Eliminating the Weakest Link Approach to Army Unit Readiness

Paul L. Goethals & Natalie M. Scala

Author Accepted Manuscript to the journal *Decision Analysis*

November 2017

Posted via INFORMS Green Option

DOI to final article: https://doi.org/10.1287/deca.2017.0366

Eliminating the Weakest Link Approach to Army Unit Readiness

Paul L. Goethals & Natalie M. Scala

Accepted to *Decision Analysis*

Abstract

One of the most difficult measurements to obtain with some level of accuracy is military readiness. While a multitude of factors exist that affect the ability of a unit to achieve success in mission, an accurate assessment of readiness is crucial and drives federal funding, defense policy, and deployment decisions. The current readiness metric for the United States Army statically assesses units on personnel, equipment on-hand, equipment readiness/serviceability, and unit training proficiency, using a weakest link approach. This leads to reporting challenges and the tendency for commanders to subjectively upgrade their units' assessments. This research proposes a metric that evaluates units with greater precision, flexibility, and robustness. By taking a decision analysis approach and using desirability functions, we are able to measure readiness based upon a set of priorities, adapting for type of mission and unit. We test our metric using notional case studies and discuss extensions to other branches of the United States military and beyond.

Keywords: readiness; applications: military; policy; simulation; multi-response optimization

1. Introduction

Given the high demand for United States military forces in today's security environment, military readiness is a frequent topic of interest in United States defense news. The significance of what defines a military unit as being "ready" cannot be overstated; the concept drives a countless number of critical functions and processes. Readiness is directly tied to the level of federal funding required for the development and sustainment of military forces. The benchmarks for readiness also influence defense policy regarding the assignment of personnel and equipment, as well as the maintenance of that military equipment. Moreover, the decision to deploy military forces into a combat zone or peacekeeping mission is undeniably linked to the understanding that these forces are manned, trained, and equipped at the appropriate level. In summary, inaccuracies in measuring military readiness not only affect monetary costs but also equate in some way to survival risk and the probability of successfully executing a mission.

The task of measuring readiness, however, is not trivial in nature. Military units vary in a wide number of attributes, such as size, cost, structure, personnel or equipment composition, and specialty. They may naturally be positioned at the forefront of combat, such as with infantry or armor units, or in more protected areas, such as with field hospital or hazardous material disposal units. In addition, military units may be called upon to perform a wide range of tasks, from conventional attacks to humanitarian assistance or peacekeeping operations. Moreover, a number of researchers point out that *defining readiness* is also inherently complicated (Harrison 2014; Bipartisan Policy Center 2017; Rumbaugh 2017). In order to model these complexities imposed by the readiness problem and develop meaningful metrics that evaluate all relevant information, approaches from decision analysis are needed.

Prior to designing a metric, it is important to identify the qualities deemed most critical to the measurement of military readiness. Perhaps one of the most important requisites is that the metric rating be *responsive* enough to negative or positive fluctuations in readiness attributes. This enables one to more easily compare and contrast among units and also prevents a situation where dissimilar metrics have the same overall influence on readiness. At the same time, the metric should be *robust* enough to these same changes; that is, small changes in readiness attributes should not result in large measurement rating shifts. In addition, a readiness metric should enable observation of short or long-term trends in time as a means to monitor the effects of available resources on the ability to deploy. This feature also promotes greater predictive capability and facilitates improved awareness of impacts to readiness by more intangible concepts like modernization or equipment aging. Furthermore, it is preferable for a military readiness metric to be *flexible* in terms of weighting influential attributes, accounting for different types of units with varying missions or objectives. A metric with this property would enable senior leaders to alter, or manage, a unit's readiness priorities, based upon its structure, purpose, and deployment posture over time. Finally, a quality commonly sought among researchers of military readiness metrics is increased *objectivity* (Moore, et al. 1991; Orlansky, et al. 1997; Junor 2017); the intent is to remove as much bias as possible, so that output may not be negatively or positively skewed in any way.

The five armed services of the United States military – Army, Navy, Air Force, Marine Corps, and Coast Guard – evaluate the readiness of their systems in slightly different ways. For this paper, our research focus is with regard to Army readiness. Although we illustrate the intricacies of our approach based upon the one service, the philosophy can be extended to the other United States military armed services. In order to establish a foundation for this research, we will first describe the evolution of readiness metrics and a detailed account of the Army's current measure. Then, upon providing an argument for modifying the current metric, an alternative index is presented and evaluated using a simulation methodology. A graphical comparison of current and proposed metrics is also performed to illustrate measurement output. Finally, extensions of this research are offered in the conclusion of this paper. To facilitate an understanding of our research methodology, a table to define notation is included as Appendix A.

2. Background

We first describe the United States Army's current readiness system under the Defense Readiness Reporting System (DRRS) and then provide a policy background of the history of readiness along with related research in this area.

2.1 The Current Army Readiness Metric

For United States Army units, Army Regulation (AR) 220-1 provides the authoritative policy for reporting readiness requirements. The Commander's Unit Status Report (CUSR), which is generally submitted monthly by Army battalions, separate companies / detachments, and higher or equivalent-size units, consists of unit administrative information, readiness measurements, and commander comments. The primary basis of the readiness measurements is an assessment of a unit's core functions (C-level rating), a quantitative measure which is the lowest overall rating of four subordinate category assessments in personnel (P), equipment on-hand (S), equipment readiness/serviceability (R), and unit training proficiency (T). For each of these subordinate categories ($PSRT$), there are distinct quantitative bins aligned with reporting a specific rating of "1" (being most ready) to "4" (being least ready). The four readiness components, corresponding focus areas, and rating levels are outlined in Table 1.

For categories such as Personnel, where there are more than one subordinate rating focus area, the overall P-level is the lowest of these subordinate ratings. For example, if a unit has an 85% available personnel strength with 85% available Military Occupational Specialty (MOS) qualified and 70% available senior grade, that unit will report P-3 for personnel. Furthermore, if the same unit described above maintains a rating of "1" or "2" in the other categories (S, R, and T), the unit reports C-3 overall, the lowest of the $PSRT$ rating levels.

C-Level Sub-Category	Rating Focus Area	Rating Level*			
		1	2	3	4
Personnel (*P*)	Available Strength	90-100%	80-89%	70-79%	≤ 69%
	Available MOS** Qualified Strength	85-100%	75-84%	65-74%	≤ 64%
	Available Senior Grade Strength	85-100%	75-84%	65-74%	≤ 64%
	Composite Senior Grade Level	≤ 1.54	1.55 - 2.44	2.45 - 3.34	≥ 3.35
Equipment/Supplies On-Hand (*S*)	ERC A / ERC P *** Items On-Hand	90-100%	80-90%	60-79%	< 60%
Equipment Serviceability (*R*)	Fully Mission Capable Equipment	90-100%	70-89%	60-69%	< 60%
Training Proficiency (*T*)	Mission Essential Tasks Trained	85-100% (no untrained tasks)	70-84% (no untrained tasks)	55-69%	< 55%

*All percentages are rounded to the nearest whole number
**MOS refers to Military Occupational Specialty, the specific job for which a soldier is trained
***ERC refers to the Equipment Readiness Code, whereby ERC P are pacing items (critical equipment items) and ERC A are mission essential items (calculations are based upon ERC P alone and ERC A/P combined for 21 items or more per line item number)

Table 1: *PSRT* Reporting Level Constraints (AR 220-1 2010)

While the lowest of the individual *PSRT* levels cannot be altered and defines the overall *C*-rating, the commander is given the flexibility to perform a “subjective downgrade” or “subjective upgrade” of the overall rating. The intent of this added feature is to align the resulting overall *C*-level with the commander’s subjective assessment of his or her unit’s capability. To assist the commander in this process, a qualitative scale is established that should coincide with a unit’s assessed *C*-level of 1, 2, 3, or 4. Generally, this scale aligns the status of resources and training with the ability of the unit to accomplish its mission, whereby there is either no impact (*C*-1), some impact (*C*-2), significant impact (*C*-3), or the complete inability to accomplish the unit mission without additional resources and training (*C*-4). If the commander disagrees with the assessed *C*-level, he or she has the opportunity to subjectively upgrade or downgrade the overall *C*-level readiness rating. The range of the upgrade or downgrade depends on the echelon of the unit and the approval authority of the commander. For example, at the battalion level, a commander can only upgrade or downgrade the *C*-rating one level. In order to upgrade or downgrade the *C*-rating two levels, approval of the next higher echelon (brigade) must be obtained.

Each subordinate assessment addresses a specific component of a unit's ability to execute its mission. Detailed descriptions of these assessments and how the *PSRT* rating levels are calculated are provided below.

(i) Personnel (*P*) Data. The *P*-level is calculated to be the least rating of three subordinate categories: Available Strength (*AS*), Available MOS Qualified Strength (*AQS*), and the Available Senior Grade (*ASG*) Composite Level. The required strength for each of these categories is governed by the Headquarters, Department of the Army (HQDA) in its authorization document, the Modified Table of Organization and Equipment (MTOE). *AS* is computed by taking the total number of available soldiers in the unit and dividing by its authorized strength. By the definitions given in AR 220-1, if a soldier is considered "deployable", then he/she is also considered "available." For *AQS*, a percentage fill is calculated for every duty position by rank, the lowest percentage being matched against Table 1 to determine the overall *AQS* rating. And, for the *ASG* composite level, a percentage fill is calculated by taking the available personnel and dividing by the authorized strength in each of five categories of officers: junior Non-Commissioned Officers (NCOs) (Army ranks E5-E6), senior NCOs (Army ranks E7-E9), warrant officers (WO), junior officers (Army ranks O1-O3), and senior officers (Army ranks O4-O6). The overall composite level is calculated as an aggregate of the five categories based upon the values in Table 1.

(ii) Equipment On-Hand (*S*) Data. For each piece of equipment in a unit, a percentage fill using on-hand numbers is calculated. The overall *S* level is computed based upon two sets of equipment: (i) pacing (P) items, which consist of the primary equipment items a unit requires in order to perform its mission, and (ii) A-coded equipment, the required secondary equipment to support mission completion. Using the MTOE to identify equipment authorization, two sub-categories are established: one containing all pacing items and A-coded equipment together (*S:AP*) and the other containing solely pacing items (*S:P*). The lowest fill percentage within each of these subcategories is used to determine the overall *S*-level using Table 1.

