Measuring Readiness and Sustainment Within Analysis of Alternatives in Military Systems Acquisition

Author(s): Boaz Golany and Moshe Kress

Source: *Military Operations Research*, 2020, Vol. 25, No. 4 (2020), pp. 63-77
Published by: Military Operations Research Society
Stable URL: https://www.jstor.org/stable/10.2307/26957616

Measuring Readiness and Sustainment Within Analysis of Alternatives in Military Systems Acquisition

Boaz Golany[1] and Moshe Kress[2]

[1]Technion, Israel, golany@ie.technion.ac.il
[2]Naval Postgraduate School, mkress@nps.edu

ABSTRACT

Analysis of alternatives (AoA) is a crucial stage in the process of acquiring new systems for the Department of Defense. AoA is essentially a multicriteria decision process that involves several stakeholders. An AoA is an analytical comparison of the operational effectiveness, cost, and risks of proposed materiel solutions to gaps and shortfalls in operational capability. Two of the three criteria—effectiveness (what can the system do and how its capabilities fit the operational requirements) and comparative costs of potentially viable solutions—draw the most attention in such studies. The third set, risks, which typically draws somewhat less attention in an AoA than the other two, is concerned with the technical, operational, and programmatic implications for each alternative. More precisely, we describe operational risks as the long-term readiness, sustainment, and logistics requirements for the specified alternative. In this paper, we study in detail the content of this set of criteria, focusing on the factors that affect the long-term viability and usefulness of an alternative, and propose data envelopment analysis as the analytic framework for evaluating alternatives with respect to the criteria in this set.

INTRODUCTION

The Department of Defense (DoD) Acquisition System comprises three interconnected stages that start with specifying requirements—a procedure called Joint Capabilities Integration and Development System (JCIDS). The second stage, called acquisition process, determines appropriate materiel solutions for the requirements. The third stage is concerned with funding and financial-controlling activities contained in the planning, programming, and budgeting execution process. Most of the decisions that have long-term sustainment, readiness, and logistics implications are taken at the second stage, where materiel choices are made. The overarching process dominating this stage is the analysis of alternatives (AoA) that, in general, trades off the effectiveness of a materiel solution with its risks and costs. DoD Instructions 5000.02, which provides policy guidelines and regulations for managing the acquisi-

https://doi.org/10.5711/1082598325463

APPLICATION AREAS:
Readiness and Training, Acquisition, Methods

OR METHODS:
Decision Analysis

tion system, states that, "The AoA assesses potential materiel solutions that could satisfy validated capability requirement(s) documented in the Initial Capabilities Document, and supports a decision on the most cost-effective solution to meeting the validated capability requirement(s). In developing feasible alternatives, the AoA will identify a wide range of solutions that have a reasonable likelihood of providing the needed capability" (Department of Defense, 2020). In general, the instructions are pushing for the AoA to be conducted earlier in the acquisition process.

The AoA is essentially a large-scale multicriteria decision analysis (MCDA) problem that involves multiple stakeholders and many uncertainties (Kress and Morgan, 2018). The set of criteria used in evaluating alternatives, and their weights or importance, depend, among others, on the technological maturity of the alternatives. For example, the risk associated with acquiring an off-the-shelf system may in some cases be considerably lower than the risk in developing a new system but in other cases using such components that were designed for other purposes may actually add risks to the new system. Thus, the risk criterion, with all its derivatives, is less prominent for the former than the latter. In this study, we focus on systems that are still in various stages of development, which means that the AoA process is typically not a one-shot decision event, but rather a sequence of decisions marked by achievements. In these settings, the AoA starts off with a set of potential alternatives being developed as prototypes. Additional data is collected over time and this enables analysts to revisit earlier analysis outcomes with better insights. Prior to making decisions that commit the resources needed for development, leading eventually to production and fielding, analysts solve additional MCDA problems to inform decision makers on which specific product or design concepts will be pursued, and whether resources will be committed to mature technology and/or mitigate any risks in research and development. The process culminates in a preferred alternative.

Several MCDA methods have been proposed in the literature as the platform for AoA studies. Many of those methods belong to the compensatory family of parametric methods based on the ability of the analysts to determine values of alternatives and weights (importance) of criteria, leading to a weighted average score of each alternative. An interesting application of these classical methods in the context of realigning US Army bases is given in Ewing et al. (2006). This family of parametric compensatory methods is very broad and includes models with a wide variety of features and properties. A notable example of this family of methods is the analytic hierarchy process (AHP), where values and weights are determined via pairwise comparisons (see Saaty, 2008). However, AHP, with its specific features associated with eigenvalues, was highly criticized as several of its basic concepts lie on shaky mathematical foundations (e.g., see Barzilai and Golany, 1994).

Outranking is another family of MCDA methods. A representative method in this family is elimination and choice expressing reality (ELECTRE), where alternatives are gradually discarded on the basis of outranking rules (Roy, 1968). However, ELECTRE and other similar methods, were criticized for their use of arbitrary rules and veto thresholds.

The main purpose of this paper is to develop a new paradigm, grounded on the well-established theory of data envelopment analysis (Charnes et al., 1978) for studying and promoting awareness to logistic considerations, readiness implications, and total ownership cost during the acquisition process while applying suitable quantitative tools to handle uncertainties involved in some of the parameters. The unique feature of DEA, which stands out compared to other MCDA models, is that it is a nonparametric method. DEA avoids the pitfalls of previous methods mentioned previously by letting the method choose an individual set of weights for each alternative through solving a series of linear programming problems, one for each alternative. DEA evaluates the efficiency of an alternative, compared to others, in the most equitable way and thus enables decision makers to focus on the most efficient alternatives for in-depth analysis.

