Navy Personnel Research and Development Center
San Diego, California 92152-7250 TR-98-3 June 1998

A Performance-Based Approach to Human Resource Readiness

B. Charles Tatum
Gerald J. Laabs
Delbert M. Nebeker

NPRDC-TR-98-3 June 1998

A Performance-Based Approach to Human Resource Readiness

B. Charles Tatum
Navy Personnel Research and Development Center

Gerald J. Laabs
Naval Air Warfare Center

Delbert M. Nebeker
California School of Professional Psychology

Reviewed and approved by
Thomas A. Blanco
Director
Personnel and Organizational Assessment

Released by
W. M. Keeney
Commander, U.S. Navy
Commanding Officer
and
Murray Rowe
Technical Director

Approved for public release;
distribution is unlimited

Navy Personnel Research and Development Center
53335 Ryne Road
San Diego, California 92152-7250

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
June 1998

3. REPORT TYPE AND DATE COVERED
October 1994 – December 1997

4. TITLE AND SUBTITLE
A Performance-Based Approach to Human Resource Readiness

5. FUNDING NUMBERS
Program Element: 0602233N
Work Unit: RM33M20

6. AUTHOR(S)
B. Charles Tatum, Gerald J. Laabs, Delbert M. Nebeker

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Navy Personnel Research and Development Center
53335 Ryne Rd.
San Diego, CA 92152-7250

8. PERFORMING ORGANIZATION AGENCY REPORT NUMBER
NPRDC-TR-98-3

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217-5000

10. SPONSORING/MONITORING

11. SUPPLEMENTARY NOTES
Functional Area: Readiness Assessment Technology
Product Line: Readiness Assessment Technology
Effort: Readiness Assessment Technology

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

12B. DISTRIBUTION CODE
A

13. ABSTRACT *(Maximum 200 words)*

Efforts to accurately assess and report naval readiness have been ongoing for over 40 years. Despite some technical improvements over the years, the Navy's system for assessing the readiness of its forces has not progressed much beyond taking static snapshots of current levels of resources in the areas of personnel, material, logistics, and training. Recently, readiness assessment has come under the microscopes of several agencies (e.g., GAO, CBO), and the reports have not been complimentary. The major problems with measuring and reporting readiness today are inaccuracy, lack of comprehensibility, and the inability to forecast trends. This report represents the initial stages of a long term program to improve how the Navy measures and reports its readiness status. The present study (1) discusses the nature of readiness assessment in the context of measurement theory, (2) presents a conceptual model of the readiness process, (3) reports on preliminary findings on linking predictor variables to readiness performance, and (4) discusses the probable impact of adopting three new analytical tools (structural equation modeling, artificial neural networks, and multi-level analyses) to the study of readiness. The report concludes that progress can be made toward more accurate and comprehensive readiness assessment, which is vital to our national interests. The research strategy recommended will lead to a better understanding and prediction of the military capability of our armed forces.

14. SUBJECT TERMS
Readiness, measures of effectiveness, organizational effectiveness, measurement theory, structural equation modeling, artificial neural networks, multi-level analysis

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

Foreword

The Navy, along with the other services, requires accurate and up-to-date assessments of military readiness. This report describes the preliminary stages of developing an assessment technology. These stages include (1) a comprehensive definition of readiness, (2) a conceptual model of readiness, and (3) some preliminary results and a discussion of future research strategies for improving readiness assessment.

The effort was an interdepartmental effort at the Navy Personnel Research and Development Center (NPRDC), and represents the combined talents of research psychologists, statisticians, operations research analysts, computer specialists, and active-duty Navy personnel. The project was supported by the Office of Naval Research as part of an applied research program entitled "Readiness Assessment Technology" (PE0602233N, Project RM33M20, Task 20). The project is part of a series of readiness related research and development efforts at the Navy Personnel Research and Development Center that include: (1) two basic research (6.1) projects (Resource Model of Readiness [PE0601153N] and Group Norms for Performance [PE0601152N]), (2) two recently completed applied research (6.2) projects (Enlisted Requirements Model [PE62936N] and Enlisted Training Readiness Model [PE62966N), (3) an advanced technology (6.3) project sponsored by PERS-2 completed in FY95 (Human Resource Readiness workshops [PE0603707N]), and (4) a continuation of the FY95 advanced technology project supported by PERS-2 for FY96 (Human Resource Readiness Focus Groups [PE0605152N]). In addition, the effort is also supported by two basic research (6.1) projects supported by the Office of Naval Research. These 6.1 projects (conducted at Carnegie Mellon University and the University of California, San Diego) are applying advanced data analytic techniques on readiness data supplied by NPRDC.

The point of contact for this effort is Dr. B. Charles Tatum, Personnel and Organizational Assessment Department, Code 122, Navy Personnel Research and Development Center, San Diego, CA 92152-7250, (619) 553-7955.

W. M. KEENEY
Commander, U.S. Navy
Commanding Officer

MURRAY W. ROWE
Technical Director

Summary

Problem

The recent Department of the Navy downsizing has raised the specter of the "hollow force" that plagued the military services in the late 1970s and early 1980s. Nevertheless, several recent reports by the Congressional Budget Office (CBO) and the Defense Science Board (DSB), addressing the issue of our country's military capability, have concluded that the readiness of the Armed Forces is adequate and high relative to historical levels. This conclusion, however, may be based on a seriously flawed set of measures, as noted in several of these reports. To get an accurate picture of Naval warfighting capabilities, the Navy needs a much improved assessment technology. "Readiness" is defined as the ability of individuals, teams, and units to deploy quickly and perform as they were designed. This report will focus on personnel and training readiness, two areas in which measurement poses serious difficulties.

Background

Two events in the collective memories of the U.S. Armed Service stand out as reminders of how important it is to know the state of military readiness. The beginning of World War II and the post Viet Nam War era of the late 1970s and early 1980s represent periods in which military leaders and political policy makers grossly misjudged our nation's readiness to fight and win wars. Pearl Harbor and Desert One (the failed Iranian hostage rescue) were two tragic examples of these miscalculations. Pearl Harbor is a lesson in the importance of preparing for the unexpected. Desert One is a reminder of the importance of accurate information on how well units can perform in actual combat conditions. The situation is much the same today. Measures of readiness have numerous problems (see reports by the CBO, 1994, and General Accounting Office (GAO), 1994), but the primary problem is that we do not have a set of performance-based measures. Without such measures, the military cannot accurately assess what its units can do, and cannot make reasonable forecasts of future capabilities. If the military had a more sophisticated means of assessing current readiness and forecasting future readiness trends, planners and policy makers could do a better job of managing precious resources and anticipating problems. Calls from many corners have been issued to improve the way we measure and report readiness (CBO, DSB, GAO, Chief of Naval Operations (CNO). This report begins to answer those calls.

