IDA

INSTITUTE FOR DEFENSE ANALYSES

# Evidence on the Effectiveness of the Defense Acquisition Board Process

David L. McNicol, Project Leader
Brandon R. Gould
Linda Wu
Eugenia Fuchs
Patricia F. Bronson, Task Leader

May 2013
Approved for public release; distribution is unlimited.
IDA Paper P-5011
Log: H 13-000529
Copy

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

*The Institute for Defense Analyses is a non-profit corporation that operates three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.*

**About This Publication**

This work was conducted by the Institute for Defense Analyses (IDA) under contract DASW01-04-C-0003, AY-7-322306, "WSARA 2009: Root Cause Analysis of Programs in Nunn-McCurdy Breach and Performance Assessment Methods and Analyses: Organizational Conflicts;" and AY-7-357802, "Root Cause Anaylsis and Performance Assessment Methods and Analyses: Organizational Conflicts," for the Director, Performance Assessments and Root Cause Analyses (D,PARCA). The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

**Acknowledgments**

Thank you to Prashant R. Patel, Brian Q. Rieksts, and David A. Sparrow for performing technical review of this document.

**Copyright**

© 2013 Institute for Defense Analyses 4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

INSTITUTE FOR DEFENSE ANALYSES

IDA Paper P-5011

# Evidence on the Effectiveness of the Defense Acquisition Board Process

David L. McNicol, Project Leader
Brandon R. Gould
Linda Wu
Eugenia Fuchs
Patricia F. Bronson, Task Leader

# Executive Summary

This study is motivated by the questions of how effective the Defense Acquisition Board (DAB) process is and, to the extent it is ineffective, why and what can be done to reduce the particular shortfalls. Apart from its brevity, this study differs from previous studies in two main respects. First, it is concerned only with the DAB process, not the wider Department of Defense (DoD) acquisition process; in fact, it focuses on just one of the DAB's roles: prompting Major Defense Acquisition Programs (MDAPs) to be completed on schedule and within budget and with the performance specified in the program's approved baseline. Second, this study attempts to use data on acquisition program outcomes to study the effectiveness of the DAB, and is in this respect a pioneering effort.

## Basis of the Study

In its program oversight role, the DAB exists to ensure that programs are properly structured and have made sufficient technical progress to move forward. Of course, unpredictable events play some role in the outcomes of all programs—and a large role in some. Granted that, an important part of the rationale for the substantial commitments of resources made to the DAB process is that the DAB can identify and correct problems with the structure of acquisition programs and their management. The other side of this coin is that a DAB failure to cull problems and intervene as necessary tends to result in substandard program outcomes in terms of cost, schedule, and performance.

The core of this study is comparisons of different categories of MDAPs in terms of cost, schedule, and performance metrics. We show that there are consistent and substantial differences in the outcomes of the categories of programs compared and infer from that the existence of significant limitations on the effectiveness of the DAB process.

## Joint and Single Service New Starts

The first comparison is single-Service MDAPs that are new starts (that is, an entirely new aircraft, such as the F-22) and joint new starts (such as the F-35). The table below presents the values of the eight metrics we used for these two groups of programs.

**Values of the Metrics for Single-Service and Joint New Starts**

| | Metric | Single-Service New Starts | Joint New Starts |
|---|---|---|---|
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | 35% (80) | 42% (19) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | 20% (12 of 60) | 32% (9 of 28) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | **40%** (32 of 80) | **63%** (12 of 19) |
| 4 | Program Cancellation Rate | **19%** (33 of 170) | **34%** (15 of 44) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | **28%** (47 of 170) | **57%** (25 of 44) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity[b] | 37% (19 of 52) | 48% (11 of 23) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | **17%** (20) | **36%** (7) |
| 8 | Percentage of programs with at least one Operational Test failure | **47%** (23 of 49) | **77%** (10 of 13) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last Selected Acquisition Report (SAR) for the program divided by the total procurement reported in the first SAR.

Joint new starts are less successful than single-Service new starts as measured by each of the eight metrics. Whether the differences between joint and single-Service new starts are large enough to show the existence of limits on the effectiveness of the DAB process is a matter of judgment. The striking differences (highlighted in yellow) are the percentage of programs with an estimated in Quantity Normalized Program Acquisition Unit Cost (PAUC) growth of ≥ 30 percent (metric 3); the cancellation rate (metric 4); the rate of cancellation, truncation, or reorganization (metric 5); the slip in the Engineering and Manufacturing Development (EMD) schedule (metric 7); and the proportion of programs that had at least one Operational Test failure (metric 8). The value of each of these metrics is at least half again as large for joint program new starts as for single-Service new starts. We believe these data do clearly show that there is a binding limit on the effectiveness of the DAB process. When over a period of years, about 57 percent of joint programs and 28 percent of single-Service new starts are cancelled, truncated, or reorganized, it is difficult to see how this conclusion could be avoided.

## Single Service Variant, Modification, or Remanufacturing (VMR) during Cold War and Post-Cold War

The VMR category includes programs that (1) acquire a variant of an existing system, (2) modify a fielded system, or (3) remanufacture some or all of the units of a system that has been fielded. These sorts of programs usually are not simple on an absolute scale. All of the VMRs in our sample are MDAPs, and most involve a considerable amount of development. The table below presents the values of the metrics for single-Service VMRs for the 1970–1988 and post-1988 periods.

**Value of the Metrics for Single-Service VMRs, Single-Service New Starts, and Joint New Starts for 1970–1988 and Post-1988**

| | Metric | Single-Service VMR | |
|---|---|---|---|
| | | 1970–1988 | 1989– |
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | **12%** (22) | **45%** (19) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | N/A (0 of 3) | 14% (4 of 28) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | **18%** (4 of 22) | **58%** (11 of 19) |
| 4 | Program Cancellation Rate | 11% (4 of 36) | 13% (4 of 32) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | 14% (5 of 36) | 19% (6 of 32) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity[b] | **100%** (3 of 3) | **36%** (10 of 28) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | N/A | 26% ( 16) |
| 8 | Percentage of programs with at least one Operational Test failure | **13%** (1 of 8) | **40%** (6 of 15) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last SAR for the program divided by the total procurement reported in the first SAR.

The two periods correspond roughly to Cold War and post-Cold War. DoD procurement funding, which peaked in 1986, fell to just under half its 1989 level by 1995 (in constant dollars). With the decrease in the overall DoD budget, and in particular the sharp drop in funds allocated to acquisition, major single-Service new starts were discouraged and the Services were strongly encouraged to instead pursue VMR programs. Previously, the choice between a VMR and a new start was to a greater extent one left to the Services.

There clearly was a substantial change between the two periods in the outcomes for single-Service VMR programs. By the two cost growth metrics (1 and 3) and the performance on operational testing (metric 8), VMR programs did worse, by a factor of more than three, post-1988 than they did during 1970–1988. They fared much better at buying at least the Milestone (MS) B baseline quantity (metric 6), by a factor of nearly three. (Note, however, that we have data on schedule for only three single-Service VMRs for 1970–1988.) A reasonable interpretation of these data is that during 1989–2001, post-MS B, the Services tended to sharply move VMR programs towards the greater capability that they preferred and which a new start would provide. The DAB evidently was not effective in fully picking up on this change and responding to it.

The body of the paper considers whether these observations reflect a change in the rigor of the Office of the Secretary of Defense (OSD) process for oversight of MDAPs, rather than a change in the Services' incentives. The paper presents evidence (Table 6, page 18) that the observed differences do primarily reflect changes in the Services' incentives.

## Conclusions

This paper uses two comparisons to identify situations in which the OSD acquisition oversight process has notably been less than fully successful. The differences in both cases were consistent, and those on which we rested our conclusion were large—differences of nearly a factor of 50 percent between single-Service and joint new starts and of about a factor of three for single-Service VMRs during the two periods. The OSD acquisition oversight process clearly fell somewhat short in dealing with the problems presented by joint new starts and by single-Service VMRs post-1988.

We attribute these failings to the intermittent and apparently reactive character of the DAB's oversight between milestone reviews. The analysis that reaches this conclusion starts with the question: What feature(s) do joint new starts and single-Service VMRs launched after 1988 both have that single-Service new starts and single-Service VMRs during 1970–1988 do not? A feature most joint new starts and post-1988 single-Service VMR programs share is guidance about which the relevant Services were not enthusiastic. Single-Service new starts may get guidance from a DAB milestone review that they would prefer not to have. The issue is then one of degree. We believe that it is a reasonable judgment that joint new starts and post-1988 single-Service VMRs operate far more often than single-Service new starts on guidance that was of major importance to the specification of the program's content. If this is granted, the comparison of the metrics for single-Service and joint new starts and single-Service VMRs in the Cold War and post-Cold War periods points to limits of the OSD acquisition oversight process in enforcing programmatic guidance.

The core argument is simple: The DAB did not provide post-milestone oversight in precisely those circumstances in which it is most needed—when one or more of the Services receive guidance to do something other than what it would prefer. We see no reason to doubt the premise that most (not all) of the joint new starts and post-1988 single-Service VMRs entailed guidance the Services involved would rather not have had. The assertion that OSD oversight of MDAPs between milestone reviews has, over the years of our data, generally been intermittent and reactive is one that we believe would be generally accepted. These are institutional facts that make our conclusion plausible. Beyond that, we also have provided circumstantial evidence that points to insufficient DAB oversight between milestone reviews as the key failing in the DAB process with respect to unsatisfactory outcomes we observed.

# Contents

A. Introduction ..... 1
B. Basis of the Study ..... 2
C. Categories, Metrics, and Data ..... 8
  1. Categories ..... 8
  2. Metrics ..... 9
  3. Data ..... 11
D. Single-Service and Joint New Starts ..... 12
E. Single-Service VMR Programs in Two Periods ..... 15
F. What Can We Say About Where the DAB Process Falls Short? ..... 19
G. Concluding Observation ..... 21
Appendix A. Data Sources and Methodology ..... A-1
Appendix B. Values of the Metrics for 1970–1988 and Post-1988 for VMRs, Single-Service New Starts, and Joint New Starts ..... B-1
Illustrations ..... C-1
References ..... D-1
Abbreviations ..... E-1

## A. Introduction

This study is motivated by the questions of how effective the Defense Acquisition Board (DAB) process is and, to the extent it is ineffective, why and what can be done to reduce the particular shortfalls.

The closest thing we have to a quantitative analysis of the defense acquisition process is the cost growth literature that began in the early 1970s.[1] These studies sought to discover the extent to which various attributes of major Department of Defense (DoD) weapon system acquisition programs are statistically associated with cost growth. While not always stated clearly, the point was to provide information to help design improvements to acquisition policies. The implicit assumption was that unacceptably high cost growth (and by implication schedule growth and performance shortfalls) were due mainly to defects in how acquisition programs were structured which, once identified, the DAB could preclude at milestone reviews.

The cost growth literature provided only a few narrow conclusions about how acquisition programs should be structured. It is mainly through accumulated experience, case studies, surveys of experience on acquisition programs, and countless symposia and workshops on acquisition problems, that it has become the conventional wisdom that "We (DoD) know how to do an acquisition program, but we sometimes fail to do it." A reasonable question to ask is then: Why does the DoD acquisition process with significant frequency fail to do what "we know how to do"?