READINESS, SUSTAINMENT, AND LOGISTICS

The *DoD Dictionary of Military and Associated Terms* (Department of Defense, 2019) defines sustainment as the "provision of logistics and personnel services required to maintain and prolong operations until successful mission accomplishment" (JP 3-0). Readiness is defined as the "ability of military forces to fight and meet the demands of assigned missions" (JP 1). Logistics is "planning and executing the movement and support of forces" (JP 4-0). Notice that the term logistics is included in the definition of sustainment. Thus, henceforth we will focus on the two factors: readiness and sustainment.

Sustainment and readiness are terms that typically apply to the entire military force. They are considered two of the four pillars of military capability (Moore et al., 1991). Measures of sustainment and readiness describe features such as resources availability, level of training, and percentage of platforms that are mission ready at any time (Rich et al., 1987).

In this study, we consider these terms in a more restricted way, as they reflect the virtues and shortcomings of acquisition alternatives.

In the context of AoA (see, e.g., Headquarters Air Force, 2016), the relative value of a certain (new) acquisition alternative (compared to other alternatives) is based on three dimensions, each a combination of several criteria: (a) effectiveness during operations, (b) cost, and (c) readiness and sustainment (R&S) attributes. The R&S dimension accounts for all the factors not directly related to the first two, such as risks, dependencies on other systems, and peripheral requirements. The three dimensions are not orthogonal. For example, sustainment involves (lifecycle) cost considerations, and effectiveness is affected by readiness.

Effectiveness of an alternative is measured through the gaps, if they exist, between the performances of each alternative and the operational requirements that were set during the JCIDS. Lifecycle cost (sometimes referred to as total ownership cost) is measured by accumulating the various expected costs directly associated with each alternative. Both dimensions have been studied in depth and there exist proven methodologies that address them in AoA studies (e.g., combat simulations and cost estimation methods). In this paper, we focus on the third dimension, R&S, which, we believe, demands as much attention as the other two dimensions. Failing to devote this attention, especially during the early stages of the AoA, may potentially lead to choosing an alternative that is most cost-effective from the point of view of the first two dimensions, but may become economically burdensome and eventually even impossible to operate in the long run.

We next review some articles that highlight challenges in identifying the components of readiness and sustainability, and the complexities in analyzing them.

Challenges in Evaluating Readiness and Sustainability

Several important issues related to the evaluation of readiness and sustainability in military context have been studied in recent years. (Canician and Daniels, 2018) emphasize the tradeoffs in the iron triangle of readiness, modernization, and force structure that military services must consider when making procurement decisions. They argue that equating readiness with military capability obscures the critical elements of training/maintenance/equipment and personnel fill in readiness. In our paper, we indeed adopt a holistic approach that includes all these elements in the procedures we suggest for measuring readiness. They further argue that the standard DoD measurement procedure of readiness, known as the Defense Readiness Reporting System, is inadequate as it focuses only on inputs and doesn't account for outputs. Again, our proposed model considers simultaneously inputs and outputs.

Harrison (2014) highlights the differences in the definition of readiness at the strategic, operational, and tactical levels. He discusses the components that define readiness and emphasize, as we do, that understanding how best to resource readiness requires better metrics and a better understanding of the relationships between inputs (resources) and outputs (readiness).

The *Defense Acquisition Guidebook* (Defense Acquisition University, 2019) revision views sustainment as a distributed and long-term activity that requires the alignment of multiple and disparate stakeholders. It emphasizes the need to carry out sustainment planning from the earliest stages of the system's life cycle.

Robbins et al. (2019) describe the longstanding shortfall in the Weapon Systems Support Program performance. They state that a primary reason for the shortfall is the failure to use tools and methodologies for accurately determining true critical weapon system readiness drivers and differentiating these items from others. Our paper addresses exactly this issue by identifying and analyzing the readiness drivers.

Doheney et al. (2019) argue that the US military currently uses readiness reports to communicate the cumulative effects of details on its "big things." However, the word "readiness," like the word "risk," is inconsistently defined, and discussions and planning around this topic are often ambiguous. They claim that standard representations of uncertainty now make it possible to roll up analytical models into larger models and call the services to embrace such representations and further develop them in the context of readiness accounting. These advancements could move military decision makers from ambiguous estimates of "ready or not?" to unambiguously calculating, quantifying, and communicating "how ready for what?" Our paper uses standard probability measures to assess risk and uncertainty and embed these estimates in a larger model that quantifies the degrees of readiness and sustainability.

Next, we describe the content of the R&S dimension and discuss its criticality.

Readiness

Readiness of an acquisition system has three aspects: technological, technical, and functional. Although technological readiness describes the state of a system still being developed and the projected time of completing its production and testing, technical and functional readiness relate to the actual operation of the system when it is fully operational and already deployed.

Technological Readiness. The technological aspect of readiness describes the maturity of the technology, and the production and testing capabilities needed for the newly developed system. As mentioned earlier, this type of readiness only applies to the development phase of the system; it describes how close the system is to becoming fully operational. Analysts have used several scales to evaluate this aspect. Back in the 1980s, the National Aeronautics and Space Administration introduced the Technology Readiness Level (TRL) as a tool to assess the maturity of particular technologies that it planned to embed in new systems. The TRL is a nine-level scale that ranges from "Basic Principles Observed and Reported" (Level 1) to "Actual System Proven through Successful Mission Operations" (Level 9). According to the DoD, a weapon system program cannot include a technology whose TRL is lower than level 7 (Government Accountability Office, 1999). Sauser et al. (2008) combine the TRL, which is purely technology oriented, with a system-integration measure—the Integration/Interface Readiness Level (IRL)—and produce the System Readiness Level (SRL), which is a more holistic engineering measure for the technological maturity of a system. Atwater and Uzdzinski (2014) generalized the SRL into a holistic view of system maturity that includes functionality (i.e., man-machine interface).

Technological readiness is tied to the risk criteria in the AoA process. The risks associated with an alternative that is still in the technological development stages reflect the uncertainties regarding the feasibility that the alternative will be mission ready on time and within budget. Thus, during the AoA and the subsequent development and production phases, the technological readiness of a system can be defined by a value proportional to the assessed probability that the system will be completed on time and within budget.

