day operations," may regard appointees' visions of their program as unrealistic (Parchem & Gowing, 2009, p. 1). According to Parchem and Gowing (2009), the collision of idealism and practicality "can cause the actual productivity of the organization to come to an abrupt halt" (p. 1).

Frequent changes in senior leadership (political and military) can also lead to significant changes in an organization's priorities, goals, and strategies. These changes can also significantly impact relationships with partnering organizations. At the program level, the lack of sustained leadership often contributes to program delays and setbacks, which can create tension among stakeholders. Frequent leadership turnover can also insulate and strengthen the existing organizational culture. Long-term or permanent employees may be reluctant to participate in organizational change initiatives that significantly change their day-to-day responsibilities when the leaders who initiated these changes are not present to see them through. Consistent senior leadership support provides a good indicator of program success.

Program Managers

Program managers of successful programs tend to share key attributes, such as experience, leadership continuity, and communication skills that facilitated open and honest decision-making. Their programs established sound, knowledge-based business plans before starting development and then executed those plans using disciplined approaches (GAO, 2010c). They also pursued incremental acquisition strategies and leveraged mature technologies, both of which are important leading indicators of program performance in and of themselves. They were able to invest in early planning and systems engineering, and made trade-offs to close gaps between customer needs and available resources to arrive at a set of requirements that could be developed within cost and schedule targets. After approval, the programs resisted new requirements and maintained stable funding.

Such was the case with the Joint Direct Attack Munitions (JDAM) program. In the early 1990s, Congress authorized the DoD to launch a small number of pilot programs that, to the extent possible, would rely on standard business practices. These programs were known as Defense Acquisition Pilot Programs (DAPP). Program managers were afforded considerable autonomy and were able to bypass several of the more lengthy documentation and reporting requirements. The JDAM program was among the first of these pilot programs. Its program manager, Terry Little, took full advantage of the flexibility he was granted. Little conducted a two-week training on how to work in a more commercialized environment. Little made it clear to his team that he would not tolerate the old way of doing business on this project (Ingols & Brem, 1998).

In 1999, during Operation Allied Force (NATO operations in Yugoslavia), U.S. bombers launched over 600 JDAMs with 96 percent reliability and hitting 87percent of intended targets (Myers, 2002). Over time, as technology improved, the Air Force and Navy acquired updated versions with enhanced guidance technology that could be used on newer aircraft. Today, the average per-unit production cost, adjusted for inflation, remains about the same (GlobalSecurity.org, 2011). It is clear that the firm price ceiling, demanded by the Air Force chief of staff and achieved by Little and his team, in addition to the accelerated acquisition plan and the use of off-the-shelf components and other commercial practices, resulted in significant cost savings and allowed the Air Force to acquire the required quantity of weapons.

These practices are in contrast to prevailing pressures in the DoD that force programs to compete for funds by exaggerating achievable capabilities, underestimating costs, and assuming optimistic delivery dates. Program managers of stable and successful programs are able to make knowledge-based, disciplined decisions from the start and resist pressure to overreach or add requirements because of this strong institutional support (GAO, 2010c).

In addition to support from the top, program managers from successful programs tended to have similar attributes for success such as experience, leadership continuity, and communication skills that facilitated open and honest decision-making. These program managers were empowered to make good decisions, allowing them to be accountable for the success or failure of the program. Having skilled, experienced program managers is a key indicator for a successful program.

Experienced Supporting Organization

The DoD's programs are often large and complex, and, as a result, there is a need for a team of professionals with diverse backgrounds and skills to manage them. The DoD's acquisition workforce is often lacking in numbers, skills, and experience. Between 1990 and 2000, the acquisition workforce was cut by a total of 60 percent. As the DoD's budgets grew significantly during the first decade of the 21st century, the acquisition workforce failed to keep pace. Further, within the military, there has been a long-time belief that acquisition and contracting work was a mere administrative function, which ultimately contributed to a general disinterest in these career paths across all services. In some cases, individuals without proper training and/or experience are given acquisition positions (most notably those undertaking contingency contracting responsibilities and, more recently, those assigned as contracting officer's representatives). In fact, in 2013, approximately 55 percent of the total DoD acquisition workforce had less than five years of experience—most of the recent hiring had been of "interns."

Moreover, in the past, the DoD was at the cutting edge of technology, leading the innovation in jet engines, space, and microelectronics; however, during the last few decades, with the growing commercial importance of information technologies, the private sector has taken the lead. And, although the DoD's older employees may have extensive acquisition experience, their technical skills frequently have not kept up with the rapidly evolving information technology.

Accordingly, these legacy employees are significantly less likely than their private-sector counterparts to have the requisite skills for the DoD's current complex requirements.

We note that the use of any cost control approach requires a trained and experienced acquisition workforce. The workforce must have sufficient understanding of industry behavior and incentives in order to achieve the desired results. Current shortages of experienced acquisitions personnel are of particular concern. For obvious reasons, the DoD can rely on contractors only up to a certain point; that is, the DoD cannot outsource program management, or management and oversight of systems engineering, and expect to acquire efficient, affordable systems (National Research Council, 2008).

Recruiting qualified, experienced systems engineers is a challenge not only for the DoD, but for industry, too. The problem is twofold. First, the production of systems engineers by U.S. universities has increased very slowly over the past decade, despite increased demand, growing salaries, and other incentives. Second, formal knowledge of the systems engineering discipline only goes so far; to be successful within the discipline, one must also have specific domain experience (National Research Council, 2008, p. 9).

