No. 42
September 2021

The Mitchell Forum

From EMD to Milestone C and Beyond: Common Issues that Affect Developmental Programs Transitioning into Production

Lt Gen Mark D. Shackelford, USAF (Ret.)

About the Forum

The Forum presents innovative concepts and thought-provoking insight from aerospace experts here in the United States and across the globe. As a means to afford publishing opportunities for thoughtful perspectives, Mitchell Institute's Forum provides high visibility to writing efforts spanning issues from technology and operational concepts to defense policy and strategy.

The views expressed in this series are those of the authors alone, and do not necessarily represent the views of the Mitchell Institute for Aerospace Studies, nor do they necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

Introduction

Major defense acquisition programs are infamous for failing to deliver new weapon systems on time and on budget. In fact, these failures occur so often, that most senior government officials are resigned to the fact that any government acquisition program, especially an aircraft development program, is destined to be delivered late and way over budget. Despite the perennial effort to reform acquisition, results have continued to fall short of expectations. Efforts are often focused on determining why a particular program failed, but there are few studies if any that analyze performance across many programs. Looking across multiple cases could help determine systemic or shared issues, identify lessons learned, and potentially point to prudent courses of action.

To address this analytic gap, this analysis looks across eight major aircraft acquisition programs and finds four primary reasons why developmental programs struggle to deliver on cost and schedule. Many of the issues identified in this paper are well-known, but some are not. What sets this research apart from previous work is the breadth of the assessed programs, as well as the methodology for comparison. The eight programs were assessed at the same decision point, Milestone C; this is when they transitioned out of development and into production.

For major aircraft programs, Milestone C is the gateway review that approves a program to transition from the engineering and manufacturing development into the low-rate production and deployment phase. This critical decision determines that the program is ready for the production of a fully representative aircraft whose design satisfies the military's operational requirements. Milestone C is best understood as a process, rather than a single event. Major aircraft programs rarely get canceled as a result of this process.[1] Instead, programs that struggle through the process are procured later than intended (KC-46 and F-35) and in smaller quantities than originally intended (F-22 and B-2). Other programs that alternatively transitioned with comparative success (F-18E/F) offer yet more insight. This paper offers some reasons for the variance in performance between major aircraft programs.

This study found four common issues evident during the transition from development to production and deployment. While these issues were not always sufficient to fully explain why a given program struggled, they were present across a majority of the case studies and often in combination. The first and most common issue was poor or ineffective communication, including either a lack of transparency or a breakdown in the relationships between the various government and contractor stakeholders on a program. Poor communication constrained the program manager's ability to manage risks during the latter part of Engineering and Manufacturing Development, and especially the transition to production. The second issue programs experienced was significant requirements changes and related funding instability during engineering and manufacturing development. This issue generated cost increases and schedule delays on programs during Low-Rate Initial Production (LRIP). A third issue was a lack of production-representative test assets and insufficient testing prior to a Milestone C decision. This issue prevented the discovery of deficiencies that could only be identified by testing fully integrated and mission-capable aircraft in realistic environments. It was often compounded by aggressive test and evaluation schedules that tended to falter when problems were eventually discovered. The final issue could be summed up as poor leadership decisions concerning program organization and management, on both the government and contractor's side, that adversely affected the contractor's ability to effectively function as a prime integrator. These decisions include company reorganization, geographic relocation of program worksites, inadequate management of subcontractors, and poor allocation of personnel and resources.

Prior studies have identified some of these issues, particularly requirements changes, but the observations on program transparency and communications between major stakeholders are important additions to the subject area. Discussions of risk management and defense acquisition reform have often focused on the type of contract used in development or on the importance of certain reviews and acquisition milestones. While many of the issues explored in this review are tangible, the less-tangible aspects of the programs, namely how various public and private stakeholders interacted with each other during development and early production, were also important in explaining the outcomes.

An Overview of Milestone C and the Programs Surveyed in the Research

Defense Department acquisition proceeds in a series of phases. Generically, these are Material Solution Analysis, Technology Maturation and Risk Reduction (TMRR), Engineering and Manufacturing Development (EMD), Production and Deployment (PD), and Operations and Support (OS). Milestones mark the decision to move from one phase to another. Milestone C marks the decision to begin production and is the presumed prerequisite for entering Low-Rate Initial Production (LRIP), operational testing and evaluation, and finally Full-Rate Production (FRP).[2] In principle, Milestone C affirms that the design is stable and will meet requirements, system software is mature, no significant manufacturing risks exist, costs and schedule are well-understood, production capabilities are ready, sustainment plans are in place, and the first production lot is fully funded.[3] In practice, although Milestone C affirms production readiness, programs often experience significant issues after the Milestone C decision that contribute to cost and schedule overruns for the EMD phase. Often, significant development work then occurs during LRIP. An examination of programs that struggled to make the transition to FRP highlights some common issues that caused them to struggle to meet schedule, cost, and performance requirements.

This study reviewed the acquisition case histories of eight programs of varying types from the early 1980s to the 2010s. Of them, the B-1B bomber, the A-12 attack aircraft, the C-17 transport, the F-22 fighter, the B-2 bomber, the F-35 fighter, and the KC-46 tanker are examples of programs that struggled or failed to move seamlessly from EMD to Full-Rate Production. The issues experienced by each of the programs are listed in Table 1. It should be noted at the outset that, with the exception of the A-12, all the programs surveyed have produced or are on track to produce superior aircraft with strong performance records. However, nearly all struggled to produce their aircraft within cost and schedule constraints set for the program. The only exception was the F/A-18E/F, which successfully and smoothly transitioned to FRP despite some technical problems. While the F/A-18 E/F was also a derivative platform with lower technical risk, it also had notably effective program management.

Table 1: A List of Programs Surveyed for this Paper and the Issues Identified in Each

Program Name and Type	Poor Communications, Transparency, and Relationships	Requirements and Funding Instability	Lack of Representative Test Assets and Insufficient Testing	Poor Management Decisions
B-1B (USAF Bomber)		X	X	X
A-12 (USN Carrier Attack)	X		X	X
C-17 (USAF Transport)	X	X	X	X
F-22 (USAF Fighter)	X	X	X	X
B-2 (USAF Bomber)		X	X	X
F-35 (USAF/USN/USMC Multi-Role Fighter)	X	X	X	X
KC-46 (USAF Tanker)	X		X	X
F/A-18 E/F (USN Carrier Fighter)				

Source: Mark D. Shackelford and Mitchell Institute

This survey includes both Air Force and Navy programs, as well as differing types of aircraft procured under different contract types. Additionally, these programs span different strategic periods over the last forty years (Cold War, the 1990s Peace Dividend, the Global War on Terror, and the threat of new great power competition). However, with the exception of the canceled A-12, they all have transitioned from development to either LRIP or FRP, although the KC-46 has still not been approved for FRP, and the F-35 has not formally passed Milestone C or entered FRP. This isolates a period of comparison for all of the programs during their transition from EMD through Milestone C or LRIP and toward FRP.

An important caveat is that the majority of our surveyed programs began engineering and manufacturing development at or before the year 2000. These programs predated reforms that have occurred since then, including the Weapons Systems Acquisition Reform Act (WSARA) of 2009 and the Better Buying Power initiatives of the 2010s. It is reasonable to contend that the surveyed programs might have benefitted from recent knowledge-based acquisition practices, including "should-cost" assessments and earlier system design reviews, which have become more prevalent since the enactment of WSARA. Our single post-WSARA example, the KC-46, has been able to control the governmental risk from cost overruns due to the use of a firm-fixed-price contract with a well-capitalized contractor able to absorb the additional cost. This has not, however, kept the KC-46 from suffering from schedule overrun, which has operational and financial implications for the government customer.

Likewise, while WSARA reforms and knowledge-based acquisition have had some empirical success in reducing cost and schedule growth, a 2020 GAO assessment of acquisition reform noted that "Many MDAPs [major defense acquisition programs] continue to move forward without the benefit of knowledge at key acquisition points" and that cost growth and schedule slips (averaging two years) remain a significant problem.[4] Our study supports continuing many best practices from the post-WSARA period, but goes further in stressing that transparency, regular communication, and trust between contractors and government are key to ensuring that knowledge-based acquisition practices achieve their desired effect in major aircraft programs. Indeed, the relative success of the F/A-18 E/F program, which predates these reforms, exemplifies many acquisitions best practices and, notably, good communications.

Issue One: Ineffective communication, lack of transparency, and lack of trust between stakeholders

The first and most important observation from these studies was that active stakeholder communication and transparency were the key to managing program risks throughout development and early production. This is vital because program stakeholders are constantly trying to meet cost, schedule, and performance targets on a wholly new aircraft. New-development aircraft specified for superior or cutting-edge performance are fundamentally risk-prone programs, even if the design or various subsystems are theoretically mature. If all stakeholders are not actively establishing a realistic assessment of the progress and risks on the program, or if they have a relationship which inhibits their ability to actively problem-solve, then those risks may be realized. Three illustrative examples are the C-17, A-12, and KC-46, each of which suffered from poor communications, transparency, and lack of trust between the contractor and the government. The F-22 program managers stressed good communication and transparency, but did not manage to achieve it uniformly. The F/A-18 E/F, which had good teamwork and an effective communication structure, did well in this area compared to most other programs surveyed.