(iii) Equipment Serviceability (*R*) Data. Given the total time in a reporting period (typically thirty days), the number of available days is computed for each individual pacing item and piece of A-coded equipment. Available equipment refers to those items that are not down for maintenance or repair. Similar to the data for the *S* level, two subcategories are established, one containing all pacing items and A-coded equipment together (*R:AP*) and the other containing solely pacing items (*R:P*). The lowest availability percentage within each of these subcategories is used to determine the overall *R*-level using Table 1.

(iv) Training Proficiency (*T*). The unit commander makes a subjective assessment of the unit's ability to perform its fundamental doctrinal tasks. Each task is assigned a *T*, *P*, or *U* status, depending on whether the unit is "trained," "needs practice," or is "untrained," respectively. If no task is assessed as *U*, then the percentage of trained tasks by Table 1 is used to determine the overall *T*-level. If one or more of the tasks are assessed as *U* and remaining tasks are assessed at least 55% being *T*, the overall *T*-level is "3". Otherwise, the *T*-level is reported as "4".

The current readiness system defined in AR 220-1 is a culmination of many revisions of the policy, including revisions in 2001, 2003, and two revisions in 2006. However, researchers since the 1960s have examined the readiness philosophy and assessments. Over time, updates to the readiness methodology have been proposed, and these approaches can be organized into types. We now summarize this background and literature below. Note that some researchers have examined specific aspects of a unit's readiness program, such as manpower in terms of an assignment problem or training in terms of a scheduling problem. Because this research methodology is more closely associated with readiness indices, the review is restricted to this topic. Furthermore, note that the current metric does not necessarily encapsulate all previous research initiatives.

2.2 Previous Research Initiatives in Unit Readiness

While each branch of service reports under the overarching framework of the DRRS, the basis of determining the overall readiness rating is slightly different. Generally, all of the services consider the categories of personnel, equipment, and training in calculating an overall readiness value; in addition, they enable the commander to subjectively upgrade or downgrade an overall rating. For specialized equipment such as aircraft in the Navy or Air Force, greater emphasis is given toward the proficiency of pilots. Ordinance and available facilities are also identified as pillars for the Navy and Air Force in developing their assessment. With subtle differences in generating an overall readiness rating, it is not surprising that previous research efforts have primarily investigated only one of the five service branches in their assessment.

Prior to 1980, a number of different techniques were investigated for military readiness index models. Gaver and Mazumdar (1967), Tolins (1968), and Greenberg (1972) used Markov chains to identify the probabilities of unit components transitioning from one readiness state to another. Lockman, et al. (1970) proposed using multiple regression and principal component analysis to relate the four readiness components of personnel, equipment, supply, and training to performance scores for naval destroyers. Kaplan (1972 1975) and Barish and Ehrenfeld (1975) related readiness in terms of unit output using a production function model. The idea of applying *C* ratings to components of a unit using a "1," "2," "3," and "4" scale, where the overall rating is related to the "weakest link," is first documented specifically for naval units in the 1960s (Frank, et al. 1968). Known as the METRI project, the method involved assigning a *C*-1 rating to a value of zero, a *C*-4 rating to a value of one, and the remaining values p_1 and p_2, where $0 < p_1, p_2 < 1$, to *C*-2 and *C*-3, respectively. In this case, the values p_1 and p_2 were arbitrarily selected numerical scores to reflect the perceived state of readiness corresponding to a *C*-2 and *C*-3 rating. A weighted average of various sub-resource areas (communication, personnel, weapon systems, etc.) would then be taken to obtain an overall readiness rating.

Most models of the late 20th century sought relationships between readiness and either the ability to maintain operations for an extended duration of time or the ability to deploy quickly. Shisko and Paulson (1982) utilized a simulation to study the effects of varying levels of manpower, equipment, fuel, spare parts, ammunition, etc., on a combined arms brigade. In their research, a continuous readiness index was proposed, outlined in terms of a comparison between a unit and some other reference organization having pre-defined capabilities. Given that $x_1, x_2, \ldots, x_n$ is a set of resources (manpower, equipment, munitions, etc.) and $q(x_1, x_2, \ldots x_n; m)$ represents a rate of decline for an element of size *m*, Shisko and Paulson (1982) calculated the overall rating *R* for the *k*th unit (in comparison to the *p*th unit benchmark) as:

$$R^k = \frac{\int_{\tau}^{\tau+T} q^k(x_1, x_2, \ldots, x_n; m)\, d\tau}{\int_{\tau}^{\tau+T} q^p(x_1, x_2, \ldots, x_n; m)\, d\tau}, \tag{1}$$

where τ is the starting time, and *T* is the time horizon for the measurement.

Several other researchers, such as Moore, et al. (1995), Betts (1995), and Raffensperger and Schrage (1997), incorporated time as a key component in measuring readiness. The goal was to facilitate a comparison of units and enable incorporation of costs associated with train-up or train-down time. Then, an optimization scheme based upon a given pre-deployment window and a coordinated readiness training cycle could be incorporated.

Jareb, et al. (1999) proposed a readiness index based upon a standardized percent change, calculated by dividing the change in an indicator value by the average change for all of the measured time periods. A weighted average was then employed to consider differences among personnel, equipment on-hand, and equipment serviceability in various ground units.

Most recently, and in line with the latest AR 220-1 revision, readiness has been considered in terms of capability metrics. Randazzo-Matsel (2009) reviewed a vast number of studies by the Center for Naval Analyses examining various metrics. Some of the metrics looked into specific areas such as personnel quality indices, where time within a unit, education experience, and the number of unit legal cases may be considered in measuring readiness (Stafford and Moore 2004; Davis, et al. 2008). Grier et al. (2012) more recently examined measuring readiness in terms of human capacity for personnel quality. A number of other researchers recently looked distinctly at improving ways to report equipment on-hand or serviceability (Peltz, et al. 2002; Menko 2004; Kaczynski and Foote 2005; Nickel, et al. 2008).

Finally, increasing training proficiency in units has received strong focus from some researchers. Cannon-Bowers, et al. (1995) examined individual-level traits such as motivation, experience, and attitude, in order to determine the effectiveness of training outcomes. Orlansky, et al. (1997) identified a list of unit events and factors, such as exercises, optempo, and live-fire range qualifications, for which they determined were ideal indicators of training readiness.

Frank, et al. (2007) prescribed a methodology used by the Federal Highway Association to link data on field experience and training to support resource allocation decision-making.

3. **Motivation**

Even with the extensive research and periodic reviews of AR 220-1, sustained criticism of the readiness policy remains. Most of this criticism is from non-partisan United States departments and agencies. For example, the United States General Accounting Office (GAO) implied that "the formal reporting system is overly optimistic in its readiness assessments, and questions can be legitimately raised about its credibility" (US GAO 1997). At that time, commanding officers, for a number of different reasons, were reporting higher states of readiness than what was perceived by subordinate unit leaders or soldiers. An earlier unpublished Army War College study on reporting readiness in 1976 found the ratings were just as inflated and field commanders had little trust in the system (Galloucis 2003).

In an effort to improve the readiness reporting process, the National Defense Authorization Act for Fiscal Year 2000 required the Department of Defense (DoD) to establish a new comprehensive reporting system that would be initially implemented in 2000 and was to gain full system capability by 2007 (US GAO 2003). In 2002, the DoD officially established the DRRS, which is still in use at the time of this research. However, the system has also received scrutiny from government officials (US GAO 2011), with instances where personnel and equipment levels were incorrectly reported in Army and Marine units. As a result, one of the most recent government reports recommends that the Department of Defense establish readiness metrics that can be used to evaluate whether outcomes are being met (US GAO 2016).

Through periodic updates to AR 220-1, the United States Army has sought a number of changes to readiness reporting, with multiple revisions in the 21st century (Headquarters 2001, 2003, March 2006, December 2006, 2010). Despite these refinements in the policy, how the Army readiness rating is calculated has not changed. It uses the same system implemented in the 1970s, relating the weakest of the *PSRT* ratings to an overall *C*-rating. Some researchers at RAND have alluded to this fact and thus have promoted only incremental improvements in the concept design for a military readiness metric (Moore, et al. 1991). A more recent RAND study suggests that the Army's current readiness reporting system should be adapted only slightly to assist more with posturing the future force (Pernin, et al. 2013). Our position is similar in this regard; that is, many of the deficiencies in the current metric may be remedied with a different model formulation, rather than considering a significant culture shift in measuring readiness, such as with alternative attributes or sub-categories. Specifically, the current system framework lacks precision and flexibility, masks preparedness, struggles to provide real-time value, and is absent of robustness in generating solutions, challenges which we now outline.

3.1 *Lack of Precision*.

As noted in Table 1, the ratings "1" to "4" depend on various ranges for the personnel, equipment on-hand, equipment serviceability, and training subcategories. These ranges are extremely broad when describing readiness. For example, a tank battalion with one tank and a tank battalion with half of its tanks would receive the same rating for equipment on-hand. This deficiency, which was noted in 1998 and 2003 with the inspection of the Army's unit status reporting system (US GAO 2003), seems to violate the quality of responsiveness preferred in a metric.

3.2 *Lack of Flexibility in Measurement*

The current readiness measure assumes that the personnel, equipment on-hand, equipment serviceability, and training status across different type units are all of equal weight.

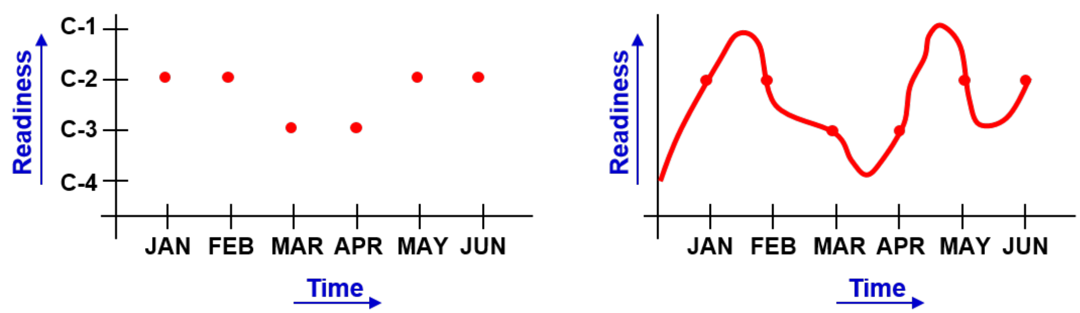
For a light infantry battalion, whose fighting element is the soldier, the personnel component would have more priority for readiness than the equipment serviceability category, as this translates to increased warfighting capability. In contrast, for a tank battalion, whose fighting element is the tank, equipment serviceability would likely have greater weight in unit readiness as it relates to capability. Current indices fail to take the priorities into account for which these subcomponents influence unit readiness. Furthermore, the current measure assumes that all personnel or equipment within each unit category are of equal weight. For instance, an infantry battalion without the commanding officer (or any officer qualified to fill it) would be considered just as ready as an infantry battalion without its supply clerk. A tank battalion without one of its machine guns would be considered just as ready as a tank battalion without one of its armored personnel carriers (provided the authorizations are the same).

