One of the primary responsibilities of the Under Secretary of Defense for Acquisition and Sustainment (USD[A&S]) is to ensure the health of the overall defense acquisition system (DAS)—distinct from the health of any particular acquisition program, portfolio, or pathway. USD(A&S) can bolster the health of the DAS by developing and promulgating sound acquisition policy that improves the function and operation of the DAS at the enterprise level. The premise of this report—and the premise of the leadership of the Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD[A&S])—is that acquisition policymaking should be data driven. However, there are limitations to relying on empirical (e.g., historical) data to guide acquisition policy. In light of these limitations, we argue that acquisition policymaking in general should be evidence based, in recognition of a wider variety of analytic tools that can be brought to bear on acquisition policy questions. This report, intended for acquisition professionals, summarizes the case for a broader evidence base and then focuses on one specific tool that we suggest might add analytic value: policy gaming.

Policy gaming can be used to generate observations about how stakeholders might change their decisionmaking and behavior in light of changes in policy. Because the strengths and limitations of games differ from those of traditional tools for acquisition analysis, we argue that games complement...
the existing portfolio of analytic approaches. In this report, we describe an acquisition policy game that RAND Corporation researchers developed to enrich the available evidence base to support acquisition policymaking, summarize insights from a prototype game focused on Middle-Tier Acquisition (MTA) policy, and speculate about how such games might yield broader insights into the ways new acquisition policies could affect the DAS.

In practice, an empirical data–driven approach to acquisition policymaking entails at least four assumptions, which cannot always be met. The first assumption is that data on the historical performance of the DAS are generalizable to the new policy under exploration—which they must be, in order to be relevant to evaluating the effect of a new acquisition policy. In many cases, this assumption might be valid, especially for incremental changes to acquisition policy. For example, a change in the dollar-value threshold that triggers increased oversight by the Office of the Secretary of Defense might be justified by historical data demonstrating that programs that reach that threshold entail greater risk of unfavorable outcomes. However, the generalizability assumption might break down for non-incremental changes in acquisition policy or technology that, in effect, change the paradigm of defense acquisition and render the historical data less relevant. For example, a new policy to govern the MTA pathway might not benefit from an empirical approach simply because historical program data do not reflect a middle-tier process. For another example, a policy to govern the acquisition of large commercial constellations of smaller satellites in low Earth orbit might not benefit from historical data on space acquisition, which have relied on the designing and building of small numbers of relatively expensive military satellites. An empirical approach to acquisition policymaking applies only in cases in which the data are generalizable, which would seem to be cases involving relatively incremental changes to acquisition policy.

The second assumption is that current conditions will hold in the future, but the reality is that they might not hold, which could place new pressures on the DAS. For example, potential funding shortfalls in either the near or the far term can adversely affect program or portfolio performance. The emergence of new technology or new threats creates opportunities and challenges that might not have a good analogue in historical program data. An analytic approach that focuses only on current conditions risks missing external forces that have been shown to affect the health of the DAS.

Thirdly, an empirical approach to developing and establishing acquisition policy assumes that the data are sufficiently rich to permit establishment

**The Strengths and Limitations of Empirical Data for Setting Acquisition Policy**

One approach to setting acquisition policy is the use of empirical data to assess the outcomes of the DAS. Under this approach, acquisition policy should be revised periodically on the basis of feedback on how well the DAS is actually performing under current conditions. Conceptually, there is no disputing the value of this approach, and this approach reflects traditional analytic approaches associated with performance evaluation of organizations. USD(A&S) has prioritized this approach, as evidenced by the establishment of the Office of Acquisition Enablers and ongoing work. However, historical performance evaluation faces challenges in providing insight into the impacts of changes in the policy or execution environment.

---

**Abbreviations**

- **A&S**  Acquisition and Sustainment
- **ACAT**  Acquisition Category
- **DAS**  defense acquisition system
- **DoD**  U.S. Department of Defense
- **MTA**  Middle-Tier Acquisition
- **OUSD(A&S)**  Office of the Under Secretary of Defense for Acquisition and Sustainment
- **USD(A&S)**  Under Secretary of Defense for Acquisition and Sustainment
of a causal effect of acquisition policy on acquisition outcomes. This discussion could devolve into a technical matter of statistics, but a few general points can be made. There are perhaps on the order of 300 Acquisition Category (ACAT)–1 acquisition programs for which there are high-quality historical data. In some cases, explanatory variables of interest might never have been collected, severely constraining analysis. Examples include technical maturity of specific subsystems critical to performance (which affects mostly older historical programs), experience base and tenure of the program management team, and industrial base considerations (structure, number of firms with specific capabilities). These data gaps can sometimes be filled, but such efforts are labor intensive and tend to take time to accomplish. The nature of these data contrasts with the requirements of many modern methods of statistics and machine learning, where data sets exceeding millions, if not billions, of items are needed to develop models involving dozens or hundreds of variables.