The C-17 program suffered from poor communications and transparency on many different levels. First, it was insufficiently transparent to top management.[5] Multiple reorganizations and the mass firings of middle management from 1988 to 1993 severely hindered internal company trust and further reduced program transparency to top management. The prime contractor's top management did not communicate effectively with subcontractors to ensure requirements were understood, and had insufficient visibility into the progress and problems of the subcontractors.[6] Additionally, program software development from 1986 to 1990 was deemed not sufficiently transparent to government oversight organizations.[7] In 1993, the Defense Science Board review of the C-17 program noted that "teamwork, trust, open communication, accountability, and responsibility" had broken down, and the government and the company no longer trusted each other. Furthermore, the report plainly said that "**An Effective Communication System is Nonexistent** [emphasis in the original]. There is no integrated management information system to provide an effective means of management communication within MDC or between MDC and the Government."[8] In the end, it required a major overhaul of contractor-government relations to change this. Famously, Under Secretary of Defense John Deutch put the C-17 "on probation" with two years to fix "technical and financial problems," which was a cover for a major improvement in working relationships. The program established a service center to maintain communications between suppliers and government organizations, and it created a formal structure for reporting problems and the progress of solutions. Both the government and contractor worked hard to re-establish trust and good communications between the Air Force and the company.[9] One article notes that "timely communication that clearly articulated the issues, concerns, and problems proved essential to the C-17 program's turnaround."[10]

The A-12 program had an inherent problem with communication and transparency due to its extreme secrecy. In particular, security concerns made the program's cost, schedule, and technical progress opaque to the Office of the Secretary of Defense and higher levels in the Department of the Navy. Senior officials "relied, in large measure, upon the representations of the [Navy Program Manager] and the contractor team regarding cost, schedule, and technical risk."[11] For a considerable period, the contractor provided overly optimistic cost and schedule estimates to the government in order to ensure that progress payments were received. Security concerns interfered with normal higher-level oversight of contractor performance that would have contradicted these optimistic estimates.[12] The Navy program manager was technically highly educated, made frequent visits to contractor facilities, and received ample information from the contractors, but reported optimistic estimates of cost, schedule, and performance up the chain of command.[13] The program was transparent to him, but he shielded it because he

Figure 1: Headline on the front page of the *New York Times* announcing the cancelation of the A-12. The program had suffered from major breakdowns in communication and transparency as both the Navy and contractors attempted to protect the program, despite indications that it could not meet cost, schedule, or performance targets. When these problems belatedly came to the attention of the Office of the Secretary of Defense, Secretary of Defense Dick Cheney canceled the program.

Source: Eric Schmitt, "Pentagon Scraps $57 Billion Order for Attack Plane," *New York Times*, January 8, 1991, Section A, Page 1

PENTAGON SCRAPS $57 BILLION ORDER FOR ATTACK PLANE

DECISION A SURPRISE

Cheney Points to Delays and Cost Overruns on Navy's Stealth Jets

By ERIC SCHMITT

Special to The New York Times

WASHINGTON, Jan. 7 — Defense Secretary Dick Cheney today canceled plans to build a new Navy attack plane, the largest weapons program ever terminated by the Pentagon.

The decision, which surprised military analysts, was announced in a Pentagon statement that said the builders of the $57 billion A-12 warplane program had defaulted on the program by failing to "design, develop, fabricate, assemble and test the A-12 aircraft within the contract schedule."

The contractors are the McDonnell Douglas Corporation and the General Dynamics Corporation, both based in St. Louis. Both denied that they defaulted on the contract and said they would challenge the ruling. Moments after the announcement, General Dynamics said it would begin laying off 4,000 workers in Fort Worth and Tulsa, Okla. McDonnell Douglas said as many as 5,000 workers could lose their jobs in the next few days. [Page A18.]

Signal to the Military

Mr. Cheney's decision to cancel the radar-evading stealth warplane, the Navy's top-priority aircraft, sent a strong signal to both the military industry and to the armed services that he would tolerate neither delays and cost overruns nor attempts to conceal such problems. [Page A11.]

"This program cannot be sustained unless I ask Congress for more money and bail the contractors out," Mr. Cheney said in a statement. "But I have made the decision that I will not do that. No one can tell me exactly how much more it will cost to keep this program going.

"I do not believe a bailout is in the national interest. If we cannot spend the taxpayers' money wisely, we will not spend it."

Defense Secretary Dick Cheney

Insurance Limit On Bank Funds Seems In Doubt

By STEPHEN LABATON

Special to The New York Times

wanted it to continue. On the A-12, the relationship was dictated by fear of failure and incentives to obscure the program's actual status, and the program managers were thus unable to effectively defend the program when the extent of these problems reached the Secretary of Defense's desk.

KC-46 is another example in which the relationship between the government and contractor—in this case a distant one—had an impact on program results. As former Under Secretary for Acquisition, Technology, and Logistics Frank Kendall has argued, the government had a "hands off" approach to the KC-46, in which they gave "no direction to the contractor as to how to perform" on the firm-fixed-price contract for development. The government did this because "any change or interference would have 'broken' the contract and at least have led to claims for consideration for the contractor due to government directed changes."[14] This "hands off" attitude likely contributed to an under-appreciation of some of the risks in the program. A DOD Inspector General report on the KC-46 has recently shown that government program officials did not undertake a maturity review of the aerial refueling system even though it had changed substantially during the evolution of the design and incorporated newer and inherently riskier technology. They were aware of the newer system. This awareness, however, did not prompt them to insist on any review or demonstration of the new technology prior to developmental testing because it was not statutorily required.[15]

Once problems became more apparent, a relationship of mistrust complicated efforts to solve them. One GAO report indicated that communications were open, as Air Force "program managers and engineers have been involved in almost daily discussions with the contractor to make design tradeoffs."[16] However, the report conveys the adversarial tone of these meetings, recommending that government officials should employ "strong negotiating skills to protect the government's interest during these daily negotiations where design tradeoffs are made."[17] Public statements made in 2018 by the Secretary of the Air Force expressed evident frustration with the contractor who, in their words, appeared "much more focused on their commercial activity than on getting this right."[18] The effort to fix problems on the platform has carried on for several years, with evidence of a difficult relationship. In early 2020, the Chief of Staff of the Air Force expressed disappointment that the Air Force lacked "a relationship of trust and confidence" with its contractor.[19] Recent statements suggest the relationship has improved, but the program remains substantially behind prior schedules.

The F-22 is a case in which, despite efforts to achieve open communications and transparency, a breakdown still occurred in the program's ability to manage risk due to an inability to track or fully control risky areas. The Air Force and OSD had good visibility into the F-22 program and had an incentive to monitor it closely given constant criticism from the GAO, Congress, and the press. The program employed the Earned Value Management System (EVMS) that provided data on technical, cost, and schedule progress and forwarded this data in monthly Cost Performance Reports to the government.[20] An Air Force officer assigned to the F-22 program argued in 1999 that there was "full teamwork" between the Air Force and the contractor, and that all members of the team were tied together with an effective electronic communication system.[21] Senior Air Force and OSD officials monitored F-22 cost reports on a weekly or monthly basis, and certainly could not credibly claim that they were unaware of the program's difficulties.[22]

Despite open communications, there was at least one documented project management opacity in a risk-prone, high-consequence area: software. A 2004 GAO report noted that "program and contractor officials were unable to provide metrics for sufficient management visibility over the overall progress of the software."[23] Software spiral development was also undertaken on undefinitized contract awards, a response to the program's rolling baseline and often-changing requirements. EVMS tracking of software development was therefore difficult to do with any confidence, as the full schedule and cost figures were not definitively established.[24] The problem with a rolling baseline and changing requirements also had a communications aspect to it. Due to a desire to meet customer demands and protect the program, the contractor often agreed to make changes to the aircraft design or features without full consultation with the F-22 System Program Office.

F-22 experienced another challenge to stakeholder communication and trust in the waning years of its EMD phase that was influenced by two circumstances. First, Air Force leadership reacted to criticism of the air superiority-focused F-22 as being irrelevant to the post-9/11 conflict in Southwest Asia where air assets were heavily tasked for close air support. This led to requirements changes to add air-to-ground capability and the short-lived effort to redesignate the aircraft as the F/A-22. Second, EMD flight testing, already under intense scrutiny as the perceived cause of program delays, uncovered a structural issue labeled "vertical fin buffet." Investigating this issue added considerably to the remaining test points that were being tracked closely by Pentagon operational test officials. These issues, combined with a mismatch between remaining EMD content and management reserve, led to a deteriorating relationship between senior Air Force leaders who wanted EMD to finish quickly, the System Program Office that wanted to ensure all EMD content was completed, and the contractor who wanted to keep the program sold. After leadership changes at the System Program Office and contractor, and after a Red Team review of the program, the Air Force added $876M and extended EMD by 4 months in late 2002.[25]

In contrast, the F/A-18E/F program maintained excellent communication and transparency between the contractor and the government. This reflected, at least in part, the lessons the Navy and the contractor had learned from the then-recently canceled A-12 program. The Navy established a system to monitor cost and schedule performance closely, which included daily calls between the Navy program office and the contractor's program manager, as well as weekly teleconferences that included representatives from the major subcontractors. The Navy set up a data line that enabled the program office to access the same cost and performance data as the contractors.[26] While both the F-22 and F-18 E/F programs used Integrated Product Teams (IPT) to enable better program transparency of communication on technical issues, the latter additionally benefitted by consciously learning from the A-12 debacle. Veterans of the F/A-18 E/F project credited the close communication and working relationship, as well as the IPT structure, with their ability to tackle technical issues and ensure program stability effectively.[27] Unlike the programs described above, the F/A-18E/F stayed on cost and on schedule, and there were no major challenges to stress the relationship between the contractor and the customer.