There have been many high-level government commissions and a few academic studies directed to identifying improved defense acquisition policies and processes.[2] These works typically ground their cases for reform on poor outcomes of major defense acquisition programs—high cost growth, for example; or in problematic features of the acquisition process—the time required for a milestone review, for example; or what is perceived to be broad flaws in results obtained by the acquisition process—for example, failures to incorporate the best available technologies in defense systems. To our knowledge, none of these studies provides a close analysis of the DAB process. Rather, their prescriptions typically embody policy positions on what the DoD acquisition process should accomplish and a vision of what process will best support that result.

---

[1] References to major studies in the literature can be found in David L. McNicol, *Cost Growth in Major Weapon Procurement Programs*, IDA Paper P-3832 (Nonstandard), Second edition (Alexandria, VA: Institute for Defense Analyses, 2005).

[2] Perhaps the best known, chaired by David Packard and often referred to as the Packard Commission report, is *A Quest for Excellence: Final Report of the President's Blue Ribbon Commission on Defense Management* (Washington, DC: June 1986).

Apart from its brevity, this study differs from previous studies in two main respects. First, it is concerned only with the DAB process, not the wider DoD acquisition process; in fact, it focuses on just on one of the DAB's roles: prompting Major Defense Acquisition Programs (MDAPs) to be completed on schedule and within budget and with the performance specified in the program's approved baseline. Second, this study attempts to use data on acquisition program outcomes to study the effectiveness of the DAB, and is in this respect a pioneering effort.

## B. Basis of the Study

The core of this study is a comparison of different categories of MDAPs in terms of cost, schedule, and performance metrics. The categories, metrics, and data used are described in Section C. This section is concerned with the basis for drawing inferences about the effectiveness of the DAB from such comparisons.

The definitive guidance on the goals of the acquisition process is contained in Department of Defense Directive 5000.01, which includes the following among a list of specific policies:

> *4.3.4. Discipline. PMs shall manage programs consistent with statute and the regulatory requirements specified in this Directive and in reference (b). Every PM shall establish program goals for the minimum number of cost, schedule, and performance parameters that describe the program over its life cycle. Approved program baseline parameters shall serve as control objectives. PMs shall identify deviations from approved acquisition program baseline parameters and exit criteria.*

Although this language does not apply only to the DAB, it warrants the assumption that one of the DAB's main goals is maintaining a discipline in major acquisition programs. This interpretation is buttressed by an accumulation over the past quarter-century of high-level committee reports, policy guidance by the Secretary of Defense, and statutes.

In its program oversight role, the DAB exists to ensure that programs are properly structured and have made sufficient technical progress to move forward. Of course, unpredictable events play some role in the outcomes of all programs—and a large role in some.[3] Granted that, the rationale for the substantial commitments of resources made to the DAB process is that the DAB can identify and correct problems with the structure of acquisition programs and their management. The other side of this coin is that a DAB

---

[3] The root cause analyses done for the Office of Program Assessments and Root Cause Analyses (PARCA) potentially provide some insight into the role of unforeseeable events in the outcomes of major DoD acquisition programs. See Irv Blickstein et al., *Root Cause Analyses of Nunn-McCurdy Breaches*, Vol. 1 (2011) and Vol. 2 (2012) (Santa Monica, CA: The RAND Corporation) and Patricia Bronson et al., *Root Cause Analyses of Programs in Nunn-McCurdy Breach*, Volume I, IDA Paper P-4935 (Alexandria, VA: Institute for Defense Analyses) (forthcoming).

failure to cull problems and intervene as necessary tends to result in substandard program outcomes in terms of cost, schedule, and performance.

Although the usage is common, it is somewhat simplistic to speak of problems with MDAPs in the context of a DAB review. In some cases, "difference of perspective" is closer to the mark, as the Component sponsoring the program may well differ with OSD on the urgency of procuring the systems, the uncertainties presented by the program, the costs of the program, and the advantages and disadvantages of various options. Consequently, what OSD may see as a problematic feature of the program—use of relatively immature technologies, for example—the Component may see as a prudent balancing of considerations. In these situations, it is probably best to see the DAB not as "correcting problems" but as attempting to adjust the program's risk profile to reflect that of the Secretary and Deputy Secretary of Defense.

These comments nonetheless characterize the role of the DAB as what might be called quality insurance. In this role, the **output** of the DAB is problems (with major acquisition programs) avoided or solved. The **outcome** of the DAB is gauged by the outcomes of major acquisition programs, in particular, by the extent to which they are completed on schedule, within budget, and provide the performance specified in the program's baseline.

The DAB's output at Milestone Reviews, especially Milestone (MS) B reviews, is produced through two distinct activities:

- The various staff elements within the DAB process review programs, identify features of those programs as problematical, provide evidence to justify their finding, and point to corrective actions.
- The Milestone Decision Authority (MDA), with support from his or her principal subordinates, decides on what changes to the program to direct.

The results of the staff processes are ordinarily contained in memoranda or briefings for the head of the Overarching Integrated Product Team that manages the program's progress through the review process or for the MDA.[4] The MDA's direction for corrective actions is provided by the Acquisition Decision Memorandum for the program.

There typically is no regularly scheduled point for a DAB review between milestones. Rather, a review of the program occurs when the MDA decides it is warranted, which is most likely to happen when the Component sponsoring the program or an Office of the Secretary of Defense (OSD) staff member identifies a major problem

---

[4] Since 1988, the MDA for most MDAPs has been what is now the Under Secretary of Defense (Acquisition, Technology and Logistics) (USD(AT&L)).

with the program that the Component cannot or has not resolved.[5] It is important to note the lack of an ongoing oversight role for the DAB between milestones, as it plays a significant role in the interpretation of the quantitative results presented in Section E.

A great deal of data exist on the outcomes of MDAPs, but there seems to be no satisfactory way to combine these data into a useful overall measure of the DAB's effectiveness in its program oversight role. In contrast, there is no readily available data on DAB outputs, but, as will be shown in a moment, there is a conceptually satisfactory definition of the DAB's effectiveness in producing its output. The choice then apparently is between ample data but no good definition of the DAB's effectiveness, and a conceptually satisfactory definition but no data.

This impasse can be avoided if it is accepted that a systematic and sustained difference in outcome measures between two categories of MDAPs reveals a limitation of the DAB's effectiveness. The reader is asked to provisionally accept this assumption, which is discussed further as the section proceeds.

Figure 1 illustrates for a rudimentary case a construct that links the outputs of the DAB to observed program outcomes. Figure 1 takes program difficulty to be the average number of significant problems with a particular category of acquisition programs.[6] Using this metaphor, the three categories illustrated are assumed to have 5, 10, and 20 problems respectively. Figure 1 also assumes that there is a limitation on the average number of problems the DAB can resolve for any one MDAP. The constraint is assumed to be MAX = 6.

In the case shown in Figure 1, the constraint is binding, and Category 3 programs on average have worse outcomes than those of Category 2 (or Category 1). If, instead, the capacity constraint is not binding, the DAB has sufficient capacity to maintain cost, schedule, and program discipline on even the most difficult categories of programs (e.g., MAX = 20). Some MDAPs will have worse outcomes than others simply because of random events, but if the DAB capacity constraint is not binding, we would not expect to observe consistent and substantial differences in the outcomes between categories.

---

[5] The Defense Acquisitions Executive Summary (DAES) process provides the USD(AT&L) with summary-level information for each MDAP at least once a quarter. When a DAES review uncovers a major problem with a program the program may be examined further through a DAB program review or some other less formal process. Since 2009, the OUSD(AT&L) has made concerted efforts to improve the utility of the DAES process, which the Weapon Systems Acquisition Reform Act of 2009 (WSARA 2009) encouraged.

[6] If this approach were used in an analysis of a set of MDAPs—which we do not necessarily recommend—individual problems probably would be weighted to reflect their severity. In effect, one problem (e.g., an unrealistic development schedule) would be picked as the unit of measure and other problems stated as multiples of it.

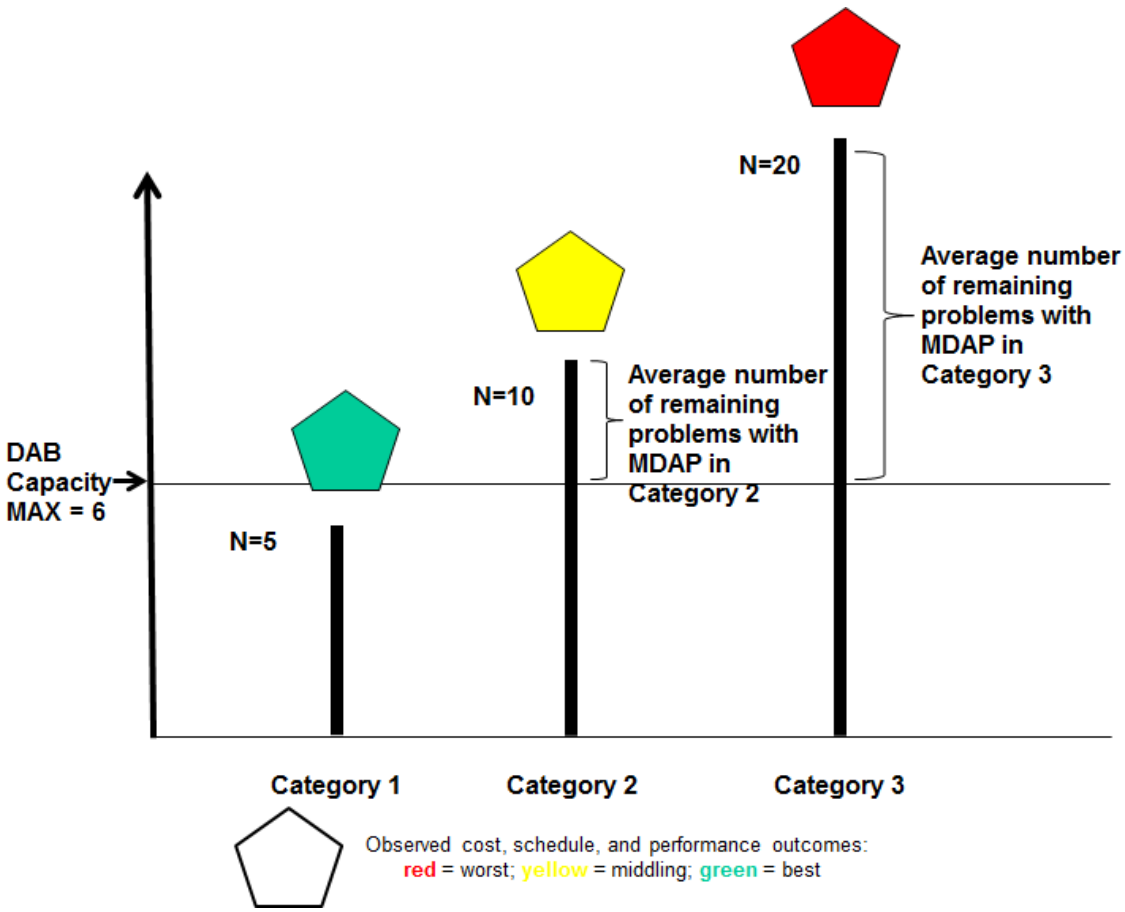


**Figure 1. Potential Consequences of a Limitation on the Effectiveness of the DAB Process**

In this construct, the effectiveness of the DAB in its program oversight role is measured by the percentage of total problems that it resolves. Given the simple assumptions of Figure 1, for Category 1 the efficiency of the DAB is 100 percent. The DAB's efficiency is 60 percent for Category 2, as it reduces the number of significant problems from 10 to 6. The DAB's output efficiency in the third category falls to just 30 percent.