Even if the effects of acquisition policy were large enough to be measurable with data sets of smaller size, it might be difficult to tease out the effects of acquisition policy from the effects of myriad other factors that are changing simultaneously. For example, large, complex programs tend to have several program managers, so no single manager is with the program for more than one major milestone. Other senior officials also might have short tenures relative to the life cycle of a major acquisition program. Rapid turnover introduces the possibility that leadership changes, not policy, are driving outcomes. Although such details can be traced in careful study of specific cases to determine the relative importance of different potential causes, such an approach further reduces the number of examples, increasing the generalizability problem discussed earlier. In short, limitations in empirical data reduce our confidence in what can be learned from these data.

There is also a fourth practical assumption—that the cost and risks of making a change to acquisition policy justify an empirical (e.g., historical data) analytic approach. Here, too, incremental changes to acquisition policy seem to make this case; it is hard to imagine dire consequences from a change to the dollar threshold that determines the level of oversight a program receives. However, less incremental changes to acquisition policy might come with unintended consequences that one might prefer to discover prior to implementation. Examples of such changes include the widespread use of other transaction authority or the creation of new acquisition pathways like middle tier. If the U.S. Department of Defense (DoD) adopts a “let’s just try” approach to acquisition policy—experimenting with new policies—those risks might manifest.

In summary, the empirical approach to setting acquisition policy is important and valuable, but it faces fundamental and practical limits. The empirical approach appears to apply most naturally to relatively incremental changes to acquisition policy for which the historical data are generalizable and the effects of acquisition policy on acquisition outcomes are large enough to be detected with relatively small data sets and many explanatory variables.

---

Even if the effects of acquisition policy were large enough to be measurable with data sets of smaller size, it might be difficult to tease out the effects of acquisition policy from the effects of myriad other factors that are changing simultaneously.
There is no disputing the appeal of benchmarking DoD practices against those of the commercial world by using commercial data, but significant differences between DoD and the private sector limit the applicability of those lessons.

Other Approaches to Setting Acquisition Policy

In addition to the use of historical empirical data, there are two other approaches traditionally used to support acquisition analysis and policymaking, as evidenced by the Section 809 Panel’s 2019 report—arguably the most comprehensive recent examination of the DAS. These approaches are analyses using commercial data and expert judgment.

There is no disputing the appeal of benchmarking DoD practices against those of the commercial world by using commercial data, but significant differences between DoD and the private sector limit the applicability of those lessons. For instance, DoD’s experience with performance-based logistics, outsourcing, contracting strategy, and managing the vendor base demonstrates that successful commercial practices often produce different results when applied to a government environment.1

Another option is expert judgment, which has proven critical to devising such nontraditional strategies as the MTA pathway, the creation of the Strategic Capabilities Office, the use of other transaction authority as an alternative to Federal Acquisition Regulation–based contracting, and a host of other innovative approaches to acquisition. However, expert judgment applies best to cases in which the experts have sufficient experience to render a judgment. Expert judgement also can be unstructured and untraceable—it often takes the form of heuristics that have grown from years of experience but are difficult to explain to an outsider, unless the logic behind the decision is elicited explicitly. But is it possible to leverage expert judgment in ways that can mitigate some of these limitations? Many elements of DoD are regularly charged with considering the future performance of new policies, strategies, and standard operating procedures, all of which might offer potential methods. One tool for evaluating potential policies that has been a mainstay of other defense communities is gaming.

In summary, although we applaud the current focus on improving the state of data-based policymaking, we are concerned that data that can answer key questions about the future health of the acquisition system are not always available. Put simply, historical data are not always sufficient to understand current and future conditions in which policies must be implemented. At the same time, traditional alternatives to empirical analysis, such as comparison with commercial examples and expert judgment, have long-recognized shortcomings. We suggest that tapping into a wider variety of evidence-based analysis, starting with gaming, might complement these approaches.