Solutions for greater transparency and communication between government and contractors not only require improved data flows and consultation but also the fostering of trust between private and public parties. The goal is to achieve a program that is data-informed and data-validated, yet flexible in dealing with problems before they become crises. For example, for projects above a certain dollar threshold, DOD has long mandated an Earned Value Management System (EVMS) that collects data and tracks project progress in terms of cost and schedule. Despite the improvements EVMS has offered for transparency, cultural problems constrain its effectiveness. In the case studies, sometimes decision-makers didn't fully trust the data. Government program managers, for example, often disregarded EVMS data on F-22 software development because they lacked confidence in its timeliness or reliability.[28] Even when reliable data is available, a culture of mistrust or contractual rigidity creates disincentives to use the data to improve the performance or outcome of a program. On the A-12, for example, EVMS data might have demonstrated much earlier in development that program expectations were unrealistic and the program should be restructured. However, the culture of secrecy around the program and the pressures of the firm-fixed-price contract reinforced a culture of non-transparency that ultimately led to program cancelation.[29] Similarly, the government's rigidity on the KC-46's firm-fixed-price contract constrained problem solving when unexpected issues arose. That program is arguably only sustainable because the contractor is able to absorb the cost overruns.

Avoiding such failures of transparency requires a culture of trust, backed up with reliable and accessible data. The best example from the surveyed programs was the F/A-18 E/F, which succeeded where others failed. As one former OSD acquisitions analyst and consultant describes it, this was due to a "combination of teamwork and enlightened EVM implementation—a conscious effort to share information openly between supplier and customer, supported by an effective IPT structure."[30] That communication allowed the program to manage risk. Of course, the F/A-18 E/F was a lower-risk derivative of an established system, and the program involved personnel who had a long history of cooperation on the F/A-18A/B/C/D. The workshare arrangements between prime and subcontractors likewise built on this experience. In cases without this unique setup, contractors and government personnel must establish trust on the program by regularly communicating, being transparent, and, crucially, by sharing risk.

Sharing program risk is fundamental to building a culture of trust. With this trust, all program personnel can better harness the data made available by EVMS and, increasingly, new digital modelling tools and software. As it stands, contract structures largely govern the apportionment of risk. There is an active debate on whether contracting vehicles like firm-fixed-price, which pushes more risk onto the contractor, or cost-plus fee, in which the government incurs additional risk, are superior for delivering programs at cost, on schedule, and able to meet the desired performance. This apportionment of risk as either public or private for developmental programs clearly aligns incentives for government and contractor, but in ways that do not always enable both parties to problem solve. For example, additional investments in testing and test facilities are difficult for a contractor to justify under a firm-fixed-price contract because they reduce profit margins, even if these investments might allow them to identify problems earlier or permit additional maturation of major subsystems (like a flying testbed for avionics).

A promising principle for future contracting would be akin to "Active Contract Management," a concept from Harvard's Government Lab. Rather than simply monitoring a contractor's performance, a governmental procuring authority incentivizes the contractor to collect and share mutually accessible data on program progress to highlight risk areas or potential blind spots.[31] A core tenet of Active Contract Management is to foster a collaborative management culture and create a sense of shared accountability for program outcomes. Managers on both sides work collaboratively to investigate areas of potential concern or improvement, and to make adjustments to their contracting approach where necessary, including additional contracting actions or mixing contracting approaches.[32] The expectation that the contractor and government will operate with the same data and common operating picture and that the program will be flexible enough to make alterations should enable both the government and contractor to feel justified in sharing and incurring risks that lead to better program outcomes. Regular and open communication reinforces and enables this trust.

As Former Under Secretary Kendall has written, delivering programs that execute well and deliver value across the lifecycle requires industry and government program managers to be more than spectators in risk management.[33] While many of the activities devised to reduce risk are supposed to occur before a development contract is signed, the case studies show that it is vital for transparency and communication to continue throughout EMD and into production in order to deal with the unforeseen and unexpected.

Issue Two: Requirement and Funding Instability during Development

The second issue common to programs that struggled to transition into production was requirement or funding instability. This manifested in three major ways: disruptive alterations to program requirements, truncations to planned program procurement, and changes to program funding involving delays, timeline changes, or outright cuts. Looking at the first of these, many studies and acquisition best practices have noted that setting unclear or unrealistic

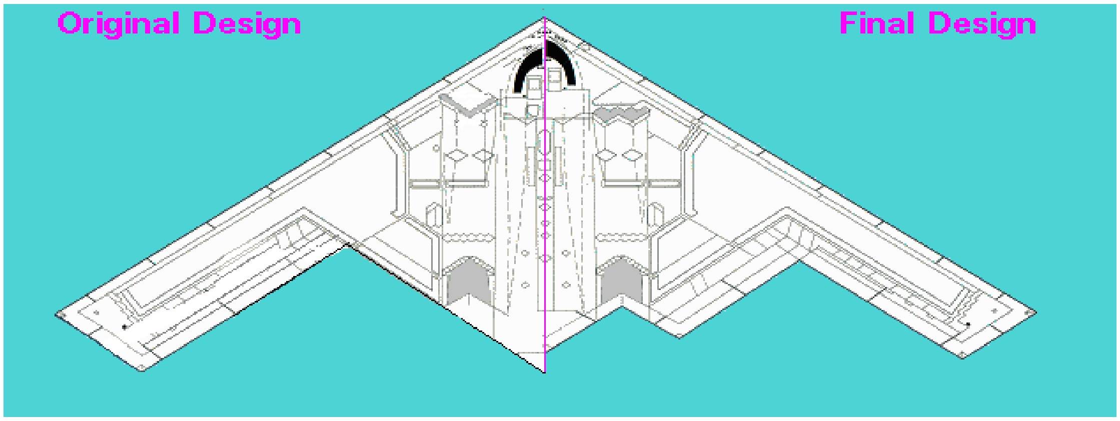

Figure 2: The B-2 underwent a major design change during EMD and received an additional low-altitude requirement that added considerable cost to the program and contributed to schedule overruns. The requirement for B-2's produced fell from over 130 to only 21, driving up unit costs considerably.

Source: Northrop Grumman design, Rebecca Grant, *B-2: The Spirit of Innovation* (Falls Church, VA: Northrop Grumman, 2013), p. 60.

requirements at the beginning of a program, or making major requirements changes during EMD, can cause costs to increase and the schedule to slip.[34] A significant alteration to the defined operational profile of an aircraft can cause serious problems, as on F-22 and B-2, by fundamentally changing the base design of the aircraft. Changes to key performance parameters or subsystems were notable issues for the B-1 and C-17 because they added additional complexity not fully appreciated at the time development began. The second, truncations of a program's total aircraft requirement, also affected nearly every program surveyed. Finally, most of these programs were also often affected by the third: unsteady funding in the form of sporadic funding cuts or perennial underfunding for programs with unrealistic cost estimates.

It's worth noting that requirements and funding instability are not always connected, although the case studies showed that funding cuts often ensued when a program with unstable requirements ran over cost and schedule. The fundamental connection between them is that both disrupt the core assumptions that underline developing a new defense system and executing a complex program to produce, test, procure, and field it. Unstable inputs are unlikely to produce consistent outputs in terms of meeting schedule and cost targets.

Disruptive Alterations to Program Requirements

A major role change or additional mission capability added during EMD was the case in two of the programs. For example, the primary mission of the F-22 repeatedly fluctuated between air superiority aircraft and ground attack, which have very different performance requirements. Moreover, F-22 low observability, performance, and avionics requirements changed, and initial weight requirements were set unrealistically low.[35] Similarly, the B-2 bomber was originally envisioned as a high-altitude penetrator, but in 1981 the Air Force insisted that the aircraft have a low-altitude capability. This required a new planform, different flight controls and subsystems, and a stronger, heavier structure, all of which added schedule and cost risks in early production.[36] In a slightly different sense, the F-35 suffered from having to reconcile the divergent requirements of the Air Force, Navy, and Marines while maintaining maximum commonality. The program also had to reconcile the competing demands of stealth, supersonic flight, and STOVL capability. This increased complexity and risk and led to cost increases and schedule delays. Each service continually altered design requirements, and every change affected all three variants of the aircraft.[37]

Major changes to key performance parameters impacted subcomponents and added complexity and risk by diluting the value of data gathered from prototypes and demonstrators. The B-1B and C-17 both had significant alterations to their requirements when their programs were re-launched after years-long delays. B-1B performance requirements changed dramatically from those of the B-1A; "The B-1B was expected to fly lower, have a longer range, carry more weapons, and be stealthy."[38] This increased the weight of the B-1B and demanded changes to the aircraft structure and the development of new materials and avionics. The C-17 payload requirement changed five times during its development, which in turn affected the wings, engines, and cargo flooring. The flight controls were changed from manual control to "fly-by-wire" control. In addition, mission computer software requirements were poorly and incompletely defined when the program went into EMD.[39] All these contributed to the cascading series of problems the program continued to face after LRIP and first flight.