Figure 1 should not be interpreted as a parable of "good guys" and "bad guys." The point of Figure 1 is to illustrate how what we see—program outcomes, represented by the colored pentagons—is linked to what we would like to investigate—whether the effectiveness of the DAB's ability to resolve problems with MDAPs is significantly limited. The claim is that if we compare outcome metrics for two categories of programs, 2 and 3, of which Category 3 is known to be more problematic, and find that the outcome metrics are consistently and substantially worse for Category 3, we can infer that the DAB is not fully successfully in resolving the more challenging problems presented on average by programs in that category. That is, the comparison as described reveals a comparative ineffectiveness of the DAB.

This proposition is sufficiently startling to consider objections to it before going any further. We are aware of four. The first, stated as a question is: Why don't differences in the difficulty of various categories of MDAPs provide a sufficient explanation of persistent and substantial differences in outcomes? Our response to this question is that, in fact, we do not see the outcomes of the program presented to the DAB; rather, we see the outcomes of programs as they have been influenced by the DAB process. Looking back at Figure 1, there would be no difference in average outcomes across program categories if all categories were equally difficult. There also would be no difference if the capacity of the DAB process was sufficiently large (at least 20 in Figure 1.) It is the combination of difference in the difficulty of various categories of programs and a binding constraint on DAB capacity that produces consistent and substantial differences in outcomes across program categories.[7]

Second: What about factors not represented in Figure 1? In particular, are there other OSD-level processes that in practice block the DAB process or override its results? Note that the problem would have to be a factor that had a differential effect on one category of MDAPs compared to others. We point out in Section D below that OSD sponsorship of many joint programs may well present such a case.

The third objection is really a call for an examination of the "capacity" of the DAB process. The capacity of the DAB in Figure 1 was identified simply as the number of problems in the way a proposed program is structured that the DAB process could resolve. A DAB review for any one MDAP requires a great deal of effort by highly specialized people, access to data, other resources (e.g., travel funds), cooperation across office lines, and support from senior officials—any of which could be a binding constraint. There also are certainly more subtle institutional, legal, and political constraints on the process and the MDA. In practice, the capacity of the DAB process is multi-dimensional and involves the ability to both recognize problematic features of programs and resolve them. This study, however, does not attempt to explore the nature of constraints on the capacity of the DAB process. The question asked is whether that capacity falls seriously short of being able to handle the DAB's program oversight problems, and that is a matter to be put to the data.

Fourth, the DAB process may make mistakes—it may occasionally "fix" something that is not broken or adopt an ineffective fix. As with the second objection, such mistakes would be a problem, in drawing an inference about the capacity of the DAB process, only if they have a differential effect on one category of MDAPs compared to others. This

---

[7] Arguing that some types of programs (joint new starts, for example) are more prone to bad accidents than other program categories does not escape this conclusion. It would be necessary to assert further that the DAB did not notice the pattern or, if it did, that it lacked the tools to ameliorate it. Either of these amounts to evoking a form of constraint on the DAB's ability to deal with problems.

would be a situation in which (repeating the language from the introduction), DoD really does not know how to do a certain category of programs. We are not aware of such a case but cannot rule it out *a priori*.

In much the same vein as these objections, we need to bear in mind that the capacity of the DAB process and the degree of difficulty in the different program categories both may change appreciably with time. This problem is not likely to be serious when comparing two program categories for the same block of years, because the direction of the bias can be gauged.[8] It is much more likely to be serious when comparing a single class of programs in two time periods.

Case-by-case, we need to keep in mind the second objection (other processes) and the fourth (systematic mistakes). Setting those cautions to one side, the validity of an inference from outcomes metrics about relative DAB effectiveness depends on the four propositions embedded in Figure 1:

- Some MDAPs have problems in how they are structured or how they handle issues that arise during execution.
- Problems not resolved at Milestone Reviews or handled expeditiously and effectively during execution are a major—probably the largest—source of poor cost, schedule, and performance outcomes.
- Some categories of MDAPs on average are more problematic than others.
- There are limits on the DAB's ability to resolve problems with MDAPs, both at Milestone Reviews and during execution.

Each of the first three of these propositions is commonplace. Viewed from the perspective of a MS B DAB, among the problems that not infrequently come up are:

- The development schedule is unreasonably optimistic.
- Key technologies are insufficiently mature.
- The cost estimate is unrealistic.
- The contract type or some contract provisions are unsuitable.
- There is too much concurrence between development and production.

A list of problems that not infrequently arise in execution could also readily be offered. It is widely accepted—and there is much evidence in support of this belief—that much of the observed cost growth, schedule slips, and performance shortfalls result from flaws in

---

[8] If the capacity of the DAB is increasing over time more rapidly than is program difficulty, the overall result for a period of years will understate the current effectiveness of the DAB. Conversely, if difficulty is increasing more rapidly, the result for the whole period overstates the current effectiveness of the DAB.

the program approved at MS B or that arise in execution. Finally, it would commonly be accepted that, for example, joint programs are more problematic than single-Service programs. In short, in three of its basic aspects, Figure 1 is a simplified and unvarnished version of the conventional wisdom on links between program characteristics and cost growth. The only new element is the possibility, to be tested in what follows, of a binding constraint on the effectiveness of the DAB process.

## C. Categories, Metrics, and Data

This section provides a brief overview of the categories (of MDAPs), metrics, and data that we used. A more extensive discussion of these topics appears in Appendix A.

### 1. Categories

We began with a list of all of the 309 MDAPs that filed at least one Selected Acquisition Report (SAR) during the years 1970–2012. This list was drawn from the Defense Acquisition Management Information Retrieval (DAMIR) system. We found several problems with this list. For example, a number of programs appeared on the list (at different times) under two or even three names. After correcting for this and other problems with the original list, 291 distinct MDAPs useful for analysis remained.

The study uses two sets of distinctions: (1) single-Service versus joint; and (2) new starts versus programs that develop and procure a variant of an existing system, modify an existing system, or remanufacture existing systems. For brevity, we refer to the latter category as VMR programs. We include in the joint category only programs that were joint during at least a significant portion of the development phase. Apart from that, we used a very broad definition of joint programs. Most of the programs in our joint category are joint programs in the conventional sense that they are undertaken by two or more of the Services. We also included in the joint category programs for which OSD is the sponsor and which involved at least one Service, programs that involved at least one Service and one or more domestic agencies, and programs that involved one or more of the Services and at least one foreign nation.

The number of MDAPs in each of the categories is shown in Table 1.

**Table 1. Number of MDAPs in Each Category**

| | New Start | VMR | Total |
|---|---|---|---|
| Single-Service | 170 | 68 | **238** |
| Joint | 44 | 9 | **53** |
| **Total** | **214** | **77** | **291** |

## 2. Metrics

Table 2 identifies the eight metrics used in this study to gauge the effectiveness of the DAB process in establishing program discipline.

**Table 2. Metrics Used to Gauge Effectiveness of DAB Process**

| | | |
|---|---|---|
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | Cost |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (1997–2012) | Cost |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | Cost |
| 4 | Program Cancellation Rate | Quantity |
| 5 | Percentage of programs cancelled, truncated, or reorganized | Quantity |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity[b] | Quantity |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | Schedule |
| 8 | Percentage of programs with at least one Operational Test failure | Performance |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last SAR for the program divided by the total procurement reported in the first SAR.

The first metric is a measure of cost growth. *Acquisition cost* is defined as the sum of development and procurement cost; Program Acquisition Unit Cost (PAUC) is that sum divided by the quantity purchased. We use PAUC estimated on the assumption that the total number of systems procured is that in the approved MS B baseline.[9] Note that the metric used is not PAUC itself, but the growth in PAUC relative to the PAUC implicit in the approved MS B baseline.

Metrics 2 and 3 provide measures of the percentage of programs that exhibit particularly large growth in PAUC. This is relevant because historically a disproportionately large fraction of total cost growth is accounted for by a relatively small number of programs whose costs showed exceptionally large growth.[10] Both

[9] If the total quantity specified in an MDAP's MS B baseline is procured, computation of the normalized PAUC is just a matter of dividing the acquisition cost by the MS B baseline quantity. In the many cases in which the quantity purchased differs from that specified in the MS B plan, the normalized PAUC used rests on an estimate of what the acquisition costs would have been, had the MS B quantity been procured. IDA did not make these estimates; rather, as is noted further below, they were reflected in the cost growth data we used.

[10] See McNicol, IDA Paper P-3832, 18.

Metrics 2 and 3 rely on one of the Nunn-McCurdy Act's definitions of a "significant" breach in the baseline unit costs. In particular, a significant breach of the PAUC baseline is one in which PAUC has increased by at least 30 percent. Significant breaches must be reported to the Congress. The actions required for a "critical" PAUC breach—growth of 50 percent or more—are considerably more demanding.

Metric 2 is simply the fraction of MDAPs in a group (e.g., joint programs) that incurred at least one Nunn-McCurdy breach (significant or critical). One major problem with this metric is that, prior to the passage of the National Defense Authorization Act of 2006, Nunn-McCurdy reporting was against a program's current baseline, and changes in baselines sometimes delayed or entirely avoided the need to report a breach. Metric 3 ameliorates this problem by doing the comparisons with quantity procured normalized to the MS B baseline—it is the fraction of programs that would have reported at least one Nunn-McCurdy breach, computed on the assumption that the MS B baseline quantity is procured.

The next three metrics characterize the success of MDAPs in terms of quantity rather than cost. Metric 4 is the program cancellation rate; that is, the fraction of some group of programs, all of which had passed MS B, that were cancelled before Low Rate Initial Production (LRIP) was completed. Metric 5 extends metric 4 to include also programs whose procurement is truncated—that is, stopped far short of the quantity anticipated at MS B—or reorganized, which ordinarily involves dropping a major part of the program approved at MS B. Truncation, and often reorganization, of a program is effectively a partial cancellation of it. Adding truncations and reorganizations to cancellations yields a measure of the number of programs that fell well short of expectations for the program at MS B. Metric 6 is simply the percentage of MDAPs in the category in question that procured less than the MS B baseline quantity.

Metric 7 compares the MS C date projected at MS B to the actual MS C date. If, for example, at MS B it was projected that MS C would occur in two years and it actually occurred three years after MS B, metric 7 would have a value of:

$$50 \text{ percent} = \frac{3-2}{2} \times 100 \text{ percent.}$$

This is a very rough measure of the slip in the Engineering and Manufacturing Development (EMD) schedule.