As mentioned previously, the technological readiness of a system has to do with projecting the completion time of the underlying research, development and production project, and the total cost of the system. The SRL, data from similar past projects, and most of all, inputs from subject matter experts are used to obtain the relative technological readiness values of the alternatives, using techniques similar to the one described in Kress and Morgan (2018).

Finally, note that technological readiness changes dynamically during the acquisition process. It is updated periodically as the project advances; the closer is the project to completion, the more accurate is the estimate of this type of readiness.

Technical Readiness. A system is technically ready if all its components are in a perfect working condition. Thus, technical readiness (TR) has to do with the reliability and maintainability of the system and its components. Technical readiness can be measured by the probability that the system is technically mission ready at any point in time during its life cycle. Technical readiness is closely tied with sustainment, which is discussed later. Technical readiness is also related to the nature of the missions planned for the system and particularly to the typical alert time the forces are expected to have before launching a mission. For example, missile interception systems are expected to become ready within seconds of an alert whereas submarine missions will typically have alert times of days or even weeks.

Three main factors determine the technical readiness of a system: (a) reliability, which is measured by the mean time between failures (MTBF), (b) the effort needed to fix the system, which is measured by the average service time, and (c) the availability of spare parts needed for its maintenance and repair. Note that factor (b) depends to large extent on factor (c); availability, or lack thereof, of repair parts affect the total service time of a system.

In principle, factor (a)—MTBF—is measurable; the MTBF of each component in the system, along with the associated reliability graph of the system, can project the system-wide MTBF. The problem is that the MTBF figures for the various components are only estimates, based on limited data and simulations. The actual MTBF of the system cannot be robustly estimated until the system is fully operational for some time and enough failure data is collected. Similarly, to factor (a), the impact of factor (b)—service time—is practically unknown until the system is operational for some time and enough service time data are collected. The fact that data regarding the first two factors are unavailable at the time when the AoA is initiated does not mean that they should be ignored. Technical details, comparisons with similar systems, meticulously designed simulations, and early testing of critical components should be utilized for evaluating these two important factors of technical readiness. The third factor is the availability of repair parts, which is affected by the number, complexity, variability, and cost of the system's components, as well as by the robustness of the supply chain that provides the repair parts. These characteristics could be evaluated well in advance and factored in quite smoothly into the evaluation of technical readiness. For example, the lead time for delivering a certain repair part, a parameter that could be estimated from the features of its supply chain, will be used to estimate total repair time. The impact of unavailable repair parts is determined by design factors such as redundancy of subsystems and components. The modularity of subsystems affects the repair time (factor (b)). Plug-and-play–type components obviously need less service time than components that require installing, reconfiguration, and adjustments.

It is quite unlikely to have a reliable TR measure during the early phases of the AoA process. There will not be enough data to support it. However, as the system advances in its development process, more knowledge and experience are accumulated, and the TR measure—combining MTBF, service time, and repair parts lead time—gets updated accordingly. We discuss the specific features of possible TR measures later.

Functional Readiness. A system that is fully developed and technically ready for operations is not necessarily that fully operational. For a system to function effectively, one may need to ensure the functionality of other, supporting or peripheral, systems. Full functionality will also require that certain types of operators will be available to run it, and specific elements of infrastructure will be in place to support it. First, a system, as advanced and sophisticated as it may be, needs people to operate, control, maintain, and utilize its outputs. These operators, controllers, support personnel, and users need to be trained and available for their respective tasks. Shortage in any of the required manpower capabilities and expertise needed for the system may render the system nonoperational. Arguably, the impact of unavailable personnel varies among the tasks. A system that needs four operators during a shift may be operational, albeit less effectively, with just three operators. But, a repair technician who is not available when the system is down can be detrimental. A system may also need peripheral support such as protected environment, robust supply chain of resources, and access to communication networks. It may also depend on the operations of other systems, where failure in one or more of them may render our system nonoperational even if all other technical and functional factors are in perfect condition.

As mentioned earlier, we note the difference between the first type of readiness (technological readiness) and the other two. Whereas the first type applies to the development/production stage of the alternative, the last two refer to the readiness of the system once it is deployable and ready for operation. In the following, when we refer to readiness as *operational readiness*, we restrict the definition to only the technical and functional aspects. The technological aspect does not apply to operational readiness and therefore should be treated separately.

Both technical and functional readiness can be further broken down into a vector of subfactors that define it in greater granularity. Combining all these factors and subfactors into a single operational readiness measure will require some weighting methods, a topic that will be discussed later.

Sustainment

In a nutshell, following acquisition, total ownership cost is the cost of sustainment, that is, the cost of maintaining a system in an adequate operational condition. Technical and functional readiness are contingent on sustainment, which encompasses all the materiel and services needed for the effective and prolonged operation of a system so that it satisfies the missions for which it is designed. There are three facets to sustainment: supplies, facilities, and personnel. Each facet requires efforts and resources to be an effective enabler of sustainment.

Supplies. Vehicles need fuel and repair parts, weapons need ammunition, source of energy, and repair parts, and operators of systems need food, water, and other personal supplies. Systems with low TR scores may require extra supplies to handle more frequent repairs. The quantity and diversity of the supply items needed to operate the system affect the economic burden on sustaining the system, and the availability of these resources affects the operational readiness of the system. Other supply-related factors that affect the economic burden are transportation and storage costs of these supply items. The responsiveness and reliability of the supply chain affect availability of supplies and thus the operational readiness.

Facilities. Defense systems need storage, maintenance, and support facilities. Advanced weapons and C2 systems may also require expensive training and simulation facilities. Systems with low TR scores may require more extensive maintenance facilities than systems with high TR scores. The size, quality, and fitness of a facility will affect the readiness of the system that relies on them. For example, if a certain system requires certain environmental conditions, say, low temperature, for operating properly, then the quality and reliability of the air condition capabilities in the facility are crucial for making the system mission ready.