A good staffing strategy summarizes approaches to identify, attract, and retain a qualified and diverse pool to meet current, ongoing, and future staffing needs. Key civilian staff, such as senior engineers, often serve many years in the program office and provide continuity and information necessary for knowledge-based decision making. The GAO (2010b) noted that the continuity of key civil service and contractor personnel has proven very beneficial because several other personnel have left the program due to military deployments and reassignments. Long-term staff planning, undertaken in conjunction with acquisition goals, secures the skills and abilities needed to achieve those goals. To help fill the shortages, programs have often turned to systems engineering and technical assistance (SETA) contractors to cover the shortages. But, even when they are used, the DoD must be able to manage and oversee the SETA providers.

A critical assessment of the education, experience, and quality of the supporting staff can provide a good indication of a program's performance. For instance, a simple chart indicating the number of unfilled position vacancies would highlight potential risk to program cost and schedule. Program offices should identify and track critical positions, as well as the total actual head count for both quantity and quality (education and experience). This indicates that the program has the minimum staffing required to execute the planned program. If there are staffing shortfalls, the program can anticipate future challenges.

Progress Indicators

Once the program is under way, management should rely on indicators that signal impending difficulties, technical or otherwise. Leading indicators include

- requirements volatility,
- contract changes,
- budget stability,
- funding flexibility, and
- manufacturing readiness.

Requirements Volatility

During program implementation, ineffective control of requirements changes leads to cost growth and program instabilities. One indicator that should be monitored is requirements volatility (i.e., adding, deleting, and modifying of a system's requirements during the development process). These requirements changes generally have an impact on several constituent systems. More problematic still is that the precise nature of the impact often cannot be anticipated (from a technical, schedule, or cost point of view). At best, several subsystems must be modified to compensate for, or otherwise facilitate, the modifications to other sub-systems as they occur. Of course, each time a modification is made, thorough simulation and testing is required. At worst, if the change is fully integrated, serious system-level challenges may result.

The problem is two-fold. On the one hand, the process by which requirements are generated and approved may not fully consider the impact to the development program's cost and schedule. High levels of requirements volatility extend development, and, as a result, long-duration programs are viewed as works in progress that often fail to deliver the initially envisioned functionality. On the other hand, ignoring requests for necessary requirements changes early in a program can cost significantly more to remedy, once the system has been fielded. Consequently, failure to aggressively monitor and manage a system's requirements can increase the development time and cost.

Take the case of the Global Hawk, an unmanned aerial vehicle (UAV) that has been used extensively in both Iraq and Afghanistan. The program, initiated in 1995, was designed to undergo multiple blocks of development; the most important goal of each block was to remain within the cost requirement of $10 million per unit and to keep the program on schedule (Gansler, Lucyshyn, & Spiers, 2008). Following the operational success of the first iteration (the RQ-4A), the Air Force decided to design a new, larger, and more capable variant of the Global Hawk, known as the RQ-4B. Originally, the RQ-4B components were to be 90 percent compatible with the A model. But in an effort to expand the UAV's capabilities, the Air Force altered the requirements to produce a significantly larger B variant. The B variant, as designed, would carry a 50 percent larger payload, fly two hours longer, and retain the approximate 10,000 nm range.

These seemingly marginal requirement shifts necessitated major reengineering. The development of the RQ-4B project was to be funded with the original budget for the 4A; however, the Air Force removed cost as a requirement, relegating it to a consideration (Gansler & Lucyshyn, 2013). Many independent commentators have regarded the Global Hawk RQ-4A program as a great success. However, the restructured Global Hawk program has faced significant cost and schedule difficulties (Gansler & Lucyshyn, 2013).

Developing an indicator that measures the rate of change of requirements over time can help forecast future program challenges (e.g., a surge in requirement changes could indicate potential risk to design, which in turn could impact cost and schedule), as well as identify problems with the requirements generation process.

Contract Changes

Because the DoD's needs can change, acquisition contractors are required to accept contract change orders that alter specifications and other contract terms. Of course, they must also price these changes. The complexity of contracts—which can involve a variety of people with diverse backgrounds from different functional areas on both the government and contractor sides—can lead to misinterpretations and miscommunications of requirements and administrative issues that do not become evident until the contract is under way.

However, contract changes are often the source of a significant number of disputes between contracting parties. Changing a contract too frequently complicates the DoD's acquisition efforts, especially if the contract is large and complex. Tracking these changes can provide the DoD with an indication of program stability.

Budget Stability

Funding instability continues to drive up costs and delays the eventual fielding of new systems (S. Rep. No. 109-069, 2005). Successful acquisition programs require accurate planning and stable budgets. Unfortunately, within the DoD, this stability rarely exists. The instability arises when Congress moves funds from specific program elements for non-programmatic reasons. The services have, at times, also moved procurement funds to pay military personnel and operations and maintenance bills, which have combined to create a root cause for program instability (DoD, 2006).

When the actual funding is less than the planned funding, work must be delayed or deferred, resulting in program disruption. Budget reallocations and shortfalls result in the purchase of reduced quantities and/or programs that are extended beyond initial schedule estimates. The end result is short-term savings—but the price is long-term cost and schedule growth. Further, variability between annual budget predictions and the ultimate budget authority makes program planning difficult.

As mentioned, DoD programs often must reduce planned quantities in order to stay within their planned budgets. The Air Force's air superiority fighter, the F-22 Raptor (as previously discussed), suffered this fate. As costs increased, quantities were reduced, causing program costs (adjusted for quantity) to increase, which, in turn, triggered further reductions in quantity.