The last of the eight metrics is the percentage of programs in a category that experienced at least one operational test failure. This metric is not in itself a measure of the level of performance eventually attained by the system procured. It is, instead, better understood as a measure of the percentage of the systems in the category that encountered serious trouble in demonstrating that the system is effective and suitable for its intended uses.

## 3. Data

As has already been noted, the data used in this study are described in Appendix A. This section is limited to a brief discussion of our sample sizes and what we can readily say about the uncertainty in each of the metrics used.

The number of MDAPs in each of our four categories for each metric is provided in Table 3. These data reveal two concerns. First, the number of programs in some cells is notably small. In particular, for each of metrics 2 and 6 we have only one MDAP that is a joint VMR program, and only two joint VMR programs for metrics 7 and 8.

**Table 3. Number of MDAPs Used in Computing the Metrics**

| Metric | Limiting Factor on Sample Size | New Start/ VMR | Single-Service | Joint |
|---|---|---|---|---|
| 1 and 3 | Number of MDAPs in the PA&E[a] cost growth database | New Start | 80 | 19 |
| | | VMR | 41 | 6 |
| 2 | Number of MDAPs that filed a SAR at least once for Dec. 1997–Sept. 2012 | New Start | 60 | 28 |
| | | VMR | 31 | 1 |
| 4 and 5 | Number of MDAPs that passed Milestone B 1969–2004 inclusive | New Start | 170 | 44 |
| | | VMR | 68 | 9 |
| 6 | Number of MDAPs in which first and last SARs reported the total quantity to be procured | New Start | 52 | 23 |
| | | VMR | 31 | 1 |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | New Start | 20 | 7 |
| | | VMR | 16 | 0 |
| 8 | Number of MDAPs that went through Operational Test and Evaluation, 1984–2011 | New Start | 49 | 13 |
| | | VMR | 22 | 2 |

[a] Office of Program Analysis and Evaluation

Second, the metrics are not computed using the same set of MDAPs. For example, metric 1 (growth in quantity normalized PAUC) for single-Service new starts is computed for the 80 MDAPs in this category in the cost growth database we used.[11] The data for metric 2 (the fraction of MDAPs incurring at least one significant Nunn-McCurdy breach) is drawn from DAMIR, which provides the relevant information for 60 single-Service new starts. Similar statements hold for the other metrics, except 4 and 5, which are computed for the entire set of MDAPs for the years 1970–2012 inclusive. Given the data available, this is not a problem that can be avoided, but it does make it

[11] We used a database compiled over the course of about 15 years by the Office of Program Analysis and Evaluation (PA&E), now the Office of Cost Assessment and Program Evaluation (CAPE). The database was last updated in 2005 and CAPE no longer maintains it.

imperative to pay attention to the possibility that the sample size is too small to give a good estimate of some of the metrics.

The second source of error is in the numbers that go into the computations. The relevant data are:

- The number of programs that reported at least one Nunn-McCurdy breach during 1997–2012,
- The number of programs with at least one operational test failure during 1984–2011,
- The total program quantity estimate reported in the first and last SARs for various MDAPs, and
- The estimates in the PA&E cost growth database.

The first two of these are matters of record; for example, we know for certain how many MDAPs incurred at least one significant Nunn-McCurdy breach during 1997–2012. Total quantities reported in a final SAR are not always in fact procured, so there is more uncertainty in metric 6, which reflects those data. The most uncertain part of the data is the PA&E cost growth estimates. Based on knowledge of the underlying data and how they were used, we are confident the PA&E estimates are generally reasonable.

Our classification of MDAPs into the four categories is also a potential source of error. These categories seem clear but are in fact not in all cases. For example:

- A program may initially be joint, and then become single-Service when one of the original partners drops out.
- Some "variants" of a previous system are effectively complete redesigns, that is, a new start.

Appendix A provides a more detailed discussion of the issues the study encountered in classifying MDAPs into these categories and identifies the classification rules we used. In broad terms, the classifications were done "by hand" using the best information we could acquire within the time and resources available for this study. In most cases we felt confident in our classifications. There were, however, some cases in which we had to make a judgment based on incomplete information or in which the program did not fit neatly into our categories; e.g., the program was clearly joint during part of its history and clearly single-Service at others. Our judgment is that, while we very probably do have some errors in our classifications, they are reasonably reliable.

## D. Single-Service and Joint New Starts

We start with a comparison of single-Service new starts and joint new starts. About 60 percent (170 of 291) of MDAPs in our set of programs are single-Service new starts.

The sample also includes 44 joint new starts. The values of the metrics for these two categories are presented in Table 4.

**Table 4. Values of the Metrics for Single Service and Joint New Starts**

| | Metric | Single-Service New Starts | Joint New Starts |
|---|---|---|---|
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | 35% (80) | 42% ( 19) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | 20% (12 of 60) | 32% (9 of 28) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | **40%** (32 of 80) | **63%** (12 of 19) |
| 4 | Program Cancellation Rate | **19%** (33 of 170) | **34%** (15 of 44) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | **28%** (47 of 170) | **57%** (25 of 44) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity[b] | 37% (19 of 52) | 48% (11 of 23) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | **17%** (20) | **36%** (7) |
| 8 | Percentage of programs with at least one Operational Test failure | **47%** (23 of 49) | **77%** (10 of 13) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last SAR for the program divided by the total procurement reported in the first SAR.

Single-Service new starts are usually what people have in mind when they speak in general terms of the DoD acquisition program. To our knowledge, no previous study has provided such a complete set of program output metrics for single-Service new starts (or any other large set of MDAPs). In this respect, the data in Table 4 are novel.

The values of the metrics, however, by and large will not surprise anyone who has watched the acquisition process and has some familiarity with the literature. The average growth in PAUC (with the caveats noted in the table) for single-Service new starts is 35 percent, and about one-fifth of single-Service new starts incur at least one significant Nunn-McCurdy breach. About one single-Service new start in five is cancelled, 37 percent eventually procure less than the quantity planned at MS B, and nearly half initially fail at least one operational test. These metrics could with equal validity be read from a "glass half full" perspective: 60 percent of single-Service new start MDAPs show unit cost growth of less than 30 percent (metric 3), are not cancelled, do not incur a Nunn-McCurdy breach, and—although a significant number initially encounter a

problem with operational tests—most go on to procure at least the quantity in the MS B baseline.

Involving two or more Services and/or other agencies as full partners in a program is problematic in itself because it requires maintaining a workable agreement among participants who place different weights on various aspects of the system's performance, as well as on cost and schedule. Joint programs also may not compete very effectively at the Service level for funding, personnel, and access to program support activities. Similarly, a single Service or Agency may provide more effective oversight of ongoing major programs than does a partnership of two or more Services/Agencies. Accordingly, we expect joint new starts to perform worse than single-Service new starts on all of our metrics, and they do.

Some joint programs are initiated by the Services or Agencies involved and are voluntary partnerships. Other joint programs are in effect initiated and sponsored by OSD, or the White House or the Congress, although the program may be managed by one of the participating Services as an executive agent or may have a joint program office in which each of the participating Services is engaged. The difference between these two subcategories is potentially significant because the DAB process is not well structured to provide oversight of programs that originate outside the usual acquisition process. Unfortunately, we do not have a readily available marker of the joint programs that are Service-initiated.

Joint new starts are less successful than single-Service new starts by each of the eight metrics. Whether the differences between joint and single-Service new starts (shown in Table 4) are large enough to show the existence of limits on the effectiveness of the DAB process is a matter of judgment. The striking differences (highlighted in the table) are the percentage of programs with an estimated Quantity Normalized PAUC growth of ≥30 percent (metric 3); the cancellation rate (metric 4); the rate of cancellation, truncation, or reorganization (metric 5); the slip in the EMD schedule (metric 7); and the proportion of programs that had at least one operational test failure (metric 8). The value of each of these metrics is at least half again as large for joint program new starts as for single-Service new starts. Moreover, looking ahead, a difference of roughly this magnitude appears in each sub-period we consider below.

We believe these data do clearly show that there is a binding limit on the effectiveness of the DAB process. Figure 2 is a graph of the values for metric 5, the percentage of programs cancelled, truncated, or reorganized. When over a period of years, about 57 percent of joint programs and 28 percent of single-Service new starts are cancelled, truncated, or reorganized, it is difficult to see how this conclusion could be avoided.

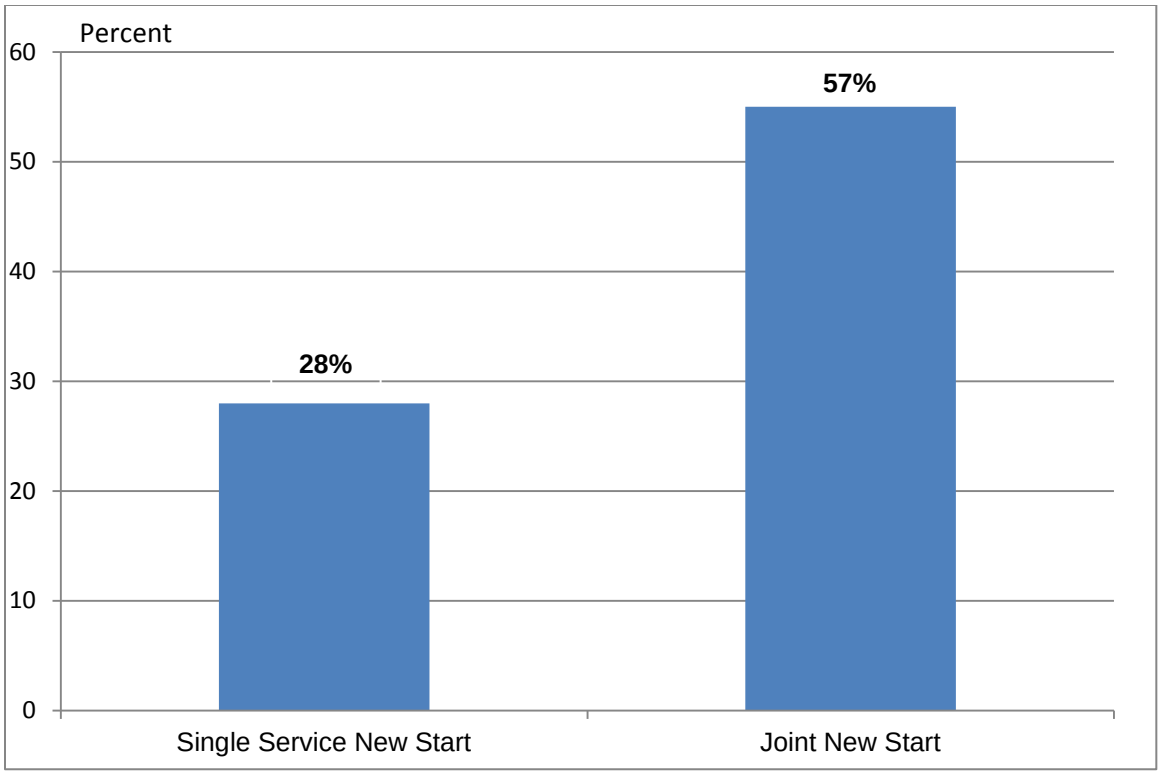


**Figure 2. Percentages of Programs Cancelled, Truncated, or Reorganized**

## E. Single-Service VMR Programs in Two Periods

As was noted earlier, the VMR category includes programs that (1) acquire a variant of an existing system, (2) modify a fielded system, or (3) remanufacture some or all of the units of a system that has been fielded. These sorts of programs usually are not simple on an absolute scale. Our sample includes a total of 68 single-Service VMR programs. Of these, 36 were initiated during the years 1970–1988, and 32 were initiated post-1988. All of the VMRs in our sample are MDAPs and most involve a considerable amount of development. However, the programs in the VMR category tend to be less ambitious than new starts.