Personnel. All systems, as advanced as they may be, need humans to operate, control, and maintain them. Personnel with a variety of skills and trainings need to be available for those tasks. If, for example, a system requires a 24/7 human controller, then sustaining proper readiness will entail at least three qualified personnel, operating in eight-hour shifts, to keep the system mission ready.

R&S FACTORS

Based on the earlier discussion, we identify the following main factors that affect the R&S dimension:

- Mean time between failures (MTBF). This is one of the most significant factors affecting technical readiness. The complexity of the system, and the reliability of each of its components, determine the failure rate of the system, that is, the probability the system is technically fit at any given moment. Arguably, this parameter depends on the alternative's regular service and preventive maintenance schedule. This schedule, measured by the mean time between service (MTBS), is inversely related to the MTBF; smaller MTBS will increase the MTBF of the system and thus enhance technical readiness, at an increased cost rate.

- Mean time between services (MTBS). Each alternative system comes with instructions concerning regular service schedule and preventive maintenance actions. MTBS is measured by the frequency of such actions, as specified by the manufacturer.
- Repair time and service time. The system is down while it is in (scheduled) service or (unscheduled) repair following a failure. Obviously, during these down times the system is inoperable. The length of a down time depends on the complexity of the system and the availability of resources, such as personnel, facilities, tools, and spare parts. For example, a modular system that facilitates a plug-and -play–type repair technique would require less repair effort and therefore experience less down time than a nonmodular system. Note that repair and service are actions that only apply to technical readiness, as defined earlier.
- Repair cost and service cost. These are the costs for maintaining the system's technical readiness. These costs include spare parts, tools, infrastructure, and personnel. These costs can be reasonably estimated from analyzing the components of the system and from the manufacturer's specifications regarding service and preventive maintenance. A possible service cost is derived from designated uptime. Certain systems require frequent uptime to maintain their lifetime expectancy whereas others can stay dormant for longer times. The larger the uptime requirement is, the more difficult it is to sustain the system.
- Setup time and cost. A system may be, by design, in a cold operational stand-by condition (e.g., a system that is only activated in an emergency). The setup time and cost that brings the system into a fully operational state is a crucial aspect in measuring readiness (time) and sustainment (cost).
- Interdependency. Dependency on other systems makes an alternative more vulnerable to failure and potentially more disruptive when failed than an alternative, which operates as standalone. Examples include: (1) a vehicle that needs to be transported by other means to the area of deployment; (2) a moving platform (aerial, ground, or sea) that depends on satellite availability for its navigation; (3) a system that requires an extensive and expensive training facility to become operational; and (4) a sensor that is connected to an elaborate command-and-control system.
- Personnel. Any system requires operators, controllers, and technicians. Finding qualified and skilled personnel, training them, and then retaining them is always challenging. Ceteris paribus, a simpler alternative to operate and maintain, which requires a few easily trained operators and technicians, is preferred to a more complex alternative, which requires highly skilled and trained personnel.
- Supply chain. The availability of spare parts that facilitate technical readiness depends on an efficient and robust supply chain. Supply chain is also one of the principal means for making a system functionally ready. A vehicle needs fuel, a sensor requires electrical power, and a weapon will not operate without ammunition. The type of supplies (size, weight, scarcity, fragility, handling requirements), and the frequency at which they are needed, affect the cost of sustainment and the length and robustness of the supply chain. Part of the supply chain is the logistic tail of the deployable system. Certain systems require a large, expensive, or difficult-to-maintain tail to ensure their functionality (e.g., a convoy of supply trucks) whereas others require little or even negligible tails (sometime referred to as deploy-and-forget systems). Obviously, larger logistic tail requirements imply more difficulties in sustaining a system. An alternative that requires more frequent deliveries of expensive supplies by a more fragile supply chain is inferior to a more self-sustained alternative, which is supported by a simpler supply chain. There are several factors that determine the vulnerability of a supply chain—e.g., single source versus multiple sources, geographical distances between the nodes in the chain, the required transportation means, their availability, and their robustness to environmental conditions such as weather and terrain.

We note that these factors are not necessarily independent. For example, the effect of the supply chain on the overall R&S rating of an alternative depends on the repair and service costs; lower demand for spare parts makes a supply chain less crucial for the sustainment of the alternative. To avoid dealing with such dependencies, we will define measures for meta-factors, which combine similar factors into (relative) measures.

MEASURING THE R&S FACTORS

Recall that this study is concerned with the R&S dimension in the context of AoA where relative evaluations, rather than absolute ones, are sufficient. This observation is important because many of the aforementioned factors, and the meta-factors defined later, are not easily measurable. This phenomenon is rather common in MCDA problems and analysts apply value or utility functions to combine such measures. For example, one may apply ordinal preferences, such as Likert scale (Allen and Seaman, 2007), and combine them, in some consistent way, with measurable factors to produce an overall ranking of the alternatives with respect to the R&S dimension.

Next, we describe measures for evaluating the various facets of the R&S dimension.