In all of the above programs, changed or unclear requirements compelled major redesign work that increased costs and caused delays. In contrast, the Navy set clear requirements for the F/A-18E/F and controlled risk by dividing development into two blocks. The first block relied on legacy and mature avionics, while the second integrated advanced, but technologically riskier, avionics. The program defined these blocks with a realistic assessment of risk, then strictly controlled the requirements in each. This contributed to stability in the program's cost and schedule.[40] Stable requirements, however, are not a panacea. KC-46 requirements were unchanged from 2011 onward, but this did not prevent cost overruns and program delays.[41]

Truncations to Planned Program Procurement

Reductions in procurement quantity increase unit cost, especially if they occur in the early stages of production. Unit cost falls with the quantity produced due to increased workforce proficiency over time, the ability to purchase raw materials more cheaply in greater quantity, and other factors. Likewise, truncations to total buy create risk for the long-lead investments that companies have to make to ramp up production to expected numbers. Nearly every program surveyed experienced a drop in projected aircraft buy. The Air Force intended to buy 240 B-1As but ended up with 100 B-1Bs. B-2 planned procurement dropped from 132 to 75, then 20. In some cases, truncation was also a reaction to troubled program performance. The Air Force planned to buy 750 F-22s, but over time reduced this to 648, then 339, then 277, and finally 187 after nearly a decade in development and a Nunn-McCurdy Breach. Similarly, the Air Force initially proposed to buy 210 C-17s but cut this to 120, then to 40 when the troubled program was threatened with premature truncation. Later, on the basis of improved performance, C-17 orders resumed, and the Air Force bought 223. The Navy proposed buying 858 A-12s but reduced this to 620 in 1990 prior to program cancelation in 1991.

Changes or Cuts to Program Funding

Throughout the case studies, funding instability adversely affected program objectives, cost, and schedule.[42] The F-22 program experienced major funding instability after it entered EMD. Congressional skepticism of the program's slow start led it to cut EMD funding and cap total expenditures on EMD and production.[43] The Secretary of the Air Force (SECAF) argued in 1996 that "small funding cuts have disproportionately large program impacts"—each dollar cut required spending two or three dollars in the future. The SECAF further stated that funding cuts since 1991 "have caused EMD costs to increase approximately $2B" and IOC to slip 32 months.[44] Due to a decision to halt development in order to study alternatives after awarding a development contract, the C-17 program was starved for funds from 1981 to 1985, and its schedule thus slipped five years.[45] The C-17 then received only 77 percent of the funds requested for procurement from 1987 to 1995, compounding its performance issues.[46]

Looking at the root causes of underfunding stresses the importance of proper cost estimates in ensuring adequate funding for the project in question. Over-optimistic cost estimates can mean that even "stable" funding is not sufficient. The A-12 program received all the funds the Navy requested, but the Navy and the contractors severely underestimated the amount of funding required.[47] The B-2 enjoyed ample funding until 1989, though specifics remain classified. It did have stable funding from 1990 through 1996, though at a far lower level than originally planned.[48] The B-1 had considerable funding instability before 1980, but exceptional funding stability in the 1980s thanks to strong administration support and the use of multi-year procurement contracts in LRIP.[49] The F/A-18E/F program had very stable funding, with only two minor Congressional reductions from 1990 to 2002.[50] The F-35 had very stable funding from 2002 to 2010, but this did not prevent two Nunn-McCurdy cost breaches.[51]

Issue Three: Production-representative test aircraft not available prior to Milestone C

The third issue common to programs that struggle to move into FRP is the unavailability of production-representative test assets and an over-aggressive test schedule. Developmental and operational testing seeks to evaluate the effectiveness, safety, survivability, reliability, and maintainability of a system in a realistic environment prior to Full-Rate Production. A Test and Evaluation Master Plan is required before entering Milestone B, and the results of both developmental and initial operational test and evaluation are used to support a Full-Rate Production decision. Developmental and early operational testing ideally occurs on "production-representative" systems that use the same parts, processes, and software intended for use on production aircraft. Limited developmental stage testing, including testing with unrepresentative assets, contributes to cost overruns and schedule delays. Conversely, realistic testing on a fully integrated system often uncovers problems not revealed on the ground or in laboratory simulations, especially electronic subcomponent or software immaturity. The issue is not that problems are discovered in testing—after all, that is what testing is for. However, the incentive exists to begin building the production aircraft quickly because that is the phase where contractors can earn greater profits, even though the major problems are not fully understood and solved. This then requires early production aircraft to undergo retrofits and requires additional concurrent redesign work to occur even as planes are rolling off the assembly line. In all cases surveyed, save F/A-18 E/F, the amount of testing done at or before an LRIP decision with a fully representative test asset was minimal, and the A-12 program never successfully assembled or tested an airframe at all. In many cases, an aggressive or overconfident test and certification schedule amplified the impact of these problems.

The B-1B is a case in point. Three B-1A aircraft were flight tested in the early 1970s, though defensive avionics were not tested. A fourth test aircraft with defensive avionics was being built when the program was canceled in 1977.[52] When the program restarted, Defense Department officials asserted that the B-1B had 70 percent commonality with the B-1A, and therefore B-1A testing provided grounds for confidence that B-1B cost and schedule goals could be met.[53] But they also argued that the B-1B would be better than the B-1A due to improved avionics, lower radar cross-section, and the ability to carry cruise missiles.[54] The B-1B was significantly heavier than the B-1A and would fly lower, requiring changes to aircraft structure, avionics, and landing gear. Many B-1B subsystems had been used on other aircraft, but they had never been integrated on a single aircraft. The B-1B test schedule was quite aggressive due to the emphasis on overall program schedule, and the aircraft production contract was signed well before the redesigned aircraft flew for the first time. The Reagan Administration was publicly committed to an IOC of 1986 and to producing 100 bombers by 1988. The test plan envisioned using two modified B-1As (clearly not "production-representative" aircraft) from 1983 to 1986 and one B-1B from March 1985 until IOC in June 1986.[55] Offensive and defensive avionics would not be tested until production of over 18 aircraft had begun. The production decision was not tied to the results of EMD or testing; LRIP and multiyear procurement contracts were awarded years before first flight.[56] The test program allotted no time for fixing any problems identified in tests.[57] Testing revealed fuel leaks, problems with offensive and defensive avionics, problems with the terrain-following equipment, and engine stalls. Nevertheless, the Air Force declared that the B-1B had attained full operational capability in September of 1986, and it acquired 100 bombers by April 1988.[58] The program met its cost and schedule goals but did not meet performance requirements. To correct deficiencies, the B-1B needed extensive retrofitting, which took place from 1988 into the early 1990s.

The C-17 test program was aggressive because the overall program was behind schedule, it was over budget, and the contractor was "in the red" on a fixed-price contract. Moreover, the company and the Air Force believed the YC-15 prototypes built in the 1970s had already demonstrated many of the technologies needed to reduce C-17 program risk.[59] Alas, the C-17 was much larger than the YC-15 and incorporated new technologies like "fly-by-wire" and the Head-Up Display (HUD). Thus, YC-15 tests were of limited relevance. The final test program took about twice the time originally planned.[60] The program entered LRIP in January 1989, before the first flight in September 1991. Testing began in 1991 and was supposed to end in December 1993, but it lasted until June 1995. Testing was inefficient because test aircraft were delivered late, had poor availability, and had hardware and software problems.[61] Even as the schedule was re-baselined, testing also discovered many deficiencies. The company had trimmed its test workforce, forced the remainder to work overtime, and put them under heavy pressure to move forward regardless of test outcomes. This created friction with government test personnel and hindered proper analysis of any given test. Insufficient time was allotted to fix problems that emerged during tests and then evaluate the changes in further tests. The fast pace of testing also contributed to the large number of serious mishaps during testing.[62] It was only after the program was reorganized and the government allocated additional funds for testing that the C-17 successfully completed its required testing and was approved for FRP in November 1995.

Figure 3: The C-17's first flight in 1991 was already well over a year behind schedule. While it performed well on its maiden voyage, outstanding issues with flight controls and software dogged the plane through an additional three years of developmental testing. Many of these issues could be traced to the decision to integrate new displays and "fly-by-wire" controls during development, which added significant complexity and risk to the aircraft.