Table 5 reports the metrics for the single-Service VMRs and repeats from Table 4 the value of the metrics for single-Service and joint new starts.

**Table 5. Values of the Metrics for the Single Service VMRs, Single Service New Starts, and Joint New Starts**

| | Metric | Single-Service VMRs | Single-Service New Starts | Joint New Starts |
|---|---|---|---|---|
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | 27% (41) | 35% (80) | 42% (19) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | 13% (4 of 31) | 20% (12 of 60) | 32% (9 of 28) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | 37% (15 of 41) | 40% (32 of 80) | 63% (12 of 19) |
| 4 | Program Cancellation Rate | 12% (8 of 68) | 19% (33 of 170) | 34% (15 of 44) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | 16% (11 of 68) | 28% (47 of 170) | 57% (25 of 44) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity [b] | 42% (13 of 31) | 37% (19 of 52) | 48% (11 of 23) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | 26% (16) | 17% (20) | 36% (7) |
| 8 | Percentage of programs with at least one Operational Test failure | 30% (7 of 23) | 47% (23 of 49) | 77% (10 of 13) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last SAR for the program divided by the total procurement reported in the first SAR.

Most of the metrics fall into the pattern that would be expected; that is, for most metrics, single-Service VMRs fare better than single-Service new starts, and single-Service new starts do better than joint new starts. There are two exceptions in the table (highlighted):

- Metric 6: Proportionately fewer single-Service VMR programs eventually procure at least the MS B baseline total quantity than do single-Service new starts; and
- Metric 7: VMRs on average had longer stretches in the EMD phase.

A plausible explanation for the first of these observations is that systems already fielded are better substitutes for the VMR programs than they are for the new starts. This consideration, on close examination, also leads to an explanation of why the average slip in EMD is larger for single-Service VMR programs than it is for single-Service new starts.

The larger slip in EMD for VMR programs leads us to consider the possibility that around the end of the Cold War, the Services' incentives for undertaking a VMR program changed because of changes in DoD funding and a related change in acquisition policy. DoD procurement funding, which peaked in 1986, fell to just under half its 1989 level by 1995 (in constant dollars).[12] With the decrease in the overall DoD budget, and in particular the sharp drop in funds allocated to acquisition, major single-Service new starts were discouraged. Instead, Components, especially the Services, found it much less difficult to gain approval in the Program/Budget processes for comparatively less ambitious VMR programs.

Table 6 presents the values of the metrics for the three program groups for two periods: 1970–1988 and post-1988. Appendix B is a rearrangement of Table 6 in which the top level grouping of the metrics is by period rather than by program category.

Looking at the two columns on the left of Table 6, we clearly see a substantial change between the two periods in the outcomes for single-Service VMR programs. By two cost growth metrics (1 and 3) and the performance on operational testing (metric 8), VMR programs did worse by a factor of more than three post-1988 than they did during 1970–1988. They did much better at buying at least the MS B baseline quantity (metric 6), again by about a factor of three. Note, however, that we have quantity data for only three single-Service VMRs for the earlier period; we have no schedule data for the earlier period (metric 7). A reasonable interpretation of these results is that during 1989–2001, after obtaining MS B approval, the Services tended to sharply move VMR programs towards the greater capability that they preferred, and which a new start would provide.[13]

---

[12] Office of the Under Secretary of Defense (Comptroller), National Defense Budget Estimates for FY 2012, March 2011, Table 6–8, 126–7.

[13] Further discussion of this point can be found in McNicol, IDA Paper P-3832., 34–35.

**Table 6. Value of the Metrics for Single-Service VMRs, Single-Service New Starts, and Joint New Starts for 1970–1988 and Post-1988**

| | Metric | Single-Service VMR | | Single-Service New Starts | | Joint New Starts | |
|---|---|---|---|---|---|---|---|
| | | 1970–1988 | 1989– | 1970–1988 | 1989– | 1970–1988 | 1989– |
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | 12% (22) | 45% (19) | 34% (60) | 38% (20) | 52% (8) | 35% (11) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | N/A (0 of 3) | 14% (4 of 28) | N/A (0 of 18) | 31% (13 of 42) | N/A (0 of 3) | 36% (9 of 25) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | 18% (4 of 22) | 58% (11 of 19) | 38% (23 of 60) | 45% (9 of 20) | 63% (5 of 8) | 64% (7 of 11) |
| 4 | Program Cancellation Rate | 11% (4 of 36) | 13% (4 of 32) | 20% (23 of 117) | 19% (10 of 53) | 47% (7 of 15) | 25% (7 of 28) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | 14% (5 of 36) | 19% (6 of 32) | 26% (30 of 117) | 32% (17 of 53) | 47% (7 of 15) | 61% (17 of 28) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity[b] | 100% (3 of 3) | 36% (10 of 28) | 41% (7 of 17) | 34% (12 of 35) | 50% (2 of 4) | 47% (9 of 19) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | N/A | 26% (16) | 3% (4) | 20% (16) | 0% (1) | 42% (6) |
| 8 | Percentage of programs with at least one Operational Test failure | 13% (1 of 8) | 40% (6 of 15) | 43% (13 of 30) | 53% (10 of 19) | 100% (4 of 4) | 67% (6 of 9) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last SAR for the program divided by the total procurement reported in the first SAR.

Does this (as we have suggested) reflect a change in the Services' incentives? Or does it reflect a change in the rigor of the OSD process for oversight of MDAPs? The second possibility cannot be dismissed out of hand for two reasons. First, in 1988, with the enactment of statutory changes recommended by the Packard Commission, the earlier Defense Systems Acquisition Review Council (DSARC) was replaced by the DAB.[14] Second, during the years 1993–2000, concerted attempts were made to reform the acquisition process, and this in part was understood to mean reducing the degree of intrusiveness of OSD oversight of major acquisition programs.

Any major change in the rigor of the OSD review process presumably would be revealed not just in the outcomes for VMR programs, but also in the outcomes for single-Service new starts and joint new starts. This is not what we see in the data. Looking back at Table 6, there are no large differences in the values of the metrics for single-Service new starts between the two periods. Moreover, performance was slightly better on two of the eight metrics (4 and 6) post-1988. Joint new starts did much worse on Nunn-McCurdy breaches (metric 2) and slips in EMD (metric 7) post-1988, but somewhat better on four metrics (1, 4, 6, and 8). We conclude, then, that what we are seeing in the metrics for single-Service VMRs for the two periods is a shift in the Services' intentions for this category of programs.

The metrics indicate that post-1988, after gaining MS B approval, the scopes of many VMRs were expanded, and they became more like new starts. This explains not only the VMR metrics that were worse post-1988, but the one (metric 6) that was better. (If the VMR became more like a new start, it would compete better with the relevant fielded system, which suggests an improvement in metric 6.) The DAB process evidently was not effective in fully picking up on these changes and responding to them.

## F. What Can We Say About Where the DAB Process Falls Short?

An important conclusion on resources and resource allocation is implicit in Figure 1 and Figure 2. These two charts suggest that the DAB's resources should be increased or might be better allocated to milestone reviews of the more challenging types of programs. Action on this observation probably would not be simply a matter of hiring more people or assigning more or more experienced people to the more challenging types of programs. Rather, the observation suggests careful thought about when and how DAB resources are applied to the entire sequence of activities from pre-MS A through MS C.

The data considered above also provide a hint regarding one respect in which the DAB process falls short in ways that can lead to unsatisfactory cost, schedule, or

---

[14] *A Quest for Excellence: Final Report of the President's Blue Ribbon Commission on Defense Management*.

performance outcomes. This hint comes from asking: What feature(s) do joint new starts for the entire sample and single-Service VMRs launched after 1988 have that single-Service new starts and single-Service VMRs from 1970–1988 do not? It seems unlikely that any such feature or features could be found in program content. In fact, in terms of program content, the major distinction is between new starts and VMRs. It also seems unlikely that the discriminating feature could be found in acquisition strategy, management arrangements, and contract types.

A feature most joint new starts and post-1988 single-Service VMR programs have in common is guidance about which the relevant Services were not enthusiastic. Single-Service new starts also may receive guidance from a DAB milestone review that they would prefer not to have. The issue is then one of degree. We believe that it is a reasonable judgment that joint new starts and post-1988 single-Service VMRs operated far more often than single-Service new starts on guidance that was of major importance to the specification of the program's content. If this is granted, the comparison of the metrics for single-Service and joint new starts and single-Service VMRs in the pre- and post-Cold War periods points to limits of the OSD acquisition oversight process in enforcing programmatic guidance.

Milestone reviews—especially the MS B review—and the periods leading up to them are a major opportunity for ensuring that the program approved after the DAB review meets OSD guidance. For joint programs, this would include attention to two "lessons learned" from past programs:

- Be sure that the program management structure and the reporting chains are satisfactory to each of the Services involved and to the MDA, and
- Achieve a consensus among the participating Services or Agencies and the MDA on requirements.

For post-1988 VMRs, the scope of the program and the associated cost estimates would have been carefully examined.

A major program, of course, does not just unfold in a predestined way after it has received MS B authority. There inevitably are unexpected technical difficulties, manufacturing problems, needs to make budget adjustments, test issues, and issues that arise from related programs, among others. Responding to these events requires decisions by the program manager and, in major cases, the people to whom the program manager reports. The response often is a change in program content and, absent guidance from the MDA or people acting on the MDA's authority, the Service or Services involved will make these decisions in accordance with their own doctrines and preferences.

As was noted in Section B above (p. 4, footnote 5), the DAES process exists to keep the USD(AT&L) apprised of the status of MDAPs, and the USD(AT&L) can intervene in

a program between milestone reviews. These reviews generally do not happen on a regular schedule, but only when a need for one has been recognized, which all too often, is after a major problem has arisen.

We find no evidence that most joint new starts or post-1988 single-Service VMR programs received any continuing special attention from the DAB process post MS B. Rather, most apparently received only episodic DAB oversight outside of milestone reviews. Evidently, the DAB process was not constituted to recognize that continuing oversight is needed when the Services or Agencies involved in a program receive guidance that runs contrary to their preferences, understood broadly to include their organizational views of missions, strategies, risks, costs, and other considerations that go into the decisions that shape an acquisition program.

## G. Concluding Observation

This paper has compared two acquisition categories in terms of output metrics to identify situations in which the OSD acquisition oversight process has notably been less than fully successful:

- Joint new starts; and
- Single-Service VMRs initiated post-1988.