In order to prevent a vicious cycle wherein reductions in quantity lead to increases in program costs, programs should ensure that there is an adequate management reserve (MR) budget. The MR budget is typically used by contractor program managers to cover unknown problems that arise during development that fall under the scope of work. Unlike the F-22 program's MR budget of only 2 percent (which it exhausted within the first year of engineering and manufacturing development [EMD]), the F-18 program's MR budget was 10 percent. According to a 2005 RAND report, the F-18's MR budget contributed directly to the program's budget stability and prolonged success (Younossi, Stem, Lorell, & Lussier, 2005).

Technical complexity should inform the amount of the MR budget. While the MR budget can be expressed as a percentage of the total budget, uncertainty analysis can also be used to determine the probability that the cost of work will be less than or equal to its budget. Goldberg and Weber (1998) termed this calculation the "probability of success." Christensen and Templin (2000) noted that the amount of an MR budget can be identified at any desired probability of success specified by project management. The MR budget, its amount, and how it was determined should be viewed as a leading indicator by program personnel.

Funding Flexibility

In general, there is no single funding source for large, complex, and/or joint (multiservice) systems (this is particularly true for systems-of-systems). Rather, funding is programmed through the individual services or through individual program offices for the individual system. As a result, there is no advocate for joint or, as the case may be, enterprise-wide capabilities. As a result, contracts are generally written that do not adequately specify how the individual products are going to be integrated and tested with other elements.

Even in instances when a central authority is tasked with allocating funds among different systems, funding rarely is reallocated in response to changes brought on by the system's evolution. Once the funding is allocated, individual program offices intend to use it and often go to considerable lengths to justify their expenses in the event that their funding levels are jeopardized. Program managers must treat funding stability as a leading indicator. In the event that a reliable funding source is unavailable, programs should be assigned a higher level of risk.

Take, for instance, the Coast Guard's recent modernization effort, Project Deepwater, which sought to develop multiple integrated assets (ships, cutters, airplanes, satellites, etc.) to replace an aging fleet. The project relied on yearly appropriations from Congress for funding. Imprecise price estimates, coupled with miscalculated bids—companies tend to underestimate cost in initial bids in order to win a contract—resulted in inadequate funding. The lack of funding resulted in significant delays, which increased costs further. In addition, budgetary trade-offs during development were rare, especially after the program's reorganization, because funds were designated to each asset rather than to the system as a whole, encouraging the optimization of each asset. Indeed, no system came in under budget; rather, most required additional funding.

The trend toward participating in multinational programs (e.g., the F-35 Joint Strike Fighter) has also greatly compounded the funding problem. Different countries, contributing different amounts, have their own relative priorities and different budget cycles. Exchange variations, different acquisition policies, and languages add even more complexity.

Manufacturing Readiness

The success of acquisition programs also requires that manufacturing capability be managed effectively. The GAO (2009) found that a lack of manufacturing knowledge at key decision points largely leads to cost growth and schedule slippages in major DoD acquisition programs. Manufacturing readiness levels (MRLs) were designed to assess manufacturing readiness and manage manufacturing risk in acquisition. They were developed by a joint DoD/industry working group under the sponsorship of the Joint Defense Manufacturing Technology Panel (JDMTP) and introduced to the defense community in 2005. DoDI 5000.02 requires evaluations of manufacturing status and risks as part of defense acquisition programs. It establishes target maturity criteria for measuring risks associated with manufacturing processes at Milestones A, B, and C, and Full-Rate Production.

The DoDI elaborated that one of the purposes of the EMD Phase is to "develop an affordable and executable manufacturing process." Further, the system capability and manufacturing process demonstration shows "that system production can be supported by demonstrated manufacturing processes"; and the EMD Phase ends when "manufacturing processes have been effectively demonstrated in a pilot line environment." The two entrance criteria for the Production and

Deployment Phase are “no significant manufacturing risks” and “manufacturing processes [are] under control.”

Manufacturing readiness was intended to go hand in hand with technology readiness. Manufacturing readiness is governed by technology readiness and design stability because manufacturing processes are not able to mature until the product technology and product design are stable. The intent was to create a measurement scale that would be similar to the TRLs. Thus, MRLs were designed with a numbering system to be roughly congruent with comparable levels of TRLs. The MRLs were developed to include a nominal level of technology readiness as a prerequisite for each level of manufacturing readiness (DoD, 2009).

Generally, MRLs serve three purposes: (1) to define current level of manufacturing maturity, (2) to identify maturity shortfalls and associated costs and risks, and (3) to provide the basis for manufacturing maturation and risk management. The GAO (2009) recommended that weapons programs make use of MRLs. One possibility, in this regard, is incorporating the MRL scale into leading indicators that are measured and monitored across time in order to mitigate program risk.

There are ten MRLs (numbered 1 through 10; see Table 4) that are correlated to the nine TRLs in use. MRL 10 measures aspects of lean practices and continuous improvement for systems in production.