3.3 *Masking Preparedness*

Another deficiency noted in the current readiness system metric is the fact that the "weakest link" methodology for selecting each rating masks the true value of the readiness measure. In the case of masking, sub-category attributes with similar and dissimilar ratings have the same overall influence on unit readiness. For instance, if Unit 1 is *P*-4 in personnel and *S*-4 in equipment on-hand, the overall rating is *C*-4, the same rating as Unit 2 that is *S*-4 in equipment on-hand and *P*-1in personnel. As a result of selecting the lowest overall sub-component level, the overall status of personnel or equipment in Unit 1 becomes secondary. Intuition would argue that Unit 1 is less ready than Unit 2. Trunkey (2013) noted this deficiency, stating that "scores may be misleading because they are based on broad measurements that can mask underlying problems in a critical area." The paper then reported the fact that individuals frequently interpret a *C*-rating that has not changed from one month to another as to mean that the level of readiness is unchanged, even though the lowest scoring attribute may not be the same. This deficiency in the current metric is one of the primary causes for generating ratings that lack robustness or sensitivity to sub-category assessments.

3.4 *Real-Time Value*

Because the current readiness status is reported for one day of the month, understanding how a unit progresses to become more or less ready is difficult. A continuous measure would facilitate better prediction and enable observation of trends or seasonal patterns. The current readiness measure may be completely masking the true readiness just by the nature of when readiness is reported. For example, Figure 1 shows that readiness may rise and fall throughout the month, but a single static report per month would not capture these changes. This has been noted as a deficiency by researchers using time-based models (Moore, et al. 1991; Betts 1995; Raffensperger and Schrage 1997), as well as those formally investigating the process (CBO 1994; GAO 2011).



(i) (ii)

Figure 1: Illustration of the current readiness measure (i) versus possible actual readiness measured continuously (ii)

3.5 *Objectivity*

Another contentious aspect of the current readiness system is its use of the "subjective upgrade" or "subjective downgrade" of the overall rating. As discussed, after the overall *C*-

rating is calculated, the Commander may upgrade or downgrade the rating in line with his or her overall judgment of the unit's capability. Reports on trends in military readiness suggest that subjective *downgrades* rarely occur unless there are very special circumstances; rather, when a decision is made to alter the objective result, the subjective upgrade to a higher capability rating is almost always observed (US GAO 1996). Many congressional members have expressed concern at the ability of Commanders to change the rating if the data and statistics suggest otherwise (Committee on Armed Services 2010).

Several researchers have also theorized on the Commander's use of the subjective upgrade. Betts (1995) detailed two investigations, noting that incoming commanders will have a tendency to report low so that they may show progress in command, and more senior commanders will report high so as to imply success. The author also noted a 1983 congressional investigation where all instances of readiness increase over a twelve-month period were a result of the commander's subjective upgrade. A more recent investigation noted that just before a unit deploys, at a time when readiness ratings would be most informative, the incentive is clearly for commanding officers to rate the unit fully ready (CBO 2011).

In contrast, this research theorizes that Commanders tend to subjectively upgrade a status primarily due to the nature of the metric in selecting the lowest possible rating. The concept of the "weakest link" approach is documented widely in the decision analysis literature, whereby the ability to evaluate tradeoffs between attributes or subcategories is prevented (Otto and Antonsson 1991; Yoon and Hwang 1995; Seppala, et al. 2002; Matheson 2013). Such a construct typically leads to solutions that are very responsive to small changes in the constraints. For instance, suppose that a unit, which is at the highest level of readiness among all *C* rating subcomponents (*P*, *S*, *R*, and *T*) has 6 senior NCOs (E7-E9) assigned to its ranks. If one senior

NCO becomes non-available for any reason (leave, duty profile, legal status, etc.), then the unit is automatically *P*-2 (hence *C*-2 overall). Moreover, if an additional senior NCO becomes non-available (4/6 = 67%), the unit reports *P*-3 (hence *C*-3 overall). It is very likely that the Commander would subjectively upgrade the overall measure to a higher rating due to the negligible effect of one or two individuals on the overall readiness of a 150-person (or more) organization. In summary, we argue that the current metric violates the preferred property of robustness in providing acceptable solutions. In any unit, there are numerous cases just like these examples that may warrant the Commander's upgrade.

3.6 Research Position

Therefore, the intent of this research is to develop a readiness index that (i) adds greater precision in its measurement, (ii) eliminates the masking of preparedness, (iii) increases flexibility in its capacity for measuring different types of units under varying conditions, and (iv) is more robust to small adjustments in sub-categories. Although the number of different indices proposed in the literature suggests that measuring readiness may be inherently complex, this research places greater importance on measurement accuracy. Following a brief description of the algorithm in Section 4, case studies are examined in Section 5 using a simulation methodology in *Mathematica* to compare the current and proposed metrics, validating our approach.

4. Model Development

In general, many different optimization techniques exist that are quantitative in nature. The readiness problem involves measuring multiple components and subcomponents into a single rating that accurately and quantitatively measures a unit's ability to execute its mission. One multi-response optimization technique that is widely used and popular for its simplicity is the desirability function approach. The method was introduced by Harrington (1965) and later modified by both Derringer and Suich (1980) and Kim and Lin (2000). It is a univariate utility function (Harrington 1965; Jiang, Murphy, and Tsui 2006). This method involves transforming each response *y* into a desirability value, where these values are bounded from 0 to 1. Within these bounds, 0 represents unacceptable or no quality and 1 represents no additional value with further improvement. The scale is dimensionless, and the desirability *d* can be evaluated for each attribute in the overall model. In this case, the desirability, or utility, is derived from attributes that are quality properties or components of the system or product being measured. Then a weighted geometric mean of the individual desirability values is computed, resulting in a composite desirability value *D*. The weights reflect the relative importance of each attribute or quality property to the overall model. Suppose that there are *k* different responses of interest. For the *i*th response, three constants are identified for the range of possible values: L_i, a lower value, U_i, an upper value, and τ_i, the target value. Then, depending on whether the objective is to minimize or maximize the response, the individual desirability values are calculated using Equation (2):

Objective: Minimize *y* Objective: Maximize *y*

$$d(y_i) = \begin{cases} 0 & \text{if } y_i > U_i \\ \left(\dfrac{y_i - U_i}{\tau_i - U_i}\right)^{s_i} & \text{if } \tau_i \le y_i \le U_i, \\ 1 & \text{if } y_i < \tau_i \end{cases} \qquad d(y_i) = \begin{cases} 0 & \text{if } y_i < L_i \\ \left(\dfrac{y_i - L_i}{\tau_i - L_i}\right)^{s_i} & \text{if } L_i \le y_i \le \tau_i, \\ 1 & \text{if } y_i > \tau_i \end{cases} \tag{2}$$

where s_i represents the shape parameter for the *i*th response, a pre-defined setting designed to indicate the priority established for obtaining the target value τ_i. Finally, given weights for each response $w_1, w_2, \ldots, w_k$, the composite desirability index is determined using the weighted geometric mean calculation in Equation (3), as developed by Derringer (1994).

$$D = \left[[d(y_1)]^{w_1} \cdot [d(y_2)]^{w_2} \cdots [d(y_k)]^{w_k} \right]^{1/\sum_{i=1}^{k} w_i} \tag{3}$$

Individual and composite desirability values equal to one are considered most desirable, whereas values equal to zero are considered least desirable. Further details on the theory of desirability functions can be found in NIST (2013). We apply a weighted geometric mean approach to the readiness problem using a model analogous to the multi-response framework proposed by Harrington (1965). The fact that the current Army metric utilizes sub-component ratings that are already mapped to the [0, 1] interval further aligns the appropriateness of the calculation to the readiness problem.

4.1 Application to Readiness

In this research, the basis of the desirability function approach will be used to measure the individual *P, S, R,* and *T* subcategories of readiness. By definition, the desirability function approach is a utility-based method that involves measuring value or worth of a component or response, whereby the functional parameters or weights are subjective in nature and can be assessed using elicitation approaches, group judgment, or other standard methods common in utility applications (Kim and Lin 2000; Jiang, Murphy, and Tsui 2006). Using a construct similar to Harrington's (1965) desirability function with a target, τ, equivalent to 1.00 or 100% for maximizing readiness, consider a tiered framework for Table 1 that incorporates the weighted geometric mean with weights w in its formulation, as a means to determine the overall rating C:

$$\text{(Tier 1)} \quad C = \left(P^{w_1} S^{w_2} R^{w_3} T^{w_4} \right)^{\sum_{i=1}^{4} w_i} \text{, where} \tag{4}$$

$$\text{(Tier 2)} \quad P = \left(AS^{w_5} AQS^{w_6} ASG^{w_7} \right)^{\sum_{j=5}^{7} w_j}, \quad S = \left(S{:}AP^{w_8} S{:}P^{w_9} \right)^{\sum_{k=8}^{9} w_k}, \quad R = \left(R{:}AP^{w_{10}} R{:}P^{w_{11}} \right)^{\sum_{m=10}^{11} w_m},$$

$$\text{(Tier 3)} \quad \text{and } ASG = \left(ASG_{\text{E5-E6}}^{w_{12}} ASG_{\text{E7-E9}}^{w_{13}} ASG_{\text{WO}}^{w_{14}} ASG_{\text{O1-O3}}^{w_{15}} ASG_{\text{O4-O6}}^{w_{16}} \right)^{\sum_{n=12}^{16} w_n}$$

Unlike the current Army readiness metric, the output for the proposed metric is a decimal between 0 and 1. In general, some degree of measurement error may exist in any metric; for this reason, the output should not be viewed as an exact quantity where small differences reflect meaningful changes. Rather, the value in obtaining a decimal form as an output versus an overall *C* rating is in the transition of the proposed metric to a continuous real-time index for evaluating trends over time. While the current metric could be continuously calculated, the output across four categories (*C*-1 to *C*-4) greatly limits the ability of any trending related analysis.