Gaming and How It Informs Other Areas of Defense Policy

Looking to other areas of defense policy, we find that games can provide useful evidence about proposed policies by providing a sandbox to observe decision-making. In particular, games appear to be valuable in cases where relevant real-world data are not available because the new policy or other condition of interest has not yet occurred. Other conditions that make games particularly appropriate are policies that touch multiple stakeholders with distinct objectives
and policies that would elicit revolutionary, rather than evolutionary, changes. This feature suggests that a particularly relevant application of gaming to acquisition is to generate evidence when considering the implementation of a new policy that represents a break from historical practice across DoD. In other words, games are a means of generating evidence-based policies when traditional approaches fall short.

Although games can vary in their appearance, ranging from groups of individuals sitting around a table discussing a briefing to more-structured boards and tokens to computerized systems, games for policy share some common elements. Games consist of human players in a competitive environment making decisions using a defined process and experiencing projected consequences of their choices. In effect, games allow players to practice decisions in a potential future environment, using processes and tools that might not be in place today. As a result, games are a way of eliciting expert judgment in a group format that is structured by both a specific process and the need to make defined decisions. Former Deputy Secretary of Defense Robert Work and former Vice Chairman of the Joint Chiefs Gen. Paul Selva summed up the value of the method when they stated that “wargaming is one of the most effective means available to offer senior leaders a glimpse of future conflict, however incomplete. Wargames provide opportunities to test new ideas and explore the art of the possible” (Work and Selva, 2015).

Looking across the defense enterprise, there are five purposes for which games are frequently used (Bartels, 2020):

- **Understanding a system.** The game allows different stakeholders to walk through the decisionmaking process in a low-risk “sandbox” to better understand how the process might work in practice.
- **Exploring decisionmaking under alternative conditions.** The game allows for structured comparison of decisionmaking under different contexts.
- **Sparking innovative ideas for new solutions to emerging problems.** Stakeholders might suggest new processes to speed and smooth decisionmaking across offices, and processes can then be further analyzed prior to being incorporated into new guidance.
- **Evaluating proposed policies.** Although such analysis is unlikely to be directly predictive, it can help compare the strengths and weaknesses of different approaches, identifying potential pitfalls before a policy is implemented.
- **Educating and communicating.** Games can explain how decisions are made about a specific problem by providing players with the experience of making choices and seeing the potential outcomes of their actions.

Each of these purposes might require somewhat different designs to achieve the intended outcomes, and the designs might look quite different from each other in practice, thus requiring custom efforts rather than an off-the-shelf analytic product.

However, regardless of the purpose, all games share a common benefit—they allow researchers and policymakers to watch how decisions are made under conditions that do not exist today. Furthermore, because groups of players are involved, games provide a forum for multiple experts to debate a concrete problem and make specific decisions in settings where researchers and policymakers can observe and interrogate the decisionmaking process to unpack heuristics and weigh competing judgments.

**Games can explain how decisions are made about a specific problem by providing players with the experience of making choices and seeing the potential outcomes of their actions.**
Despite its many benefits, gaming has important limitations, many of which are shared with other methods. The most fundamental limitation is that games are models, not reality. Thus, what is observed is inherently artificial, and analysis of gaming results must always explore how the differences between the real world and the world of the game might skew results. Although the traceability of games allows for some additional types of quality control compared with complex computerized models, care is still required in their analysis. In addition, games are not a good approach to test the effects of minor policy changes or estimate point solutions; instead, they are more appropriate for understanding broad implications of large changes. Importantly, the strengths and limitations of games differ from those of traditional tools for acquisition analysis, suggesting that games might complement the existing portfolio of approaches.

Prototyping an Acquisition Policy Game

To explore how games might inform our understanding of acquisition policy, RAND researchers designed and prototyped a structured seminar-style game examining the recent innovation of the MTA pathway. The MTA pathway policy has several of the key attributes that make gaming an attractive option for analyzing it. The policy is new enough that we do not have much empirical data about its use for different types of programs or its relevance to a variety of acquisition stakeholders with different responsibilities in the process, and it represents a sizable shift from current business practices.

The RAND game consisted of a highly simplified simulation of decisionmaking about how to use the middle tier. We convened a group of players who represented the interests of key staff in the acquisition decisionmaking process, including project managers, contracting officers, technical specialists from the Acquisition and Sustainment (A&S) staff, and warfighters. Players were asked to review a series of program descriptions using a structured assessment tool, make a recommendation about whether the program should be acquired, and, if so, determine what acquisition pathway should be used: middle-tier prototype, middle-tier fielding, the traditional 5000 series, or urgent capability. We designed the program descriptions to address operationally relevant capabilities. The programs varied by several factors we believed might shape acquisition pathway decisions, including projected cost and schedule, senior leader and warfighter demand, and technology maturity and complexity. After deciding on what pathway to pursue, players assessed the risk of program failure in terms of performance, cost, and schedule.