Mean Time Between Downs

The MTBD is a combination of the MTBF and the MTBS. The mean time between failure of a system can be statistically estimated only after it has been in operation for some time and enough failure data has been collected. This is obviously not the case in an AoA setting where the alternatives are still in a development stage. One possible way to assess the MTBF is by considering how the components relate to each other in the system (i.e., in parallel or in series) and evaluating separately the reliability of each component, assuming such data is available. Integrating all this information, say, in a simulation, can produce a reasonable estimate for the MTBF. The MTBS is derived from the manufacturer recommended service and preventive maintenance schedule. This parameter should be given as part of the specification of the system. If T_S is the (deterministic) MTBS, and the failure process follows an exponential distribution with mean $\frac{1}{\lambda_F}$ (which may or may not be dependent on T_S) then, assuming a failure resets the service clock, the MTBD is given by:

$$MTBD = T_S e^{-\lambda_F T_S} + \int_0^{T_S} t\lambda_F e^{-\lambda_F t} dt.$$

Mean Down Time

The system is down while in (unscheduled) repair or (scheduled) service. If the mean repair time and the mean service time are μ_F and μ_S, respectively, then, assuming exponential distribution, the MDT is given by $MDT = \mu_F(1 - e^{-\lambda_F T_S}) + \mu_S e^{-\lambda_F T_S}$. Note that while μ_S can be directly estimated from the service and preventive maintenance specifications, μ_F is more elusive and may be estimated from simulation, similarly to the MTBF. Also note that the computation of both MTBD and MDT are easily generalized when the failure distribution is general, not necessarily exponential. Specifically, if the failure distribution has the CDF $F_F(t)$, and the down states generate a renewal process, then the MTBD and the MDT are, respectively,

$$T_S(1 - F_F(T_S)) + \int_0^{T_S} t dF_F(t) \text{ and } \mu_F F_F(T_S) + \mu_S(1 - F_F(T_S)).$$

Maintenance Cost

Maintenance cost (MC) is truly a meta-factor encompassing all the resources needed to maintain the system in operational state. These expenditures include fixed costs, denoted FMC, such as infrastructure (e.g., shops, storage facilities, labs, equipment, personnel) and variable costs covering replaceable parts, energy, and other resources needed for a specific maintenance mission. Standard practices of cost estimation (Mislick and Nussbaum, 2015) may be used to obtain estimates for the two types of cost. If the average variable maintenance cost per maintenance incident is denoted by VMC, then the average variable maintenance cost rate is $VMCR = \frac{VMC}{MTBD}$. Ignoring, for simplicity discount rates, and assuming a reference time horizon of length T time periods, the average maintenance cost per time period is:

$$MC = \frac{FMC}{T} + VMCR.$$

Operational Cost

Operational cost (OC) is relatively simple to compute because it relates to a fixed set of actions that need to be executed by the system. Such a set is typically well defined as it establishes the foundation for functional readiness. Operational cost is the cost of daily, or recurrent, operations. It can be measured by the number of operators and controllers, broken down by required skills, cost of operating facility (when applicable), and the amount and type of energy needed for the operation. If the functional characteristics of the system to be selected is such that it is dormant most of the time and is activated only when needed, then the operational cost also includes the setup cost and time required for activation. An alternative that can become active faster and at lower cost has lower operational cost than an alternative that takes time to set up. The parameter measuring operational cost is the average daily cost of operations.

Interdependency

The more a system depends on other systems and processes, the more it is vulnerable to possible breakdowns and failures of those peripheral systems and processes. Thus, such dependency leads to lower functional reliability. To capture this vulnerability, we first define, for each alternative, the set of peripheral systems and processes upon which it depends. We call it the systems' dependency set. Arguably, ceteris paribus, the larger the dependency set, the lower the functional reliability of the alternative because more things can go wrong. If an alternative is a fully standalone system, then interdependency has no effect.

We describe the state of the dependency set by a k-dimensional {0,1} vector x, where k is the cardinality of that set. If a system in the dependency set is up and running, its corresponding entry in the vector is 1, otherwise that entry is 0. For example, if k = 3, then the vector x = (1,0,1) indicates that the first and third peripheral systems in the dependency set of the alternative are up and running whereas the second system is down.

The functional readiness of the alternative depends on the state vector of its associated dependency set. We denote that effect by the function $D(x)$. For example, if the number of peripheral systems of an alternative is k = 3, then $D(1{,}1{,}1) = 1$ (no effect) and $1 = D(1{,}1{,}1) > D(1{,}1{,}0) > D(1{,}0{,}0) > D(0{,}0{,}0) \geq 0$. In other words, fewer functioning peripheral systems imply lower functional reliability of the alternative. In general, $0 \leq D(x) \leq 1$.

Let p_i, $i = 1, \ldots, k$ denote the probability that the ith peripheral system in the dependency set of the alternative is operational and functioning. Assuming independence, which in many cases is a reasonable approximation, we have that

$$p(x) = \Pr[x = (x_1, \ldots, x_k)] = \prod_{i=1}^{k} p_i^{x_i}\left(1 - p_i\right)^{1-x_i}.$$

The power set 2^x is the set of all possible realizations of the k-dimensional vector describing the state of the dependency set. We define the interdependency index of an alternative by:

$$INT = \sum_{x \in 2^x} p(x)D(x).$$

The higher the value of INT the more robust is the alternative with respect to its dependency on other systems.

Personnel

The cost of personnel (PER) is accounted for in the operational cost. There is another aspect of personnel that affects the functional and technical reliability of an alternative: the dependence on certain types of qualified personnel. The more a system relies on a large variety of skilled personnel, the more it is vulnerable to their possible absence. Thus, like the interdependency, such dependency may lead to lower functional and technical reliability. To capture the personnel vulnerability, we propose the same approach used for interdependency.

We define for each alternative the personnel dependency set, which comprises the skill set of persons needed for the operation of the system. We describe the state of the dependency set by a k-dimensional vector x of natural numbers, where k is the number of skill types (e.g., technicians of certain types or operators of different training levels) needed for operating and maintaining the system, and $x_i, i = 1, \ldots, k$, is the number of people of type i available at any given time. Let $s_i, i = 1, \ldots, k$, denote the number of people of type i required by the system at any given time. As in the interdependency case, we define $R(x) = R(x_1, \ldots, x_k)$, $x_i \leq s_i$, as the effectiveness of the system when the available personnel team is x. We have that $R(s) = R(s_1, \ldots, s_k) = 1$ and $R(x) \leq 1$. Trivial calculation show that there are $S = \prod_{i=1}^{k}(s_i + 1)$ possible profiles of personnel availability.