Each desirability function is a single response, so that four total desirability functions will be calculated per unit, one each for *P, S, R,* and *T*. By definition, these responses are aggregated by use of the geometric mean, which is more holistic than using an arithmetic average. For an arithmetic average, if any one of the four readiness sub-categories (*P*, *S*, *R*, or *T*) completely fails and the other categories rate high, the arithmetic average would not penalize the overall composite response as much as is needed. In fact, if a unit was rated as completely "ready" in the personnel, equipment on-hand, and training categories, but had all of its equipment in an unserviceable state, a metric using the arithmetic average would assess the unit at or around *C*-3 overall. This would be unacceptable, as the unit could not serve with equipment in such a state. In contrast, the geometric mean enables a composite quality calculation, highlighting very low unacceptable values by calculating less than the lowest individual utility value (Jiang, Murphy, and Tsui 2006). Thus, units that perform well or at near-mission capacity in *P, S, R,* and/or *T* will score highly, while units that have significant areas to address will score lower in the composite scale. Of note, the formulation of the proposed index results in an objective calculation that is not aligned to the weakest link, removing the need for an overall subjective upgrade or downgrade of the rating. Rather, senior decision makers establish priority with the design of the weights for the attributes or sub-categories. As mentioned previously, subjective upgrades in the current metric occur primarily due to the nature of the 'weakest link' selection of each sub-category rating, in particular with the lack of robustness in measurement. This tiered approach, which incorporates the concept of the weighted geometric mean, will generally have overall ratings at or above the 'worst-case' rating. It will also be more robust to small adjustments in sub-category values. Most importantly, as the overall result *C* for the proposed approach accounts for each value of *P*, *S*, *R*, and *T* (rather than just the lowest), it does not mask the true "preparedness" of a unit. Theoretically, we can assign equal weights to the *P, S, R,* and *T* responses but using a weighted geometric mean will enable scaling of each desirability function to reflect the importance of each response to each unit by its mission, tailored for type of unit and type of mission (e.g. combat, humanitarian, or peacekeeping operations).

4.2 Selection of the Weight Parameters

In Equation 4, the weights enable each *P*, *S*, *R*, and *T* component and subcomponent to be appropriately valued according to the unit type and mission. For example, the United States Army consists of units that have different objectives and structure. A light infantry battalion is restrictive and carries out its mission without the use of vehicles, whereas an armored battalion relies on the tank and accompanying vehicles for accomplishing its mission. Thus, acceptable *P, S, R,* and *T* ratings may vary but would be consistent across each type of unit. Appendix B outlines notional personnel, equipment, and training data for a light infantry battalion and an armored battalion. Note the complexity of personnel comprising the units, the differences in the equipment across both units, and the examples of required training. In order to properly and objectively assess readiness of these units, the weights assigned to each *P*, *S*, *R*, and *T* component will likely vary by type of unit and mission. It is important to note the effects of using different weighting schemes in this manner, particularly with comparisons across units or services. While the ability to prioritize sub-category ratings for any one unit translates to added flexibility (a preference for a metric), it is possible that generalities or comparisons of readiness across a range of units are misinterpreted.

Weights may be assessed via a variety of decision analysis approaches. The platinum standard of Parnell, Bresnick, Tani, and Johnson (2013) utilizes subject matter expert elicitation and is often employed in additive aggregation models. In this scenario, senior decision makers and key stakeholders are interviewed to assess model components, such as value hierarchies or weights. For a multiplicative aggregation model such as with Equation (4), the elicitation of weights is not a trivial effort; quantifying their magnitude for a variety of units within an organization as large as the Army is particularly challenging. Beyond the platinum standard, there exists many methods for calculating weights. However, many of these methods are designed for additive models and would not apply in this situation.

The key in readiness is to develop weights that accurately reflect mission, in terms of both unit type and task. For example, not only should the weights for a light infantry battalion and an armored battalion differ based on the type of unit and equipment and/or personnel it has, but also the weights should reflect the mission task. A unit called to a humanitarian mission should have different priorities than one called to a combat mission.

Due to the classified nature of readiness reports, derived weights for this model and case studies are not included here. However, a benefit of methodically deriving weights is that the weights and subsequent readiness calculations can be updated for every unit and every mission task. To illustrate our model, we employ the rank order centroid method to derive weights. Only the rank order importance of *P, S, R,* and *T* are needed to calculate weights using this method, and general order of the components can be deduced by a unit's capability, eliminating the barrier that classified data presents.

Barron (1992) proposed rank order centroids when only partial information regarding a set of attributes is known, i.e., in this case, we know only the general ranking of the components. Using Barron's construct, the *p*th most important attribute out of *r* total attributes for *C*, *P*, *S*, *R*, or *T*, would have a weight, w_p, calculated by:

$$w_p = \frac{1}{r}\sum_{i=1}^{r}\frac{1}{i} \text{ , for } p = 1,\ldots,r \text{ , with } \sum_{p=1}^{r} w_p = 1 \tag{5}$$

Rank order centroid weights are appropriate when direct weights are onerous and are particularly useful when multiple decision makers are involved (Scala and Pazour, 2016). They have been used in other defense related multi-attribute models, including Scala and Pazour (2016) and Scala, et al. (2012). The literature has shown that rank order centroids allow the weights to be derived systematically by using implicit information in the ranks, and the calculated ranks are not ad hoc, outperforming other weighting methodologies (Barron and Barrett 1996). This is key for illustrating our methodology, as although the rank order centroid weights are serving as a proxy and may not be the actual weights employed by the Army, confidence can be given to their ability to accurately demonstrate the calculation without bias or chance.

5. Case Studies

5.1 The Simulation Methodology

In order to observe the behavior of the proposed metric in comparison to the current Army readiness index, a simulation is developed in *Mathematica* using notional unit data for two different type units, a light infantry battalion and an armored battalion. The array of authorized, assigned, and available personnel and equipment is provided for various scenarios, as well as a hypothetical training status (see Appendix B). Within particular line items for personnel and equipment (sample sizes ≥ 10), a uniform distribution is established to facilitate fluctuations in assigned or available strength at specified ranges. Personnel and equipment populations with sample sizes less than 10 items are fixed at full levels in order to limit their ability to influence the general behavior of the metrics and facilitate a standard comparison. In addition, a uniform distribution is established for the training category, whereby a unit may be considered "Trained," "Needs Practice," or "Untrained" for a percentage of its tasks.

As a baseline example of how the simulation results are displayed, consider the case of the Light Infantry Battalion with all of its authorized equipment and personnel on-hand at 100% strength and with no training shortfalls (i.e. "Trained" in all tasks). The available personnel and available equipment (due to serviceability) for this battalion, however, are set to fluctuate based upon a uniform distribution at 80-100% of its assigned strength. Equal weighting across all categories and sub-categories is employed with the proposed metric (i.e. $w_1 = w_2 = w_3 = w_4$, $w_5 = w_6 = w_7$, etc.), to facilitate a comparison. With 5000 iterates, the simulation provides the baseline output at Figure 2.

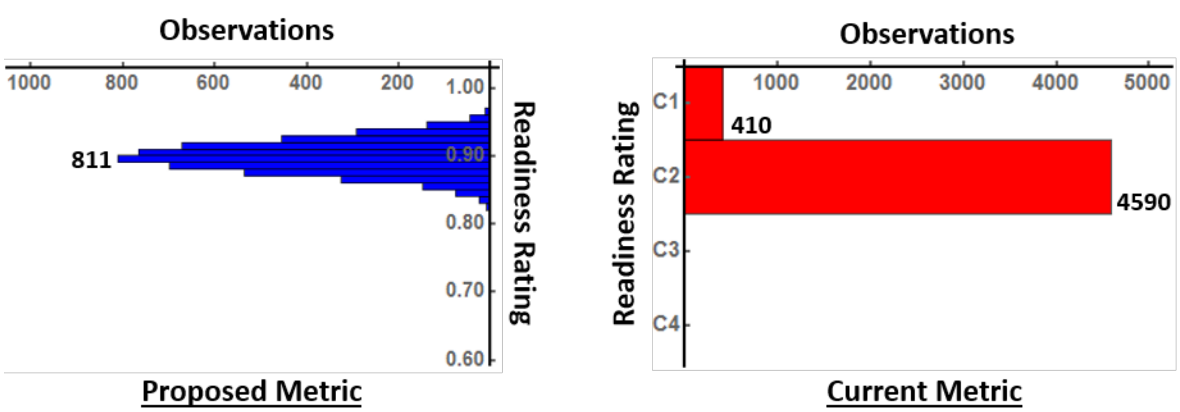


Figure 2: Proposed and Current Readiness Metric,
80-100% Availability for Both Personnel and Equipment

It is important to note that the current and proposed "Readiness Ratings" cannot be compared in an exact manner; there are slight differences that must be recognized. First, the proposed metric with decimal output is continuous in nature (binned for the histogram in Figure 2), in contrast to the discrete *C*-ratings provided by the current metric. Second, the current metric may result in an overall rating of *C*-2 based upon the lowest of three different subcategory ranges for personnel and equipment non-availability (see Table 1), namely 70-89%, 75-84%, and 80-89%. In contrast, the proposed metric considers the ratings across all of the subcategories in formulating an overall measurement. Finally, the sub-category ratings for the current metric map in different ways to the [0, 1] scale, as shown in Figure 3, which is another way to visualize the data in Table 1. For these reasons, one cannot simply suggest that a decimal value for the proposed metric is equivalent to a specific *C*-rating under the current system. Despite this fact, we can compare the general behavior of the metrics using various case study scenarios.

