CSBA

Center for Strategic and Budgetary Assessments

RETHINKING READINESS

TODD HARRISON

Rethinking Readiness

Todd Harrison

STRATEGIC STUDIES QUARTERLY

FALL 2014

VOLUME 8, NO. 3

THE CENTER FOR STRATEGIC AND BUDGETARY ASSESSMENTS

1667 K Street, NW, Suite 900
Washington, DC 20006
Tel. 202-331-7990 Fax 202-331-8019

The Center for Strategic and Budgetary Assessments (CSBA) is an independent, nonpartisan policy research institute established to promote innovative thinking and debate about national security strategy and investment options. CSBA's analysis focuses on key questions related to existing and emerging threats to U.S. national security, and its goal is to enable policymakers to make informed decisions on matters of strategy, security policy, and resource allocation.

Rethinking Readiness

Todd Harrison

In this era of austerity, the debate over the defense budget is, in many respects, a debate over readiness. Nearly every part of the defense budget is related to readiness in one form or another, whether through pay and benefits for military personnel, funding for training and maintenance, or the development and procurement of weapon systems. Over the next decade, the US military plans to spend more than $5 trillion dollars on readiness in all its forms.[1] To have an informed debate over the right level and allocation of defense spending, Congress and the nation first need a better understanding of what military readiness is and how budget decisions affect readiness.

The 2011 *National Military Strategy* defines readiness as "the ability to provide and integrate capabilities required by Combatant Commanders to execute their assigned missions."[2] The chairman of the Joint Chiefs of Staff's (CJCS) readiness system describes three levels of war-fighting readiness: strategic, operational, and tactical. The common thread in how the military defines readiness at all three levels is the ability of forces to perform the missions and tasks assigned to them.[3]

While there is broad agreement on the importance of readiness, these definitions fail to answer some basic questions. What does it mean to be ready? What are the attributes of a ready force? And how much readiness is enough? Readiness can mean the level of training or staffing of units. It can also refer to how well equipment is maintained or to the availability of supplies. It can refer to unit-level readiness or joint force readiness. It can be low or high but is rarely too high. Perhaps Richard Betts offered the best description of readiness, writing almost 20 years ago: "Although we may not know what readiness is, we know it when we see it, or, more often, when we do not see it."[4] Betts distilled the specific meaning of readiness into three fundamental questions:

Todd Harrison is the senior fellow for defense budget studies at the Center for Strategic and Budgetary Assessments. He joined CSBA from Booz Allen Hamilton, where he supported clients across the Department of Defense assessing challenges to modernization initiatives. He previously worked in the aerospace industry and served as a captain in the Air Force Reserve. Mr. Harrison is a graduate of the Massachusetts Institute of Technology with both BS and MS degrees in aeronautics and astronautics.

> *Readiness for what?* The most basic element of understanding readiness is knowing what types of wars the military must be prepared to fight. This includes potential adversaries it could face, the capabilities these adversaries are likely to possess, the conditions under which conflict may occur, and how the military plans to fight or deter such wars.
>
> *Readiness for when?* Readiness also depends on the time interval in which the military must be prepared to respond. Near-term readiness depends in part on the peacetime force posture, such as the mix of forces in the active and reserve components and the stationing of forces at home or overseas. Some conflicts could begin with little or no warning, greatly compressing required response times. Long-term readiness depends more on the capabilities the military is investing in for the future and how these capabilities will address the future threat environment.
>
> *Readiness of what?* One must also know which parts of the force must be ready, and the answers to the first two questions may vary for different parts of the force. Some elements of the force may need to be prepared for certain threats but not others. Likewise, some parts of the force may need to be ready for today's fight, some may need to prepare for tomorrow's fight, and some may need to be ready for both.[5]

The answers to all three of Betts' questions are fundamentally matters of strategy; what it means to be "ready" can only be understood in the context of one's strategy. For example, military strategy could emphasize defense in depth, mobilization, preemption, or forward defense. For the military and its civilian leaders to know if it is sufficiently ready, it must have a strategy that adequately describes what it must be ready for, when it must be ready, and what parts of the force must be ready. A strategy that does not define these attributes of readiness is, at best, incomplete.

For example, part of an overall defense strategy might be to use ground-based national missile defense forces to deter an adversary from launching ballistic missiles at the homeland. To achieve this objective, missile defense forces would need to be ready to detect, track, and launch interceptors. Since an adversary's offensive missile forces could reach targets in the United States within minutes of being launched, interceptors must be ready to respond within minutes to make a successful intercept possible. This description of the role of ground-based missile defense forces as part of an overall defense strategy answers each of Betts' questions:

- *Readiness for what?* Providing national missile defense for the homeland.

- *Readiness for when?* Within minutes of being notified.
- *Readiness of what?* Ground-based missile defense forces.

A defense strategy could also choose to take risks in near-term readiness, as the British did in the interwar period with the implementation of the "ten-year rule." In hindsight, the ten-year rule is often remembered as foolish and shortsighted because it remained in effect through 1933, and war came only six years later. However, from 1919 to 1929 the rule worked as intended and allowed Britain to reduce defense spending by cutting near-term readiness.[6] Near-term readiness, by definition, has a short shelf life. If the military is not used during the period it is kept at a high state of readiness, near-term readiness yields little value beyond its deterrent effect. Investments in long-term readiness, such as new technologies and capabilities, have the potential to yield value years or decades into the future. Of course, one never knows when threats may emerge and how much warning will be afforded—a risk inherent with any time-based strategy. A 10-year rule that is automatically extended year after year will eventually prove to be misguided.

Why "How" Matters

This article does not attempt to offer an overall strategy for the military or make recommendations for how readiness should figure into that strategy. Instead, it focuses on what to do once the questions of "readiness for what," "readiness for when," and "readiness of what" have been settled. The trillion-dollar question for defense is: How can resources be allocated most effectively to achieve the readiness required by strategy? Unlike Betts' three questions, the question of *how* to achieve readiness is fundamentally one of resource management rather than strategy and is of particular importance in an austere defense environment.

In the book *Moneyball*, Michael Lewis chronicles the story of how Billy Beane, the general manager of the Oakland Athletics, conducted a grand experiment to "rethink baseball." Knowing his team would never have the resources of wealthier teams like the New York Yankees, Beane began a systematic and scientific look for inefficiencies in baseball. By using new metrics (known as "sabermetrics" in baseball) to gauge the value of players and by understanding how these metrics contribute to winning games, Beane was able to build a roster of players and a winning record that in many ways upended the game.[7]

In defense, as in baseball, the way money is spent often matters as much as the total amount of money available. History is replete with examples of wealthier nations being defeated by more modestly resourced adversaries.[8] Understanding how best to resource readiness requires the same two things Billy Beane brought to the Oakland A's—better metrics and a better understanding of the relationship between inputs (resources) and outputs (readiness). With such an understanding, the inputs can be fine-tuned to produce a more ready and capable force for a given level of resources. The first section of this article examines the differences between measuring readiness inputs and outputs and proposes a method for developing strategy-based metrics for readiness outputs. Next, it explores methods to identify causal relationships among readiness inputs and outputs so resources can be optimized to achieve the readiness required by one's strategy. The article concludes by making specific recommendations to improve the way the US military measures and resources readiness.

Measuring Readiness: Inputs versus Outputs

> *I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter may be.*
>
> —Sir William Thompson (a.k.a. Lord Kelvin)

The way the US military thinks about readiness is driven in no small part by the way it measures readiness. Current readiness metrics focus on the inputs, such as flying hours, steaming days, tank miles, and training events. The military and Congress naturally focus on readiness inputs because they can monitor and control these directly through the budget. Readiness inputs are used as a proxy measure for the output—the ability of forces to perform the missions assigned to them. But an implicit assumption in this approach is that changes in the inputs will result in corresponding changes in the outputs. Moreover, it assumes that the target levels of inputs set by the military are optimal to achieve the types and levels of readiness required by defense strategy. When thinking about how the military can most efficiently and effectively achieve readiness, the first step is to reexamine how readiness is measured.