The personnel dependency measure is:

$$PER = \frac{1}{S}\sum_{x_1=0}^{s_1} \ldots \sum_{x_k=0}^{s_k} R(x_1, \ldots x_k).$$

An alternative that is reasonably functional with less personnel will have a higher PER score than an alternative that is sensitive to staffing.

For example, suppose $k = 3$ (three different types of personnel) and $s_1 = 2, s_2 = 1, s_3 = 1$. In other words, the system requires a team of four, say, two operators ($i = 1$), one controller ($i = 2$), and one technician ($i = 3$). We have $1 = R(2{,}1{,}1) > R(1{,}1{,}1) > R(1{,}1{,}0)$, etc. The closer the PER of a system to 1 the smaller the effect of reduced staff.

Supply Chain

Many factors affect the robustness of a supply chain (SC), for example, single source versus multiple sources of supply, geographical distances between the nodes in the chain, the type and availability of means of transportation, etc. The literature on SC has not reached a consensus on unified quantitative measures that can objectively assess the relative value of different SCs and hence we propose using an ordinal scale where the alternatives in the AoA are simply ranked by subject matter experts in the AoA team. This is the only qualitative and subjective input into the R&S part of the AoA. We assume that subject matter experts can produce such a ranking that may have ties for alternatives that are evaluated as having equally effective and robust supply chain.

Thus, the higher the rank SC of an alternative, the more robust and effective is its supply chain. Note that a higher rank implies a smaller (integer) value of SC.

Using DEA to Evaluate R&S

Recall that our goal is to evaluate the relative value of an alternative with respect to the R&S factors, not its absolute value. Suppose, for simplicity, that the R&S dimension comprises only two factors: MTBD and MC. Obviously we wish to have a reliable alternative with high MTBD and low MC. A reasonable relative measure for the alternative would be the reliability/cost ratio MTBD/MC. The higher this ratio, the better the alternative with respect to the R&S dimension.

So, if the only R&S factors were MTBD and MC, then we could easily rate the alternatives from best to worst. However, we have seven factors that affect this dimension. The challenge is how to extend the ratio idea described earlier with respect to MTBD and MC to all seven factors.

To meet this challenge, we propose to use the data envelopment analysis (DEA) methodology. Since it was first proposed in the late 1970s by Charnes et al. (1978), DEA has been applied to hundreds of application areas including several DoD-related applications such as evaluating the efficiency of air force maintenance units (Charnes et al., 1984). Other relevant applications are described in the following:

- Bowlin (1996) describes how DEA can be used to address various DoD evaluation procedures;
- Brockett et al. (1997) evaluate the efficiency of US Army recruitment units;
- Han and Sohn (2011) evaluate the performance of Korean Air Force bases;
- Falagario et al. (2012) propose DEA as a fair, equal, and transparent decision-making tool aimed at helping an awarding committee in selecting tenders that will meet governmental procurement regulations and requirements in accordance with European Union directives;
- Sutton and Dinitrov (2013) use DEA to assign sailors to tasks for the US Navy;
- Yang et al. (2016) present how DEA was adopted by the Taiwanese government as a method of choice to evaluate procurement alternatives;
- Lu et al. (2019) propose DEA to evaluate alternatives in selecting major weapon in a cost-effective manner and demonstrate it with an example of procuring fighter jets for the Singaporean government; and
- Boehmke et al. (2017) measure installation support activities in the US Air Force.

Essentially, DEA is a nonparametric methodology for comparing multiple entities, all of which use the same set of inputs (albeit, in different quantities) to produce the set of outputs (again, with different quantities). If $O_1, \ldots, O_m$ denote a vector of outputs and $I_1, \ldots, I_n$ denotes a vector of inputs then the efficiency ratio is:

$$\frac{x_1 O_1 + \ldots + x_m O_m}{y_1 I_1 + \ldots + y_n I_n}.$$

The question is what are the right weights $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$.

DEA evaluates the relative efficiency (i.e., determines the values of $x_1, \ldots, x_m$ and $y_1, \ldots, y_n$) of each alternative by solving a corresponding linear programming model that determines for the alternative the values of the aforementioned weights. The idea is to present each alternative in the best possible way while meeting some logical constraints. Normalizing the value of the highest relative efficiency to 100%, a system that receives a score that is smaller than 100% (i.e., its associated model failed to find a set of weights that will present it as 100% efficient) is by definition inefficient and DEA can assess the gap between its current performance and the efficient frontier that the methodology construct on the basis of the entities that were found to be 100% efficient. DEA was specifically designed to handle situations such as the one we face here as it is capable of handling data that is not easily converted into universal quantitative measure such as dollar. In our case, we have measures associated with time (MTBD, MDT), money (MC, OC), effectiveness (INT, PER), and an ordinal scale (SC).

In our context, we distinguish between factors for which more is better (MTBD, INT, PER) and factors for which less is better (MDT, MC, OC, and SC). Accordingly, we will consider the efficiency ratio:

$$ER = \frac{x_M MTBD + x_I INT + x_P PER}{y_D MDT + y_M MC + y_O OC + y_S SC}.$$

The presence of the ordinal factor SC necessitates a modification of the standard DEA model, as described in Cook et al. (1993).

Note that the selection of the input and output variables in this DEA framework is not rigid; it may vary according to the specific type of acquisition. For example, vehicles and weapons may have a common set of core variables but each may have additional factors, which are specific to the system under evaluation.