Source: U.S. Air Force photo

The B-2 program originally envisioned four years of flight tests using six test aircraft, but testing actually took eight years. The B-2 test program began months behind schedule due to manufacturing, software, and aircraft availability problems.[63] Testing revealed problems with radar signature, defensive avionics, the aft deck structure, mission planning software, and stealth maintainability.[64] B-2 production began before flight testing was over, so production aircraft subsequently needed costly modification to correct the problems identified in testing.[65]

Much like the B-1B and C-17, the F-22 contractor flew YF-22 prototypes in 1990 for technology demonstration and risk reduction prior to winning the development contract. The F-22, however, differed significantly from the YF-22 prototype in external geometry and avionics, diluting the representative value of the prototype. The contractor began building nine test aircraft for a planned five-year test program, but delays hindered their scheduled delivery. The F-22's first flight occurred in 1997, but it did not test a fully integrated system until the fourth test aircraft. The test program incorporated both a 757 flying test bed for avionics and software and an F-16 with flight controls modified to mimic F-22. The F-22, however, had special requirements for testing low observability, super-cruise, thrust vectoring, and advanced avionics and sensors.[66] Testing was economized due to funding shortfalls, and the reduced test program increased risk and ultimately drove up costs.[67] Test aircraft were delivered late due to design changes, parts shortages, subsystem unreadiness, and integration issues. Most significantly, the test aircraft with fully integrated avionics were delivered a year late, delaying the discovery of significant problems.[68] During testing, software locked up when radar, communications, navigation, and electronic warfare systems were used simultaneously.[69] This showed the difficulty of testing complex systems on the ground or via surrogates like the flying test bed.[70] The late delivery of fully representative test assets delayed the test program and overall program completion, driving up costs and undermining Pentagon and Congressional support. Somewhat like with the B-1B, the Air Force decided to begin producing F-22s before testing was complete and had to make costly modifications to the production aircraft after testing uncovered problems.[71]

Similar issues dogged the F-35 program. In the late 1990s, the prime contractor built two X-35 concept demonstrators. They were not production-representative—many capabilities were not demonstrated, and the design was still very incomplete. The program built eighteen test aircraft to test three different variants; the Air Force F-35A, the carrier-capable Navy F-35C, and the Short Take Off, Vertical Landing (STOVL) Marine F-35B. Late delivery of aircraft, software development problems, and the immaturity of delivered software slowed down testing. The program built and tested mission systems software in simulation laboratories and flying test beds, but this failed to reduce the need for flight testing as intended.[72] The software was not fully tested in "realistic stressing conditions," and numerous deficiencies arose when it was used in actual aircraft.[73] Operationally realistic testing also revealed many problems with F-35's logistics software.[74] Despite the discovery of a large number of deficiencies in testing, "low-rate production" moved forward.[75] Over 625 aircraft have been delivered, and they will require major, costly modifications.[76]

Even the KC-46, which the government assessed to be a low-risk program because it was a commercial-derivative aircraft, did not achieve Milestone C with a mission-capable production-ready design. The Director of Operational Test and Evaluation warned in 2012 that the KC-46 test program was excessively aggressive. In particular, the number of flight hours planned per test aircraft was almost twice the typical experience.[77] The plan devised in 2011 envisaged completing 66 percent of EMD testing before Milestone C in 2016, but the program actually only completed 30 percent, and many subsystems had only been tested in the laboratory, not on the aircraft.[78] Before Milestone C, the contractor discovered problems with insufficiently spaced and shielded wiring bundles, as well as fuel system parts that did not meet specifications, delaying first flight. During initial operational assessment, the refueling boom showed a tendency to inadvertently strike the receiver aircraft, and the boom operator had difficulty detecting such events. In addition, a problem with the refueling boom's axial loading held up Milestone C. It was resolved in a reduced series of flight tests that deviated from the test and evaluation master plan for Milestone C. Though Air Force program officials deemed the issue resolved, the reduced tests were insufficient to evaluate whether the "solution" had truly worked. Subsequently, the Air Force had to award the contractor additional money to resolve the persistent problem after the Milestone C declaration.[79] Post-Milestone C, other deficiencies came to light. There were chronic fuel leaks and cargo restraint locks that came unlocked in flight, both of which required a permanent fix to meet safety requirements.

The contractor and the Air Force had a protracted dispute over ongoing problems with the Remote Vision System which controlled the boom, a technology which worked satisfactorily in the lab but did not demonstrate operational utility in live testing. The full extent of the problems with the Remote Vision System were only discovered after several LRIP lot buys. The contractor attempted to resolve the issue with software versus hardware fixes, most likely due to the firm-fixed-price contract, and that management decision caused a much longer period of time to effectively close out the issue. The dispute was resolved when the contractor agreed to develop, at

Figure 4: A KC-46 with boom extended. The aircraft, a commercial derivative, was believed to be a low-risk acquisition but suffered from a series of issues. The Remote Vision System for the boom operator, for example, failed to work as well in real-world testing as it had in computer simulations. The full extent of these problems was only revealed after Low-Rate Initial Production was approved. A required redesign has delayed Full-Rate Production until 2024, over five years behind schedule.

Source: U.S. Air Force photo.

its expense, an improved system with new cameras, sensors, and cockpit displays.[80] The KC-46 will not officially enter FRP until 2024, after the new system is tested and deemed acceptable.[81] Throughout, as noted, the Air Force declined to suggest mitigation strategies for the issues revealed in testing, relying on the contractor for solutions and accepting schedule risk in doing so.

In contrast with the other programs, the F/A-18E/F test program was prudent in schedule. It took place over three years and employed seven production-representative test aircraft. Testing revealed challenges, including engine problems and a "wing drop" issue that required modification of the wings. But there was ample time to resolve these problems. After successful operational testing and evaluation, the aircraft entered FRP under a multi-year procurement contract.[82]

Issue Four: Poor Management Decisions by the Prime Contractor or the Government

Poor management was the final common theme among programs that struggled to transition from EMD to Full-Rate Production. In some cases, short-sighted company reorganization, geographic relocation of program assets, poor management of subcontractors, poor personnel decisions, and other questionable conduct resulted in program disruptions, increased costs, and schedule delays. In some of the case studies, government decisions also caused problems. The most common government failure was lack of effective oversight—the A-12 program is a good example—but other, sometimes policy or politically motivated decisions had negative ramifications for program stability. Often, there was some element of fault on both the government and contractor's part.

On the C-17 program, the government and the contractor repeatedly made poor decisions that brought the program close to collapse. First, the C-17 was in a state of low-level development in the early 1980s, then the government put the program on ice within weeks of selecting the contractor's design.[83] As the Air Force considered alternatives and starved the program of resources, the contractor experienced organizational restructuring and diverted labor and talent elsewhere. The C-17 team had to be reestablished when EMD officially began in 1985. Next, the prime contractor and the Air Force underestimated the difficulty of developing the avionics, flight control system, and software. The contractor lacked experience with computer software, and struggled to provide detailed requirements and useful, timely guidance to the subcontractors responsible for the flight control system and mission computer.[84] Furthermore, the contractor made basic planning errors, including inadequately staffing the C-17 program, failing to develop a systems engineering management plan, failing to develop an avionics integration test plan, and failing to adopt CAD/CAM tools.[85] Due to the declining profits of a subsidiary, in 1989 the contractor CEO imposed the "Total Quality Management System," which began with the firing of 5,200 managers and the appointment of new managers from outside the subsidiary. Over the coming year, some 17,000 workers were laid off. Workforce experience and morale collapsed, with an inevitable negative effect on the C-17.[86] With the loss of the F-22 competition, the cancelation of the A-12, and the lack of success of the MD-11 airliner, the company seemed on the verge of collapse in 1991. The company was further besmirched by allegations of misconduct that emerged in 1992. The Defense Department's Inspector General alleged that Air Force officials provided the corporation with progress payments in 1990 to ensure continued performance on the C-17, falsely giving the impression of program success. The Inspector General cited a number of officials for improper conduct, and the affair ended the careers of three general officers.[87]

As noted, undoing these decisions required an unprecedented degree of cooperation and coordination between the company and the government. With government assistance, the company brought costs under control and improved its management processes from 1994 to 1995. The program met its schedule goals and passed its reliability, maintainability, and availability evaluation, leading to a FRP decision in late 1995.[88] It was only by enabling a close working relationship and mending long-standing transparency issues that C-17 was saved from early termination.