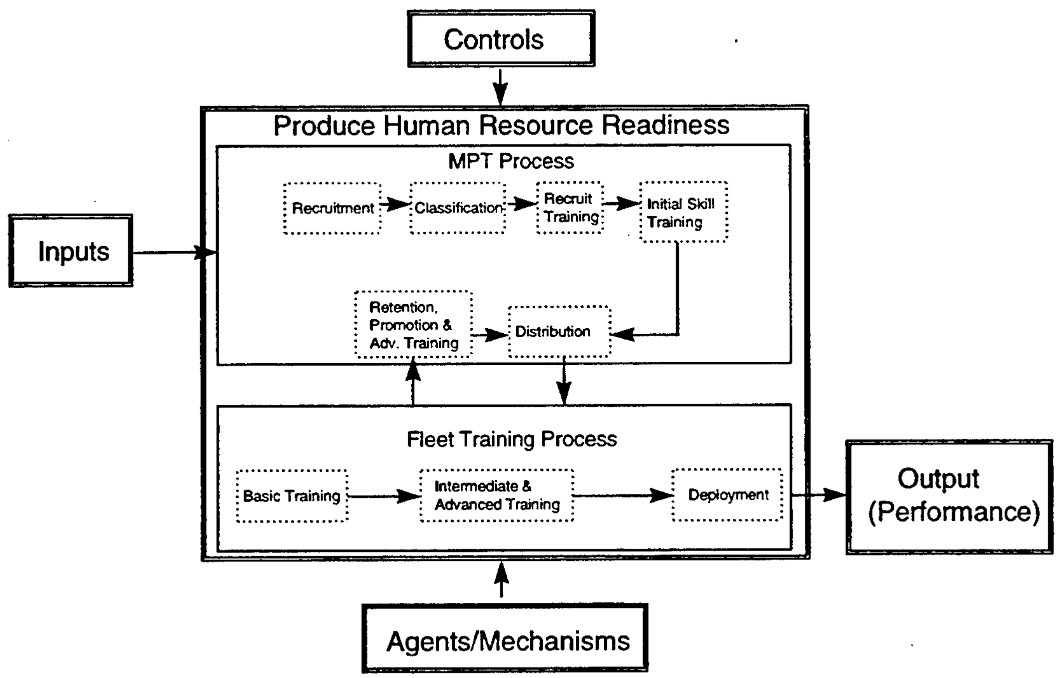
Objective

Improving readiness assessment is not simply a matter of going out and collecting a whole set of "new" measures. The assessment of the effectiveness of any organization, including the Navy, rests on a foundation of goals, definitions, assumptions, concepts, and criteria. Until this groundwork is properly laid, progress in improving the measures and indicators of readiness will be slow and laborious, and probably will not succeed. The immediate objectives of this report are to (1) establish some of the ground rules for readiness measurement, (2) provide a conceptual framework for building a readiness assessment technology, and (3) describe a preliminary analysis of existing data along with the implications of the analysis. Long-term objectives of the project include: (1) identifying ways to improve existing readiness measures, (2) finding cheaper,

less intrusive substitutes for existing measures, (3) creating new readiness measures, and (4) developing mathematical models for better resource management.

Approach

This report covers several areas that are critical to constructing a framework for future research efforts in readiness assessment. First, the report describes a conceptual model of the readiness process (shown here in abbreviated form). The model is not only helpful in integrating the many parts of readiness, it also helps to point out the different kinds of measures required in a complete assessment of readiness. Second, the report describes several empirical findings from an analysis of 16 ships conducting Anti-Submarine Warfare (ASW) exercises. Finally, the report identifies promising measurement technologies (i.e., Structural Equation Modeling, Artificial Neural Networks, and Multi-Level Analysis) that can advance our understanding of the complex set of variables that make up readiness.



Human Resources Readiness Model

Conclusion

To accomplish the goal of maintaining and improving readiness, the Navy needs a much better system of readiness assessment than is currently available. Obviously, there is no guarantee that the approach offered here, or any other approach for that matter, will significantly advance our ability to measure and report our military's readiness status. However, if we can make even incremental improvements in readiness assessment following this approach, the Navy, the military, and the Nation will be better served.

Recommendations

1. Continue ongoing development of a readiness database to be used for analyses employing new measurement technologies.

2. Explore research efforts in the areas of Structural Equation Modeling, Artificial Neural Networks, and Multi-Level Analyses to determine if these advanced techniques can improve our ability to detect readiness relationships and trends.

3. Support efforts to create new readiness measures. This may be accomplished in several ways: (a) find creative ways to use existing measures (e.g., combine existing measures in novel, empirically justifiable, ways), (b) find easily obtained substitutes for current (difficult to obtain) measures, and (c) develop new sets of measures based on new measurement technologies (e.g., expert judgment, data aggregation).

4. Expand existing efforts in mathematical modeling of the link between resources and readiness outcomes.

5. If applied research demonstrates that readiness outcomes can be more accurately predicted, and long-term trends can be reasonably assessed, the research should transition into advanced development of an integrated information system.

Table of Contents

List of Figures

Introduction and Background

> During the cold war, military readiness was a sacred cow. Thousands of Soviet tanks and nuclear warheads stood poised against the West, and the lesson of Pearl Harbor was branded on the brains of American strategists. In an atmosphere of titanic struggle, none but pacifists or radicals challenged the value of readiness, at least until they tried to figure out what it meant and how much of it was enough (Betts, 1995, p.3).