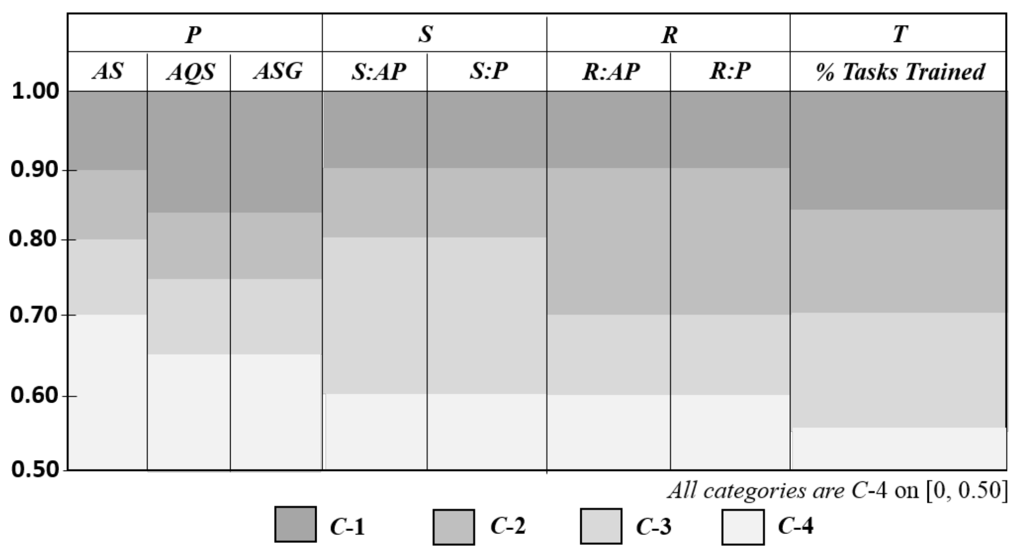


Figure 3: Current Readiness Metric, Mapping Sub-Categories to the [0, 1] Scale

Based upon the baseline results in Figure 2, some observations may be made. With the uniform distribution generating random observations for personnel and equipment levels between 80-100% of their assigned strength, we may intuitively expect ratings to fluctuate in that range with an average close to 90%, that which is depicted by the proposed metric. The worst case selection methodology of the current metric, however, results in more than 90% of the readiness ratings being at *C*-2. This is somewhat surprising, since 2 out of 3 of the Available Personnel subcategory ranges in Table 1 are constrained to facilitate a *C*-1 rating with strengths of 85% or more. This is also an argument to understand why a Commander may decide to subjectively upgrade to a higher rating; that is, the sense that the rating is underscoring the true readiness of the unit.

5.2 Case Study 1: The Light Infantry Battalion

The proposed metric is designed to focus weight in areas where priority is established. The weighting scheme may change for any number of reasons: unit type, the deployment rotation, available budget dollars, preferences at the command or Army levels, etc. For a light infantry battalion, the individual is the fighting element. Hence, the strength of any one unit would be tied most closely to the availability of personnel and the level of training they maintain. While equipment readiness is also important, it may receive less in terms of priority (budget expenditure, focus of time, etc.), when compared to personnel and training readiness. When focusing solely on personnel readiness, maintaining a strong senior leader corps of Non-Commissioned and Commissioned Officers in a unit may take priority for any one unit over the larger junior corps of soldiers. In terms of overall equipment on-hand and serviceability, pacing item readiness may take priority over the equipment at-large. Given these preferences, the following weighting scheme is developed using the rank order centroid method for the proposed metric (first set of calculations shown):

Tier 1: $P = T > R = S$. The subsequent values q_i of the rank order centroid method are:

$q_1 = \left(\frac{1}{4}\right)\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) = 0.5208\overline{3}$, $q_2 = \left(\frac{1}{4}\right)\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) = 0.2708\overline{3}$,

$q_3 = \left(\frac{1}{4}\right)\left(\frac{1}{3} + \frac{1}{4}\right) = 0.1458\overline{3}$, and $q_4 = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = 0.0625$ (the number of significant digits is extended to illustrate the q_i summing to 1). Averaging the weights corresponding to equal priorities, we have $w_1 = w_4 = 0.3958\overline{3}$ and $w_2 = w_3 = 0.1041\overline{6}$.

Tier 2: $AS > ASG > AQS$, results in $w_5 = 0.6111$, $w_7 = 0.2778$, $w_6 = 0.1111$; $S{:}P > S{:}AP$, results in $w_9 = 0.75$ and $w_8 = 0.25$; and $R{:}P > R{:}AP$, results in $w_{11} = 0.75$ and $w_{10} = 0.25$.

Tier 3: $ASG_{\text{O4-O6}} = ASG_{\text{E7-E9}} > ASG_{\text{O1-O3}} = ASG_{\text{E5-E6}}$, results in $w_{13} = w_{16} = 0.3958\overline{3}$ and $w_{12} = w_{15} = 0.1041\overline{6}$ (no warrant officers in a light infantry battalion).

5.2.1 Precision

There are two scenarios for this investigation: (i) a battalion returned from a recent deployment loses approximately 45% of its personnel in accordance with local “refit” policies, and (ii) the same battalion enters a “block leave” period where up to 75% of the battalion proceeds on leave at one time. A uniform distribution is established for the first scenario, where personnel availability fluctuates between 50-60%, with all other categories at maximum capacity (100% equipment on-hand, 100% equipment serviceability, and all training tasks “*T*”). In the second scenario, the distribution is reduced to an availability range between 20-30% of the unit personnel with the same equipment and training ratings. After 5000 iterations, the simulation provides the output at Figure 4a and Figure 4b (note the slight shift in the ordinate scale for the proposed metric when compared to the other figures). As outlined in Section 1, the more *responsive* metric is able to react to negative or positive fluctuations in readiness sub-categories.

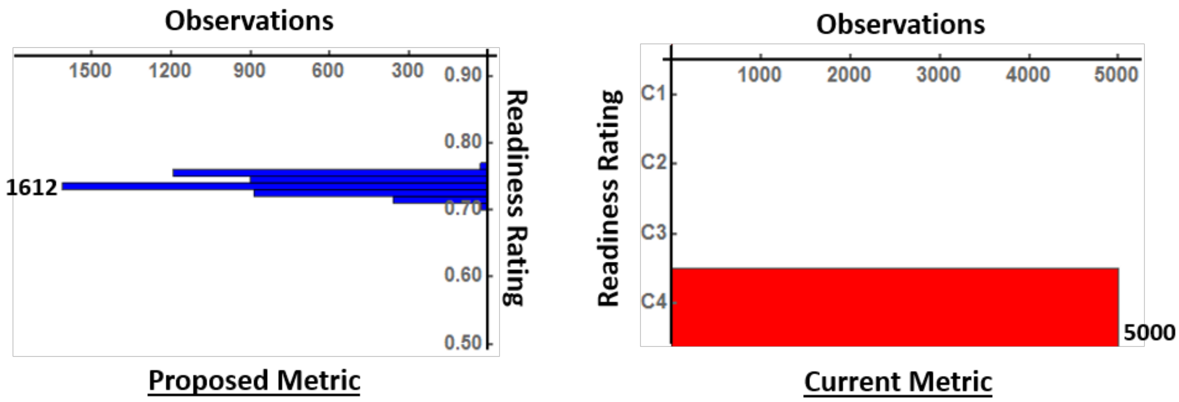


Figure 4a: Proposed and Current Readiness Metric,
50-60% Availability for Personnel; All Other Categories at 100%

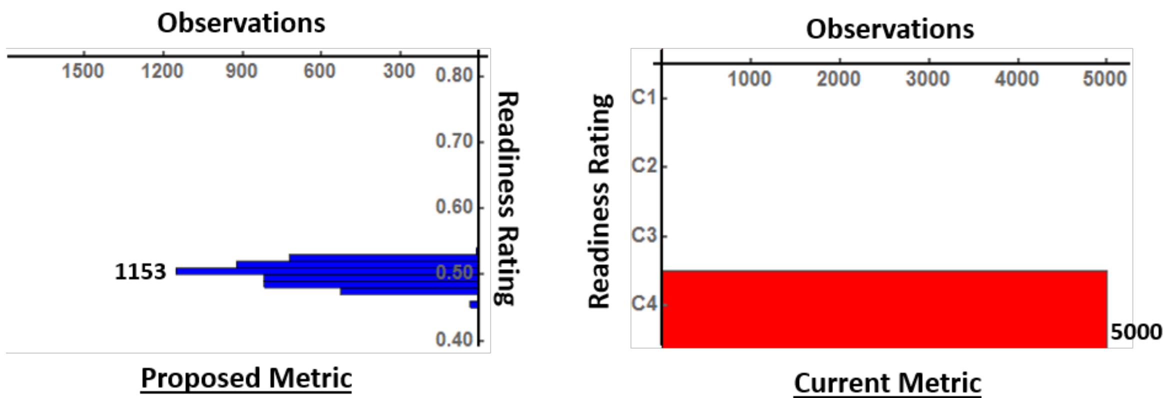


Figure 4b: Proposed and Current Readiness Metric,
20-30% Availability for Personnel; All Other Categories at 100%

In this example, the results demonstrate two points. First, only the proposed metric is able to differentiate the change in personnel readiness; the current metric results in the same rating (*C*-4) for both scenarios. This reveals the responsiveness of the proposed metric to shifts in the measurement of readiness attributes, a preferred quality of a metric. Second, it is important to note that the proposed metric takes a holistic approach where the equipment and training ratings are also considered in its final rating. Rather than consider only the lowest subcategory in its final assessment, as with the current metric, it considers all subcategories under a given weighted scheme. Hence, if the equipment and training readiness ratings also deteriorated under these settings, we would continue to observe a reduction in the overall rating of the proposed metric.

5.2.2 Robustness

Consider the battalion whose personnel availability fluctuates between 80% and 100% and with all other categories at maximum capacity (as in Section 5.2.1). Figure 5a provides the comparison of output for the current and proposed metrics after 5000 iterations, given this scenario. Then, as a follow-on scenario (Figure 5b), suppose that one additional change is made; specifically, the 12 assigned Infantry Non-Commissioned Officers at the rank of E7 are given an availability that fluctuates between 0 and 90%. If a metric is to be *robust*, small changes in readiness sub-categories should not result in large shifts to overall output.

The second scenario above demonstrates the sensitivity of the current metric; small adjustments can have a large impact on the overall readiness rating. The 12 assigned E7 Infantry Non-Commissioned Officers represent only 2% of the assigned battalion strength. Despite the fact that the assigned E6 in the battalion can fill the roles of those E7 that are non-available, the current metric rating fluctuates heavily. This is precisely the problem that can occur with the "weakest link" approach where all possible solutions lie on the boundary of the solution space, rather than inside that space. In contrast, the difference between scenarios in the measurement output for the proposed metric is negligible, demonstrating the preferred quality of robustness in generating solutions.