MRL	Definition	Description
10	Full Rate Production demonstrated and lean production practices in place	This is the highest level of production readiness. Technologies should have matured to TRL 9. This level of manufacturing is normally associated with the Production or Sustainment phases of the acquisition life cycle. Engineering/design changes are few and generally limited to quality and cost improvements. System, components or items are in full rate production and meet all engineering, performance, quality and reliability requirements. Manufacturing process capability is at the appropriate quality level.
9	Low rate production demonstrated; Capability in place to begin Full Rate Production	At this level, the system, component or item has been previously produced, is in production, or has successfully achieved low rate initial production. Technologies should have matured to TRL 9.
8	Pilot line capability demonstrated; Ready to begin Low Rate Initial Production	This level is associated with readiness for a Milestone C decision, and entry into Low Rate Initial Production (LRIP). Technologies should have matured to at least TRL 7. Detailed system design is complete and sufficiently stable to enter low rate production.
7	Capability to produce systems, subsystems, or components in a in a production representative environment	This level of manufacturing readiness is typical for the mid-point of the Engineering and Manufacturing Development (EMD) Phase leading to the Post-CDR Assessment. Technologies should be on a path to achieve TRL 7. System detailed design activity is nearing completion. Material specifications have been approved and materials are available to meet the planned pilot line build schedule. Manufacturing processes and procedures have been demonstrated in a production representative environment.
6	Capability to produce a prototype system or subsystem in a production relevant environment	This MRL is associated with readiness for a Milestone B decision to initiate an acquisition program by entering into the Engineering and Manufacturing Development (EMD) Phase of acquisition. Technologies should have matured to at least TRL 6. It is normally seen as the level of manufacturing readiness that denotes acceptance of a preliminary system design.
5	Capability to produce prototype relevant environment components in a production	This level of maturity is typical of the mid-point in the Technology Development Phase of acquisition, or in the case of key technologies, near the mid-point of an Advanced Technology Demonstration (ATD) project. Technologies should have matured to at least TRL 5. The industrial base has been assessed to identify potential manufacturing sources.
4	Capability to produce the technology in a laboratory environment	This level of maturity is typical of the mid-point in the Technology Development Phase of acquisition, or in the case of key technologies, near the mid-point of an Advanced Technology Demonstration (ATD) project. Technologies should have matured to at least TRL 5. The industrial base has been assessed to identify potential manufacturing sources.
3	Manufacturing Proof of Concept Developed	This level begins the validation of the manufacturing concepts through analytical or laboratory experiments. This level of readiness is typical of technologies in Applied Research and Advanced Development. Materials and/or processes have been characterized for manufacturability and availability but further evaluation and demonstration is required. Experimental hardware models have been developed in a laboratory environment that may possess limited functionality.
2	Manufacturing Concepts Identified	This level is characterized by describing the application of new manufacturing concepts. Applied research translates basic research into solutions for broadly defined military needs. Typically this level of readiness includes identification, paper studies and analysis of material and process approaches. An understanding of manufacturing feasibility and risk is emerging.
1	Basic Manufacturing Implications Identified	This is the lowest level of manufacturing readiness. The focus is to address manufacturing shortfalls and opportunities needed to achieve program objectives. Basic research (i.e., funded by budget activity) begins in the form of studies.

Table 4. Manufacturing Readiness Level Scale (DoD, 2010)

Evaluating Leading Indicators

Leading indicators can be analyzed both individually and collectively. Figure 5 illustrates how requirements growth might be monitored over time. As with EVM, a common program management practice described in some detail in Section II, an actual value (in this case, the number of requirements) is mapped along a planned value (the number of planned requirements). In this example, program management took corrective action in April in order to reduce requirements growth in response to a significant deviation from the planned value. Many of the indicators described above can be mapped out in a similar manner. We noted in Section II that, from a historical standpoint, EVM did not appear to significantly reduce the magnitude or frequency of cost overruns within DoD programs. This, we contend, has more to do with the quality of the information and the way the information is interpreted than with the technique itself (i.e., measuring an actual value against a planned value). Indeed, it is possible that EVM has been prone to abuse precisely because the technique itself is logically unassailable. Unscrupulous program personnel and managers unduly motivated by their career aspirations find it relatively easy to disguise problematic program data within the confines of a supposedly tried and true program management paradigm.

The leading indicators that we have proposed are less prone to manipulation in that they are discrete values that can be objectively measured. Unlike “work performed” (the typical EVM metric), the number of requirements, technical readiness levels, years of management experience, contract changes, and so forth cannot be as easily manipulated.

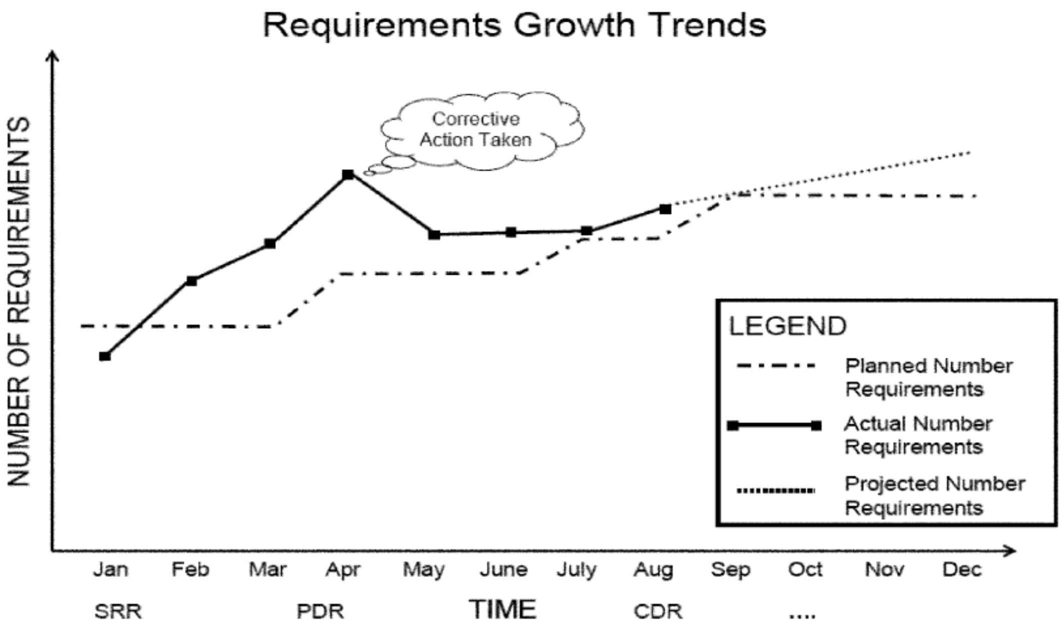

Figure 5. Requirements Growth Trends Leading Indicator (Rhodes, Valerdi, & Roedler, 2009)

Of course, there is always the danger that the leading indicators might be measured poorly or reported too late to be of use, but we contend that various safeguards can be put in place to prevent problems of this sort. For example, program managers might simply mandate that certain indicators be monitored and reported on a near-continuous basis. Further, under the leading indicators framework, there is explicit recognition that the indicators are designed to highlight deviations from planned values before these deviations become serious problems. The pressure to keep programs in the green—a common objective under EVM—is less of an issue. There is no motivation to hide problematic values, because the values are indicative of potential problems, not actual ones.