Given that the Cold War is over and the U.S. has emerged as the sole superpower in the world, one might well ask, "why is there a readiness debate?" The answer to that question is complicated, but the short answer is that policy makers do not want to repeat the mistakes of the past. The specter of our woefully unprepared military at the start of the Second World War, and the "hollow force" of the post Viet Nam War era, are legacies that contemporary decision makers ignore at their peril. The recent and rapid downsizing of the military has left many in the highest levels of government wondering whether we are on the "ragged edge" of readiness, with "pockets of unreadiness" already beginning to form (Congressional Budget Office [CBO], 1994; Defense Science Board [DSB], 1994; Pexton, 1994; Prina, 1994).

Readiness in the 1990s may still be a sacred cow, but not just pacifists and radicals are asking the hard questions and posing serious challenges. Of the many questions and challenges, several issues cry out for exploratory research. At the core of these exploratory research issues are the very important questions of (1) how is readiness defined and (2) how is readiness measured? Once the questions of definition and measurement are satisfactorily answered, other issues can be addressed. Some of the more fundamental research issues are: What are the important predictors of readiness? How should we measure these predictors? What is the nature of the relationship between predictor variables and measures of readiness? Can we develop better measures of readiness? How are resources (e.g., dollars) related to readiness?

To define readiness, we must first examine the concept of capability. The Joint Chiefs of Staff (JCS, 1986) defines capability as the ability to achieve wartime objectives. Capability is composed of four elements: force structure (the number, size, and composition of military units), modernization (the technical sophistication of the forces, weapon systems, and equipment), sustainability (the "staying power" of the forces), and readiness. Readiness and sustainability are closely related ideas. Readiness, according to the JCS definition, is the ability of forces to deploy quickly and perform initially in wartime as they were designed; whereas, sustainability is long-term readiness. Readiness and sustainability are often linked as a single concept (e.g., Moore, et al., 1991), and both are viewed as essential elements of overall military capability.

Looking more closely at the JCS definition of readiness, two important ideas emerge. First, readiness places an emphasis on deployment. This is important because readiness is often viewed as an exercise in getting resources (personnel, supplies, equipment, etc.) in place for the start of a war (e.g., the "ability to jump into action" ["Readiness at risk," 1997], or "to get every dog in the fight at the right time and place" [Rosenberger, 1995]). This emphasis on managing resources for deployment is clearly reflected in the Status of Resources and Training System (SORTS) that measures and reports manpower levels, equipment and supply status, and completed training evolutions.

The second thing to note about the JCS definition of readiness is that it also focuses on performance. Performance has not been the primary target of existing readiness assessment systems. SORTS, and other systems of monitoring readiness (e.g., Type Commanders Readiness Management System [TRMS]), mostly assess resource status; operational performance of military personnel or battle units is not adequately evaluated, especially in terms of quality of performance.

A key point to be made in this report is that resource status and performance are both important properties of readiness, and a complete understanding of readiness involves both. We limit our discussion to the human aspects of readiness, what we call "human resource readiness." As shown in Figure 1, human resource readiness combines fleet performance with personnel and training resources. From a systems point of view, performance is the output of the system, whereas resources are the inputs and throughputs (processes) of the system.

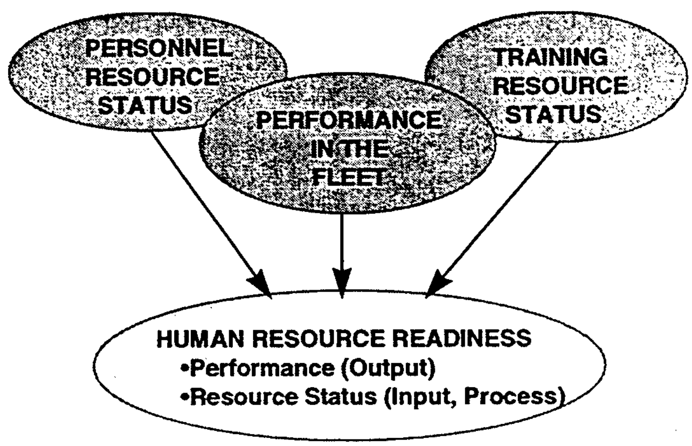


Figure 1. Human resource readiness.

Readiness is more complex than many other systems because it operates on several levels simultaneously. Naval readiness is a multi-level system that encompasses individuals, teams, and units (ships and squadrons). There are also levels above the unit (e.g., fleet, joint, and combined operations), but for the purposes of this report we limit ourselves to units (and occasionally multiple units) for the sake of simplifying an already complicated problem. At each of these levels, readiness can be described as a system of inputs, processes, and outputs (see Figure 2).

Measuring and reporting readiness is not just an interesting academic question. Readiness assessment has emerged as a critical concern for the Navy and other services. The CBO stated that "both Congress and the senior civilian leadership within DoD need access to measures of readiness that are objective, consistent over time, easily understood, and not dependent on assumptions that might be subject to even inadvertent manipulation" (p. 63). In a study prepared for the Under Secretary of Defense (Personnel & Readiness) (USD[P&R]) on training readiness (Burba, et al., 1994), the authors recommended that USD (P&R) work to "improve measurement and reporting of current training readiness" (p. S-3).