We have shown how these shortfalls reveal significant constraints on the capacity of the DAB to deal in a fully successfully way with the problems presented by joint new starts and post-1988 VMRs. This is the principal conclusion of this study. It is significant because it directs attention to the questions of just what constrains the DAB process and how tight those constraints are

Constraints on the DAB's capacity may render particular reform efforts ineffective or even unworkable. Nevertheless, this study does point to one direction for improvement. We have attributed the DAB's shortfalls in the cases examined to the intermittent and apparently reactive character of the DAB's oversight between milestone reviews. The core argument is simple: The DAB did not provide post-milestone oversight in precisely those circumstances in which it is most needed—when one or more of the Services receive guidance to do something other than what they would prefer.

This way of stating the suggestion focuses on post-milestone review activity by the DAB. The predicate of the suggestion, however, is that the Acquisition Decision Memorandum issued by the MDA after the milestone review contains clear and executable guidance. Without that, continuing oversight by the DAB is unlikely to be successful. The suggestion is then as much about the Milestone Review Process as it is about post-milestone tracking.

# Appendix A.
# Data Sources and Methodology

Section C of the main body of this paper provides an overview of the categories, metrics, and data used in this study. While not repeating the earlier material, this appendix provides some additional details.

## Population of Major Defense Acquisition Programs (MDAPs)

We started with a list of 309 programs that had filed Selected Acquisition Reports (SARs) between 1969 and 2004. This list was compiled by the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics (OUSD(AT&L)). We believe that this list is complete.

We removed from this list Major Automated Information System programs, programs that never reached Milestone (MS) B,[1] and duplicate listings. We also recombined the data for (and treated as a single MDAP under the original name) several programs that had been split into two programs. Similarly, two programs that began as separate programs and then were combined were treated as a single program throughout. We arrived at a final list of 291 MDAPs.

## Program Categories

We classified each of the 291 MDAPs as:

- Single-Service or joint; and
- New starts or variant, modification, or remanufacturing (VMR) of a system already fielded.

Single-Service new starts (the largest of the categories) did not present any substantial definitional issues. There were, however, issues with defining joint programs and VMRs.

### Joint Programs

The Department of Defense (DoD) defines joint programs as those having multi-Service or multi-Agency participation during the research and development (R&D) phase

[1] In 1998, the milestones used in the DAB process were designated A, B, and C. These correspond roughly to Milestones I, II, and III of earlier versions of the OSD process for oversight of MDAPs. We use only the post-1998 designations of the main milestones.

and/or during the procurement phase. We used a more restrictive definition, classifying as joint only programs that had significant multi-organization participation during the R&D phase. We did not classify as joint programs that had multi-organization participation only during the procurement phase.

We also classified as joint:

- Programs in which at least one of the participating parties is an international organization or a foreign nation; and
- A few programs initiated by organizations or authorities outside of DoD, such as the president or the Congress, and with crucial elements not congruent with what the Service would have preferred.

**New Start or VMR**

As noted above, a VMR program is one that develops and procures a variant or a modification of an existing system, or remanufactures an existing system. The line between a new start and a VMR is a matter of degree. Some new starts make only modest advances over the fielded system they will replace and some programs that are nominally intended to procure a variant effectively are a new start. We had no easy and reliable way to distinguish new starts from VMRs. Our approach was to make a trial classification based on the program name and, if there seemed to be reason for doubt, to consult the SAR for the program. Since the SAR described the scope of the program, that was almost always sufficient.

## Program Outcomes

Each program in our database was labeled as "successful," "cancelled," "truncated," or "reorganized." A successful program was one that had not been cancelled, truncated, or reorganized. The definitions we used are as follows:

Cancelled programs: Programs were considered cancelled if they were terminated before the end of Low Rate Initial Production (LRIP). Typically, cancelled programs have not begun to procure units to be fielded, though they may have procured a few prototype units.

Truncated programs: Programs were categorized as truncated if they were terminated after the end of LRIP; but did not procure a substantial number of units. As a rule of thumb, we would classify a program as truncated only if it procured less than half the total quantity specified in the MS B baseline.

Reorganized programs: Programs were classified as reorganized if they underwent a significant restructuring to reduce program risks. Restructuring may involve a large part of a program being dropped.

We consulted a variety of resources in applying these definitions and guidelines. The Defense Acquisition Management Information Retrieval (DAMIR) system provided the Executive Summaries in the SARs for programs filing SARs in 1997 and subsequent years. Basic program data for earlier years is readily available in *Jane's Yearbooks* made available by Jane's Information Group. We also used Government Accountability Office reports, queries to individuals or organizations with knowledge of the relevant programs, and Internet searches.

## Metrics

The paper uses eight metrics to measure differences in cost, schedule, and performance outcomes. This section reviews the data sources and methodology behind each of the eight metrics.

### Metric 1: Estimated Growth in Quantity Normalized Program Acquisition Unit Cost

Data: The source of all the cost growth data used in this study is the Office of Program Analysis and Evaluation (PA&E). The PA&E cost growth database was updated approximately annually and made available to qualified users in the defense community through about 2005.[2] We are using the most recent version of the database, which contains cost growth data for 160 programs that filed SARs between 1976 and 2004. The cost growth data include MS B estimates and cost variances for Research, Development, Test, and Evaluation (RDT&E), Procurement, and Military Construction. The variances are stated in CY 2000 dollars.

Methodology: Metric 1 is a simple average, across programs, of growth in Program Acquisition Unit Cost (PAUC). PAUC is the sum of RDT&E and Procurement cost divided by the number of units procured. We use "quantity normalized" PAUC; that is, PAUC computed on the assumption that the quantity procured is that specified in the MS B baseline.

### Metric 2: Percentage of Programs with at Least One Nunn-McCurdy Breach

Data: Data on Nunn-McCurdy breaches come from DAMIR, and only breaches against the original baseline are considered. The data are for programs filing SARs between 1997 and 2012. When the data were merged into the base dataset, the result was Nunn-McCurdy breach data for 120 programs.

---

[2] The Weapon Systems Acquisition Reform Act of 2009, among many other things, renamed PA&E the Office of Cost Assessment and Program Evaluation (CAPE). CAPE does not maintain or provide to others the PA&E cost growth database.

Methodology: Metric 2 is the number of programs within a category that experience at least one significant or critical Nunn-McCurdy breach against the original MS B baseline divided by the total number of programs in that category.

**Metric 3: Percentage of Programs with an Estimated Quantity Normalized PAUC Growth of ≥ 30%**

Data: Same as for Metric 1.

Methodology: Metric 3 is a count of the programs with RDT&E and Procurement cost growth equal to or greater than 30 percent stated as a percent of the total number of programs in the category for which cost growth data were available.

**Metric 4: Program Cancellation Rate**

Data: As described in Table 4, item 4.

Methodology: Metric 4 is simply the percentage of programs in a category that were cancelled.

**Metric 5: Percentage of Programs Cancelled, Truncated, or Reorganized**

Data: As described in Table 4, item 5.

Methodology: Metric 5 is simply the percentage of cancelled, truncated and reorganized programs in a category.

**Metric 6: Percent of Programs that Eventually Procure Less Than the Milestone B Total Quantity**

Data: The source of the quantities data is DAMIR. The data are for programs filing SARs between 1997 and 2012. When the data were merged into the base dataset, the result was data for 107 programs. The data consist of current quantities estimates and SAR baseline quantities estimates for RDT&E and Procurement.

Methodology: Metric 6 is the percentage of programs in the category that procured less than the quantity specified in their SAR MS B baseline estimates.

**Metric 7: Estimated Slip in Engineering and Manufacturing Development**

Data: The source of the schedule data is DAMIR. The data are for programs filing SARs between 1997 and 2012. When the data were merged into the base dataset, the result was a sample of 43 programs. The data consist of the SAR's Development Estimate MS B date (actual), the SAR's Development Estimate MS C date (projected), and either the SAR's Production Estimate MS C date or the most current SAR's MS C

date (actual). The small sample size is due to missing data; many programs only had one milestone date, making it impossible to calculate a schedule slip.

Methodology: Metric 7 is calculated as the difference between the actual duration of EMD and the planned duration of EMD divided by the planned duration of EMD. Actual EMD is calculated as actual MS C minus actual MS B. Planned EMD is calculated as projected MS C minus actual MS B.

$$\text{metric 7} = (\text{Actual EMD} - \text{Planned EMD}) / \text{Planned EMD}$$

**Metric 8: Percentage of Programs with At Least One Operational Test Failure**

Data: The Operational Test & Evaluation data come from the Office of Operational Test and Evaluation. The data are for programs that underwent testing between 1984 and 2011. When the data were merged into the base dataset, the result was a sample of 88 programs. Programs must have an appropriate test plan and then be found, on testing in accordance with the plan, to be operationally effective and suitable.

Methodology: Metric 8 is simply the percentage of programs within the category that did not have an appropriate test plan, or which were found on testing not to be operationally effective and/or suitable.

## Program Lists by Category

Table A-1 through Table A-4 list the programs by category and indicate those that were cancelled, truncated, or reorganized. Additional data is available on the CD provided with this paper. The tables on the CD, numbered CD-1 through CD-4 correspond with Table A-1 through Table A-4, respectively.

**Table A-1. Single-Service MDAP New Starts**

Cancelled Truncated Reorganized

| Single-Service | |
|---|---|
| New Start | |
| Program | Service |
| 5-INCH GUIDED PROJ | Navy |
| 8-INCH GUIDED PROJ | Navy |
| ABL | Air Force |
| ASAT | Air Force |
| ASDS | Navy |
| ATCCS-CHS | Army |
| B-1 | Air Force |
| CHEYENNE | Army |
| CONDOR | Navy |
| CRUSADER | Army |
| DD(X) | Navy |
| DE 1052 | Navy |
| DIVAD (SGT York) | Army |
| FAADS LOS-F-H ADATS | Army |
| FAADS NLOS | Army |
| FCS | Army |
| Follow-on Early Warning System (FEWS) | Air Force |
| HFAJ | Navy |
| LAND WARRIOR | Army |
| LGM-118A Peacekeeper Rail Garrison | Air Force |
| LLLBGK | Air Force |
| M712 CLGP (Cannon Launched) Copperhead | Army |
| PLSS | Air Force |
| RAH-66 Comanche | Army |
| RPV Aquila | Army |
| SADARM Rocket | Army |
| SCOUT | Army |
| SLAT (AQM-127A) | Navy |
| SMALL ICBM | Air Force |