The A-12 program's management strategy was largely based on mistaken assumptions about feasibility, technical risk, and programmatic structure by the contractors and the Navy. Clearly, both parties seriously underestimated the difficulty and cost of building a carrier-capable, stealth attack aircraft and chose a contracting approach that, in retrospect, added risk. The winning prime contractor team submitted a bid for a firm-fixed-price EMD contract that was under the Navy's cost cap. The losing team notably refused to accept a firm-fixed-price contract on a risky program and submitted a much higher bid that, in retrospect, was likely more realistic.[89] The winning team lacked experience in low-observable and composite technologies but believed it would have access to the technologies that had been developed for the B-2 program. Unfortunately, A-12 and B-2 program personnel lacked access to each other's programs and could not discuss them with each other.[90] As a result, the EMD program had to develop these technologies from scratch and immediately fell behind on cost and schedule. The contractors failed to report ever-increasing cost and schedule variances to the government—they made "best case projections of cost at completion" based on overly optimistic assumptions. Program management felt heavy pressure from company upper management to obtain progress payments and maximize cash flow.[91] Program secrecy interfered with normal reporting of contractor cost and schedule performance. Information was reported verbally and in general terms.[92] The contractor missed key dates for design release, tooling, assembly, and first flight due to a lack of capability and capacity for design and fabrication of components. These failures reflected "a plain lack of objectivity at the contractor team level, and wholly inadequate oversight by [the contractor's] corporate management."[93]

Yet, the poor decisions on A-12 were not purely the contractor's. As noted, Navy program managers did not maintain oversight of the program. More broadly, the entire approach on the program's structure appears remarkably short-sighted in retrospect. As a veteran of the program wrote, there was "little oversight, firm-fixed-price development, an acquisition approach that in the development phase teamed two competitors who would later compete for production, and a very aggressive schedule tied to fixed-price production options. It was a disaster waiting to happen."[94]

The principal mismanagement in the B-2 program was on the government side—namely, the Air Force changed requirements. This forced a major, very expensive and time-consuming redesign, but the Air Force insisted that the production schedule should remain unchanged. The contractor's misstep was agreeing to begin manufacturing before the aircraft design was complete, resulting in labor inefficiency, parts shortages, tooling problems, and other negative effects.[95] One possible case of contractor mismanagement was the late delivery of aircraft and the high level of manufacturing defects in aircraft delivered.[96] The GAO argued in 1990 that the contractor lacked effective production labor standards (the amount of time needed to complete a given task), the management tools needed to measure and correct labor inefficiency, and effective programs to identify and correct the causes of defects.[97] A later GAO report criticized the contractor's management of defensive avionics development. The study claimed that the contractor failed to analyze test data before continuing tests, and it had excessive confidence that upgraded computer software would solve the avionics problems.[98]

The F-22 not only struggled with unrealistic cost and schedule estimates, but also dealt with some short-sighted decisions on the part of the prime contractor and the government in regards to assembling and integrating the aircraft's many technically ambitious components. The development award distributed F-22 workshare among three locations: one subcontractor built the wings, aft fuselage, and APU in Washington; the other built the center fuselage and armament in Texas; and the prime contractor built the forward fuselage, vertical tails, flaps, and landing gear, as well as conducted the overall integration in Georgia. Avionics elements were split between the three locations even though the overall avionics system had to be highly integrated. Each team optimized its subsystem in isolation from the others and had its own personnel for activities such as finance and contracting. This was inefficient and drove up costs. Software developers in geographically dispersed locations developed their own solutions that were not standard across the program, which ultimately complicated software integration and support.[99] Lastly, the prime contractor made a short-sighted decision to move F-22 EMD program management from California to Georgia. The Georgia facility lacked in-house fighter and stealth experience, and few original program engineers and managers chose to move from California to the new location.[100]

Finally, on the KC-46 program, both the government and the contractor made poor management decisions. As noted, the government took a "hands off" approach to the development of risky new technologies and approved an overly aggressive testing schedule. The prime contractor underestimated the cost and difficulty of executing the contract and bid very aggressively to win the contract, thus exposing themselves to greater losses. KC-46 program execution suffered from a lack of basic engineering and quality discipline. This resulted in avoidable problems such as its wiring issue, fuel leaks, and production debris left in the aircraft. The contractor also failed to manage its suppliers and subcontractors effectively. For example, the supplier of the centerline drogue system and wing pods did not follow the procedures for FAA certification and later discovered a design flaw that required correction. This provoked a legal dispute and caused yet another schedule delay.[101] A decision to close the facility that performed tanker modifications and boom development may also have contributed to design problems and delays. As in the F-22 case, the work moved between states, but the majority of the experienced workforce did not move with it.[102]

On the F/A-18E/F program, the contractor was determined not to repeat the errors of the A-12. The company worked closely with the Navy to monitor cost and schedule performance and to keep requirements stable. Program workshare was divided among subcontractors who had experience on F/A-18A/B/C/D, and the program drew largely on existing suppliers. The F/A-18E/F was the pilot Navy program for the Integrated Product Team concept that facilitated effective management of the technical challenges that arose.[103] The prime contractor also maintained a substantial management reserve that was used when the program encountered technical issues, something that could not be said for many of the other programs.[104]

Conclusions

The four common issues that resulted in poor performance in our cases studies were ineffective communication and lack of trust between stakeholders; instability in requirements or funding, including disruptive alterations to program requirements, truncations to planned program procurement, and changes or cuts to program funding; a lack of production-representative test assets and overly aggressive test schedules; and poor management by the government, the contractor, or both. None of the four issues noted here is necessarily a fresh discovery or the root cause of the difficulties of any one program. These issues were, however, common between multiple programs. While all, except the A-12, eventually produced excellent aircraft with good service records (or promising future ones), the difficulty of meeting cost and schedule targets in delivering these aircraft argues that future major programs should acknowledge these issues and manage risks accordingly.

Trust and good communication are extremely important to managing unforeseen risks. Suspicious or hostile relationships within a company, between a prime and subcontractors, or between the company and the government, can destroy trust and inhibit good communications, as the C-17 and KC-46 cases show. But relationships which obscure the reality of a program's performance or do not adequately flag or mitigate risk can also be problematic, as in the A-12 case.

Repeated examples from the case studies, including purportedly "low-risk" programs that have encountered major problems, argue that a particular contracting approach like firm-fixed-price or mandating contractor data for compliance purposes is not enough to manage risk effectively. A better approach is for programs to use an "active contract management" approach: using reliable and mutually accessible data and sharing program risk throughout development and into the production phase. The goal should be to enable communication and trust between the major stakeholders. With reliable data and a sense of trust, contractor and government program managers have both the confidence to make wise investments as well as the flexibility to proactively address problems that often arise when completing development and transitioning a program to production.

Unstable requirements, unstable or insufficient funding, and reductions in procurement quantities can cause programs to struggle in the transition from development to production. On the B-1, B-2, C-17, F-35, and F-22, technical requirements were established and later changed without adequate recognition of the cost and schedule implications and risks. Inadequate funding at an early stage can cause program delay and cost increases down the line. Stable requirements and a fixed-price contract can insulate the government from cost risk but not from schedule risk, as on the KC-46.

The B-1B, B-2, C-17, F-22, F-35, and KC-46 programs clearly show the importance of having an adequate number of production-representative test aircraft available in a timely manner. They are also examples of the need to build these test aircraft in sufficient time to conduct important tests of the integrated system prior to an LRIP decision. Subcomponent and software maturity is often not definitively understood until a fully integrated system is tested. Failing to account for this risk can lead to cost and schedule overruns. This issue has only grown in importance, as additional sensors, processors, and software have become integral to modern combat aircraft. If systems move into LRIP on the basis of program schedule before sufficient testing has occurred, then design changes to correct problems, additional testing, and then expensive retrofits to production aircraft may become necessary.[105]

Contractor mismanagement, or ill-judged governmental decisions on program structure, are common characteristics of defense programs that struggle or fail to reach FRP. On the A-12, poor management and weak government oversight resulted in program cancelation. The C-17, F-22, and KC-46 programs showed the ill effects of program personnel turbulence and poor management of subcontractors or company subsidiaries. Poor engineering management and discipline, such as the quality control problems on the KC-46, also disrupted cost and schedule.

This research leads to five considerations for future MDAPs:

1. **Establish trust and good communication early.** This is extremely important to managing unforeseen risks. Having data directly accessible by the contractor and government, and presenting it regularly and in a transparent manner, is a key to managing risk and enabling trust between the government and contractor.
2. **Consider implementing Active Contract Management to foster a collaborative management culture and create a sense of shared accountability for program outcomes.** Managers on both sides work collaboratively to investigate areas of potential concern or improvement, and to make adjustments to their contracting approach where necessary, including additional contracting actions or mixing contracting approaches to better suit program objectives.
3. **Maintain stability in requirements and predictable funding.** This steadfastness is required to effectively execute a complex program to produce, test, procure and field

a major weapon system while also meeting schedule and cost targets.

4. **Ensure sufficient production-representative test assets are available and built in sufficient time to conduct important tests of the integrated system.** Developmental and early operational testing ideally occurs on "production-representative" systems that use the same parts, processes, and software intended for use on production aircraft. Testing seeks to evaluate the effectiveness, safety, survivability, reliability, and maintainability of a system in a realistic environment prior to an FRP decision.
5. **Ensure strong, experienced management by the government and contractor.** The list of poor contractor management decisions to avoid is long, but examples include, short-sighted company reorganization, geographic relocation of program assets, ineffective management of subcontractors, poor personnel decisions, and questionable conduct which yielded program disruptions, increased costs, and schedule delays. In some of the case studies, government decisions also caused problems. The most common government failure was lack of effective oversight.

These observations are made consistently through the many case studies. To make a direct policy recommendation with methods to achieve would require further research and several more acquisition professionals.

The Air Force and Navy intend to begin a number of new aircraft acquisition programs in the 2020s, including Next Generation Air Dominance (NGAD), next-generation tankers, and new ISR platforms. Given the need to modernize capabilities to deter Great Power competitors, it is vital that these programs do not struggle for years to enter production. Speed, however, will not be the only determinate of success, as tight budgets also require the DOD to be a good steward of taxpayer dollars. The Air Force and the Navy are more likely to achieve schedule and cost objectives if they take a realistic approach to the inherent risks of acquisition, rather than focusing on only those associated with one dimension like speed of fielding or cost control. Finding a successful acquisition approach that balances time, schedule, and capability risks requires government and industry to align their focus and resources and to share accountability for program outcomes. This will require program structures, data reporting, and contracts to foster transparency and trust between government and industry partners so that program leaders can proactively problem-solve and mitigate unnecessary risks. Improvements in virtual environments and automated data capture may go some way to create this desired transparency. But it is important that all parties are incentivized to communicate regularly and align their efforts closely to actively manage the challenges in their programs and successfully deliver needed capabilities to the warfighter on time and in sufficient quantity.