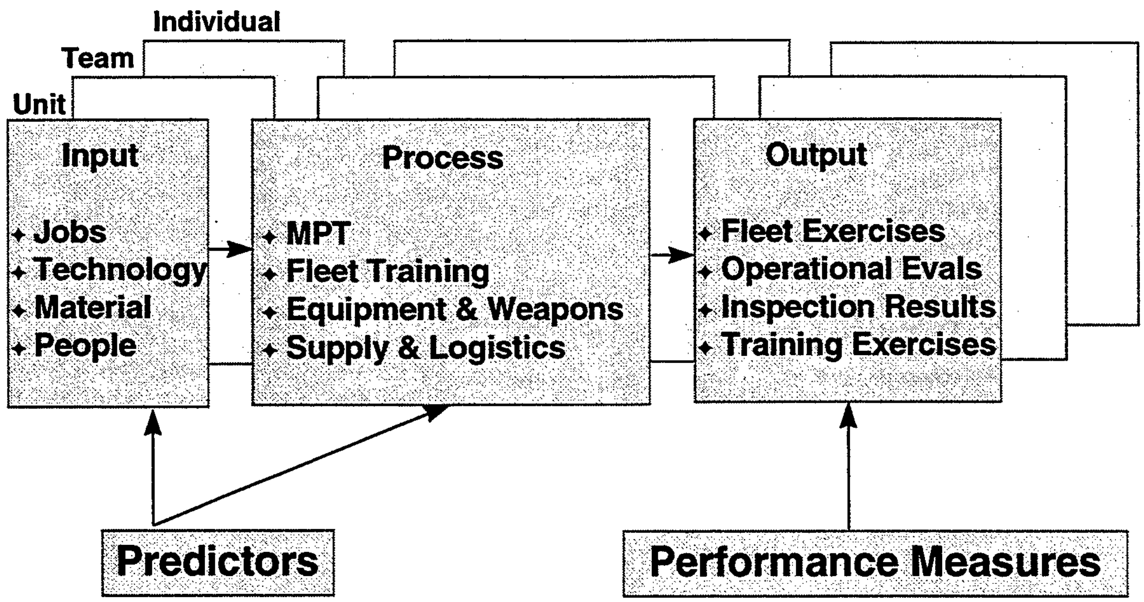


Figure 2. Readiness viewed as a multi-level system.

Accurate and stable readiness measures can serve many purposes, but perhaps the most important purpose is to predict future readiness so planners and policy makers can manage today's resources. The DSB Task Force on Readiness stated that "there currently exists a well-defined reporting system to evaluate the current readiness of combat and support units [but]. . . the Department's systems for predicting future unit readiness [are] significantly less mature and less comprehensive" (DSB, 1994). The Army and the Air Force have made significant advances in their ability to predict and forecast readiness trends (General Accounting Office [GAO], 1994). The Air Force has designed a system called ULTRA, and the Army has developed a Status Projection System. Both systems enhance the ability of these services to project and forecast their readiness status. The Navy has lagged behind these efforts. This need for predictive measures of readiness in the Navy has been recognized by the Chief of Naval Operations (CNO), who recently issued a memorandum on "Predictive Measures of Future Readiness," with specific tasking to "develop a set of measurable predictors of future readiness" (CNO, 1994).

Purpose

The purpose of this report is to address many of the important measurement and research issues that underlie readiness, and report on some preliminary data that link readiness performance to other (predictor) variables. The report is divided into three parts. The first part of the report presents a conceptual model of readiness that emphasizes the manpower, personnel, and training elements of readiness. The second part describes an analysis of performance-based readiness measures. The last part suggests some emerging analytic techniques (viz., Structural Equation Modeling, Artificial Neural Networks, and Multi-Level Analysis), that show some promise for future readiness research. The use of acronyms and abbreviations are unavoidable in any discussion of military capability and readiness. The reader is directed to the Glossary for definitions of abbreviations and acronyms.

A Conceptual Model of Readiness

Research on readiness assessment must be driven by a conceptual understanding of what readiness is and what variables impact readiness in predictable ways. Figure 2 illustrates that readiness is a multi-level system. However, there are generic processes and subprocesses that apply to all levels. So, for example, regardless of whether one is focused on the individual, team, or unit, personnel must be recruited, classified, and trained. Therefore, processes such as recruitment, classification, and training are common to all levels. The following discussion focuses on developing a conceptual model of readiness that is generic to all levels. Techniques for analyzing readiness from a multilevel perspective will be discussed below.

The generic model is structured around techniques for defining Integrated Computer Aided Manufacturing (ICAM) developed by the Air Force (D. Appleton, Co., 1993, p. 10). These ICAM definition techniques (known as IDEF, which stands for ICAM definition) have been used to depict functional requirements for many different processes (e.g., business re-engineering, data architecture, information management, job structuring) and are well suited for modeling readiness as will be shown. The essence of IDEF is shown in the somewhat whimsical example portrayed in Figure 3. An activity (e.g., producing cola) can be modeled by links to inputs and outputs, and by identifying controls that regulate (constrain) the activity, and mechanisms (agents) that provide energy to the activity (D. Appleton, Co., 1993, p. 66-67). In the example shown in Figure 3, inputs (e.g., labor, materials) feed into a production process and result in outputs (bottles of cola). In the example, the FDA and EPA act as controls on the production process and agents/mechanisms such as the MIS system and management aid the process.

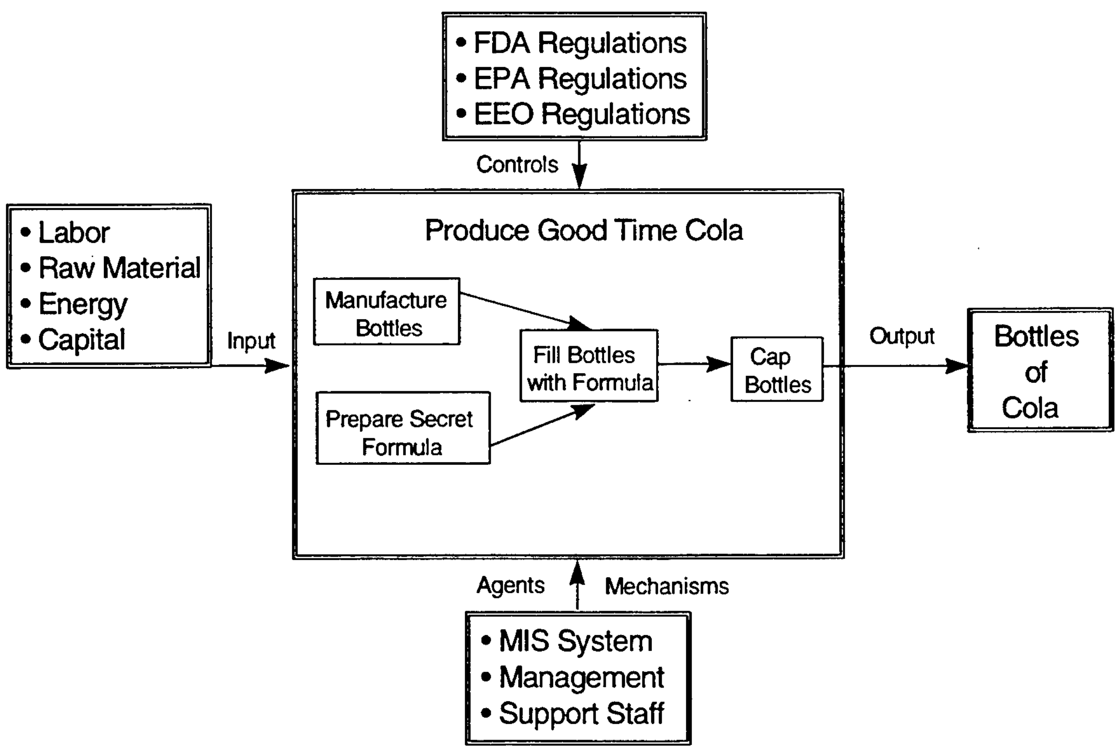


Figure 3. Example of IDEF system analysis.