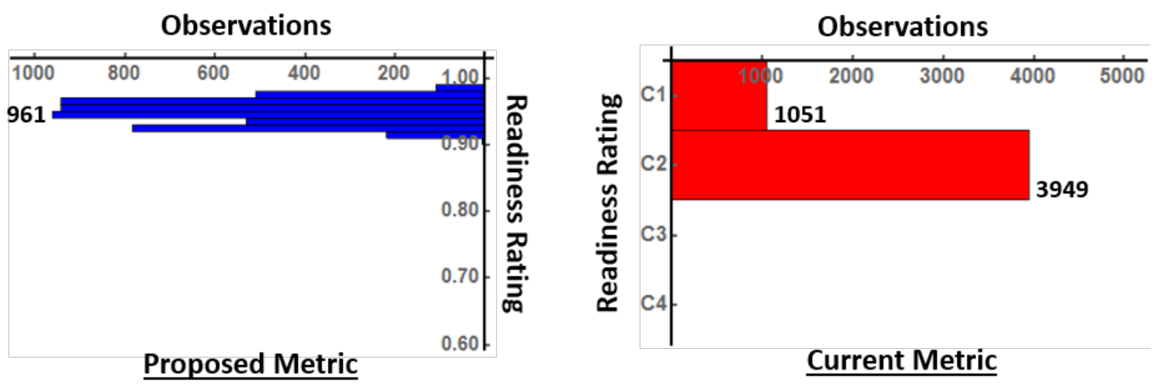


Figure 5a: Proposed and Current Readiness Metric,
80-100% Availability for Personnel; All Other Categories at 100%

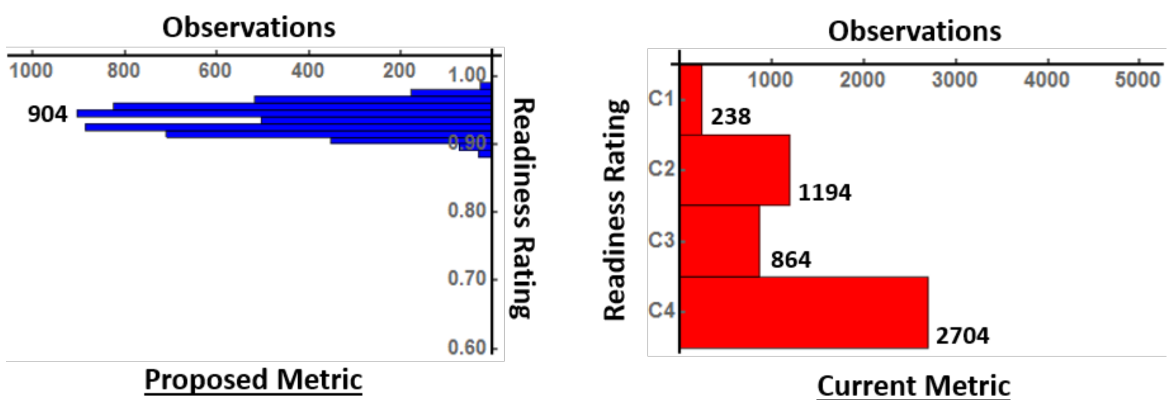


Figure 5b: Proposed and Current Readiness Metric,
Same Settings (+ 0-90% Availability for E7 Infantryman)

5.3 Case Study 2: The Armored Battalion

To our knowledge, there are no examples of military readiness metrics in the literature that capture the ability to assess different types of units based upon altering priorities. In order to fill this void and further demonstrate *flexibility* as a quality for the proposed metric, an additional case study is presented. For an armored battalion, the tank is the fighting element. Hence, the strength of any one unit is tied most closely to the availability of the tanks in terms of equipment serviceability. It is a fair assumption than any one armored battalion will have all of its tanks on-hand, whether serviceable or not, and so equipment on-hand for the battalion may be less of a priority. While personnel and training readiness are also critical to the armored battalion, they may receive less in terms of priority (budget expenditure, focus of time, etc.), when compared to equipment readiness. When focusing solely on personnel readiness, we consider availability as a priority, but equity in terms of priority across the leaders and junior soldiers for this case study. Given these preferences, the following weighting scheme is developed using the rank order centroid method for the proposed metric:

Tier 1: $R > P = T > S$. Hence, we have $w_3 = 0.5208\overline{3}$, $w_1 = w_4 = 0.208\overline{3}$, and $w_2 = 0.0625$.

Tier 2: $AS > ASG > AQS$, results in $w_5 = 0.6111$, $w_7 = 0.2778$, $w_6 = 0.1111$; $S{:}P > S{:}AP$, results in $w_9 = 0.75$ and $w_8 = 0.25$; and $R{:}P > R{:}AP$, results in $w_{11} = 0.75$ and $w_{10} = 0.25$.

Tier 3: $ASG_{\text{O4-O6}} = ASG_{\text{E7-E9}} = ASG_{\text{WO}} = ASG_{\text{O1-O3}} = ASG_{\text{E5-E6}}$, results in $w_{12} = w_{13} = w_{14} = w_{15} = w_{16} = 0.20$ (one Warrant Officer specializing in maintenance within an armored battalion).

To examine the effects of an alternative distribution on metric output, a triangular (rather than uniform) distribution is established for particular personnel and equipment line items (sample sizes ≥ 10). In addition, a triangular distribution is established for the training category, whereby a unit may be considered “Trained,” “Needs Practice,” or “Untrained” for a percentage of its tasks. Varying the distribution for the personnel, training, and equipment categories enables one to determine whether metric behavior results from the characteristics or properties of a particular distribution.

5.3.1 Objectivity

Consider the armored battalion whose personnel availability, equipment on-hand, equipment serviceability, and training tasks all fluctuate between 70 and 90%. Moreover, suppose that Commanders subjectively upgrade the current metric a percentage of the time. Figure 6a and 6b provide the comparison of output after 5000 iterations for the current metric, given the Commander performs a subjective upgrade 0% and 33% of the time, respectively. It may be more intuitive to suggest that subjective upgrades occur when only one sub-category readiness level is low in comparison to the other levels (rather than a percentage of time); for this reason, a third trial is performed using these settings with the results shown in Figure 6c. As outlined in Section 1, the preference is to remove as much bias as possible as a means to increase *objectivity*.

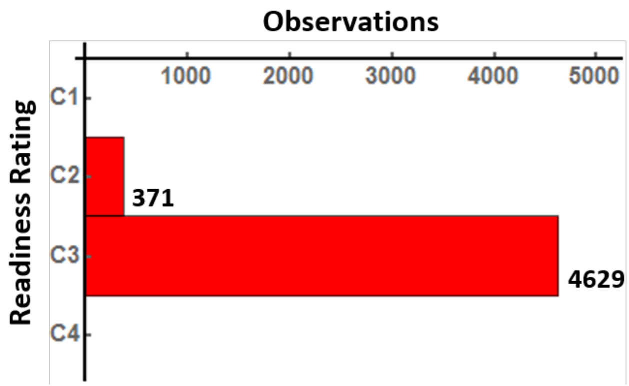


Figure 6a: Current Readiness Metric
70-90% for all Categories, 0% Subjective Upgrade

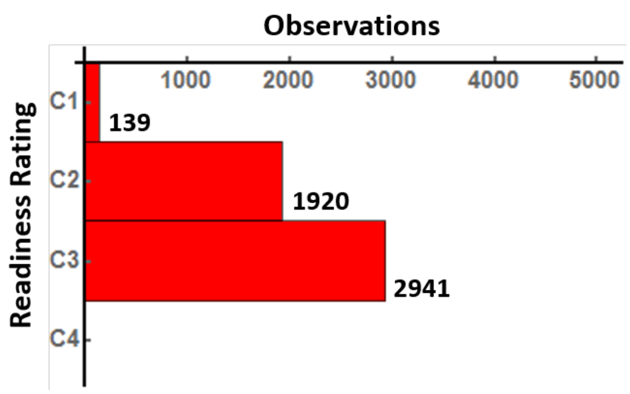


Figure 6b: Current Readiness Metric,
70-90% for all Categories, 33% Subjective Upgrade

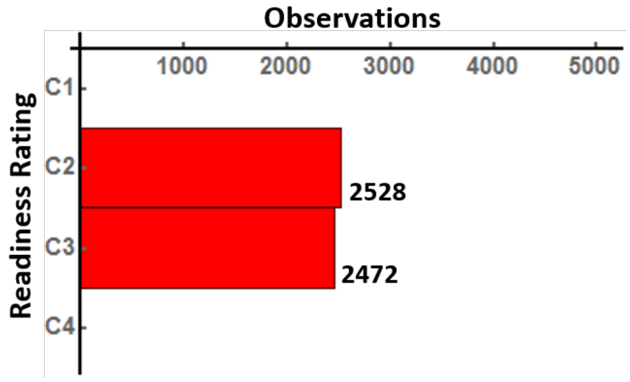


Figure 6c: Current Readiness Metric,
70-90% for all Categories, Subjective Upgrade (one sub-category lower than others)

Because the proposed metric does not utilize the concept of the subjective upgrade or downgrade, it is not shown in comparison to the current metric for this case; rather, this study is meant to examine the behavior of the current metric in its use of the subjective upgrade. With the ranges of personnel, equipment, and training fluctuating between 70-90%, and no subjective upgrade, more than 90% of the current metric observations suggest *C*-3 as the overall rating, similar to the baseline experiment in Section 5.1. Increasing the subjective upgrade to 33% of the unit status reports, we see instances where one may suggest the unit is performing at the highest readiness rating (*C*-1), despite all subcategory measures at less than optimal measurements. This provides some understanding of the effect of quantity on the degree of shift in the overall rating due to upgrades. In contrast, Figure 6c offers a qualitative aspect; given the same conditions whereby a subjective upgrade is performed only if one sub-category is lower than the others, we observe upgrades occurring roughly 40% of the time (more than in Figure 6b). Because the literature does not offer an account of the true number of upgrades, this may provide some indication as to the likelihood of upgrades occurring. Unlike the current metric, the proposed metric delivers a more objective quantity, whereby subjectivity is used to provide value, namely in the formulation of established priorities and weights for the desirability functions.

5.3.2 Flexibility

Because the proposed metric considers a unit holistically with contributions from all of the attributes rather than purely selecting the worst rating, it retains the flexibility to measure readiness based upon a set of desired weights. Given the settings described earlier, an armored battalion with a 100% equipment serviceability rating for its 80% of equipment on-hand should rate differently than the same battalion with a 80% equipment serviceability rating for its 100% of equipment on-hand. This is illustrated in Figures 7a and 7b. A more *flexible* metric is able to differentiate between units with different structure or makeup.