Figure 1 shows the game board on which players’ votes were conducted. For each program, each player placed a chip in one of the gray boxes at the top of the board to indicate their preferred pathway. For example, in the figure, the majority of players voted to put the program in the middle-tier prototyping pathway, with two dissenters voting not to acquire the program at all. Players then placed three more poker chips on the bottom of the board to indicate their assessment of the program’s risk in performance, schedule, and cost. In other words, we asked players to project where they thought the program might fail. If players opted to place the program in an MTA pathway, they then completed a worksheet that allowed them to select core elements of an acquisition strategy, including decisions regarding testing regime, contract type, and target cost and schedule. This step was designed to allow players to buy down...
some of the risks they identified in their initial analysis.

After players had selected which programs to acquire using an MTA pathway and developed an acquisition strategy for each, the game staff then projected the performance of the MTA programs two years into the future to provide feedback on program performance, schedule, and cost. To accomplish this projection, the team used an adjudication process that combined an assessment of the program risk based on program characteristics and the acquisition strategy developed by players to generate a probability distribution of potential outcomes. The researchers then injected an element of chance to reflect aspects of program success that lie outside the control of acquisition officials—including technological progress, contractor performance, and outside economic events—to determine a specific outcome for the program from the distribution. Players were updated on the status of the program and asked if they wanted to maintain the program in the MTA pathway or transition to another option. If the players elected to retain the program in MTA, they could update their choice for an acquisition strategy, triggering a repeat of the adjudication process, which provided an update on program performance at four years. Players were then asked to decide whether they believed the program would successfully meet MTA guidance to field a residual capability in five years or whether the program needed to be transitioned to an alternative pathway.

Nine RAND experts gathered for one day to play our prototype game. We drew the players from among RAND staff with diverse acquisition experience, including former project managers, former service program executive officers (PEOs), and leaders of RAND’s acquisition research programs. Over the course of the game, we asked the players to consider seven program descriptions, five of which they opted to acquire using the MTA pathway.

**Insights from the Game**

Although this game was only a prototype, it suggested several insights into MTA policy design and
of the MTA policies promulgated at the time the game was played.

- The perceived flexibility of MTA might translate into novel acquisition strategies. In several cases, the participants broke large, complex programs into more-manageable elements that could be executed within the MTA. This included designing incremental acquisition strategies or splitting programs into components or major subsystems to take advantage of more than one pathway to buy down program risk, although it was recognized that this latter approach introduces a form of integration risk. The interim policy provided little guidance on how MTA acquisition strategies should be constructed. It might also be useful to capture and document novel strategies for use by future MTA programs with similar characteristics.

- MTA might need governance to align policy implementation across DoD. Guidance on MTA was interpreted differently by different players. This included data collection and reporting, decisionmaking and internal milestones, oversight and approvals, and the firmness of MTA constraints (i.e., schedule). Policy alignment should include establishing the primary purpose of the pathway in contrast to other pathways, as well as establishing consistent roles, responsibilities, and authorities for key stakeholders (i.e., defense acquisition executive [DAE]/service acquisition executive [SAE]/milestone decision authority [MDA],...
program executive officer, program manager, functional staff (contracting, testing, engineering, etc.).

MTA has been in place for nearly two years, so individuals deeply familiar with existing policies and practices might argue that these insights are borne out by experience. However, we emphasize that observations were developed with just a day of exercises. This suggests that a gaming platform like this one could serve as a sandbox to experiment with new acquisition policies.

**Limitations**

These findings illustrate the types of evidence that a game can provide. However, it is important to note that the scope and scale of this prototype effort imposed several limitations on game findings. First, we were not able to examine service-specific implementation of MTA as part of this effort. Because of known differences in service culture and MTA implementation approach, our players’ decisions might be different from those of real decisionmakers. Second, because of limited time, the description of each program was limited to between two and three pages of text. It might be that inclusion of additional information might have introduced other key factors into the discussion or even shifted player decisions regarding pathway. A third limitation was that requirements to generate unclassified materials on operationally relevant but not currently available capabilities excluded some categories, such as commercial off-the-shelf programs, from consideration. Such limitations could be mitigated in future iterations by making adjustments to the game’s processes and materials.