Readiness Inputs

While readiness is often associated with training, key inputs to readiness also include people, equipment, supplies, and maintenance. People are an important input because a ready force requires units that are staffed with a sufficient number of skilled military personnel. Units must also have a sufficient quantity of equipment and supplies on hand, such as munitions, major weapons systems, and support equipment, and this equipment must have capabilities appropriate to the missions assigned and the threats the force is likely to face. Equipment must also be properly maintained so it will operate reliably and effectively when needed. Training is needed to ensure people know how to operate their equipment and perform the tasks assigned under realistic conditions.

Virtually every part of the defense budget contributes to readiness in one form or another, as shown in figure 1. Operation and maintenance (O&M) accounts are a central component of readiness, providing funding for training, equipment maintenance, and some supplies. Military personnel accounts fund the pay and benefits necessary to recruit and retain a sufficient number of quality people.[9] Procurement and research, development, test, and evaluation (RDT&E) accounts fund the acquisition of equipment and supplies to ensure the force is equipped with weapons that are sufficient in quantity and capability. Together, these funding streams provide the basic inputs needed to produce a ready force.

Readiness funding is especially important at present given the fiscal constraints put in place by the Budget Control Act (BCA) of 2011.[10] Due to congressional reluctance to reduce military compensation or close excess bases and facilities, the DoD will have little choice in this drawdown but to cut some combination of the size of the force (the number of people and units), the amount of training, the quantity of equipment and supplies on hand, the capabilities of the equipment it procures, and/or the maintenance of equipment. In other words, the key inputs to readiness—people, training, equipment and supplies, and maintenance—are likely to suffer.

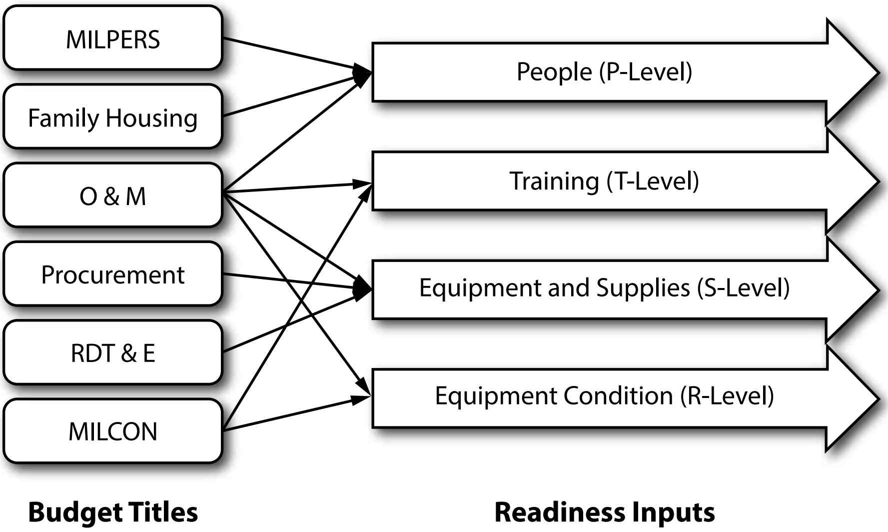

Figure 1. Mapping of budget titles to readiness inputs

The challenge for the Pentagon is to maintain balance among readiness inputs while attempting to fit within the resource constraints required by law and still support the strategy. When readiness inputs are out of balance, the result is what GEN Edward C. "Shy" Meyer famously termed a "hollow force." In his 1980 congressional testimony, General Meyer used the term to reference the inadequate number of soldiers available to fill Army divisions, but the term has since expanded in its use.[11] In a 1993 report to Congress entitled *Going Hollow: The Warnings of Our Chiefs of Staff*, Senator John McCain summarized the meaning of a "hollow force" as follows:

> Readiness is not a matter of funding operation and maintenance at the proper level. It is not a matter of funding adequate numbers of high quality personnel, it is not a matter of funding superior weapons and munitions, of funding strategic mobility and prepositioning, of funding high operating tempos, of funding realistic levels of training at every level of combat, or of funding logistics and support capabilities. Readiness is all of these things and more. A force begins to go hollow the moment [it] loses [its] overall mix of combat capabilities in any one critical area.[12]

While there is general agreement that a hollow force is one in which the inputs are out of balance, the question remains, what is the optimum balance of inputs? More specifically, how can readiness be measured so the balance of inputs can be optimized over time to achieve the highest level of readiness possible with a given set of resources?

Current Metrics

The Status of Readiness and Training System (SORTS) has been used since 1986 to report the readiness of individual units across the services. The SORTS compares the level of inputs to target amounts determined by the services. Individual units are measured on a scale of one to four (with one being the best score) in four resource areas: personnel (P-level), equipment and supplies on hand (S-level), equipment condition (R-level), and training (T-level).[13] The resource areas used in the SORTS map directly to the readiness inputs listed in figure 1. Units also report an overall score, called a C-level, which is equal to the worst score among the four resource areas; however, unit commanders have the discretion to raise the C-level by one increment if they believe it does not reflect the unit's true readiness.[14]

The SORTS scoring system is based on inputs rather than performance, as noted in the CJCS guide to the readiness system.[15] The SORTS does not attempt to measure the ability of units to carry out the missions assigned to them. Instead, resources are used as a proxy—a stand-in measure—for performance. The SORTS assumes, by definition, that if all resource areas meet their target levels then a unit will be fully ready. It further assumes that the target levels of resources set by the services are correct, both in the total level required in each resource area and in the relative weighting of resources among the four areas. Yet the target levels could be excessive, insufficient, or irrelevant to actual readiness.

An additional difficulty with the SORTS is that the target levels set for each type of input do not account for substitution effects. For example, a unit with a shortfall in its target number of personnel but with excess funding in training could potentially compensate by cross-training personnel so individuals can fill multiple jobs. In some instances, a unit could increase overall training so a smaller number of better-trained personnel could achieve a level of readiness equivalent to a larger number of lesser-trained personnel. The input measures in the SORTS, however, do not account for this possibility, with the exception of commanders using their authority to raise the overall C-level of a unit subjectively.

In 1999, the DoD began developing the Defense Readiness Reporting System (DRRS) in response to criticisms and shortfalls in the SORTS.

The DRRS was initially intended to replace the SORTS but has since been modified to include SORTS metrics and improve upon the SORTS reporting system. Perhaps the most significant difference with the DRRS is the inclusion of a commander's self-assessment of whether a unit is ready to perform the missions and tasks assigned to it on a three-level scale: yes, qualified yes, and no. The DRRS also automates the calculation of the resource-level scores used in the SORTS according to the rules and target levels established by each of the services. The results are expressed on a scale of 0–100 rather than 1–4, as is the case with the SORTS C-ratings.[16]

The DRRS attempts to measure readiness more directly by asking commanders to assess unit readiness against the list of mission essential tasks assigned to each unit. A commander's self-assessment of his or her unit's readiness "incorporates a judgment about not only the specific resources and training a unit has but also other factors, such as morale or confidence, that are not quantitatively captured in the resources and training metrics."[17] Self-assessments create an incentive for commanders to inflate unit readiness to avoid telling superiors that the unit under their command is unfit for combat.

Beginning in 1996, Congress mandated that the DoD provide quarterly reports on military readiness. The Quarterly Readiness Report to Congress is a classified report, typically hundreds of pages in length, which attempts to satisfy Congress' reporting requirements using data compiled from the services, Joint Staff, and Office of the Secretary of Defense (OSD).[18] Of the 26 specific reporting elements set by Congress, none requires the DoD to report measures of readiness *outputs*—the ability of forces to perform the missions assigned to them.