THE DEA MODEL

Let n denote the number of alternatives to be evaluated with respect to the R&S dimension. The index i, $i = 1, \ldots, n$, indicates the identity of an alternative. We solve n linear programming problems, one for each alternative. The data for the optimization model are the quantitative parameters $MTBD_i, INT_i, PER_i, MDT_i, MC_i, OC_i$, and the qualitative ordinal parameter $SC_i(k)$, $i = 1, \ldots, n$, $k = 1, \ldots, l$ where l is the number of rank positions, and

$$SC_i(k) = \begin{cases} 1 & \text{if alternative } i \text{ is ranked in the } k-th \text{ place} \\ 0 & \text{Otherwise} \end{cases}$$

The idea is as follows: each alternative, in its turn, is given the opportunity to select its coefficients such that it gets the highest possible efficiency ratio ER. It can do it as long as these best coefficients it selected, when applied to the efficiency ratio of any other alternative, does not exceed 1. Thus, the maximum possible efficiency ratio is ER = 1. Those alternatives, which reach 1, are considered R&S efficient. In addition to the normalization constraint that limits each efficiency ratio not to exceed 1, we require that all coefficients are nonnegative and the weights of the ordinal parameter (SC) adhere to the rank positions, that is, a weight of rank 1 should be lower than the weight of rank 2, and so on.

For each alternative, temporarily assigned the index 0, we solve the following linear optimization problem see Cook et al. (1993) for details:

$$\begin{aligned}
&Max\ x_M\, MTBD_0 + x_I INT_0 + x_P PER_0 \\
&st \\
&y_D MDT_0 + y_M MC_0 + y_O OC_0 + \sum_{k=1}^{l} w_k SC_0(k) = 1 \\
&x_M MTBD_i + x_I INT_i + x_P PER_i - \left(y_D MDT_i + y_M MC_i + y_O OC_i + \sum_{k=1}^{l} w_k SC_i(k) \right) \le 0,\ i = 1, \ldots, n \\
&\qquad\qquad w_{k+1} - w_k \ge \varepsilon
\end{aligned}$$

All decision variables $\geq \varepsilon$,

where ε is an arbitrary separation parameter determined by the decision maker. In the next section we illustrate this methodology on an example.

Table 1. R&S data for three alternatives.

Alternatives (systems)	MTBD (days)	INT	PER	MDT (days)	MC (K$ per day)	OC (K$ per day)	SC (ranking)
Alt1	33	0.7	0.7	3	18	5	I
Alt2	30	0.6	0.7	3	16	2	II
Alt3	20	0.6	0.5	2	15	4	III

EXAMPLE

At each milestone of the AoA more detailed and reliable information is gathered about the seven measures, MTBD, MDT, MC, OC, INT, PER, and SC. The data is entered into the DEA model and the relative standings of the alternatives are obtained with respect to the R&S dimension. Suppose there are currently three alternatives under consideration, and the R&S measures are as shown in Table 1.

We see from the data that no alternative dominates another with respect to all seven R&S measures. For example, Alt1 has the highest (best) MTBD, INT, and PER scores (it is tied with Alt2 with respect to the latter). Alt1 is also ranked highest for SC. Alt2 has the lowest OC, and Alt 3 has the lowest down time (MDT) and lowest MC. Thus, a simple inspection of the data will not reveal which alternative is more R&S efficient.

Running the DEA model presented earlier, where we select for each alternative its maximum feasible value of ε, we obtain that while the *ER* values for Alt1 and Alt2 are 1, that is, they are R&S efficient, for Alt3 $ER = 0.77$, which means that the best coefficients it could find for its data still rate it 23% lower than Alt1 and Alt2. Note that if ε is set at a value smaller than its maximum value, the differentiating power of the model decreases, that is, for a smaller ε all alternatives may be tied at the top. More on this differentiating effect in the next section.

SUMMARY

This report bridges a gap in AoA, addressing the role of examining R&S in such analyses. In the first part of the report, we define these characteristics and study their components. This study results in a set of MOEs that must be observed, and updated as new data becomes available, throughout the development process of a new system. In the second part of the report, we propose an analytic procedure, grounded in the well-established methodology of DEA to continuously assess the R&S aspects of the AoA.

Like all other quantitative methodologies, DEA has certain weaknesses that users must be aware of and be ready to address when the need arises. We describe next two such weaknesses along with recommendations on ways to overcome them.

Differentiation Power

DEA's differentiation power increases as the ratio between the number of alternatives and the number of MOEs increases. When this ratio is close to 1 (i.e., the number of alternatives is approximately the same as the number of MOEs), most alternatives are likely to be evaluated as fully efficient. In such scenarios, it is enough, for example, that the value of one of its numerator MOEs is larger than the corresponding values of all other alternatives to be evaluated as efficient. This phenomenon was discussed earlier. In the context of R&S evaluation, this scenario is quite likely as we have seven MOEs and the typical number of alternatives in an AoA study is less than 12. To overcome this difficulty, we recommend two possible remedies that have been implemented in similar situations elsewhere:

- Weight restrictions. The objective function of the model we presented above seeks to maximize the efficiency score for the alternative it evaluates. Thus, adding constraints to this model will cause a decrease in the efficiency score. The constraints should reflect qualitative assessment by the decision makers involved in the process. For example, putting some priorities on the weights (Roll et al., 1991).

- Adding alternatives. During the development of a certain alternative, there are several design and engineering options that could be explored, each generating another alternative. Also, one could add utopian alternatives, generated artificially based on experience, and use them as benchmarks.

Uniqueness

Certain alternatives may be affiliated with some attribute that doesn't exist in other alternatives and hence it makes them unique. A unique alternative is, by definition, fully efficient as we can't compare it to any other alternative. For example, suppose we evaluate the procurement of platforms that would transfer combat personnel from sea to shore and backward and that all but one of the alternatives are different kinds of vessels and only one alternative is airborne. The contractor of the airborne alternative may claim that the contractor's alternative is unique to ensure it is ranked as efficient, absent competitors. To avoid such claims, one should ensure that the MOEs are as general as possible yet relevant and useful. The way we presented the MOEs in this report is indeed quite general and we believe that in most cases it can be used as is.

REFERENCES

Allen, I. E., and Seaman, C. A. 2007. Likert Scales and Data Analyses, *Quality Progress*, Vol 40, No 7, 64–65.

Atwater, B., and Uzdzinski, J. 2014. Wholistic Sustainment Maturity: The Extension of System Readiness Methodology across All Phases of the Lifecycle of a Complex System, *Procedia Computer Science*, Vol 28, 601–609.