**Next Steps**

We offer several next steps for USD(A&S) to consider.

- **First**, look to expand beyond traditional evidence-based approaches to acquisition policy. Although empirical data are critical, they have limitations. Adding complementary methods, such as games, will build a more rigorous evidence-based policy than depending on data-driven approaches alone. Thus, we suggest that USD(A&S) fund pilot efforts using new tools, in order to better understand their utility in evaluating the health of the DAS.

- **Second**, instrument the adaptive acquisition system to enable continuous, incremental improvements to acquisition policy. The value of an empirical, data-driven approach appears indisputable. But this analysis reveals that both historical data scarcity and the existence of many explanatory variables limit the evidence-based approach. In particular, our work suggests that data-based evidence is likely to be informative when considering incremental policy changes but might be insufficient for informing larger changes in policy.

- **Third**, experiment with developing and using policy games to rapidly prototype more-revolutionary changes to acquisition policies for which an empirical approach might not apply. Evidence from our prototype effort suggests that problems involving policy implementation, developing acquisition strategies, and understanding information requirements might benefit from gaming. Games developed to inform policymakers might also have additional utility as an educational tool in DoD schoolhouses.
Notes

1 See, for example, Section 809 Panel, 2019; Klitgaard and Light, 2005; and Drezner et al., forthcoming.

2 For key texts on policy games and their uses, see Perla and McGrady, 2011; Perla, 2012; Parson, 1996; and Levine, Schelling, and Jones, 1991.

References


Drezner, Jeffrey A., Jon Schmid, Justin Grana, Megan McKernan, and Mark Ashby, Benchmarking Data Use and Analytics in Large, Complex Private Sector Organizations: Implications for Department of Defense Acquisition, Santa Monica, Calif.: RAND Corporation, RR-A225-1, forthcoming.


About the Authors

Elizabeth M. Bartels is an associate policy researcher and specialist in national security policy analysis gaming. Her work explores a wide variety of strategic and operational concerns, with a focus on games on novel topics and that integrate other analytical techniques. Other research includes work on defense planning, force development, and measures short of armed conflict. Bartels has a Ph.D. in policy analysis.

Jeffrey A. Drezner is a senior policy researcher at RAND. He has more than 35 years of experience conducting policy analysis on a wide variety of issues, including planning and program management and defense acquisition policy and reform. His research emphasizes mixed qualitative and quantitative approaches to analyze issues associated with technology development, organizational behavior, and program management. Drezner holds a Ph.D. in political science.

Joel B. Predd is a senior engineer at RAND and director of the National Security Research Division’s Acquisition and Technology Policy Center, where he oversees RAND’s portfolio of force development, technology, and acquisition research for the Office of the Secretary of Defense, the Joint Staff, the Combatant Commands, and international sponsors. Predd holds a Ph.D. in electrical engineering.
About This Report

One of the primary responsibilities of the Under Secretary of Defense for Acquisition and Sustainment (USD[A&S]) is to ensure the health of the overall defense acquisition system (DAS). USD(A&S) can bolster the health of the DAS by developing and promulgating sound acquisition policy that improves the function and operation of the DAS at the enterprise level. The premise of this report is that acquisition policymaking should be data driven. However, there are limitations to relying on empirical (e.g., historical) data to guide acquisition policy. In light of these limitations, the authors argue that acquisition policymaking should be evidence based, in recognition of a wider variety of analytic tools that can be brought to bear on acquisition policy questions. This report, intended for acquisition professionals, summarizes the case for a broader evidence base and then focuses on one specific tool that the authors suggest might add analytic value: policy gaming.

Policy gaming can be used to generate observations about how stakeholders might change their decisionmaking and behavior in light of changes in policy. Because the strengths and limitations of games differ from those of traditional tools for acquisition analysis, the authors argue that games complement the existing portfolio of analytic approaches. The authors describe a prototype game focused on Middle-Tier Acquisition (MTA) policy that RAND researchers developed to enrich the available evidence base to support acquisition policymaking, summarize insights from the game, and offer several next steps for USD(A&S) to consider.

This research was sponsored by the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics Acquisition Resources and Analysis Directorate and conducted within the Acquisition and Technology Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense intelligence enterprise.

For more information on the RAND Acquisition and Technology Policy Center, see www.rand.org/nsrd/atp or contact the director (contact information is provided on the webpage).