As shown in table 1, each of the reporting elements set by Congress relates to readiness inputs or the overall readiness ratings generated by the SORTS and DRRS. For example, four of the reporting elements relate to the level of training, but none of them requires the DoD to report on the results of that training. Many of the items have a clear connection to readiness, such as equipment availability and mission-capable rates (i.e., the percentage of time equipment is capable of being used), but some items do not have a clear connection to readiness. For example, it is not readily evident how the age of equipment affects the readiness of forces using that equipment. If the average age of main battle tanks increases by one year, does that mean the readiness of armored units has declined? Do fighter squadrons with 15-year-old aircraft perform better than squadrons with 20-year-old aircraft of the same type?

Table 1. Congressionally mandated readiness reporting requirements

Reporting Requirement	Corresponding Resource Area
Personnel status, including the extent to which personnel are in positions outside of their specialty and/or above their grade	Personnel
Historical data and projected trends in personnel strength and status	
Recruit quality	
Borrowed manpower	
Personnel stability	
Personnel morale	
Recruiting status	
Training unit readiness and proficiency	Training
Training operations tempo	
Training funding	
Training commitments and deployments	
Deployed equipment	Equipment and Supplies
Equipment availability	
Availability of ordnance and spares	
Equipment that is not mission capable	Maintenance
Age of equipment	
Condition of non-pacing items	
Maintenance backlog	
Status of prepositioned equipment	
Overall readiness rating for units rated C-3 or below for the quarter and each month of the quarter by unit designation and level of organization	Overall Readiness Ratings
Resource areas that adversely affected the readiness rating for units rated C-3 or below	
Each readiness problem and deficiency identified using internal DOD assessments	
Planned remedial actions to address readiness problems and deficiencies	
Key indicators and other relevant information related to each identified problem and deficiency	
Readiness of the National Guard to support the National Response Plan in support of civil authorities	
Reasons why the unit received a readiness rating of C-3 or below	

Source: Government Accountability Office, *Military Readiness: Opportunities Exist to Improve Completeness and Usefulness of Quarterly Reports to Congress* (Washington: GAO, 2013), 5.

A 2013 GAO report faulted the DoD for not fully and consistently complying with all of the reporting requirements set by Congress.[19] This criticism, however, misses the larger point. The reporting requirements Congress established in law do not capture readiness outputs—the ability of forces to perform the missions and tasks assigned to them. By focusing on the inputs to readiness rather than the outputs, Congress is not requesting or receiving the information it needs to make informed resource allocation decisions. It is the equivalent of judging the performance of a baseball team by the size of its payroll (an input) rather than number of games it has won (an output). Teams with a larger payroll do not necessarily win more games. While the resources available to hire more talented players can certainly affect the performance of a team, just as readiness inputs logically affect readiness outputs, many other factors can be at work as well.

The Circular Logic of Readiness

Reporting the status of readiness using inputs creates a circular chain of logic when those reports are used to justify the inputs required. In other words, the readiness reporting system is used to justify a certain level of readiness inputs, but the readiness reporting system is merely a measurement of the inputs it is used to justify. This approach implicitly assumes that outputs are directly proportional to inputs—that is, if the inputs increase, readiness will increase.

As the DoD noted in its most recent budget submission, the inputs to readiness are "non-linear variables [that] work together to produce ready forces." Numerous studies have established a nonlinear relationship between training and performance for a variety of jobs in the military, with the best correlation often being a power law or log(n) function.[20] Nonlinear systems can behave in complex and unexpected ways because the output is not directly proportional to the input. For example, increasing the flying hours of a squadron could harm its readiness if crews are forced to fly to the point of fatigue. In practice, a pilot's flying time is limited in both the military and commercial aviation because excessive flying has been shown to reduce performance and increase accident rates.[21]

Nevertheless, the DoD has continued the circular logic of using inputs to justify inputs when appealing to Congress for readiness funding. For example, in congressional testimony on the effects of sequestration on readiness, Deputy Secretary of Defense Ashton Carter testified that

"The consequences of sequestration and a lowering of the discretionary caps are serious and far-reaching. In the near-term, these reductions would create an immediate crisis in military readiness."[22] As evidence of a readiness crisis, he offered the following examples:

- The Army would have to cancel as many as five "full-spectrum training rotations" and "reduce maintenance for units that are not scheduled to deploy to Afghanistan."
- The Air Force would "be forced to cut flying hours sharply and will reduce remaining weapon system sustainment funding by about 30 percent."
- The Navy and Marine Corps would be forced to cut back on "fleet operations."[23]

In each of these examples, the specific reductions cited are reductions in the inputs to readiness. The Army would be forced to reduced training and maintenance; the Air Force would be forced to reduce flying hours and sustainment funding; and the Navy and Marine Corps would be forced to cut peacetime operations. The department is essentially arguing the obvious—a reduction in readiness inputs will result in a reduction in readiness inputs. While it is generally accepted that cuts in readiness inputs will harm readiness outputs, it is not clear how much harm would be done. By not reporting measures of readiness outputs—the ability of forces to perform the missions assigned to them—it is difficult for the DoD to make a compelling case for maintaining readiness funding.[24]

Toward Better Metrics

The revolution in baseball ushered in by Billy Beane's Oakland A's began decades earlier as an attempt to rethink baseball metrics. In 1977, Bill James, who had served briefly in the Army and was working as a night security guard, published his first book, the *Bill James Baseball Abstract*.[25] In this book, and in subsequent editions, James questioned some of the basic metrics used in baseball, such as runs batted in, errors, and batting averages, to measure a player's performance and, by extension, his value. James pointed to the inadequacy of these measures, writing that "baseball statistics are not pure accomplishments of men against other men . . . they are accomplishments of men in combination with their circumstances."[26] In his annual *Baseball Abstract*, James began to develop useful and relevant metrics to answer some of the most important questions in baseball, such as how much an individual player contributes to the overall success of his team. It was a pragmatic approach born at just the right time—skyrocketing baseball salaries meant more was at stake for the teams, and advances in computing power in the 1970s and 1980s meant that large volumes of data were easier to accumulate and analyze.[27]

In many ways, the US military may now be at a similar turning point when it comes to readiness. The cost of readiness—training, staffing, equipping, maintaining, and operating forces—has grown to the point that the DoD cannot maintain the size of force it has today with the budget constraints Congress has placed on it.[28] At the same time, advances in data networks, data storage, and sensors mean that information on the maintenance and utilization of equipment as well as the performance of personnel using this equipment can be collected, tracked, and analyzed in ways that were not possible just a few years ago. The military appears to be entering an era of increasingly constrained resources and unconstrained data, therefore, the way it measures readiness should adapt to these changing circumstances.

A "big data" approach to measuring readiness is only useful if the metrics being collected help answer important questions, such as how do readiness inputs affect the ability of forces to perform the missions assigned to them? Current metrics are not particularly useful in this respect. The SORTS method of measuring readiness is based on inputs, and thus it sheds little light on how readiness inputs affect outputs. The DRRS is largely based on input measures as well since it incorporates and aggregates SORTS data. The exception in the DRRS is the commander's self-assessment of unit readiness using a three-tier scale: yes, qualified yes, or no. At the discretion of the commander, this self-assessment may or may not be based on a unit's actual performance in operations or in training exercises.[29] At best, the DRRS offers a subjective, low-fidelity measure of readiness. Subjective self-assessments of readiness are like judging the performance of baseball players by asking their coaches how they are doing instead of keeping track of key statistics of their actual performance, like hits and on-base percentage. What the military needs is a box score for readiness—quantitative measures of the relevant performance attributes of forces.