It is also possible to aggregate leading indicators to determine a given program's "score," which might be helpful to gauge overall programmatic risk. Because the relative importance of a given indicator will likely vary across programs, each could be weighted prior to the launch of the program. The indicators could then be assessed and summed for an overall score at various points in time.

The analytic hierarchy process (AHP) is but one technique that could be used. Using a simple ordinal scale (assigning each criterion a 1, 2, 3, etc. to designate importance) allows one to rank different criteria but reveals nothing about how much more important one criterion is relative another. AHP, however, allows users to compare incommensurable elements (in this case, requirements volatility, leadership, technical maturity, etc. to the success of the program) in a rational and consistent way. Program management, along with various stakeholders, would decompose each of the chosen indicators into a hierarchy of more specific sub-indicators, each of which would then be analyzed independently, allowing meaningful numerical values to be calculated for each of the indicators. By using AHP in this way, program managers would have a better understanding of the relative importance of the chosen indicators. They might choose to use this information to define thresholds for each indicator. For instance, if AHP assigns requirements growth a low score, then the threshold might be set relatively high; that is, the actual value might be allowed to diverge significantly from the planned value before corrective action is taken and before warnings are sent up the chain of command.

Of course, program management must be cautious when subjective assessments are converted into scores. Additionally, while AHP and other weighting schemes allow program personnel, contractors, and other stakeholders to discuss and measure what were once diverse, incommensurable values in a rational way, the overall program score may nevertheless disguise problematic values. Indicators may interact in ways that are difficult to predict and there is no reason to think that a relatively important indicator's solid performance can compensate for a less important indicator's failure to remain on target. Thus, it is important to stress the importance of analyzing leading indicators both individually and collectively so as to avoid unintentionally disguising problematic values.

V. Conclusion

The success of large projects is often assessed in terms of schedule, cost, and quality—the so-called iron triangle. We have identified several predictive measures for use by program managers to ensure that programs operate within budget and schedule constraints. However, it is important to note that a project that "fails" in any (or all) of these three categories may go on to deliver large benefits to users, contractors, program personnel, and/or other stakeholders, especially as time passes. Conversely, projects that meet established requirements, and are completed on time and under budget, may fail to meet the expectations of stakeholders. Turner and Zolin (2012) pointed to large civilian projects, including the Sydney Opera House, which were "substantially late and overspent but were later perceived to be very successful" (p. 87). It is not hard to find "successful failures" of this sort among DoD programs.

Take, for example, the F-111 Aardvark ("F-111") program, begun in the 1960s. The F-111 was a multipurpose tactical fighter-bomber capable of supersonic speeds. Although their needs differed considerably, Secretary McNamara insisted that the Navy and Air Force work together to develop joint requirements to the extent possible, believing that this strategy would substantially reduce acquisition costs. However, because the aircraft made use of numerous unprecedented technologies (e.g., variable sweep-wing, which pivoted back for high-speed flight and pivoted forward for a short takeoff and landing, and a crew compartment, which, in the event of an emergency, would serve as an escape module for the two-man crew), accurate cost estimates were difficult to calculate.

More problematic still, the DoD pursued concurrent development and production of the F-111. In other words, the DoD guaranteed that the selected contractor would be paid to both develop and produce the aircraft, which, it has been argued, served as a disincentive to efficient development. In any case, costs increased quickly. Despite its controversial origins and costly procurement, the F-111 turned out to be one of the most effective all-weather interdiction aircrafts ever built. At the time, no other aircraft in the Air Force could carry out the F-111's mission, which included precise, long-distance air strikes in all-weather conditions.

The challenge, then, revolves around how the DoD should define and measure program success. Even if leading indicators are used narrowly to help predict program "performance" (i.e., how the program rates in terms of quality, cost, and schedule), with the understanding that success is more difficult to define, there is the possibility that needed programs will be in danger of cancelation. Had leading indicators been implemented in the 1960s, the F-111 may never have gotten off the ground. The underlying point is that success is often achieved in an environment that permits some degree of failure. And failures, in turn, occur in an environment that encourages moderate risk-taking.

Accordingly, programs that undertake the use of leading indicators also must consider the strategic importance of the program. In some instances, programs should take on higher levels of risk and be willing to accept moderate increases in schedule and cost. Unfortunately, such a suggestion rings hollow in an environment where most programs regularly exceed their budgets and schedules.

Weapons system cost growth can be attributed to a litany of different factors, including over-optimism, estimating errors, unrecognized technical issues, and schedule changes. Public opinion, however, is less forgiving. A major poll by the Center for Public Integrity and the Stimson Center revealed that 80 percent of Americans believe that there is "a lot of waste" in the defense budget (Mehta, 2012, p. 1). Another recent poll by Reuters and Ipsos revealed that the majority of Americans prefer cutting defense spending to reduce the federal deficit, as opposed to taking money from public retirement and health programs (Smith, 2011). Justified or not, current defense spending is still at a high level and, in light of current budgetary conditions, is likely unsustainable.