| Single-Service | |
|---|---|
| New Start | |
| Program | Service |
| T-46A Eaglet Trainer | Air Force |
| TSAT | Air Force |
| USQ-84(V) SOTAS (Target Acquisition Sys) | Army |
| XM803 TANK | Army |
| ACM | Air Force |
| ALQ-165 ASPJ (Jammer) | Navy |
| AN/BSY-2 | Navy |
| B-2A | Air Force |
| C-5A | Air Force |
| CGN-38 | Navy |
| EMSP | Navy |
| GLOBAL HAWK | Air Force |
| ROTHR | Navy |
| SAFEGUARD | Army |
| SEA LANCE | Navy |
| STINGRAY | Army |
| GPALS (SDS) | DoD |
| WIN T | Army |
| A-10 Thunderbolt | Air Force |
| ADDS EPLRS (Enhanced Pst Location Rpt Sys) | Army |
| AEHF (Adv Extremely High Freq) Satellite | Air Force |
| AESA | Navy |
| AFATDS (Adv Field Artillery Tact Data Sys) | Army |
| AGM-114 Hellfire | Army |
| AGM-158 JASSM | Air Force |
| AGM-84A Harpoon | Navy |
| AGM-86B ALCM | Air Force |
| AH-64 Apache | Army |
| AN/BQQ-5 | Navy |
| AN/BSY-1 | Navy |
| AN/SQQ-89 | Navy |
| AOE 6 | Navy |
| ARC-210 SINCGARS Radio | Army |
| ASAS BLK II/III | Army |
| ATACMS Blk I (APAM) /Blk II/IIA | Army |
| ATCCS CSSCS | Army |
| ATCCS FAAD C2I | Army |
| ATCCS MCS (Maneuver Ctrl Sys) | Army |
| C-17A Globemaster | Air Force |
| CAPTOR (MK 60 MINE) | Navy |

| Single-Service | |
|---|---|
| New Start | |
| Program | Service |
| CEC (Coop Engagement Capability) | Navy |
| CG 47 Aegis Cruiser | Navy |
| CH-47 Chinook | Army |
| CH-53 Super Stallion & MH-53 Sea Dragon | Navy |
| CIS (MK XV IFF) | Air Force |
| CMU | Air Force |
| COBRA JUDY REPLACEMENT | Navy |
| CSRL (Rotary Launcher) | Air Force |
| CVN 21 | Navy |
| CVN 68 CLASS (Nimitz CVN) | Navy |
| DD 963 | Navy |
| DDG-51 Burke | Navy |
| DMSP (Meteorological Satellite) | Air Force |
| DSP | Air Force |
| EELV (Atlas V & Delta IV) | Air Force |
| EFV (Formerly AAAV) | Navy |
| F/A-18 E/F Super Hornet | Navy |
| F/A-18 Hornet | Navy |
| F-111A/D/E/F | Air Force |
| F-14A | Navy |
| F-14D Tomcat | Navy |
| F-15 Eagle | Air Force |
| F-16 Falcon | Air Force |
| F-22 ATF | Air Force |
| FAADS LOS-R Avenger | Army |
| FBCB2 | Army |
| FDS | Navy |
| FFG-7 | Navy |
| FIM-92 Stinger Missile | Army |
| FMTV (Family Med Tact Vehicles) | Army |
| FPS-118 OTH-B (Radar) | Air Force |
| GBS (Global Broadcast Service) | Air Force |
| GMLRS (Guided Multi Launch Rocket Sys) | Army |
| GPS NAVSTAR | Air Force |
| HAWK IMPROVED | Army |
| HIMARS | Army |
| IUS (Inertial Upper Stage) | Air Force |
| KC-10A | Air Force |
| KIOWA WARRIOR | Army |
| LANCE | Army |

| Single-Service | |
|---|---|
| New Start | |
| Program | Service |
| LANTIRN (Low Alt Nav & Targeting Sys) | Air Force |
| LCAC (Landing Craft Air Cushion) | Navy |
| LCS | Navy |
| LGM-118A Peacekeeper | Air Force |
| LHA | Navy |
| LHD 1 Amphibious Assault Ship | Navy |
| LPD 17 Amphibious Transport Dock Ship | Navy |
| LSD 41 Whidbey Island | Navy |
| M1 Abrams Tank | Army |
| M198 155MM Howitzer | Army |
| M2/M3 Bradley FVS | Army |
| M26 MLRS (Mult Launch Rocket Sys) | Army |
| M47 Dragon Guided Missile | Army |
| M551 (AR/AAV) | Army |
| MCM 1 | Navy |
| MILSTAR | Air Force |
| MIM-104 Patriot Guided Missile System | Army |
| MK 48 ADCAP | Navy |
| MK 48 TORPEDO | Navy |
| MK 50 TORPEDO | Navy |
| MLRS-TGW | Army |
| MMA | Navy |
| MP RTIP | Air Force |
| MSE (Mobile Subscriber Equipment) | Army |
| MUOS | Navy |
| NESP | Navy |
| NTW | Navy |
| PHALANX CIWS (MK-15) | Navy |
| PLS FHTV (Palletized Load System) | Army |
| POSEIDON | Navy |
| ROLAND | Army |
| S-3A | DoD |
| SBIRS (Space Based IR Sensor) High | Air Force |
| SDB | Air Force |
| SHILLELAGH MISSILE | Army |
| SM-6 | Navy |
| SMART-T (Secure Mobile Terminal) | Army |
| SPARROW (AIM-7F) (AF) | Air Force |
| SQR-19 TACTAS | Navy |
| SRAM | Air Force |

| Single-Service | |
|---|---|
| New Start | |
| Program | Service |
| SSN 637 | Navy |
| SSN 688 Los Angeles | Navy |
| SSN 774 Virginia Class New Attack Sub | Navy |
| STINGER RMP | Army |
| SURTASS | Navy |
| T-45 Goshawk Training System | Navy |
| TACFIRE | Army |
| T-AGOS | Navy |
| T-AKE | Navy |
| T-AKR 295 Strategic Sealift | Navy |
| T-AO 187 Oiler | Navy |
| Titan IV ELV (Expend Launch Veh) | Air Force |
| TOW | Army |
| TRIDENT SUB | Navy |
| TRI-TAC | Air Force |
| TTC-39 Nodal Comm Switch | Army |
| UGM-133A Trident II Missile | Navy |
| UH-60A Blackhawk | Army |
| UHF Follow-On Comm Satellite Sys | Navy |
| VAST | Navy |
| WGS (Wideband Gapfiller Sat) | Air Force |

**Table A-2. Joint MDAP New Starts**

Cancelled Truncated Reorganized

| Joint | |
|---|---|
| New Start | |
| Program | Service* |
| ACS | Army |
| ATARS | Air Force |
| D-HLH | DoD |
| I-S/A AMPE | Air Force |
| JOINT COMMON MISSILE | Army |
| JSIMS | Army |
| JTUAV Short Range Hunter | Joint |
| MEADS | Army |
| NATO AAWS | Navy |
| NATO PHM Pegasus Class | Navy |
| NAVY Area TBMD | Navy |
| NPOESS | Air Force |
| TACIT RAINBOW | Air Force |
| THAAD | MDA |
| WAM (WWMCCS/WIS) | DoD |
| EXCALIBUR | Army |
| SCAMP (BLOCK II) | Army |
| ATIRCM/CMWS | Army |
| BMDS | DoD |
| CHEM DEMIL (LEGACY) | Army |
| JTRS CLUSTER 1 | Army |
| JTRS CLUSTER 5 | Army |
| JTRS WAVEFORM | DoD |
| NMD | DoD |
| T-6A JPATS (Jt Prmy AC Training Sys) | Air Force |
| AGM-88 HARM USAF/USN | Air Force/Navy |
| AMRAAM USN/USAF | Air Force |
| CHEM DEMIL | Army |
| CHEM DEMIL-ACWA | Army |
| CHEM DEMIL-CMA NEWPORT | Army |
| FGM-148A Javeline AAWS-M | Army |
| JDAM (Jt Direct Attack Munition) | Joint |
| JSF (F-35) | Air Force |
| JSOW | Navy |
| JSTARS GSM | Army |
| JSTARS USAF | Air Force |

| Joint | |
|---|---|
| New Start | |
| Program | Service* |
| JTIDS (USAF/US Army) | Air Force |
| MAVERICK (LASER) | Air Force |
| MIDS LVT (Low Vol Terminal) | Navy |
| MIM-104 Patriot PAC-3 (Pat Adv Capability) | Army |
| NAS (National Airspace System) | Air Force |
| SYQ-23 JSIPS (Jt Ser Imagery Proc Sys) | Navy |
| TRN-45 MMLS Ground Components | Air Force |
| V-22 Osprey | Navy |

*The Service indicated is the lead Service.

**Table A-3. Single-Service MDAP VMRs**

Cancelled Truncated

| Single-Service VMR | |
|---|---|
| **Program** | **Service** |
| A-6E/F Intruder | Navy |
| AGM-131A SRAM II (Short Range Msl) | Air Force |
| ATACMS P3I (BAT) | Army |
| AV-8A | Navy |
| B-1 CMUP-DSUP | Air Force |
| C-130 AMP | Air Force |
| HHD-60 | DoD |
| P-7A (LRAACA) | Navy |
| C-130H | Air Force |
| FB-111A | Air Force |
| PATRIOT P31 | DoD |
| A-7D | Air Force |
| A-7E | Navy |
| AEGIS MK-7 | Navy |
| AGM-114K Hellfire Longbow | Army |
| AH-64D Apache Airframe | Army |
| AIM-54C Phoenix Missile | Navy |
| AIM-9X Sidewinder | Navy |
| AV-8B Harrier | Navy |
| AV-8B Harrier Remanufacture | Navy |
| B-1 CMUP-COMPUTER UPGRADE | Air Force |
| B-1 CMUP-JDAM | Air Force |
| B-1B CMUP (Conventional Mission Upgrade Prgm) | Air Force |
| B-1B Lancer | Air Force |
| B-2 RMP | Air Force |
| B-52 OAS/CMI MODS | Air Force |
| BATTLESHIP REACT | Navy |
| BGM-109G Tomahawk GLCM | Air Force |
| C-130J Hercules | Air Force |
| C-5 RERP | Air Force |
| C-5B Galaxy | Air Force |
| CBU-97B SFW (Sensor Fuzed Weapon) | Air Force |
| CH-47F (Improved Cargo Helio) | Army |
| E-2C | Navy |
| E-2C REPRODUCTION | Navy |
| E-2D ADVANCED HAWKEYE | Navy |

| Single-Service | |
|---|---|
| VMR | |
| Program | Service |
| E-3A Sentry AWACS | Air Force |
| E-4 AABNCP NEACP | Air Force |
| E-6A TACAMO | Navy |
| EA-18G | Navy |
| EA-6B Prowler Adv CAP | Navy |
| EF-111A TJS | Air Force |
| F-5E Tiger | Air Force |
| H-1 UPGRADES | Navy |
| KC-135R Stratotanker | Air Force |
| LGM-30 Minuteman III GRP | Air Force |
| LGM-30 Minuteman III PRP | Air Force |
| LSD 41 Whidbey Island Cargo Variant | Navy |
| M1A2 Abrams Tank Upgrade | Army |
| M2/M3 Bradley FVS Upgrade | Army |
| M6042 TANK | Army |
| MAVERICK (TV) | Air Force |
| MH-60R Strikehawk | Navy |
| MH-60S (Formerly CH-60S) | Navy |
| MINUTEMAN II | Air Force |
| MINUTEMAN III | Air Force |
| MLV III | Air Force |
| P-3C | Navy |
| PERSHING II | Army |
| RGM-109 Tomahawk BIP (Baseline Imp Prgm) | Navy |
| RIM-67 Standard Missile II | Navy |
| SH-60B LAMPS Mk III | Navy |
| SH-60F CV Helo | Navy |
| SSGN (OHIO CLASS CONVERSION) | Navy |
| Stryker | Army |
| TOW 2 | Army |
| UGM-109E Tactical Tomahawk | Navy |
| UH-60M Blackhawk Upgrade | Army |

**Table A-4. Joint MDAP VMRs**

Truncated

| Joint VMR | |
|---|---|
| **Program** | **Service*** |
| LAV (US Army/USN) | DoD |
| AGM-65D Maverick IR | Air Force |
| AIM-7M Sparrow (USAF/USN) | Air Force/Navy |
| AIM-9L Sidewinder | Air Force/Navy |
| AIM-9M Sidewinder | Joint |
| DSCS II | Navy |
| DSCS-III (Def Sat Comm Sys) | Air Force |
| E-3 Sentry AWACS RSIP | Air Force |
| SPARROW (AIM-7E) (AF) | Air Force |

*The Service indicated is the lead Service.