Strategy-Based Metrics

Betts' questions of readiness for what, readiness for when, and readiness of what should drive the development of strategically relevant readiness metrics. Readiness metrics should measure the ability of forces to perform the missions and tasks assigned to them by the strategy. Each unit's mission essential task list (METL) specifies the tasks it is expected to perform as part of its core capabilities and to support top-priority plans and named operations in the strategy. As the strategy changes and evolves, these tasks can change, and readiness metrics should adapt as well.

As an example of how readiness metrics should flow from overall strategy, consider a tactical fighter unit. A hypothetical strategy could require the military to be prepared to fight a major theater war, which could include attacking military targets defended by enemy fighters and ground-based air defenses. A mission for an Air Force fighter squadron under this strategy could include being able to penetrate enemy defenses, deliver the attack, and recover to base, all at a specified level of proficiency. Mission essential tasks could include air-to-air combat to defeat enemy fighter defenses and low-altitude navigation and bomb delivery to avoid ground-based defenses. Key readiness metrics for a fighter squadron with these assigned tasks should therefore include measures of how effective fighter crews are in air-to-air combat and low-altitude bombing. Measuring bomb delivery ability is relatively straightforward and quantifiable using bomb miss distances in training exercises. Air-to-air combat skills are more difficult to quantify objectively, but the performance of fighter crews can be measured in simulated air-to-air missile launches during combat training missions using recordings of aircraft heads-up display data to assess whether a missile launch was within established launch parameters.[30]

The purpose of strategy-based metrics is to measure how capable units are in performing the mission essential tasks assigned to them by the strategy. These metrics should provide a greater level of fidelity than the simple yes, no, or maybe self-assessments used in the DRRS and should be based on objective measures whenever possible. Metrics must also adapt over time as strategy, technology, and forces change. As Barry Watt and James Roche have noted, "technological and other changes can erode the appropriateness of the criteria by which we have become accustomed to assessing a given category of weapons or forces."[31]

An unavoidable challenge in developing readiness metrics is that one must quantify what in many cases is a combination of quantitative and qualitative factors. The resources Congress appropriates for defense each year are inherently quantified in dollars. Since the purpose of measuring readiness is to understand how these inputs can be best allocated to achieve a desired output—the trillion dollar question—one must have a quantifiable measure of that output, even if it is merely a subjective assessment on a numeric scale. If one cannot quantify the output and know if it has increased or decreased by some amount, then one cannot know if the inputs applied are sufficient or insufficient. As Lord Kelvin adroitly noted, "when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind."[32] To be useful in understanding how readiness inputs affect readiness outputs, metrics should enumerate the *degree* of one's ability to perform assigned tasks.

The level at which readiness data is collected (individual, unit, combined unit, or joint force) should be whatever level is relevant to the strategy and most practical for collection. For some parts of the force structure, such as a fighter pilot's ability to hit targets, individual readiness may be important and measurable. In contrast, it may only make sense to measure readiness at the unit or combined unit level for combined arms operations.

Performance scores from training events are one potential source of readiness data since these events should already be testing the missions and tasks assigned to particular units, including all of the supporting tasks needed to accomplish assigned missions. In many cases, the services already conduct the necessary testing as part of routine training exercises—they merely need to record, aggregate, and report the scores. Units also routinely engage in competitions to test their skills against other units. In Air Force fighter squadron bombing competitions, for example, units from across the force compete in various bombing categories.[33] Rather than reporting readiness inputs, like flying hours and maintenance levels, the DoD should be reporting readiness outputs, like average bomb miss distances.

Periodic testing of individual and unit-level proficiency will also need to be conducted independent of major training events and competitions because the very act of being measured can alter one's performance—a phenomenon known as the observer effect. For example, in preparation for a major training exercise, such as an Army National Training Center rotation, units often increase their level of readiness by increasing training, reassigning personnel to fill vacancies, and taking equipment and supplies from other units. Units that have recently been through a major exercise may also experience a post-exercise reduction in readiness due to the stress and fatigue of the exercise itself and the loss of personnel and equipment temporarily loaned to the unit. To account for the observer effect—or even better, to measure the observer effect—units should periodically be tested at random with minimal notification given and restrictions placed on what they can borrow from other units.

Despite one's best attempts, readiness metrics will never be perfect measures. Measuring the performance of forces in realistic combat scenarios is not a perfect substitute for performance in actual combat. In war, outcomes are not solely determined by the readiness of US forces, or more specifically, the performance of forces in the mission essential tasks assigned to them. In air-to-air combat, for example, success may be a function of many factors beyond the pilot, such as jamming provided by other aircraft or the capabilities of enemy air defenses. In addition, forces are also often tasked in wartime to perform missions they were not designed to perform and for which may not have been trained. Combat outcomes can only be assessed through actual warfare, and readiness is just one of many contributing factors. The best that one can achieve with peacetime readiness assessments is an *approximation of performance* short of actual military operations.

For these reasons, readiness metrics will never be a perfect predictor of how forces will perform in actual combat. Rather, the goal should be to develop metrics that come closer to measuring the relevant performance characteristics of the force and to continue improving and refining these metrics over time. The most important criteria for readiness metrics are that they should (1) measure outputs rather than inputs, (2) be linked to the strategy, (3) be quantifiable, and (4) avoid subjective assessments (particularly self-assessments) where possible. Most importantly, readiness metrics should be developed that help answer the trillion-dollar question: how can the military most effectively achieve the readiness required by its strategy?

Why Outputs Matter: A Case Study in Air-to-Air Combat Skills

In a 1999 RAND report, Dr. John Stillion examined the effects of training and experience (readiness inputs) on the ability of fighter pilots to perform certain mission essential tasks (readiness outputs). One of the specific areas examined was air-to-air combat skills, which is perhaps the most mentally and physically demanding skill required of fighter pilots. As Stillion notes, air-to-air combat can be compared to "simultaneously playing the piccolo, driving a formula-one race car, and bench pressing 200 pounds."[34]

The data analyzed in the study included 137 simulated air-to-air missile launches recorded during training missions for a particular fighter squadron from 1 October 1997 to 28 February 1998. Videos of the pilots' heads-up displays were used to determine whether the launching and target aircraft were within the proper parameters (e.g., range, velocity, angle, etc.) for a missile to be effective. Of the 137 shots recorded, 19 were assessed to be invalid, meaning the missile would have likely missed its target.[35]

Stillion identified a statistically significant correlation between the number of days since a pilot's most recent practice missile launch and the probability of a valid shot. Interestingly, the analysis did not find a statistically significant correlation between the performance of pilots who were instructors versus non-instructors—a proxy measure for experience and overall skill level. What seemed to matter most was recent practice in air-to-air combat. The analysis showed a "strong logarithmic relationship between the probability a pilot launches an out of parameters simulated air-to-air missile shot and the number of days since he last exercised his air-to-air combat skills." Pilots who had practiced simulated missile launches within the past 10 days had an average hit rate of 93 percent, whereas pilots whose most recent practice was a month or longer had an average hit rate of 78 percent.[36]

The type of flying performed—specifically, the amount of practice in air-to-air combat—appeared to affect air-to-air combat skills more than the total number of hours flown or experience. Pilots who had longer gaps between simulated missile launches were still flying, perhaps more hours in some instances, but these hours did not include practice in air-to-air combat. Moreover, the analysis found that air-to-air combat practice must be relatively recent to have a significant impact on performance. Overall experience levels and practice conducted more than 30 days prior did not appear to affect performance as much as practice within the past 10 days.

This type of analysis is only possible when readiness metrics are used that measure the actual performance of forces. If flying hours were used as the measure of readiness instead of missile hit rates, then readiness in this case study would have appeared to increase when more hours were flown and decrease when fewer hours were flown. It would not have revealed that readiness for air-to-air combat depends on a specific type of flying and how recently that training was conducted. Using inputs as readiness metrics can obscure the true readiness of forces to perform the missions and tasks assigned to them.