Today, the military is often unable to acquire weapons systems in the intended quantities because of program cost growth. The DoD has reduced its orders of F-22s and F-35s by hundreds of aircraft. Reductions of this sort will lead some to believe that our military is underprepared to face threats to our national security or, perhaps, that the need for the specified capability was exaggerated to begin with. Given the current polling data, it appears that many are likely to believe that the need was exaggerated, which increases perceptions of waste and ineptitude and, in turn, exerts greater downward pressure on the defense budget. Sooner or later, this sequence of events will leave our military without the adequate resources to counter serious threats. The DoD must improve program cost control so that the services can acquire sufficient quantities of essential systems, improving public opinion and enabling our men and women in uniform to successfully carry out their missions. Implementing leading indicators across defense programs can help the DoD to meet this goal.

Reference List

Acquisition Weapons Systems: Hearings before the Subcommittee on Economy in Government, House Joint Economic Committee, 91st Cong. 1st session, Part 1, 29–31 (1969, December).

Aerospace Industries Association. (2007). *Overcoming barriers to public service*. Retrieved from http://www.aia-aerospace.org/assets/report_barriers.pdf

Arena, M., Leonard, R., Murray, S., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs*. Santa Monica, CA: RAND.

Army Safeguard Office. (1972). *Should cost/will cost/must cost: A theory on the cause of cost growth.* Arlington, VA: Author.

Brown, A., Stricklan, L., & Babich, D. (2006). *Implementing a GPS waveform under the software communication architecture.* Colorado Springs, CO: NAVSYS Corporation.

Card, D. (2008). *Leading indicators for project management.* Retrieved from http://www.dtic.mil/ndia/2008cmmi/Track5B/Thursday/AM/6898card.pdf

Carter, A. (2011, April 22). *Implementation of will-cost and should-cost management* [Memorandum]. Washington, DC: Office of the Under Secretary of Defense.

Cassata, D. (2013, April 8). As military health care costs soar, Congress rebuffs Pentagon attempts to charge higher fees. The Associated Press. Retrieved from http://talkingpointsmemo.com/news/as-military-health-care-costs-soar-congress-rebuffs-pentagon-attempts-to-charge-higher-fees.php

Christensen, D., & Templin, C. (2000). An analysis of management reserve budget on defense acquisition contracts. *Acquisition Review Quarterly, 7*, 191–208.

CIOinsight. (2005, October 22). How to lie with earned value. Retrieved from http://www.cioinsight.com/c/a/Past-Opinions/How-to-Lie-with-Earned-Value

Cook, D. (2013, April 13). 5 bad habits for the Pentagon to fix. *The Christian Science Monitor*. Retrieved from http://www.csmonitor.com/USA/USA-Update/2013/0430/Top-general-5-bad-habits-for-the-Pentagon-to-fix-video

Cooper, R., & Chew, B. (1996, January–February). Target costing lets customers, not the product, set the price. *Harvard Business Review*, 88–97.

Davies, R., & Woods, R. (2011, February 11). *Implementing will-cost & should-cost management*. Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=444823

Department of Defense (DoD). (1999). *Defense acquisition deskbook*. Retrieved from http://deskbook.osd.mil/deskbook.html

Department of Defense (DoD). (2006). *Defense Acquisition Performance Assessment Report.* Retrieved from https://acc.dau.mil/CommunityBrowser.aspx?id=18554

Department of Defense (DoD). (2010). *Manufacturing readiness level deskbook.* Retrieved from http://www.dodmrl.com/MRL_Deskbook_v1.pdf

Department of Defense (DoD). (2011). *Technology Readiness Assessment (TRA) guidance*. Retrieved from http://www.acq.osd.mil/chieftechnologist/publications/docs/TRA2011.pdf

Department of Defense Authorization Act of 2005, Pub. L. No. 109-163 (2005).

Ellram, L. (2006, February). The implementation of target costing in the United States: Theory vs. practice. *The Journal of Supply Chain Management*, *42*(1), 13–26.

Gansler, J., & Lucyshyn, W. (2013). *Cost as a military requirement.* Monterey, CA: Naval Postgraduate School.

Gansler, J., Lucyshyn, W., & Rigilano, J. (2011). *The Joint Tactical Radio System: Lessons learned and the way forward*. Monterey, CA: Naval Postgraduate School.

Gansler, J., Lucyshyn, W., & Spiers, A. (2008). *Using spiral development to reduce acquisition cycle times.* Monterey, CA: Naval Postgraduate School.

Gansler, J., Lucyshyn, W., & Varettoni, W. (2010). *Acquisition of Mine-Resistant, Ambush-Protected (MRAP) vehicles: A case study*. Monterey, CA: Naval Postgraduate School.

GlobalSecurity.org. (2011). Joint Direct Attack Munition (JDAM) history. Retrieved from http://www.globalsecurity.org/military/systems/munitions/jdam-history.htm

Goldberg, M., & Weber, C. (1998). *Evaluation of the risk analysis and cost management (RACM) model* (IDA Paper P-3388). Alexandria, VA: Institute for Defense Analysis.

Gordon, L. (2000). *Managerial accounting: Concepts and empirical evidence* (5th ed.). New York, NY: McGraw-Hill.

Gove, R. (2007). *Development of an integration ontology for systems operational effectiveness* (Master's thesis). Hoboken, NJ: Stevens Institute of Technology.

Gove, R., Sauser, B., & Ramirez-Marquez, J. (2010). *Integration maturity metrics: Development of an integration readiness level*. Hoboken, NJ: Stevens Institute of Technology, School of Systems and Enterprises.

Government Accountability Office (GAO). (1999a). *Best practices: Better management of technology development can improve weapon system outcomes* (GAO/NSIAD-99-162). Washington, DC: Author.