In a similar fashion, Naval readiness is like a production process (Burba, et al., 1994) as shown in Figure 4. At least four components (personnel, training, equipment, and supplies) combine to produce a Naval force that is ready to fight. These components make up the SORTS reports that are submitted by all the Armed Forces to the Joint Chiefs of Staff. Although there is some debate over whether these are independent processes (Hammon & Horowitz, 1987, Horowitz, 1986), it is generally agreed that all four components are vital to readiness. Figure 4 captures these components as processes that form the basis for operational readiness. Manpower, Personnel, and Training (MPT) is a process that involves personnel recruitment, school house training, and assignment, among others. Fleet Training involves the training that occurs once personnel are assigned to fleet units. The Equipment and Weapons process includes the procurement and maintenance of weapons systems. Finally, Supply and Logistics refers to the process of keeping units well supplied with munitions and goods. These four processes, as shown in Figure 4, are fed by inputs (people, jobs, etc.) and result in the outputs (e.g., inspections, exams, exercise scores) that are indicative of readiness. Also, as shown in Figure 4, there are controls (e.g., resource requirements, policies) and agents/mechanisms (e.g., organizations, communities) that constrain and facilitate these processes.

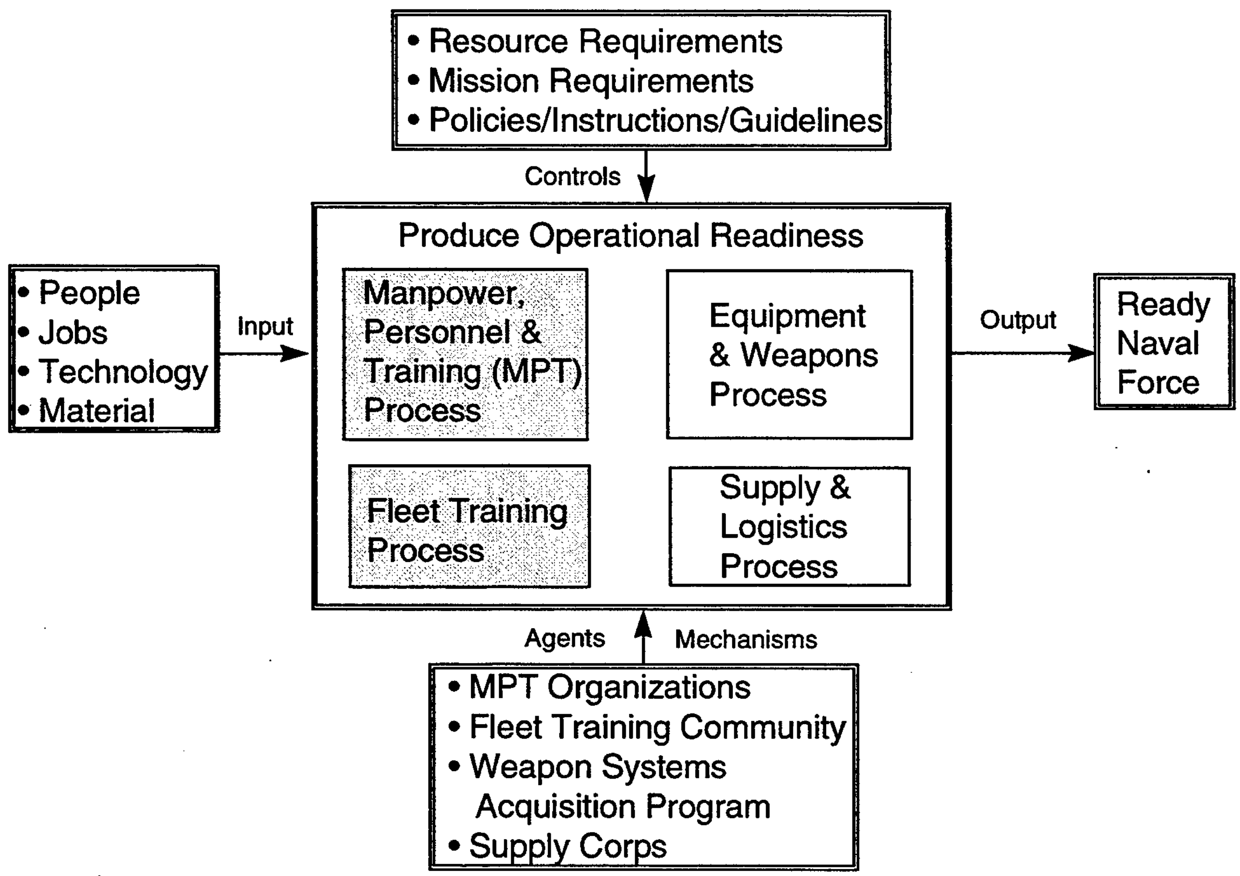


Figure 4. Generic readiness process.

As noted earlier, rather than studying all the activities involved in readiness, we have narrowed the focus of this report on human resource readiness by examining MPT and fleet training, and excluding equipment, weapons, supply, and logistics. Figure 5 presents an IDEF diagram for human resource readiness that incorporates the MPT process and the fleet training process. As shown in Figure 5, the MPT process and fleet training process are further divided into sub-processes. In addition, the controls (e.g., human resource requirements, instructions, policies) and agents/mechanism (e.g., quality of life, sponsors, models) are more specific than for the generic model in Figure 4. As with the more generic version of this IDEF diagram, the output is still a ready force as defined by various performance indicators (e.g., inspections, exams, exercise scores).