Barzilai, J., and Golany, B. 1994. AHP Rank Reversal, Normalization and Aggregation Rules, *Information Systems and Operational Research*, Vol 32, No 2, 14–20.

Boehmke, B. C., Jackson, R. A., Johnson, A. W., White, E. D., Weir, J. D., and Gallagher, M. A. 2017. Measuring US Air Force Installation Support Activities via Data Envelopment Analysis, *Military Operations Research*, Vol 22, No 1, 39–58.

Bowlin, W. F. 1996. Benchmarking and Efficient Program Delivery for the Department of Defense's Business-Like Activities, *Military Operations Research*, Vol 2, No 1, 19–30.

Brockett, P. L., Golany, B., Rousseau, J. J., Thomas, D. A., and Zhou, L. 1997. Forecasting and Allocation of US Army Recruiting Resources, *Military Operations Research*, Vol 3, No 3, 13–29.

Canician, M. F., and Daniels, S. P. 2018. The State of Military Readiness: Is There a Crisis? Center for Strategic & International Studies (April 18), https://www.csis.org/analysis/state-military-readiness-there-crisis.

Charnes, A., Clark, C. T., Cooper, W. W., and Golany, B. 1984. A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the US Air Force, *Annals of Operations Research*, Vol 2, 95–112.

Charnes, A., Cooper, W. W., and Rhodes, E. 1978. Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, Vol 2, No 6, 429–444.

Cook, W. D., Kress, M., and Seiford, L. M. 1993. On the Use of Ordinal Data in Data Envelopment Analysis, *Journal of the Operational Research Society*, Vol 44, No 2, 133–140.

Defense Acquisition University. 2019. Life Cycle Sustainment. *Defense Acquisition Guidebook*, Chapter 4.

Department of Defense. 2019. DOD Dictionary of Military and Associated Terms. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/dictionary.pdf.

Department of Defense. 2020. DOD Instruction 5000.02. https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/500002p.pdf?ver=2020-01-23-144114-093:DoD.

Doheney, S., Gray, S., McLemore, C., and Savage, S. L. 2019. Measuring Military Readiness: Open Standard Representations of Uncertainty: From "Ready or Not" to "How Ready for What," OR/MS Today (November 27), https://pubsonline.informs.org/do/10.1287/orms.2019.06.07/abs/.

Ewing, P. L., Tarantino, W., and Parnell, G. S. 2006. Use of Decision Analysis in the Army Base Realignment and Closure (BRAC) 2005 Military Value Analysis, *Decision Analysis*, Vol 3, No 1, 1–59.

Falagario, M., Sciancalepore, F., Constantino, N., and Pietroforte, R. 2012. Using a DEA-Cross Efficiency Approach in Public Procurement Tenders, *European Journal of Operations Research*, Vol 218, No 2, 523–529.

Government Accountability Office. 1999. Better Management of Technology Development Can Improve Weapon System. *The Defense Acquisition System*, United States General Accounting Office, GAO/NSIAD-99-162.

Headquarters Air Force. 2016. *Analysis of Alternatives (AoA) Handbook*. Headquarters Air Force HAF/A5R-OAS, Office of Aerospace Studies, Kirtland AFB, NM.

Han, H. K., and Sohn, S. Y. 2011. DEA Application to Grouping Airforce Bases, *Military Operations Research*, Vol 16, No 2, 31–40.

Harrison, T. 2014. Rethinking Readiness. *Strategic Studies Quarterl*, Vol 8, No 3, 38–68.

Kress, M., and Morgan, B. 2018. *A Robust Framework for Analyzing Acquisition Alternatives.* Naval Postgraduate School, Monterey, CA.

Lu, W. M., Kweh, Q. L., Nourani, M., and Shih, J. M. 2019. Major Weapons Procurement: An Efficiency-Based Approach for the Selection of Fighter Jets, *Managerial and Decision Economics*, Vol 41, No, No 4, 574–585.

Mislick, G. K., and D. A. Nussbaum. 2015. *Cost Estimation: Methods and Tools.* Wiley.

Moore, S. C., Stockfisch, J. A., Goldberg, M. S., and Holroyd, S. 1991. *Measuring Military Readiness and Sustainability.* RAND Corporation, Santa Monica, CA.

Rich, M., Cohen, I. K., and Pyles, R. A. 1987. *Recent Progress in Assessing the Readiness and Sustainability of Combat Forces.* RAND Corporation, Santa Monica, CA.

Robbins, M., Broyles, J. R., Giradini, J., Abel, K. V., and Boren, P. 2019. Improving DoD's Weapon System Support Program: A Critical Readiness Driver Approach. Report, RAND Corporation, Santa Monica, CA.

Roll, Y., Cook, W. D., and Golany, B. 1991. Controlling Factor Weights in Data Envelopment Analysis, *IIE Transactions*, Vol 23, No 1, 2–9.

Roy, B. 1968. Classement et choix en présence de points de vue multiples (la méthode ELECTRE). *La Revue d'Informatique et de Recherche Opérationelle*, Vol 2, No V1, 57–75.

Saaty, T.L. 2008. *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World.* RWS Publications.

Sauser, B. J., Ramirez-Marquez, J. E., Davanandham, H., and DiMarzio, D. 2008. A System Maturity Index for the Systems Engineering Life Cycle, *International Journal of Industrial and Systems Engineering*, Vol 3, No 6, 673–691.

Sutton, W., and Dinitrov, S. 2013. The US Navy Explores Detailing Cost Reduction Using Data Envelopment Analysis, *European Journal of Operational Research*, Vol 227, No 1, 166–173.

Yang, J. B., Wang, H. H., Wang, W. C., and Ma, S. M. 2016. Using Data Envelopment Analysis to Support Best-Value Contractor Selection, *Journal of Civil Engineering and Management*, Vol 22, No 2, 199–209.