Government Accountability Office (GAO). (1999b). *DOD training can do more to help weapon system programs implement best practices* (GAO-99-206). Washington, DC: Author.

Government Accountability Office (GAO). (2001). *Tactical aircraft: Continuing difficulty keeping F-22 production costs within the congressional limitation* (GAO-010782). Washington, DC: Author.

Government Accountability Office (GAO). (2002). *DOD faces challenges in implementing best practices* (GAO-020469T). Washington, DC: Author.

Government Accountability Office (GAO). (2004). *Best practices: Using a knowledge-based approach to improve weapon acquisition* (GAO-04-386SP). Washington, DC: Author.

Government Accountability Office (GAO). (2005). *Better Support of Weapon System Program Managers Needed to Improve Outcome* (GAO-06-110). Washington, DC: Author.

Government Accountability Office (GAO). (2008). *Defense acquisitions: Better weapon program outcomes require discipline, accountability, and fundamental changes in the acquisition environment* (GAO-08-728T). Washington, DC: Author.

Government Accountability Office (GAO). (2009). *Defense acquisitions: Assessment of selected weapon programs* (GAO-09-326SP). Washington, DC: Author.

Government Accountability Office (GAO). (2010a). *Best practices: DOD can achieve better outcomes by standardizing the way manufacturing risks are managed* (GAO-10-439). Washington, DC: Author.

Government Accountability Office (GAO). (2010b). *Defense acquisitions: Assessments of selected weapon programs* (GAO-10-388SP). Washington, DC: Author.

Government Accountability Office (GAO). (2010c). *Defense acquisitions: Strong leadership is key to planning and executing stable weapon programs* (GAO-10-522). Washington, DC: Author.

Government Accountability Office (GAO). (2010d). *Information technology: Agencies need to improve the implementation and use of earned value techniques to help manage major system acquisitions* (GAO-10-2). Washington, DC: Author.

Government Accountability Office (GAO). (2011). *Defense acquisitions: Assessments of selected weapon programs* (GAO-11-233SP). Washington, DC: Author.

Government Accountability Office (GAO). (2013). *Assessments of selected weapon programs* (GAO-13-294SP). Washington, DC: Author.

Henderson, R. M., & Clark, K. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly, 35*(1): 9–30.

Hough, P. (1992). *Pitfalls in calculating cost growth from selected acquisition reports.* Santa Monica, CA: RAND.

Ingols, C., & Brem, L. (1998, July). *Implementing acquisition reform: A case study on Joint Direct Attack Munitions*. Fort Belvoir, VA: Defense Systems Management College.

Kaye, M., Sobota, M., Graham, D., & Gotwald, A. (2000). Cost as an independent variable: Principles and implementation. *Acquisition Review Quarterly*. Retrieved from http://www.dau.mil/pubscats/PubsCats/AR%20Journal/arq2000/kaye.pdf

Keating, C., Rogers, R., Unal, R., Dryer, D., Sousa-Poza, A., Safford, R., Peterson, W., & Rabadi, G. (2003). System of systems engineering. *Engineering Management Journal*, *15*(3), 36–45.

Kluge, J. (1997). Reducing the cost of goods sold: Role of complexity, design, relationships. *The McKinsey Quarterly,* 2.

Land, J. G. (1997, March–April). *Differences in philosophy—Design to cost vs. cost as an independent variable.* Retrieved from Defense Acquisition University website: http://www.dau.mil/pubscats/PubsCats/PM/articles97/land.pdf

Major Weapons Systems Reform Act of 2009, Public Law 111 – 23.

Mehta, A. (2012). Public overwhelmingly supports large defense spending cuts. Retrieved from The Center for Public Integrity website: http://www.publicintegrity.org/2012/05/10/8856/public-overwhelmingly-supports-large-defense-spending-cuts

Mihm, J. (2010). Incentives in new product development projects and the role of target costing. *Management Science*, *56*(8), 1324–1344.

Mosher, D. (2000). Understanding the extraordinary cost growth of missile defense. *Arms Control Today*: 9-15.

Myers, D. (2002, Fall). Acquisition reform—Inside the silver bullet: A comparative analysis—JDAM versus F-22. *Acquisitions Review Quarterly*, 312–322.

Nambisan, S. (2002). Complementary product integration by high-technology new ventures: The role of initial technology strategy. *Management Science, 48*(3), 382–398.

National Research Council. (2008). *Pre-Milestone A and early-phase systems engineering: A retrospective review and benefits for future Air Force systems.* Retrieved from http://www.incose.org/chesapek/Docs/2010/Presentations/Early-Phase_Sys_Engr-AirForceStudy-NRC_2008.pdf

Newnes, L., Mileham, A., Cheung, W., Marsh, R., Lanham, J., Saravi, M., & Bradbery, R. (2008). Predicting the whole-life cost of a product at the conceptual design stage. *Journal of Engineering Design*, *19*(2), 99–112.

Office of the Under Secretary of Defense for Acquisition, Technology, & Logistics (OUSD[AT&L]). (2008, December 8). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

Osborn, Kris. (2007). Gates urges ramping up MRAP acquisition. *Army Times*. Retrieved from http://www.armytimes.com/news/2007/05/defense_mrap_070509/

Parchem, A., & Gowing, M. (2009, May). Improving government: Accelerating the impact of new political-career executive teams. *Federal News Radio*. Retrieved from http://www.federalnewsradio.com/?nid=&sid=1678133

Petraeus, D. (2010, January). Adaptive, responsive, and speedy acquistions. *Defense AT&L*, 2–10.