Resourcing Readiness: An Experimental Approach

> *Human knowledge and human power meet in one; for where the cause is not known the effect cannot be produced.*
>
> —Sir Francis Bacon

In a 2011 study, the Congressional Budget Office (CBO) concluded that the DoD has not been able to identify a clear link between readiness spending and actual readiness, noting that "the military's current measures of readiness are not readily applicable to such analyses, and there are some concerns about the quality of its assessments of readiness."[37] Without a firm grasp of the causal relationships between inputs and outputs, readiness may be underfunded, overfunded, or out of balance and exacerbated by sequestration. The result could be a hollow force, or worse, a hollow force masked as a ready force. Disentangling the many cause-and-effect relationships among readiness inputs and outputs is a challenging task requiring the tools of science.

The Military as a Complex System

The military can be viewed as a complex system that in peacetime transforms resources (inputs) into ready forces (outputs). Because this transformation involves a set of interactions among the inputs, many of which may be nonlinear, the output can at times appear random or unexpected. An additional complexity is that the system itself is dynamic. The rules by which it is governed are constantly changing as technology, threats, operational concepts, and the military itself change.

For example, unmanned systems such as the RQ-4 Global Hawk can loiter for longer periods than manned aircraft, with flights lasting 32 hours or longer. This capability enables new missions for the military, such as the ability to provide continuous surveillance over wide swaths of territory that would not be logistically feasible with manned aircraft. The relationship between readiness inputs and outputs is fundamentally different for unmanned systems because simulators can provide realistic training for pilots, reducing the need for costly training flights. This reduces not only the cost of training, but also the number of platforms procured for training and the number of support personnel needed to operate and maintain training platforms. The readiness costs of using an unmanned system, once these training and personnel savings are factored in, can be half that of a comparable manned system.[38]

Moreover, with increasing levels of automation, a single operator can control (or monitor) multiple unmanned systems simultaneously, further reducing personnel requirements and the associated training pipeline. Fully autonomous systems and the robotics revolution taking hold in the military have the potential to flip the notion of readiness on its head. While humans require regular practice to maintain certain skills, once software is written and tested, it does not need recurring practice because its abilities do not degrade with time.

Like many complex, dynamic systems, military readiness does not readily lend itself to simple models. As Betts noted, "good models of operational readiness are difficult to formulate because their subject is in large part an ecological phenomenon, a jumble of vectors whose interdependencies are hard to trace or isolate."[39] This makes it difficult to establish causal relationships between the resources allocated to readiness and the performance of forces. While models are a useful tool in understanding readiness, models alone are insufficient to capture such a complex and ever-changing system.

Fortunately, unraveling the many causal relationships among readiness inputs and outputs does not require an understanding of the precise interactions that occur within the military system. Like many complex systems, these internal mechanics can be regarded as what is commonly referred to as a "black box."[40] One does not need to know what goes on within the black box to develop a functional understanding of how it transforms inputs into outputs.

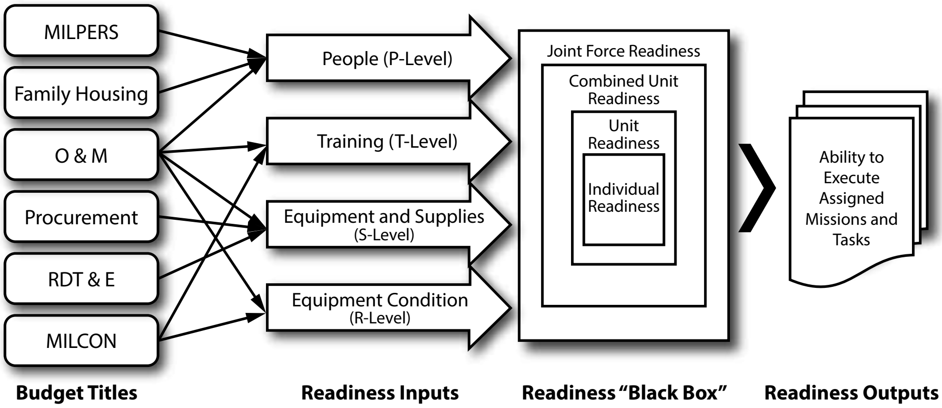

Figure 2. Military readiness as a "black box"

Consider how the knowledge of other complex systems, such as the human body, has advanced without understanding the precise mechanisms by which these systems work. Aspirin, one of the most commonly used drugs in the world, was developed thousands of years ago without any real understanding of how or why it relieved pain, fevers, and inflammation. Salicylic acid, the active ingredient in aspirin, can be derived from the bark of willow trees. The earliest known reference to the use of willow bark for medicinal purposes is a stone tablet from the Ur Dynasty in Mesopotamia dating to around 3000 BC. The same tablet also included references to using magic and spells as cures for common illnesses.[41]

Aspirin was discovered through 5,000 years of trial and error. People looked for an input to the human body that would produce a desired output: relief from pain, fever, and inflammation. The human body was treated as a black box, only knowing that a certain input produced a certain output. The mere fact that willow bark produced a desirable effect meant that its use was passed down from one generation to the next, while other treatments that did not work, such as magic and spells, were eventually abandoned. The active ingredient in willow bark was not isolated and synthesized until the nineteenth century AD, and even then, its makers did not understand how or why it worked. Only in the 1970s did scientists begin to unravel the precise chemical pathways through which aspirin interacts with the human body to produce its desired effects.[42]

A "New Method" for Resourcing Readiness

Nearly 400 years ago, Sir Frances Bacon challenged other scholars to apply a more rigorous approach to developing theories for complex systems, such as nature. As Jim Manzi notes in his book, *Uncontrolled*, Bacon recognized that nature is "extraordinarily complicated as compared to human mental capacities," and that "humans tend to over interpret data into unreliable patterns and therefore leap to faulty conclusions."[43] Perhaps Bacon's most important insight was that scientists should focus their energies on developing practical rules that approximate how the world works rather than trying to discern philosophical truth.[44] Others built upon Bacon's revolutionary ideas to develop what is known as the scientific method—a process that effectively compresses the amount of time it takes to determine cause-and-effect relationships. Instead of taking thousands of years of trial and error to determine the efficacy of medicines like aspirin, it now takes only a few years of randomized controlled trials.

Experiments have also proved valuable in understanding social phenomena, such as how humans make decisions. In the 2012 election, the Obama campaign used randomized controlled experiments to test the effectiveness of everything from phone scripts and flyers to the subject lines of e-mails seeking donations.[45] Companies test marketing strategies by conducting experiments so they can more accurately predict customer behavior and fine-tune their messaging and targeting of customers. For example, Target has developed the ability to identify when women are pregnant in their second trimester and send them relevant coupons based on changes in their shopping behavior.[46] Experiments were also used to measure the effectiveness of a counterinsurgency program in Afghanistan, the National Solidarity Program. Villages were randomly selected to participate in the program, and the experiments showed it was effective in reducing the level of violence, but only in villages that were relatively peaceful already.[47]

Understanding a complex technical and dynamic social system like the military requires an iterative and dynamic process much like Bacon's *Novum Organum* ("New Method"). This process, as applied to military readiness, is shown in figure 3. It begins with the collection of relevant readiness metrics. Statistical analysis of this data is used to identify trends and correlations, which form the basis for building theories and associated models of readiness. These theories yield hypotheses for how the system works, which are tested through experiments that control for other variables not being tested. The results of these experiments produce new data to update readiness theories and models and produce more hypotheses for testing. The following section describes how such an iterative process can be used to build a better understanding of the causal links among readiness inputs and outputs.