The priority of weight established for equipment serviceability (~0.52) over equipment on-hand (~0.06) should naturally increase the overall rating when transitioning from one scenario to another (Figure 7a to 7b). Only the proposed metric is able to differentiate between the fully equipped battalion with a less-than-optimal serviceability rating and the battalion with a perfect serviceability rating but without all of its equipment. This point demonstrates the preferred quality of *flexibility* in the proposed metric.

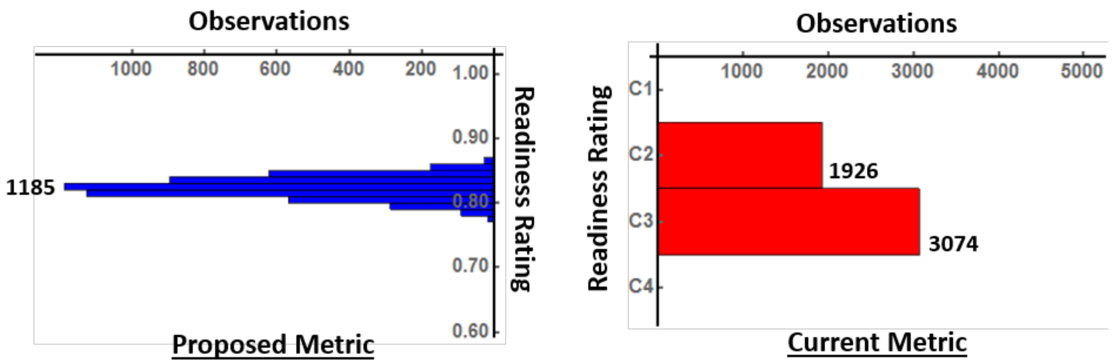


Figure 7a: Proposed and Current Readiness Metric
70-90% for Personnel and Training Categories,
100% Equipment On-Hand, 80% Equipment Serviceability

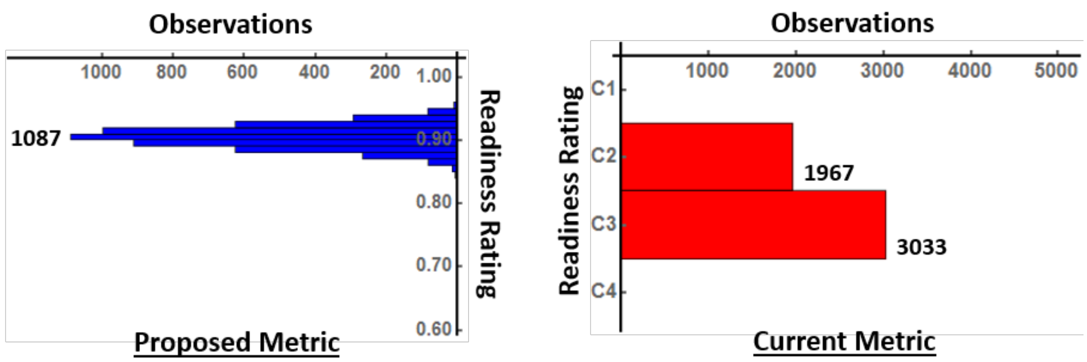


Figure 7b: Proposed and Current Readiness Metric
70-90% for Personnel and Training Categories,
80% Equipment On-Hand, 100% Equipment Serviceability

6. Conclusion

Without question, the warfighting readiness domain is characterized by complex and dynamic features. Different types of units with varying levels of strength and capability are called upon to perform diverse missions under varying circumstances. Unit expectations are changing over time. Deploying units are outfitted and trained in tasks that support success in combat, while redeploying units may recycle their personnel and reinitiate their training programs. Alongside these deployable elements are units primarily serving in a garrison environment with dissimilar objectives and roles. Despite these facts, one metric under a specific and fixed set of rules dictates the readiness rating for these units.

In order to account for the numerous factors that may affect a unit's posture, we designed a metric with a layered weighting scheme using the desirability function approach as a basis to capture the Army's preferences or priorities for a unit at any one time. The metric was constructed to provide continuous (rather than discrete) scoring on a full 0-100 grade scale; as such, we may examine trends more readily and prevent situations where readiness may be masked unintentionally. The resulting measure is also a composite analytic assessing all of the unit's features, rather than focusing on the lowest measure among several readiness attributes. Moreover, it is robust to small adjustments in its sub-category assessments, and it provides increased flexibility and precision, when compared to the current readiness metric. With this greater precision, situations where general subjective upgrades or downgrades are needed may then be removed from the assessment process. Generally speaking, the goal of developing a readiness metric is to understand if and where units need improvement intervention as well as to comprehend their ability to execute the mission. Contrary to traditional decision analysis models, a choice of the most ready units is not being made, and readiness between two units is not directly compared.

While this paper focused on a comparison of the proposed and current readiness models, there are a number of extensions for future research that should be considered with implementation. First, a comparison can be made between the proposed model and other additive or multiplicative functional forms found in the decision analysis literature to evaluate and assess responsiveness, robustness, flexibility, and objectivity. Another significant research area supporting this effort is the need to validate a formal weighting scheme for various types of units and missions. These schemes and validation can be derived from many sources of data, including in-depth surveys, elicitation techniques, or historical evidence. Axioms may also be developed, as a readiness measure does not conform to the axioms of a common preference model. A desirability function would not be the only model to satisfy such axioms, but the function is straightforward to use and has precedent in the literature. Finally, and perhaps most importantly, a framework for daily readiness must be introduced, whereby the nature of the proposed metric can be evaluated in a simulation environment. When the extension to a continuous measure is made, the true predictive power of a metric is realized.

Given the frequent demand on units in today's warfighting environment, readiness is considered one of the Army's highest priorities. Secretary of Defense Jim Mattis, like many leaders before him, announced guidance on rebuilding the U.S. Armed Forces, citing "Improve Warfighting Readiness" as the highest priority for his tenure (OSD 2017). In order to validate improvements in Army warfighting readiness associated with the Secretary of Defense's guidance, increased precision in measuring strength and capability are necessary. Our ability to measure readiness correlates in some way to the allocation of defense dollars for the military, the military's ability to secure or defend our national interests, and in many instances, the lives of soldiers.

Acknowledgements: The views expressed in this paper are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense. This research was supported by the Towson University College of Business and Economics Summer Grant Program.

References

Barish NN, Ehrenfeld S (1975) *Utilities Estimated from Actual Decisions in Readiness Measurement* (Graduate School of Public Administration, New York University, New York, New York).

Barron FH (1992) Selecting a best multi-attribute alternative with partial information about attribute weights. *Acta Psychologica* 80:91-103.

Barron FH, Barrett BE (1996) Decision quality using ranked attribute weights. *Management Science* 42(11):1515–1523.

Betts RK (1995) *Military Readiness: Concepts, Choices, Consequences* (The Brookings Institution, Washington, D.C.)

Bipartisan Policy Center (2017) The building blocks of a ready military: people, funding, tempo. (Task Force on Defense Personnel, Washington, D.C.)

Cannon-Bowers J, Salas E, Tannenbaum S, Matheiu S (1995) Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. *Military Psychology* 7(3):141-164.

Committee on Armed Services (2010) *The Current Readiness of the U.S. Forces.* (Hearing before the Subcommittee on Armed Services, U.S. Senate).

Congressional Budget Office (1994) *Trends in Selected Indicators of Military Readiness*. (Committee on Armed Services, U.S. House of Representatives, Washington, D.C.)

Congressional Budget Office (2011) *Linking the Readiness of the Armed Forces to DoD's Operation and Maintenance Spending*. (U.S. Congress, Washington, D.C.)

Davis S, Jondrow J, Dacus C (2008) *Aggregating Readiness Metrics* (Center for Naval Analyses, Research Memorandum D0017960.A2).

Derringer GC (1994) A balancing act: optimizing a product's properties. *Quality Progress* 27(6):51-58.

Derringer GC, Suich R (1980) Simultaneous optimization of several response variables. *Journal of Quality Technology* 12:214-219.

Frank G, Hubal R, O'Bea M (2007) Using competency definitions to adapt training for mission success. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*.

Frank SA, Gruttke WB, Marlow WH, Mathis SJ (1968) *Readiness Measurements Via Subresource C-Ratings*. (George Washington University, Washington, D.C.)

Galloucis MS (2003) *The Army's Strategic Readiness System (SRS): The Road to Improved Readiness or Just Another Bright Idea?* (National War College, National Defense University, Fort Lesley J. McNair, Washington, D.C.)

Gaver DP, Mazumdar M (1967) Statistical estimation in a problem of system reliability. *Naval Research Logistics Quarterly* 14:473-488.

Grier RA, Fletcher RA, Morrison JE (2012) Defining and measuring military cognitive readiness. *Proceedings of the Human Factors and Ergonomics Society Conference* 56(1):435-437.

Greenberg I (1972) *On the Determination of the "Closeness" to Complete Readiness of Dynamic Readiness* (Graduate School of Public Administration, New York University, New York, New York).

Harrington E (1965) The desirability function. *Industrial Quality Control* 21: 494-498.

Harrison T (2014) Rethinking readiness. United States Air Force: *Strategic Studies Quarterly, Fall 2014*: 38-68.

Headquarters, Department of the Army (2001) *Unit Status Reporting* (Army Regulation 220-1, Washington, D.C.)

Headquarters, Department of the Army (2003) *Unit Status Reporting* (Army Regulation 220-1, Washington, D.C.)

Headquarters, Department of the Army (March, 2006) *Unit Status Reporting* (Army Regulation 220-1, Washington, D.C.)

Headquarters, Department of the Army (December, 2006) *Unit Status Reporting* (Army Regulation 220-1, Washington, D.C.)

Headquarters, Department of the Army (2010) *Army Unit Status Reporting and Force Registration - Consolidated Policies* (Army Regulation 220-1, Washington, D.C.)

Jareb AM, Roepke DM, Damm EM, Pham HT, Macilvaine ME (1999) *Marine Corps Resources to Readiness Study* (Center for Naval Analyses, Alexandria, Virginia).

Jiang W, Murphy T, Tsui KL (2006) Statistical methods for quality and productivity improvement. *Springer Handbook of Engineering Statistics* 173-192.

Junor LJ (2017) *Managing military readiness*. (Institute for National Strategic Studies, National Defense University Press, Washington, D.C.).

Kaczynski B, Foote BL (2005) *A Capabilities Based Readiness Metric* (Operations Research Center of Excellence, United States Military Academy, West Point, New York).