Endnotes

1 For reasons beyond the scope of this paper, since 2000, the Army canceled many major acquisition programs, including the Future Combat System, the Comanche helicopter, the Crusader howitzer, the Armed Reconnaissance Helicopter, and the Aerial Common Sensor. The Air Force canceled a number of programs before they reached Milestone C, including the Multi-Sensor Command and Control Aircraft, the CSAR-X helicopter, and the Transformational SATCOM satellite. The Navy truncated the DDG-1000 program after building three ships. With the exception of the A-12 and F/A-18 E/F, the case studies in this paper examine Air Force fixed-wing, manned aircraft. Further studies could examine Army programs, such as the Comanche helicopter; other Navy aircraft, such as the E-2D; unmanned programs, such as Global Hawk; or shipbuilding programs, such as the DDG-1000. Such further analysis could reveal the validity and limitations of the methodology used here.

2 Some of the programs surveyed were procured prior to the introduction of Milestone C into procurement frameworks; prior to 2003 the transition from development to production occurred at Milestone III, often separated into Milestone IIIa (Low-Rate Production) and Milestone IIIb (Full-Rate Production). For the purposes of this paper, Milestone C is synonymous with the period of time from the latter part of EMD until Full-Rate Production contract is awarded, as this decision is usually the truer indication that a program has ceased "development" and entered full rate production.

3 Department of Defense Instruction 5000.02, August 10, 2017, p. 29.

4 Government Accountability Office (GAO), *Defense Acquisitions Annual Assessment: Drive to Deliver Capabilities Faster Increases Importance of Program Knowledge and Consistent Data for Oversight* (Washington, DC: GAO, June 2020), pp. 1–2.

5 Betty R. Kennedy, *Globemaster III: Acquiring the C-17* (Scott AFB: Air Mobility Command Office of History, 2004), pp. 113, 126.

6 Department of Defense, Office of the Inspector General (DODIG), "Selected Acquisition Actions on the C-17 Aircraft," November 2, 1990, pp. 19, 23.

7 Ibid, pp. 13–16.

8 Office of the Under Secretary of Defense for Acquisition and Technology, "Report of the Defense Science Board Task Force on C-17 Review," December 1993, p. 9.

9 Ibid, pp. 207–208.

10 Randy Davis, Bill Phillips, and Bud Vasquez, "The Phoenix Rises," *Defense Acquisition Review Journal*, July 2018, p. 216.

11 Chester Paul Beach Jr., "A-12 Administrative Inquiry," memorandum to the Secretary of the Navy, (Washington DC: Office of the Under Secretary of Defense for Acquisitions, November 28, 1990), p. 23.

12 Ibid., pp. 4–6.

13 Ibid., pp. 26–27.

14 Frank Kendall, "KC-46 Tanker Problems Call For A Rethink of Fixed-Price Development Programs," *Forbes*, March 31, 2020.

15 DODIG, *Evaluation of the Air Force Systems Engineering Processes Used in the Development of the Refueling Boom for the KC-46A Tanker*, DODIG-2021-088 (Washington, DC: DOD, May 21, 2021), pp. 11–13.

16 GAO, *KC-46 Tanker Modernization* (Washington, DC: GAO, June 2019), p. 29.

17 Ibid. Staff with "strong cost estimating skills" and expertise in "data rights" supported the negotiations.

18 Valerie Insinna, "Boeing pushes back on the KC-46 program's bad reputation with the Air Force," *Defense News*, May 4, 2018.

19 "Failed to provide refueling tankers, USAF warns Boeing it's not happy," *Bloomberg*, January 18, 2020.

20 Obaid Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs* (Santa Monica, CA: RAND, 2005), p. 47.

21 Michael D. Williams, *Acquisition for the 21st Century* (Washington, DC: NDU Press, 1999), pp. 100–127. See especially pp. 114–115, 123, 126.

22 House of Representatives, Committee on Government Reform, "F-22 Cost Controls: How Realistic Are Production Cost Reduction Plan Estimates?" August 2, 2001.

23 GAO, *Defense Acquisitions: Stronger Management Practices Are Needed to Improve DOD's Software-Intensive Weapon Acquisitions* (Washington, DC: GAO, March 1, 2004), p. 23.

24 John C. Dibert and John C. Velez, "An Analysis of Earned Value Management Implementation within the F-22 System Program Office's Software Development," MBA Professional Report, Naval Postgraduate School, December 2006, pp. 80–81, 84–85.

25 Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs*, p. 49.

26 Ibid., p. 22.

27 Ibid., pp. 25–26.

28 A survey of participants conducted for an MBA thesis from the Naval Postgraduate School concluded, "There were indications that EVM is used more as a tool for receiving and reporting cost and schedule status versus using it as a tool for proactively managing the development effort." Dibert and Velez, "An Analysis of Earned Value Management Implementation within the F-22 System Program Office's Software Development," pp.102–103.

29 Wayne Abba spent 30 years in DOD acquisition before retiring as a consultant specializing in EVM. Wayne Abba, "The Evolution of Earned Value Management," Defense AT&L, March–April 2017, p. 4.

30 Ibid., p. 4.

31 "Active Contract Management: How Governments Can Collaborate More Effectively with Social Service Providers to Achieve Better Results," Harvard Kennedy School Government Performance Lab, pp. 3–5.

32 Ibid, pp. 5–6.

33 Frank Kendall, "Risk and Risk Mitigation—Don't Be a Spectator," *Defense AT&L*, January–February 2015, p. 5.

34 See Gene Porter et al., *The Major Causes of Cost Growth in Defense Acquisition,* Vol. II (Alexandria, VA: Institute for Defense Analyses, 2009), pp. 42–44; and Mark A. Lorell et al., *Program Characteristics That Contribute to Cost Growth* (Santa Monica, CA: RAND 2017), pp. 9–10.

35 Joshua McConnell, Jason Sickler, and Trent Yang, *The "Problems Behind the Problems": Systems Engineering and Management Risk Factors in Acquisition Programs* (Alexandria, VA: IDA, 2004), pp. V-1–V-12.

36 Rebecca Grant, *B-2: The Spirit of Innovation* (Falls Church, VA: Northrop Grumman, 2013), pp. 55–62.

37 Irv Blickstein et al., *Root Cause Analyses of Nunn-McCurdy Breaches*, Vol. 1 (Santa Monica, CA: RAND, 2011), p. 57.

38 Susan J. Bodilly, *Case Study of Risk Management in the USAF B-1B Bomber Program* (Santa Monica, CA: RAND, 1993), pp. 21–22.

39 A. Lee Battershell, *The DOD C-17 versus the Boeing 777* (Washington, DC: National Defense University, 1999), pp. 72–74.

40 Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs*, pp. 21–27.

41 Any changes in KC-46 requirements require high-level Air Force approval. GAO, *KC-46 Tanker Modernization*, p. 28.

42 An article from 1983 laments the negative effects of funding instability in the 1970s. John T. Correll, "The Costly Alternative to Controlling Cost," *Air Force Magazine*, June 1983.

43 Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs*, pp. 60–65.

44 House Committee on National Security, National Defense Authorization Act for FY 1997, March 13, 1996, p. 448. See also Lt Gen Richard E. Hawley's testimony, House of Representatives, Committee on Appropriations, Subcommittee on National Security, Part 4, May 10, 1995, pp. 219–220, 243.

45 Battershell, *The DOD C-17 versus the Boeing 777*, p. 85.

46 Ibid.

47 Beach, "A-12 Administrative Inquiry," p. 26. The Navy rejected Northrop's A-12 bid as too expensive. In retrospect, that bid, informed by their experience with stealth and composites, was far more realistic.

48 In the 1980s, "he B-2 program had famously proceeded without the need for cost-cutting." Grant, *B-2: The Spirit of Innovation*, p. 73. The estimated cost of the redesign for low-altitude flight alone was $2 billion. Grant, *B-2: The Spirit of Innovation*, p. 60. In 1996, the GAO claimed "Congress has provided the Air Force with almost 96 percent of the $44,785 million total estimated cost to complete the 21 B-2 aircraft program through fiscal year 1996." GAO, *B-2 Bomber: Status of Efforts to Acquire 21 Operational Aircraft* (Washington, DC: GAO, October 1996), p. 14.

49 Susan Bodilly, Frank Camm, and Richard Pei, *Analysis of Air Force Aircraft Multiyear Procurements with Implications for the B-2* (Santa Monica, CA: RAND, 1991), pp. 18–19.

50 Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs*, pp. 65–66.

51 Blickstein et al., *Root Cause Analyses of Nunn-McCurdy Breaches*, p. 52.

52 The fourth aircraft would have been "production-representative" and was intended to become an operational aircraft after testing was complete. See "Defense - B-1 Bomber" from the Ron Nessen Papers at the Gerald R. Ford Presidential Library.