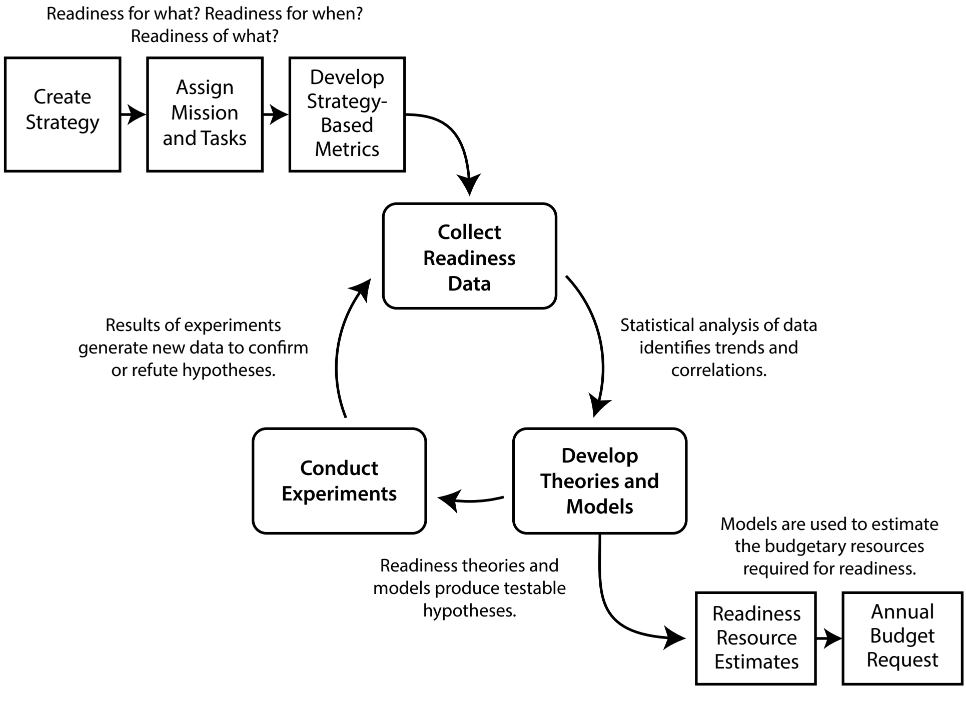

Figure 3. Proposed method for linking readiness inputs to outputs

Collect Readiness Data. At the most basic level, understanding the relationship between readiness inputs and outputs begins with data collection. For data to be useful, they must be based on relevant metrics. The current understanding of readiness is limited by the lack of strategy-based metrics collected, aggregated, and reported on a widespread basis. Many modern weapon systems automatically record a tremendous amount of data for maintenance and training purposes. This data could be repurposed to measure operator and system performance, and software could be modified to collect additional data if needed.

The collection of data over time allows for statistical analysis and the observation of natural experiments. Natural experiments occur when one or more inputs are varied in part of the force for reasons unrelated to the collection of data. This allows for a comparison between units whose inputs were altered and those whose inputs were not. Natural experiments are not a substitute for randomized controlled experiments, because the reason units are selected to have their inputs changed may be due to factors that bias the outcome. For example, if a service has a shortfall in training money it may deliberately chose to cut training for units that are already in a depressed state of readiness to protect the readiness of its top-tier units.

Develop Theories and Models. Statistical analysis of readiness data can identify correlations and form a preliminary assessment of which variables seem to matter and which do not. Statistical analysis can also show which inputs are most closely correlated with which types of readiness outputs and quantify the sensitivity of outputs to changes in the inputs. Through inductive reasoning, many specific observations can be used to build a broader and more generalized theory of readiness.

A readiness theory is a set of basic ideas and principles for relating readiness inputs to outputs—an intellectual framework for thinking about the problem. A readiness model puts these ideas into practice by codifying them in formal mathematical relationships. Because of the complexity involved, it is impractical and indeed unnecessary to build a single theory or model of readiness that encompasses all parts of the force. The structure, capabilities, and resources of units vary widely across the military, and the missions assigned to units can differ considerably. Different types of units require different theories and models of readiness.

The difficulty in using historical data as a basis for generating theories and models is the problem of counterfactuals. Natural experiments and historical data reveal what happened under a particular set of circumstances. This type of data cannot reveal what would have happened had the inputs or circumstances been different. Other variables not being measured or controlled for—known as hidden conditionals—could be the actual cause of any observed correlation.[48] More observations collected over time can build stronger correlations and suggest modifications to readiness theories and models, but correlations alone cannot establish a causal link between readiness inputs and outputs. Moreover, correlations can be misleading when used outside the bounds of previous observations.

Despite these limitations, readiness theories and models are important for two reasons. First, models provide a practical way of estimating the resources required using the best available understanding of readiness. The military must submit an annual budget request each year, and nearly every part of the budget is related in some way to readiness. Readiness models provide an imperfect but rational basis for estimating what resources are needed and in what proportions. However, readiness models should always be held as provisional and subject to revision as additional data become available.

Readiness theories are important because they provide a broad framework for understanding how the system behaves, from which specific, testable hypotheses can be generated. A testable hypothesis is a predictive statement specific enough that it is possible to design an experiment that proves it wrong—it is falsifiable. Strictly speaking, experiments can never prove a hypothesis and its associated theory true. Rather, experiments can show that a theory has passed numerous falsification tests. The lack of falsifying evidence builds confidence that a theory is true and is therefore more likely to be reliable in practice.[49]

Conduct Experiments. The third component of the process, and arguably the most important, is to test specific hypotheses using randomized controlled experiments in which inputs are varied and the resulting outputs are measured. Controlled experiments, as opposed to uncontrolled, assign part of the subjects being tested to a control group in which the inputs are not varied. Control groups are essential for understanding the counterfactual of what would have happened had the inputs not been changed. Randomization is important because randomly assigning units to the test and control groups helps isolate the effects of hidden conditionals that could bias the results. Blinding prevents those being tested and those assessing the results from introducing their own biases to the experiment. In a double-blind experiment, for example, neither the subjects nor those running the experiment know which subjects fall into the test and control groups. Multiple independent experiments can also help reduce the chance of hidden conditionals or biases affecting the results.[50]

The process of testing readiness theories creates a self-correcting feedback loop to continually refine and update one's understanding of readiness, including the readiness metrics being used. If prior assumptions about the relationships among readiness inputs and outputs are correct, these assumptions should stand up to the scrutiny of rigorous experiments. One may find, however, that these assumptions do not hold true or that the readiness output metrics initially selected are not appropriate measures for the types of readiness required by one's strategy. When experiments yield results inconsistent with current readiness theories, they should not be regarded as failures. Rather, the goal of conducting experiments is to find such counterexamples so theories, models, and metrics can be revised to reflect reality more accurately. As Manzi notes, the process is referred to as "trial and error" not "trial and success."[51]

Comparison to Current Method for Resourcing Readiness

Figure 4 shows the current method for resourcing readiness in the same framework as the new method proposed in this article. As shown in the gray-shaded areas, the current method for resourcing readiness lacks two key components: strategy-based metrics and experiments to test hypotheses.