Rhodes, D., Valerdi, R. & Roedler, G. (2009) Systems engineering leading indicators: Assessing effectiveness of programmatic and technical performance, *Systems Engineering, 12*(1), 21–35.

Rush, B. (1997). Cost as an independent variable: Concepts and risks. *Acquisitions Review Quarterly*, 161–172.

S. Rep. No. 109-069 (2005).

Sauser, B., Ramirez-Marquez, J., Henry, D., & DiMarzio, D. (2008). A system maturity index for the systems engineering life cycle. *International Journal of Industrial and Systems Engineering, 3*(6), 673–691.

Shults, J. (1976). *Case study: Lockheed and the C-5A.* Maxwell Air Force Base, AL: Air War College.

Sledge, N. (2012, September). Pentagon procurement reforms face slim chance of success. *National Defense*. Retrieved from http://digital.nationaldefensemagazine.org/i/80621/17

Smith, D. (2011, March 9). Public prefers cutting defense spending: Reuters/Ipsos poll. Retrieved from http://www.reuters.com/article/2011/03/09/us-usa-budget-poll-idUSTRE7286DW20110309

Taleb, N. (2007). *The Black Swan: The impact of the highly improbable.* New York, NY: Random House.

Turner, R., & Zolin, R. (2012, September 7). Forecasting success on large projects: Developing reliable scales to predict multiple perspectives by multiple stakeholders over multiple time frames. *Project Management Journal, 43*(5), 87–99.

Williams, M., ed. (2002). *Superfighters: The Next Generation of Combat Aircraft*. London: AIRtime Publishing Inc.

Younossi, O., Stem, D., Lorell, M., & Lussier, F. (2005). *Lessons learned from the F/A-22 and F/A-18 E/F development programs.* Santa Monica, CA: RAND.

Zengin, Y., & Ada, E. (2010). Cost management through product design: Target costing approach. *International Journal of Production Research, 48*(19), 5593–5611.

Acknowledgements

This research was sponsored by the Naval Postgraduate School, and we are especially grateful for the support and encouragement provided by Rear Admiral Jim Greene (USN, Ret.) and Keith Snider. We would also like to acknowledge our colleague John Rigilano for reviewing and editing the draft, and Jiahuan Lu, a doctoral candidate of public policy at the Maryland School of Public Policy, whose research contributed to this report. Finally, we would like to thank our co-worker, Caroline Dawn Pulliam, for her assistance with the planning and coordination of this study.

About the Authors

Jacques S. Gansler

The Honorable Jacques S. Gansler, former Under Secretary of Defense for Acquisition, Technology, and Logistics, is a professor and holds the Roger C. Lipitz Chair in Public Policy and Private Enterprise in the School of Public Policy, University of Maryland; he is also Director of the Center for Public Policy and Private Enterprise. As the third-ranking civilian at the Pentagon from 1997–2001, Dr. Gansler was responsible for all research and development, acquisition reform, logistics, advance technology, environmental security, defense industry, and numerous other security programs. Before joining the Clinton Administration, Dr. Gansler held a variety of positions in government and the private sector, including deputy assistant secretary of defense (material acquisition), assistant director of defense research and engineering (electronics), senior vice president at TASC, vice president of ITT, and engineering and management positions with Singer and Raytheon Corporations.

Throughout his career, Dr. Gansler has written, published, testified, and taught on subjects related to his work. He is the author of five books and over 100 articles. His most recent book is *Democracy's Arsenal: Creating a 21st Century Defense Industry* (MIT Press, 2011).

In 2007, Dr. Gansler served as the chair of the Secretary of the Army's Commission on Contracting and Program Management for Army Expeditionary Forces. He is a member of the Defense Science Board and the Government Accountability Office (GAO) Advisory Board. He is also a member of the National Academy of Engineering and a fellow of the National Academy of Public Administration. Additionally, he is the Glenn L. Martin Institute Fellow of Engineering at the A. James Clarke School of Engineering, an affiliate faculty member at the Robert H. Smith School of Business, and a senior fellow at the James MacGregor Burns Academy of Leadership (all at the University of Maryland). From 2003–2004, Dr. Gansler served as interim dean of the School of Public Policy at the University of Maryland, and from 2004–2006, he served as the vice president for research at the University of Maryland.

William Lucyshyn

William Lucyshyn is director of research and senior research scholar at the Center for Public Policy and Private Enterprise in the School of Public Policy at the University of Maryland. In this position, he directs research on critical policy issues related to the increasingly complex problems associated with improving public-sector management and operations and with how government works with private enterprise.

His current projects include modernizing government supply-chain management, identifying government sourcing and acquisition best practices, and analyzing Department of Defense business modernization and transformation. Previously, Mr. Lucyshyn served as a program manager and the principal technical advisor to the Director of the Defense Advanced Research Projects Agency (DARPA) on the identification, selection, research, development, and prototype production of advanced technology projects.

Prior to joining DARPA, Mr. Lucyshyn completed a 25-year career in the U.S. Air Force. Mr. Lucyshyn received his bachelor's degree in engineering science from the City University of New York and earned his master's degree in nuclear engineering from the Air Force Institute of Technology. He has authored numerous reports, book chapters, and journal articles.

The Center for Public Policy and Private Enterprise provides the strategic linkage between the public and private sector to develop and improve solutions to increasingly complex problems associated with the delivery of public services — a responsibility increasingly shared by both sectors. Operating at the nexus of public and private interests, the Center researches, develops, and promotes best practices; develops policy recommendations; and strives to influence senior decision-makers toward improved government and industry results. The Center for Public Policy and Private Enterprise is a research Center within the University of Maryland's School of Public Policy.

UNIVERSITY OF MARYLAND 1856