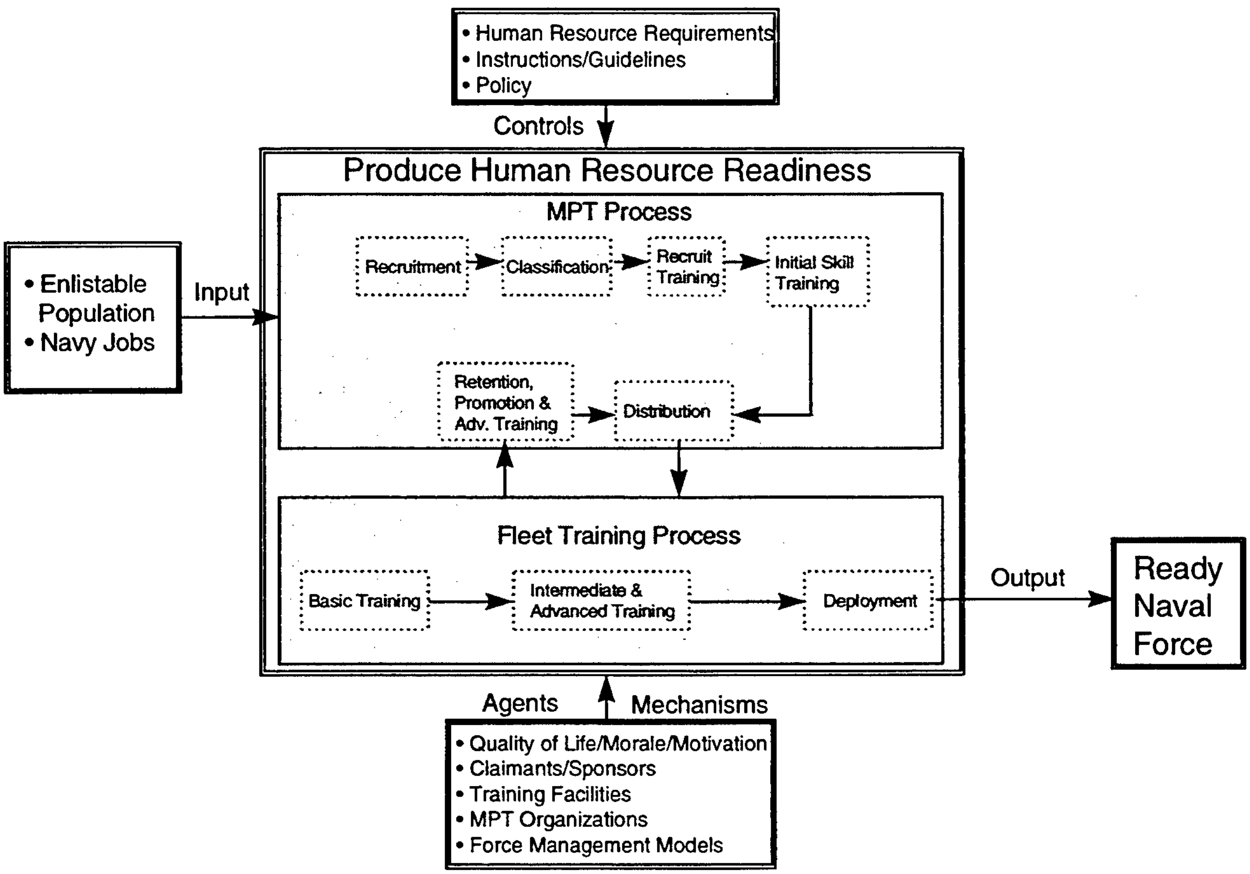


Figure 5. Human resource readiness model.

The sub-processes listed under the MPT process were identified by searching the MPT literature and by solicitation of expert judgment (Systems Research and Applications [SRA], 1995). The expert judges were personnel from the Bureau of Personnel (BUPERS), the CNO, the Chief of Naval Education and Training (CNET), the Chief of Naval Recruiting Command (CNRC), the Defense Manpower Data Center (DMDC), and the Center for Naval Analysis (CNA) who participated in a workshop during July 1995. Experts on each of the sub-processes (e.g., recruitment, classification, recruit training) provided detailed flow charts and discussed their respective sub-processes to the workshop participants. After all sub-processes were presented, the workshop participants formed into smaller groups and discussed (1) links between sub-processes, (2) feedback mechanisms, and (3) measures of effectiveness (MOEs). The sequence of sub-processes shown in Figure 5 under the MPT process is a fair representation of the force management system as it exists today. This "as is" view of the MPT process does not imply that this is the best or "ideal" system, but it does provide a starting point for modeling and measuring the current system.

Figure 6 elaborates on these MPT sub-processes and shows MOEs identified in the workshop. The MOEs shown in Figure 6 are candidates for input into a readiness database (along with other measures). Such a database is under development at the Navy Personnel Research and Development Center (NPRDC) and data are currently under analysis in an attempt to understand the nature of readiness. As discussed in more detail below, studies are underway to (1) identify the best set of existing MOEs for each sub-process (these MOEs are the predictors of readiness) and (2) relate MOEs to performance measures of readiness (e.g., propulsion exam results, fleet exercise scores). The analyses aimed at the second item (link predictors to performance measures) will form the foundation for specifying mathematical models that relate resources to readiness.

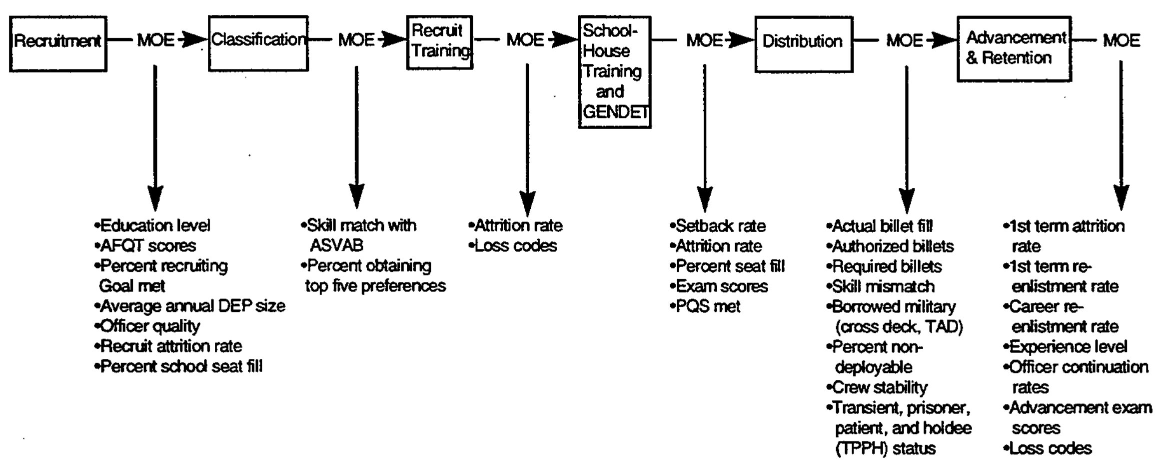


Figure 6. MPT process showing some measures of effectiveness (MOEs) at each stage.