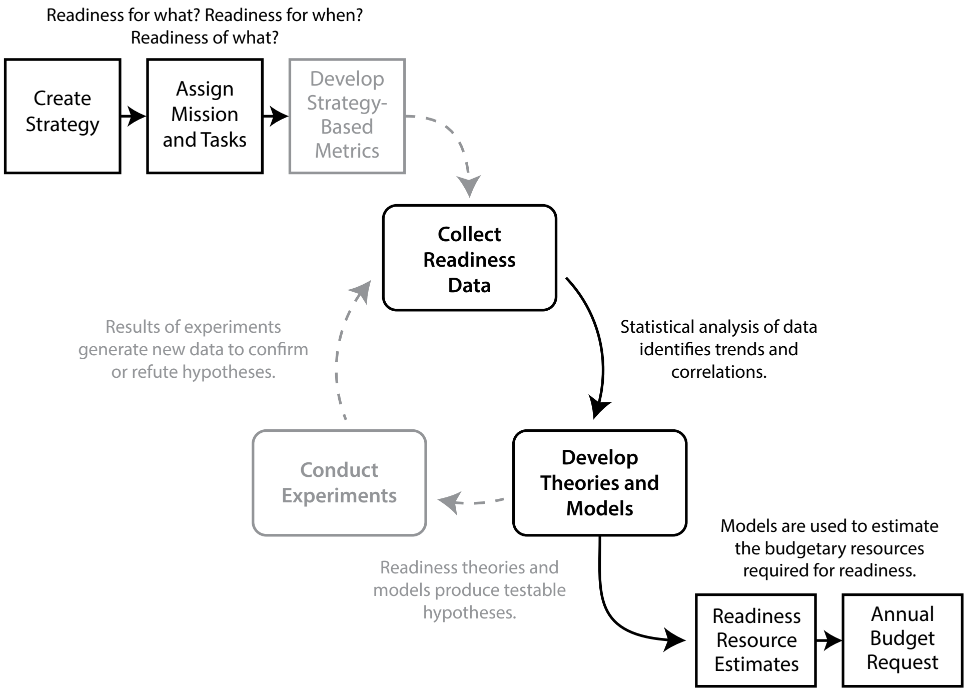

Figure 4. Gaps in current method for resourcing readiness

Strategy-based metrics are vital because they connect strategy, assigned missions, and mission essential tasks to the readiness data being collected and analyzed. Nearly all of the readiness data collected now through the SORTS and DRRS are not connected to the strategy and are in fact measures of readiness inputs rather than outputs. These metrics provide little insight into the ability of forces to perform the missions assigned to them. Without this link between strategy and readiness metrics, the military may be collecting the wrong readiness data, which in turn leads to theories and models that produce erroneous or unjustifiable resource requirements.

Without experiments, readiness models must rely on correlations identified through statistical analysis of historical data. While this analysis is useful, correlations cannot prove causal relationships because they cannot account for hidden conditionals that may be the underlying cause of any observed correlations. Controlled experiments help isolate the underlying causal relationships between inputs and outputs. Conducting such experiments is also necessary to create a self-correcting feedback loop to account for the complex and dynamic nature of the military. Without a feedback loop, the military could miss disruptive shifts in the conduct of war, such as the introduction of new technologies and operational concepts or the emergence of new threats. An experimental approach allows readiness theories and models to accommodate such changes more quickly.

The current DoD method for resourcing readiness starts with the wrong metrics, lacks experimental data to isolate causal effects, and does not have a continuous feedback loop to update and refine readiness theories and models. Without these important steps in the process, the DoD is operating with significant blind spots when it resources readiness. The military could be significantly overfunding or underfunding readiness without knowing it. Worse still, it cannot reliably predict how changes in resources will affect readiness.

The lack of good readiness data naturally makes the military resistant to changes in readiness resources. The current balance of inputs—people, training, equipment and supplies, and maintenance—was crafted through years of war-fighting experience, and these inputs appear to work, as is evident by the high performance of US forces in recent military operations. However, this does not mean the current levels of inputs are optimum or efficient. Moreover, what worked in the past may not work in the future, because future wars may be fundamentally different.

As resources become more constrained, the DoD will likely be forced out of its current balance, whether by another sequester or more deliberate changes to strategy and resources. The accumulated wisdom resident in today's military for what is needed to produce a ready force should not be dismissed or disregarded. Rather, it should be the starting point for developing a more robust and adaptive method for resourcing readiness. A more rigorous method for resourcing readiness allows for the possibility that the conventional wisdom guiding these resource decisions may be wrong, may be right but suboptimal, or may be right only under certain conditions.

Why Experiments Matter: Revisiting Air-to-Air Combat Skills

In Stillion's case study of fighter pilot air-to-air combat skills, the data collected was from a single fighter squadron (the Air Force's 4th Fighter Squadron) operating a single platform (F-16Cs) over a five-month period.[52] When the pilots went more than a month without any simulated air-to-air missile launches, it was because they were deployed to Saudi Arabia to conduct patrols of the no-fly zone in Iraq.[53] While deployed, the pilots did not practice their air-to-air combat skills; they flew racetrack patterns in the skies over Iraq.

This data was used because it was the best data available at the time and was a natural experiment worthy of analysis. The conclusions derived from this data beg several follow-on questions. Was it the lack of practice that caused pilots' combat skills to degrade or something else related to the deployment, such as a drop in morale from long family separations? Was there some other factor at work in this particular squadron that could have caused the decline in performance? Perhaps the 4th Fighter Squadron was already a low-performing squadron relative to others, and its air-to-air combat skills were fragile to begin with. Or perhaps there was a change in squadron leadership or equipment maintenance following its deployment that affected performance. It is thus possible that the pilots' skills would have declined even if they had continued practicing their air-to-air combat skills while deployed.

Historical data cannot demonstrate what would have happened had the situation been different—the counterfactual. Controlled experiments can. To test whether the observed decline in air-to-air combat skills was caused by a lack of recent practice, one could devise several experiments to increase the confidence that a lack of recent practice was the causal factor in declining performance. One experiment could randomly assign fighter pilots from multiple squadrons into test groups and a control group. The test groups would get practice at different frequencies ranging from less than 10 days between training for one group to several months between training for another group, while those assigned to the control group would continue their normal training and deployment routine. Randomly assigning pilots to test and control groups helps account for any hidden conditionals that may be at work, and using a control group provides a reference for what would have happened had no changes been made. Another experiment could randomly assign pilots deployed to the no-fly zone to a test group that receives regular air-to-air combat training while deployed or a control group that does not.

The data collected from the 4th Fighter Squadron shows that air-to-air combat skills appear to degrade quickly and to recover quickly. This suggests a "10-day rule" for air-to-air combat, akin to Britain's 10-year rule in the interwar period, and presents a testable hypothesis: As long as US forces have at least 10 days warning before engaging in air-to-air combat operations, these skills can be maintained at a low level in peacetime and ramped up quickly when conflict is imminent. If true, such an approach would save significant resources during peacetime that could be redirected to other priorities, such as equipment modernization.

While it would be foolish to adopt a 10-day rule as policy based on such a narrow dataset, experiments could be used to test this hypothesis before deciding whether to apply it widely across the force. For example, one could test whether pilots with longer gaps in air-to-air combat practice require more time and training to regain their competence. One might discover a breaking point, perhaps several months or years, at which more than 10 days of training must be conducted to regain skills. Similarly, one could test whether the overall experience of pilots (i.e. how much accumulated training they have had) affects how quickly their skills can be recovered.

Of course, experiments are not always possible, particularly large-scale, randomized controlled experiments, and performance in training is not a guarantee of performance in actual combat. Success in air-to-air combat, for example, depends on many factors outside the control of a fighter squadron, such as the effectiveness of jamming from other platforms and intelligence on enemy air defenses. While there is no perfect solution, there is certainly room for improvement in the current approach to resourcing readiness. The point is to introduce a mechanism for testing hypotheses by the best means available and refining readiness theories and models based on the results. Just as it would be wrong to assume that a 10-day rule would work without supporting experimental evidence, it is equally invalid to assume that current training regimens are necessary. Readiness theories not subject to regular testing and refinement are essentially educated guesses.

Conclusion

Even without the pressures of austere defense budgets, readiness is too important to ignore and too complex for guessing. The US military needs a more robust and adaptive approach to understanding readiness, particularly in an era of increasingly constrained resources. Such an approach begins with strategy-based metrics that identify the important readiness outputs to measure and includes controlled experiments to test hypotheses and continually update readiness models. The purpose of this approach is to answer the trillion-dollar question: how can the military most effectively achieve the readiness required by its strategy?