53 House of Representatives, Committee on Appropriations, Subcommittee on Defense, "Department of Defense Appropriations for 1982," Part 9, September 24, 1981, pp. 218, 223–224. House of Representatives, Committee on Armed Services, "Hearings on Military Posture," Part 2 of 7, October 6, 1981, p. 240.

54 "Department of Defense Appropriations for 1982," Part 9, September 24, 1981, p. 171. See also Bodilly, *Case Study of Risk Management in the USAF B-1B Bomber Program*, p. 24.

55 One of the B-1A crashed in August 1984, leaving only one B-1A for testing until the first B-1B arrived in 1985.

56 Bodilly, *Case Study of Risk Management in the USAF B-1B Bomber Program*, p. 29.

57 Ibid., p. 25. See also GAO, *The B-1 Bomber Program—A New Start* (Washington, DC: GAO, April 13, 1983), pp. 7–8.

58 Bodilly, *Case Study of Risk Management in the USAF B-1B Bomber Program*, pp. 34–35.

59 United States Senate, Committee on Appropriations, Subcommittee on Defense, "Department of Defense Appropriations for FY 1983," Part 3, June 24, 1982, pp. 146–170.

60 Major Bill Norton, "The C-17 and Fast-Paced Flight Testing," AIAA Aircraft, Technology, and Operations Forum, October 2001, p. 4. The initial plan was 5 test aircraft, 22 calendar months, 77 aircraft months, and 2,300 flight hours. In the end, testing required 6 aircraft, 39 calendar months, 135 aircraft months, and 4,096 flight hours.

61 GAO, *Status of the C-17 Development Program* (Washington, DC: GAO, March 10, 1993), p. 8.

62 Norton, "The C-17 and Fast-Paced Flight Testing," pp. 6–10.

63 GAO, *B-2 Bomber: Status of Compliance with the 1992 and 1993 Defense Authorization Act* (Washington, DC: GAO, December 1992), pp. 3–4.

64 GAO, *B-2 Bomber: Status of Cost, Development, and Production* (Washington, DC: GAO, August 1995), pp. 15–17; and GAO, *B-2 Bomber: Additional Costs to Correct Deficiencies and Make Improvements* (Washington, DC: GAO, June 1998), pp. 5–8.

65 GAO, *B-2 Bomber: Additional Costs to Correct Deficiencies and Make Improvements*, pp. 2, 10–11.

66 Bernard Fox et al., *Test and Evaluation Trends and Costs for Aircraft and Guided Weapons* (Santa Monica, CA: RAND, 2004), pp. 126–132.

67 United States Senate, Committee on Armed Services, "Department of Defense Authorization for Appropriations for Fiscal Year 2002," Part 4, July 10, 2001, p. 13.
68 GAO, *F-22 Delays Indicate Initial Production Rates Should Be Lower to Reduce Risks* (Washington, DC: GAO, March 2002), pp. 9–15.
69 Other technical challenges included vertical fin buffeting, overheating of the aircraft rear, and a weak horizontal tail that required redesign. GAO, *DOD Should Reconsider Decision to Increase F/A-22 Production Rates While Development Risks Continue* (Washington, DC: GAO, March 2003), pp. 4–6.
70 GAO, *F-22 Delays Indicate Initial Production Rates Should Be Lower to Reduce Risks*, p. 19.
71 GAO, *DOD Should Reconsider Decision to Increase F/A-22 Production Rates While Development Risks Continue*, pp. 12–14.
72 The approach to testing mission systems software is described in Greg Lemons et al., "F-35 Mission Systems Design, Development, and Verification," AIAA Aviation Forum Paper, June 25–29, 2018, pp. 1–16.
73 See "F-35 Joint Strike Fighter (JSF)," FY20 DOD Programs factsheet, pp. 20, 23–24.
74 See "F-35 Joint Strike Fighter (JSF)," FY18 DOD Programs factsheet, pp. 29–32.
75 A formal Full-Rate Production decision, originally planned for April 2012, has still not been made.
76 DOT&E questioned whether this was prudent and affordable, and whether Lockheed would have any incentive to fix problems with these aircraft that testing identified. See "F-35 Joint Strike Fighter (JSF)," FY15 DOD Programs factsheet, p. 41 of the FY15.
77 "KC-46A," FY12 Airforce Programs factsheet.
78 "KC-46A," FY16 Airforce Programs factsheet.
79 DODIG, *Evaluation of the Air Force Systems Engineering Processes Used in the Development of the Refueling Boom for the KC-46A Tanker*, pp. 21–25.
80 Brian W. Everstine, "Pegasus Prep," *Air Force Magazine*, January 2017; and Valerie Insinna, "Remember that list of KC-46 problems, Here's how they're getting fixed," *Defense News*, May 9, 2018.
81 Brian W. Everstine, "KC-46 Full-Rate Production Pushed Back to 2024," *Air Force Magazine*, June 9, 2020.
82 Christopher Bolkcom, *Military Aircraft: The F/A-18E/F Super Hornet Program: Background and Issues for Congress* (Washington, DC: CRS, July 13, 2005), pp. 1–6.
83 Due to aggressive competitor lobbying, the Pentagon studied airlift alternatives such as purchasing more C-5s or modified 747s as alternatives to the C-17.
84 DODIG, "Selected Acquisition Actions on the C-17 Aircraft," pp. 17–20.
85 Ibid., pp. 21–22. Battershell, *The DOD C-17 versus the Boeing 777*, p. 94.
86 David J. Lynch, "Turbulence Dogs Douglas Overhaul," *Chicago Tribune*, September 9, 1991. Battershell, pp. 32-33. Douglas employees joked that TQMS stood for "Time to Quit and Move to Seattle."
87 Battershell, *The DOD C-17 versus the Boeing 777*, pp. 87–91.
88 On the turnaround, see Kennedy, *Globemaster III: Acquiring the C-17*, pp. 128–161.
89 James P. Stevenson, *The $5 Billion Misunderstanding* (Annapolis, MD: Naval Institute Press, 2001), pp. 92–103.
90 Stevenson, *The $5 Billion Misunderstanding*, pp. 114–118, 160–163.
91 Beach, "A-12 Administrative Inquiry," pp. 3–4.
92 Ibid., pp. 5–6.
93 Ibid., p. 7.
94 Frank Kendall, "Adventures in Defense Acquisition," *Defense AT&L*, January–February 2017.
95 GAO, *Strategic Bombers: B-2 Program Status and Current Issues* (Washington, DC: GAO, February 1990), pp. 13–15, 19.
96 GAO, *B-2 Program: Trends in Manufacturing* (Washington, DC: GAO, July 1991), pp. 4–5, 18–21.
97 Ibid., pp. 24–27.
98 GAO, *B-2 Bomber: Additional Costs to Correct Deficiencies and Make Improvements*, p. 6.
99 Ibid., pp. 18–21. See also GAO, "Air Force F-22 Embedded Computers," Letter to the Secretary of Defense, September 23, 1994, p. 2.
100 The Marietta facility also produced the C-130J. When C-130J sales were lower than anticipated, the F-22 program had to absorb a higher share of plant overhead costs. Ibid., p. 20.
101 GAO, KC-46 Tanker Modernization (Washington, DC: GAO, April 2018), p. 10.
102 Molly McMillin, "End of an era: Boeing in final stages of leaving Wichita," *The Wichita Eagle*, July 29, 2014.
103 Younossi et al., *Lessons Learned from the F/A-22 and F/A-18E/F Development Programs*, pp. 21–27.
104 Ibid., p. 52.
105 This is not a recent or infrequent problem. See GAO, *Low Rate Initial Production Used to Buy Weapon Systems Prematurely* (Washington, DC: GAO, November 1994).

About The Mitchell Institute

The Mitchell Institute educates the general public about aerospace power's contribution to America's global interests, informs policy and budget deliberations, and cultivates the next generation of thought leaders to exploit the advantages of operating in air, space, and cyberspace.

Forum Submissions and Downloads

For more information about submitting papers or ideas to the Forum, or for media inquiries, email our publications team at forum.mitchellaerospacepower@afa.org

Copies of Forum papers can be downloaded under the publications tab on the Mitchell Institute website at https://www.mitchellaerospacepower.org

About the Author

Lt Gen Mark D. Shackelford, USAF (Ret.) served for 34 years as a fighter pilot, test pilot, program manager, program director, and staff officer. He held executive positions in missile defense, military space requirements, and weapon system acquisition. He retired in 2011 as Military Deputy, Office of the Assistant Secretary of the Air Force (Acquisition). For the last decade he has advised corporate and government executives on program strategy, capture, and execution focused on aircraft, space vehicles, and space launch.

Author's acknowledgements:

"I am indebted to Dayton Aerospace for the extensive research and analysis they provided for this project. Many of the specific details and lessons learned found in this report came directly from interviews with previous program managers and program executive officers facilitated by Dayton Aerospace. As an SPO director, I had personal experience on a couple of the programs, but I was particularly surprised to discover the commonality in challenges that all programs faced and the variety of ways in which programs tried to mitigate developmental risk. I wish I knew then, what I have learned through doing this study."

An Affiliate of the Air Force Association | www.mitchellaerospacepower.org