The sub-processes listed under Fleet Training in Figure 5 were identified from two sources. The first source was documentation of surface force training for the Atlantic and Pacific Fleets (Commander Naval Surface Force, 1995). The second source of information came from focus groups conducted by researchers at NPRDC with experts from both the Pacific and the Atlantic fleets (CINCLANTFLT, CINCPACFLT, SURFLANT, SURFPAC, AIRPAC, SUBPAC, ATGLANT, ATGPAC, ATGMIDPAC, TRALANT, and TRAPAC).[1] The different training levels (Basic, Intermediate, Advanced, and Deployment) shown in Figures 5 and 7 reflect fleet training activities conducted by the Afloat Training Groups (ATGs) that emphasizes training tailored to specific missions of each unit. Embedded within each of these more general phases are activities such as Tailored Ship Training Availabilities (TSTAs), Command Assessment

[1] To conserve space, these acronyms and abbreviations are not defined here. Please refer to the Glossary.

Readiness Training (CART), and Final Exercise Periods (FEPs). The MOEs shown in Figure 7, as with many of the MOEs in Figure 6, are potential measures to be used in future studies of readiness. The measures taken during advanced training and deployment are probably the best candidates for readiness performance measures. Wright, Crosby, and Bidgood (1993) have shown that measures taken later in the training cycle (advanced training and deployment) have stronger relationships to predictor variables (e.g., steaming days) than measures taken earlier in the cycle.

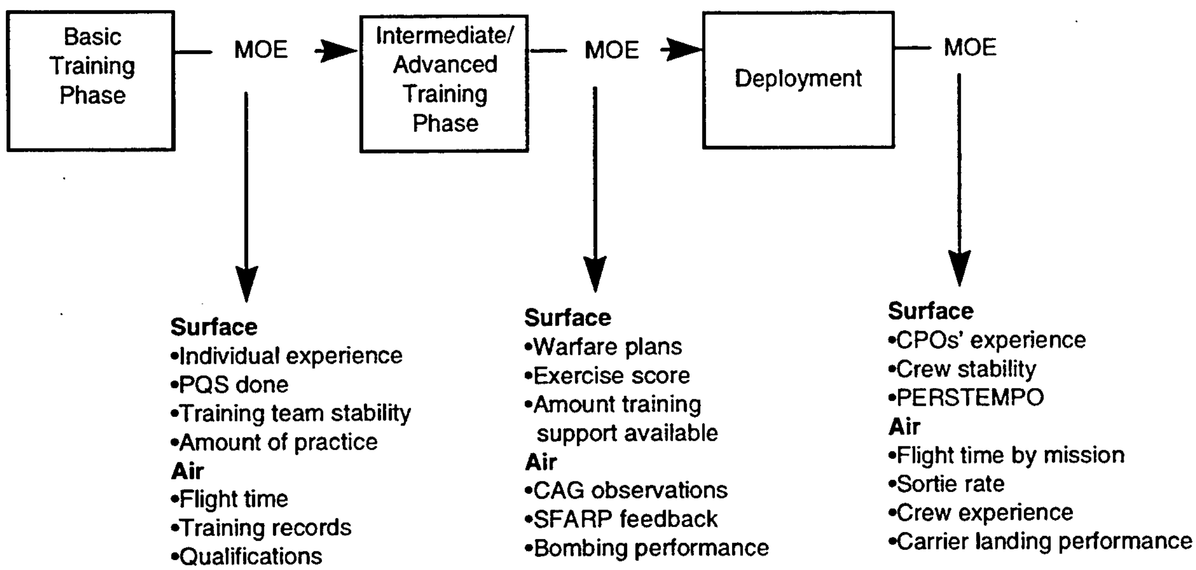


Figure 7. Fleet training process showing some measures of effectiveness (MOEs) at each stage.

Preliminary Analysis of Performance Data

Readiness assessment (measuring and reporting readiness) is fundamentally a problem in the domain of measurement theory and causal modeling. There are many approaches that could be proposed, and some of these will be discussed below. However, the approach adopted for this preliminary analysis employs the statistical technique known as linear regression analysis. Before reporting the results of this analysis, a more general discussion of the readiness assessment approach is in order. The general approach follows three basic steps: (1) develop performance-based criterion measures, (2) identify potential predictors of readiness, and (3) determine the relationships between the predictors and the criterion measures.

The first step is to develop criterion measures. Criterion measures provide the most direct assessment of how well individuals, teams, and units can perform. As noted earlier, the best criterion measures are likely to be those measures taken during the latter parts of the training cycle. It should be noted, however, that any criterion measures selected are merely surrogates of "true" readiness. Because readiness is defined as the ability to deploy and perform during wartime, it is clear that no absolute readiness indictors are obtainable during peacetime. For the purposes of our preliminary analysis, we will use expert ratings of units performing operational training exercises (e.g., torpedo exercises, tracking exercises, final evaluation problems). Unfortunately, the manner in which these expert ratings were obtained precludes an assessment of the reliability and construct validity of these data. Future studies must ensure the reliability and validity of the criterion measures as discussed in Cronbach (1960) and Nunnally and Burnstein (1994).

The second step is to find predictors of the readiness criteria. Such information allows a diagnostic look at what contributes to readiness. In some cases, it may be possible to make changes to upstream processes (e.g., classification, schoolhouse training) to influence downstream outcomes (e.g., better performance during gunfire support, improved submarine tracking). If strong predictor variables can be identified, the predictors can often be used as surrogates for the performance-based measures. This makes it possible to use simpler or more convenient indicators to assess readiness. For example, if we know from empirical studies that crew stability is a predictor of a performance-based measure of readiness (e.g., team performance in a full scale battle exercise), then it makes sense to use that indicator as a substitute for readiness status when we don't have the battle exercise data currently available. To the extent that predictor measures are available substantially ahead of the direct measures, they may also be shown to be valid forecasters of future levels of readiness.