Kaplan S (1972) *A Production Function Approach to the Measurement of Short Term Readiness of Navy Units* (Graduate School of Public Administration, New York University, New York, New York).

Kaplan S (1975) *An Approach to the Measurement of the Short Term Readiness of Military Systems* (New York University, New York, New York).

Kim KJ, Lin DKJ (2000) Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 49(3):311-325

Lockman RF, Stoloff PH, Manheimer BH, Hardgrave JR, Story WF (1970) *Navy Readiness Analysis System Methodology Study* (Institute of Naval Studies of the Center for Naval Analyses, University of Rochester, Rochester, New York).

Matheson JE (2013) Decision analysis in practice. In Encyclopedia of Operations Research and Management Science, edited by Gass S and Fu MC, (*Springer Science and Business*, New York, New York: 372-381).

Menko RH (2004) *Process Enhancement and Database Support for Vehicle Operational Readiness Reporting* (Naval Postgraduate School, Monterey, California).

Moore SC, Hanser LM, Rostker BD, Holroyd SM, Fernandez JC (1995) *A Framework for Characterization of Military Unit Training Status* (The Rand Corporation, Santa Monica, California).

Moore SC, Stockfisch JA, Goldberg MS, Holroyd SM, Hildebrandt GG (1991) *Measuring Military Readiness and Sustainability* (The RAND Corporation, Santa Monica, California).

National Institute of Standards and Technology (2013) *NIST/SEMATECH e-Handbook of Statistical Methods* (retrieved from http://www.itl.nist.gov/div898/handbook/)

Nickel RH, Goodwyn SC, Brown KS (2008) *Readiness Impact of Reduced Funding for Ship Repair Parts* (Center for Naval Analyses, Alexandria, Virginia).

Office of the Secretary of Defense (2017) *Memorandum, SUBJECT: Implementation Guidance for Budget Directives in the National Security Presidential Memorandum on Rebuilding the U.S. Armed Forces* (Washington, D.C.).

Orlansky J, Hammon C, Horowitz S (1997) *Indicators of Training Readiness* (Institute for Defense Analyses, Alexandria, Virginia).

Otto KN, Antonsson EK (1991) Trade-off strategies in engineering design. *Research in Engineering Design* 3(2): 87-103.

Parnell GS, Bresnick TA, Tani SN, Johnson ER (2013) *Handbook of Decision Analy*sis. Hoboken, New Jersey: Wiley.

Peltz E, Robbins M, Boren P, Wolff M (2002) *Diagnosing the Army's Equipment Readiness: The Equipment Downtime Analyzer* (The Rand Corporation, Santa Monica, California).

Pernin CG, Butler DM, Constant L, Geyer L, Long D, Madden D, Peters JE, Powers JD, Shurkin M (2013) Readiness Reporting for an Adaptive Army (The Rand Corporation, Santa Monica, California).

Raffensperger JF, Schrage LE (1997) A new paradigm for measuring military readiness. *Military Operations Research* 3(5):21-34.

Randazzo-Matsel A (2009) *Readiness: A Synthesis of CNA's Work for the Marine Corps* (Center for Naval Analyses, Alexandria, Virginia).

Rumbaugh R (2017) Defining readiness: background and issues for congress. (Congressional Research Service, Washington, D.C.).

Scala, NM, Kutzner, R, Buede, DM, Ciminera, C, and Bridges, A (2012) Multi-objective Decision Analysis for Workforce Planning: A Case Study. *Proceedings of the 2012 Industrial and Systems Engineering Research Conference.*

Scala, NM, Pazour, J (2016) A Value Model for Asset Tracking Technology to Support Naval Sea-Based Resupply. *Engineering Management Journal* 28(2):120-130.

Seppala J, Basson L, Norris GA (2002) Decision analysis frameworks for life-cycle impact assessment. (Massachusetts Institute of Technology and Yale University: Journal of Industrial Ecology 5(4): 45-68).

Shisko R, Paulson RM (1982) *Relating Resources to the Readiness and Sustainability of Combined Arms Units* (The Rand Corporation, Santa Monica, California).

Stafford D, Moore C (2004) *A Metric for Surge Capability: Personnel* (Center for Naval Analyses, Alexandria, Virginia.

Tolins IS (1968) *A Continuous Time Markov Process Model of Naval Operational Readiness* (George Washington University, Washington, D.C.)

Trunkey RD (2013) *Implications of the Department of Defense Readiness Reporting System* (Congressional Budget Office, Working Paper 2013-03)

United States General Accounting Office (1996) *Military Readiness: Data and Trends for January 1990 to March 1995* (Washington, D.C.)

United States General Accounting Office (1997) *Military Readiness: Improvements Still Needed in Assessing Military Readiness* (Washington, D.C.)

United States General Accounting Office (2003) *Military Readiness: New Reporting System is Intended to Address Long-Standing Problems, but Better Planning is Needed* (Washington, D.C.)

United States General Accounting Office (2011) *Military Readiness: Army and Marine Corps Reporting Provides Additional Data, but Actions Needed to Improve Consistency* (Washington, D.C.)

United States General Accounting Office (2016) *Military Readiness: DoD's Readiness Rebuilding Efforts May Be at Risk without a Comprehensive Plan* (Washington, D.C.)

Yoon KP, Hwang CL (1995) *Multiple attribute decision-making: an introduction*. (Sage University: Thousands Oaks, California).

Appendix A (Methodology Notation)

Notation	Definition
C	Overall rating for a unit's "core" functions, the lowest overall rating of four subordinate category assessments
P	Overall rating for the "personnel" category; the lowest overall rating of four subordinate personnel category assessments
S	Overall rating for the equipment on-hand (or "supply") category; the lowest overall rating of two subordinate equipment on-hand category assessments
R	Overall rating for the equipment serviceability or "readiness" category; the lowest overall rating of two subordinate equipment serviceability category assessments
T	Overall rating for the "training" proficiency category; determined by the assessment of the unit tasks (T = Trained, P = Needs Practice, and U = Untrained)
D	Composite desirability function value
w_i	The given weight for the ith response, or sub-category, where $i = 1 \ldots k$
AS	Available strength for personnel (%)
AQS	Available Military Occupational Skill (MOS) qualified strength for personnel (%)
ASG	Available senior grade composite level for personnel
$S{:}AP$	On-hand A-coded and pacing item equipment (%)
$S{:}P$	On-hand pacing item equipment (%)
$R{:}AP$	Serviceable A-coded and pacing item equipment (%)
$R{:}P$	Serviceable pacing item equipment (%)

Appendix B (Notional Unit Data)

Case I: Light Infantry Battalion

Sample Personnel Status

Specialty	Grade	Auth	Asgn	Avail
Infantryman	O5	1	1	1
	O4	1	1	1
	O3	7	7	6
	O2	24	22	21
	E9	1	1	1
	E8	4	4	3
	E7	12	11	9
	E6	38	35	33
	E5	89	75	72
	E1-E4	335	341	328
Supply	E7	1	0	0
	E6	3	2	2
	E5	3	4	4
	E1-E4	3	3	3
Chemical	O2	1	1	1
	E5	4	4	4
Intelligence	O2	1	1	1
	E5	2	2	2
Communications	E6	1	1	1
	E1-E4	4	4	4
Logistics/Maintenance/Other	E7	1	1	1
	E6	3	3	3
	E5	15	15	13
	E1-E4	13	12	11

Sample Equipment Status

Description	Auth	Asgn	% FMC
Antitank Set*	18	18	97
60mm Mortar	6	6	90
81mm Mortar	6	6	91
Machine Gun, 50-cal	15	15	93
Machine Gun, MK-19	14	14	100
Machine Gun, M249	72	70	85
Grenade Launcher, M203	65	65	95
TOW Carrier, Truck	20	20	89
Night Vision Goggles	567	564	96
Mask, Protective	585	581	93
Rifle, M16A2	367	366	91
Pistol, M9	35	35	94

*Pacing Items, **FMC = Fully Mission Capable

Sample Training Status

Mission Essential Tasks	Cdr's Assessment
Execute the Readiness SOP	Needs Practice (P)
Command and Control the Battalion	Trained (T)
Attack	Trained (T)
Defend	Trained (T)
Perform Security Opns	Trained (T)
Conduct Combat Service Spt Opns	Needs Practice (P)

Case II: Armor Battalion

Sample Personnel Status

Specialty	Grade	Auth	Asgn	Avail
Armor	O5	1	1	1
	O4	2	2	2
	O3	8	8	8
	O2	17	17	16
	E9	2	2	2
	E8	5	5	5
	E7	13	13	12
	E6	25	25	24
	E5	52	52	47
	E1-E4	164	164	158
Indirect Fire	E8	1	1	1
	E7	2	2	2
	E6	2	2	2
	E5	2	2	2
	E1-E4	40	40	37
Maintenance (Wheeled Vehicle)	W3	1	1	1
	E8	1	1	1
	E7	5	5	4
	E6	6	6	5
	E5	23	23	22
	E1-E4	34	34	30
Maintenance (Tank)	E5	4	4	4
	E1-E4	13	13	12
Fuel Handling	E6	1	1	1
	E5	10	10	10
	E1-E4	53	53	49
Transportation	E7	1	1	1
	E6	5	5	5
	E5	5	5	4
	E1-E4	21	21	20
Medical	O3	3	3	3
	O2	1	1	1
	E7	1	1	1

	E6	1	1	1
	E5	10	10	9
	E1-E4	23	23	22
Communications/Supply Food Service/Other	O3	3	3	3
	O2	2	2	2
	E7	4	4	4
	E6	9	9	8
	E5	21	21	20
	E1-E4	37	37	34

Sample Equipment Status

Description	Auth	Asgn	% FMC
Tank, M1*	44	41	88
Truck, 5-Ton Cargo	62	58	92
Truck, Fueler	16	15	76
Truck, Pump Unit	19	18	83
Armored Personnel Carrier	11	10	100
Truck, High Mobility	21	21	91
Truck, 2.5-Ton Cargo	17	15	97

Sample Training Status

Mission Essential Tasks	Cdr's Assessment
Command and Control the BN	Needs Practice (P)
Move Tactically	Trained (T)
Attack/Counterattack by Fire	Trained (T)
Assault	Trained (T)
Defend	Trained (T)
Employ Fire Support	Needs Practice (P)
Perform Combat Service Spt Opns	Needs Practice (P)