The chief recommendation of this article is that both the DoD and Congress should revisit the way readiness is measured and resourced. The DoD should use existing METLs to identify the key tasks required of each unit in support of overall defense strategy. From these key tasks, it should develop quantifiable performance measures, using objective standards where possible, and report this data both internally and to Congress. Congress should review these strategy-based metrics, determine which are most useful for oversight and resource allocation, and amend the quarterly reporting requirements in Section 482 of Title 10 of the US Code to include these metrics. Once strategy-based metrics are in place, each of the services should conduct controlled experiments to test existing hypotheses for how resources affect readiness. Moreover, the services should institutionalize the process of developing hypotheses, conducting experiments, and continually refining readiness theories and models.

Now is the perfect time to rethink how readiness is funded, because budgetary and legislative constraints are likely to force the military to cut readiness resources. Rather than making these cuts based on preconceived notions of what will have the least impact on readiness, the DoD should use this occasion to test such hypotheses and collect data. The coming readiness crisis is an opportunity for the DoD to fine-tune its understanding of readiness so it can squeeze the maximum military advantage out of each defense dollar.

For the military to rethink readiness, it must change its process, metrics, and, perhaps most critical of all, its culture. Rethinking readiness requires a culture of experimental inquiry that encourages leaders to question everything they think they know about resourcing readiness, including what factors matter, what factors do not matter, and how much readiness is enough.

Clearly, improved readiness metrics and a better understanding of the causal relationships among readiness inputs and outputs are not a cure-all for the challenges the US military faces due to sequestration. Better metrics and experimental data cannot compensate for a poor strategy, a force that is mismatched to the strategy, or resources that are insufficient to execute the strategy. A more effective allocation of resources, however, can give the United States a significant fiscal and military advantage by enabling it to afford a larger, better-trained force for a given level of resources or the same size force at a lower level of resources. Greater efficiency in the allocation of military resources can be a source of enduring strategic advantage.

If the US military does not take advantage of this opportunity to rethink how it resources readiness, an adversary may do so and use its readiness advantage to challenge the United States in peacetime competition or in actual conflict. Rethinking readiness funding is not just about efficiencies and savings; it is a matter of maintaining the US military's preeminence as the best equipped, best trained, and most highly capable force in the world. SSQ

Notes

1. If the Budget Control Act (BCA) caps currently in effect are extended through FY-2024, the DoD will receive roughly $5.5 trillion over the next decade, not including war-related supplemental funding.

2. *The National Military Strategy of the United States of America* (Washington: DoD, 2011), 20.

3. *CJCS Guide to the Chairman's Readiness System* (Washington: DoD, 15 November 2010), 1–2.

4. Richard K. Betts, *Military Readiness: Concepts, Choices, Consequences* (Washington: Brookings Institution, 1995), 5.

5. Ibid., 33.

6. Ibid., 53–58.

7. Michael Lewis, *Moneyball: The Art of Winning an Unfair Game* (New York: W. W. Norton, 2004), 74–75.

8. For example, in the American Revolution a wealthy nation, the United Kingdom, was defeated by a less wealthy adversary in the Continental Army.

9. O&M also funds the military health care system through the Defense Health Program, which supports military personnel in the force.

10. The BCA established annual limits for the base defense budget through FY-2021.

11. Congressional Budget Office, *Trends in Selected Indicators of Military Readiness, 1980 through 1993* (Washington: CBO, March 1994), 2.

12. John McCain, *Going Hollow: The Warnings of our Chiefs of Staff* (Washington: Office of Senator McCain, July 1993).

13. *CJCS Guide*, 10. A rating of five is also possible for units temporarily taken out of service for reorganization or depot maintenance.

14. R. Derek Trunkey, "Implications of the Department of Defense Readiness Reporting System," CBO Working Paper 44127, May 2013, 4.

15. *CJCS Guide*, 10.

16. Trunkey, "Implications," 7–9.

17. Ibid., 3.

18. Government Accountability Office, *Military Readiness: Opportunities Exist to Improve Completeness and Usefulness of Quarterly Reports to Congress* (Washington: GAO, July 2013), 5.

19. Ibid., 7–8.

20. John Stillion, *Blunting the Talons: The Impact of Peace Operations Deployments on USAF Fighter Crew Combat Skills* (Santa Monica, CA: RAND Corp., 1999), 25–27, 46–48.

21. See Federal Aviation Administration, *Regulatory Impact Analysis: Flightcrew Member Duty and Rest Requirements* (Washington: Department of Transportation, 18 November 2011).

22. Deputy Secretary of Defense Ashton B. Carter, "Prepared Testimony for the House Armed Services Committee," 13 February 2013, 1.

23. Ibid., 8–9.

24. Congress allowed sequestration to go into effect two weeks after Deputy Secretary Carter's testimony.

25. Lewis, *Moneyball*, 74–75.

26. Ibid., 75–76.

27. Ibid.

28. The 2012 Defense Strategic Guidance explicitly addresses the tradeoff between readiness and capacity, saying, "we will resist the temptation to sacrifice readiness in order to retain force structure." See *Sustaining U.S. Global Leadership: Priorities for 21st Century Defense* (Washington: DoD, January 2012), 7.

29. Trunkey, "Implications," 9.

30. Stillion, *Blunting the Talons*, 80–81.

31. James Roche and Barry Watts, "Choosing Analytic Measures," *Journal of Strategic Studies* 14, no. 2 (June 1991): 167–68.

32. Sir William Thomson, "Electrical Units of Measurement," *Popular Lectures and Addresses*, vol. 1 of 3 (London: Macmillan, 1889), 73–74.

33. For an example, see Christine Griffiths, "357th Fighter Squadron wins the 56th Fighter Wing 2012 Turkey Shoot," 355th Fighter Wing Public Affairs news release, Luke AFB, AZ, 30 November 2012.

34. Stillion, *Blunting the* Talons, 80.

35. Ibid., 82.

36. Ibid., 82–83.

37. *Linking the Readiness of the Armed Forces to DoD's Operation and Maintenance Spending* (Washington: CBO, April 2011), 2.

38. For a more detailed explanation of the savings from using unmanned systems, see Todd Harrison and Evan Montgomery, *Changing the Business of Defense* (Washington: CSBA, October 2011), 3–5.

39. Betts, *Military* Readiness, 88.

40. The term *black box* refers to any process or system that is hidden from an observer. The observer sees the inputs and outputs of the black box without knowing what goes on within it.

41. Diarmuid Jeffreys, *Aspirin: The Remarkable Story of a Wonder Drug* (New York: Bloomsbury, 2005), 16.

42. "100 Years of Aspirin," Aspirin Foundation, 2014, http://www.aspirin-foundation.com/what/100.html.

43. Jim Manzi, *Uncontrolled: The Surprising Payoff of Trial-and-Error for Business, Politics, and Society* (New York: Basic Books, 2012), 18.

44. Ibid., 19.

45. Sasha Issenberg, "How President Obama's Campaign Used Big Data to Rally Individual Voters," *Technology Review*, 19 December 2012, http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/.

46. Charles Duhigg, "How Companies Learn Your Secrets," *New York Times Magazine*, 16 February 2012, http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=3&_r=2&hp&pagewanted=all.

47. Todd Harrison and John Speed Meyers, *Contracting Under Fire: Lessons Learned in Wartime Contracting and Expeditionary Economics* (Washington: CSBA, 2012), 21–25.

48. Manzi, *Uncontrolled*, 27.

49. Ibid., 27–28.

50. Ibid., 30.

51. Ibid.

52. Stillion, *Blunting the Talons*, 80.

53. Ibid., 81–82.

American Revolution
War of 1812
Mexican War
American Civil War
First Barbary War
Opening of Japan
Spanish Florida
Peru
Gulf of Mexico
North Africa
Caribbean
Counter-Piracy
Undeclared Naval War
in the Atlantic