# Appendix B.
# Values of the Metrics for 1970–1988 and Post-1988 for VMRs, Single-Service New Starts, and Joint New Starts

**Table B-1. Values of the Metrics for 1970–1988 and Post-1988 for VMRs, Single-Service New Starts, and Joint New Starts**

| | | 1970–1988 | | | Post-1988 | | |
|---|---|---|---|---|---|---|---|
| | **Metric** | **Single-Service VMR** | **Single-Service New Start** | **Joint New Start** | **Single-Service VMR** | **Single-Service New Start** | **Joint New Start** |
| 1 | Estimated growth in Quantity Normalized Program Acquisition Unit Cost (PAUC)[a] | 12% (22) | 34% (60) | 52% (8) | 45% (19) | 38% (20) | 35% (11) |
| 2 | Percentage of programs with at least one Nunn-McCurdy breach (Dec. 1997–Sept. 2012) | N/A | N/A | N/A | 14% (4 of 28) | 31% (13 of 42) | 36% (9 of 25) |
| 3 | Percentage of programs with an estimated Quantity Normalized PAUC growth of ≥ 30%[a] | 18% (14 of 22) | 38% (23 of 60) | 63% (5 of 8) | 58% (11 of 19) | 45% (9 of 20) | 64% (7 of 11) |
| 4 | Program Cancellation Rate | 11 % (4 of 36) | 20% (23 of 117) | 47% (7 of 15) | 13% (4 of 32) | 19% (10 of 53) | 25% (7 of 28) |
| 5 | Percentage of programs cancelled, truncated, or reorganized | 14% (5 of 36) | 26% (30 of 117) | 47% (7 of 15) | 19% (6 of 32) | 32% (17 of 53) | 61% (17 of 28) |
| 6 | Percent of programs that eventually procure less than the Milestone B total quantity [b] | 100% (3 of 3) | 41% (7 of 17) | 50% (2 of 4) | 36% (10 of 28) | 34% (12 of 35) | 47% (9 of 19) |
| 7 | Estimated slip in Engineering and Manufacturing Development (EMD) | N/A | 3% (4) | 0% (1) | 26% (16) | 20% (16) | 42% (6) |
| 8 | Percentage of programs with at least one Operational Test failure | 13% (1 of 8) | 43% (13 of 30) | 100% (4 of 4) | 40% (6 of 15) | 53% (10 of 19) | 67% (6 of 9) |

[a] Computed for the total acquisition quantity approved at Milestone B.

[b] Approximated by the ratio of the total procurement reported in the last Selected Acquisition Report (SAR) for the program divided by the total procurement reported in the first SAR.

# Illustrations

## Figures

Figure 1. Potential Consequences of a Limitation on the Effectiveness of the DAB Process .....5
Figure 2. Percentages of Programs Cancelled, Truncated, or Reorganized.....15

## Tables

Table 1. Number of MDAPs in Each Category .....8
Table 2. Metrics Used to Gauge Effectiveness of DAB Process .....9
Table 3. Number of MDAPs Used in Computing the Metrics .....11
Table 4. Values of the Metrics for Single Service and Joint New Starts.....13
Table 5. Values of the Metrics for the Single Service VMRs Single Service New Starts, and Joint New Starts .....16
Table 6. Value of the Metrics for Single-Service VMRs, Single-Service New Starts, and Joint New Starts for 1970–1988 and Post-1988 .....18

# References


Blickstein, Irv, Michael Boito, Jeffrey Drezner, James Dryden, Kenneth Horn, James G. Kallimani, Martin C. Libicki, Megan McKernan, Roger C. Molander, Charles Nemfakos, Chad J. R. Ohlandt, Caroline Reilly, Rena Rudavsky, Jerry M. Sollinger, Katharine Watkins Webb, and Carolyn Wong. *Root Cause Analyses of Nunn-McCurdy Breaches*, Vol. 1. Santa Monica, CA: RAND Corporation, 2011.

Blickstein, Irv, Jeffrey Drezner, Martin C. Libicki, Brian McInnis, Megan McKernan, Charles Nemfakos, Jerry M. Sollinger, and Carolyn Wong. *Root Cause Analyses of Nunn-McCurdy Breaches*, Vol. 2. Santa Monica, CA: RAND Corporation, 2012.

Bronson, Patricia F., Scot A. Arnold, John W. Bailey, Harold S. Balaban, Ji S. Byun, Harley A. Cloud, Gregory A. Davis, Jeffrey N. Fox, Thomas P. Frazier, Alexander O. Gallo, Matthew W. Gonwa, Bruce R. Harmon, John R. Hiller, Paul M. Kodzwa, Tzee-Nan K. Lo, Caolionn L. O'Connell, Prashant R. Patel, Andrew S. Rehwinkel, Colin D. Sullivan, and David M. Tate. *Root Cause Analyses of Programs in Nunn-McCurdy Breach*, Volume I. IDA Paper P-4935. Alexandria, VA: Institute for Defense Analyses (forthcoming).

Coulam, Robert F. *Illusions of Choice: The F-111 and the Problem of Weapons Acquisition Reform*. Princeton, NJ: Princeton University Press, 1977.

Harris, Howard and Mark Lewis. "Proposed Leadership Structure for Joint Acquisition Programs." *Defense Acquisition Research Journal* 19, No. 1 (January 2012), 33–52.

Johnson, Dana J., Gregory H. Hilgenberg, and Liam P. Sarsfield. "Policy Issues and Challenges for Interagency Space Systems Acquisition." Santa Monica, CA: RAND Corporation, 2001.

McNicol, David L. *Cost Growth in Major Weapon Procurement Programs*. IDA Paper P-3832. Alexandria, VA: Institute for Defense Analyses, 2004.

Office of the Under Secretary of Defense (Comptroller). National Defense Budget Estimates for FY 2012. March 2011.

Packard Commission. *A Quest for Excellence: Final Report of the President's Blue Ribbon Commission on Defense Management*. Washington, DC: June 1986.

United States General Accounting Office. DOD Acquisition: Information on Joint Major Programs. Washington, DC, July 1989.

# Abbreviations

| | |
|---|---|
| CAPE | Office of Cost Assessment and Program Evaluation |
| DAB | Defense Acquisition Board |
| DAES | Defense Acquisition Executive Summary |
| DAMIR | Defense Acquisition Management Information Retrieval |
| DoD | Department of Defense |
| DoDD | Department of Defense Directive |
| DSARC | Defense Systems Acquisition Review Council |
| EMD | Engineering and Manufacturing Development |
| IDA | Institute for Defense Analyses |
| LRIP | Low Rate Initial Production |
| MDA | Milestone Decision Authority |
| MDAP | Major Defense Acquisition Program |
| MS B | Milestone B |
| MS C | Milestone C |
| OSD | Office of the Secretary of Defense |
| PA&E | Office of Program Analysis and Evaluation (predecessor of CAPE) |
| PARCA | Office of Program Assessments and Root Cause Analyses |
| PAUC | Program Acquisition Unit Cost |
| PM | Program Manager |
| PNO | Program Number |
| R&D | Research and Development |
| RDT&E | Research, Development, Test and Evaluation |
| SAR | Selected Acquisition Report |
| USD(AT&L) | Under Secretary of Defense (Acquisition, Technology and Logistics) |
| VMR | Variant, Modification, or Remanufacturing |
| WSARA | Weapon Systems Acquisition Reform Act of 2009 |

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

The public reporting burden for this 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 the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 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 any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.
**PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

**1. REPORT DATE** *(DD-MM-YYYY)*
xx-05-2013

**2. REPORT TYPE**
Final

**3. DATES COVERED** *(From - To)*

**4. TITLE AND SUBTITLE**
Evidence on the Effectiveness of the Defense Acquisition Board Process

**5a. CONTRACT NUMBER**
DASW01-04-C-0003

**5b. GRANT NUMBER**

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
McNicol, David, L.
Gould, Brandon, R.
Wu, Linda
Fuchs, Eugenia
Bronson, Patricia F.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**
AY-7-322306 and AY-7-357802

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

**8. PERFORMING ORGANIZATION REPORT NUMBER**
IDA Paper P-5011

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
Director, Performance Assessments and Root Cause Analyses
Office of the Assistant Secretary of Defense for Acquisition
3620 Defense Pentagon
Room 3C889A, The Pentagon
Washington, DC 20301-3620

**10. SPONSOR/MONITOR'S ACRONYM(S)**
D,PARCA

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

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

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

This study is concerned primarily with the effectiveness of the Defense Acquisition Board (DAB) process in assuring that Major Defense Acquisition Programs are completed within cost and on schedule and perform as anticipated. It begins by articulating how inferences about the effectiveness of the DAB can be drawn from comparisons of the outcomes of groups of programs. The paper presents two comparisons: (1) single-Service new starts—that is, programs to acquire a system of a new design—with joint new starts; and (2) single-Service programs launched during the Cold War to acquire variants or modifications of, or to remanufacture (VMR) systems already fielded, with such programs initiated post-Cold War. The comparisons were made using eight metrics. The differences in the metrics were consistent and those on which the paper's conclusions rest were large. The Office of the Secretary of Defense acquisition oversight process clearly fell somewhat short in dealing with the problems presented by joint new starts and by single-Service VMRs post-Cold War. The paper concludes by noting that these shortfalls seem to reflect a lack of sufficient oversight by the DAB in circumstances in which oversight is needed after a program has been given authority to proceed.

**15. SUBJECT TERMS**
Acquisition Programs, Defense Acquisition Board, DAB, joint programs, Major Defense Acquisition Programs, MDAPs, acquisition program outcome metrics, cost growth, schedule slips, Nunn-McCurdy Breach, Operational Test and Evaluation, program cancellation

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | Bliss, Gary, R. |
| Unclassified | Unclassified | Unclassified | Same as Report | 51 | 19b. TELEPHONE NUMBER *(Include area code)* (571) 256-0646 |

Reset

**Standard Form 298** (Rev. 8/98)
Prescribed by ANSI Std. Z39.18