Additionally, strong predictors can be used to decompose readiness into its contributing elements. For example, the number of days a unit spends in a particular Anti-Air Warfare training program might be expected to influence readiness levels. Data on the number of days of training and a valid and reliable measure of their readiness levels can be gathered for many different units. The relationship between days of training and the criterion readiness can then be calculated. This information can be used to establish targets for training days to maintain a desired level of readiness. Other resources or activities can be evaluated in a similar fashion.

The number of potential predictors of readiness seems limitless. The GAO (1994) and others have identified several hundred possible measures without even proposing any new ones. As predictors of readiness are identified, the reliability and validity of the measures must be established in ways similar to those used to develop criterion measures. The dilemma is knowing which potential predictors of readiness are worth the time and effort to measure and collect. Expert opinion is the starting point. However, rather than using experts to make final judgments of what relationships exist between predictors and criterion measures, the judgments are used to screen or narrow the field for empirical analyses and create a model or models of structural relationships between exogenous variables (variables that begin the causal sequence) and endogenous variables (variables causally linked within a model). Such a model is shown in Figure 5. As noted earlier, Figure 5 was derived from the results of a workshop that used subject matter experts in the manpower, personnel, and training areas.

The first two steps laid the groundwork for this final step. The first step develops the performance-based (criterion) measures of readiness. The second step identifies high priority predictors and evaluates these predictors for reliability and construct validity. Also, the second step develops a set of theoretical structural relations. The third step empirically tests whether the predictors are related to the criteria. The MOEs shown in Figures 6 and 7 represent potential predictors of readiness based on the informed judgments of the experts. The actual relationships between these predictors and the criterion measures of readiness must be established empirically.

One approach to empirical testing the structural relations shown in Figure 5 is standard linear-regression (Cohen & Cohen, 1983, Pedhazur, 1997). The standard regression approach applies statistical modeling to determine covariation among sets of measured variables (e.g., predictor variables and performance variables). A regression analysis of a set of readiness data follows.

As stated earlier, we obtained performance-based readiness measures from expert judges rating training exercises. The raters were four, former Navy personnel from DYNCORP, the primary support contractor for conducting ASW training exercises for CINCPACFLT. The experts (who had 75 years of collective Navy experience) conducted a wide array of evaluations (e.g., torpedo exercises, submarine tracking, final battle problems) for 16 different ships. Each ship received a rating by each expert on a 0-4 scale (0 = unsatisfactory, 4 = outstanding) and the four ratings were averaged across the four experts (see Table 1).

Table 1

Ship, Class, and Average Exercise Score

Ship	Class	Average Score
1	Cruiser	1.31
2	Destroyer	3.02
3	Destroyer	2.23
4	Destroyer	2.40
5	Destroyer	3.13
6	Frigate	2.29
7	Cruiser	1.13
8	Destroyer	.95
9	Cruiser	2.75
10	Frigate	2.43
11	Frigate	1.77
12	Frigate	2.40
13	Destroyer	1.75
14	Cruiser	2.82
15	Cruiser	2.10
16	Frigate	2.63

Several things are clear from Table 1. First, the experts were not too lenient nor too strict in their ratings (the average rating was 2.19, which is about the midpoint of the 0-4 scale). Second, the raters were willing to use the full range of the scale (the lowest rating was .95 and the highest was 3.13). Third, the ratings were pretty evenly distributed across different ship classes (the average ratings for the destroyers, frigates, and cruisers were 2.25, and 2.30, and 2.02 respectively). Despite the fact that we do not have reliability and validity statistics on these ratings, the statistics we do have reflect reasonably good measurement properties.

Ideally, the predictor variables selected for this analysis should be a representative sample from each process shown in Figures 6 and 7. Unfortunately, the NPRDC readiness database is not sufficiently developed to yield such a sample. For the present analysis, only the processes of recruitment, skill training, distribution, and operational training were selected, and MOEs from each of these processes were identified. Figure 8 shows the MOEs used as predictor variables, along with a sample of exercises used to asses performance. The predictor variables cover a broad range of measures in the areas of ability, skill level, billet fill, and practice.

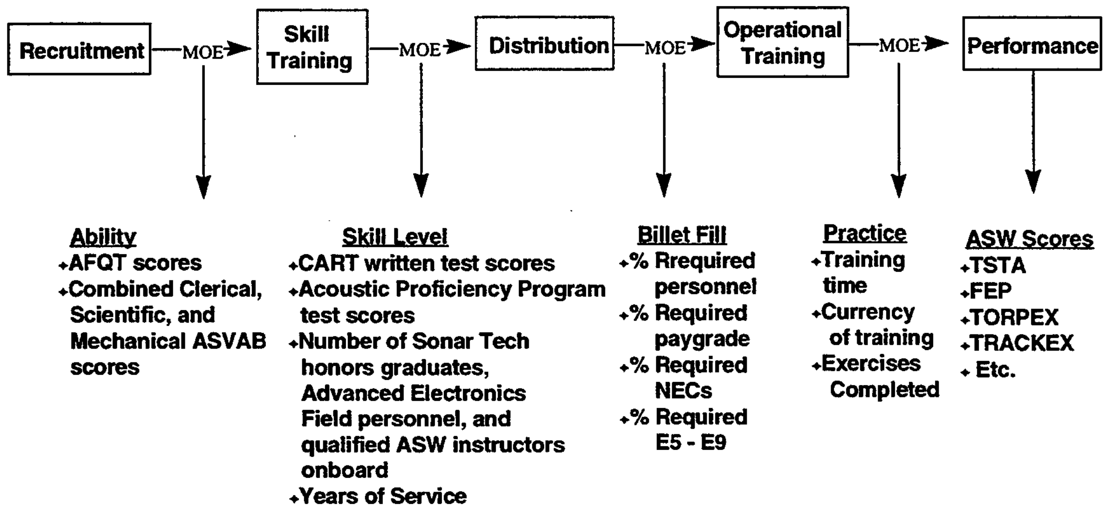
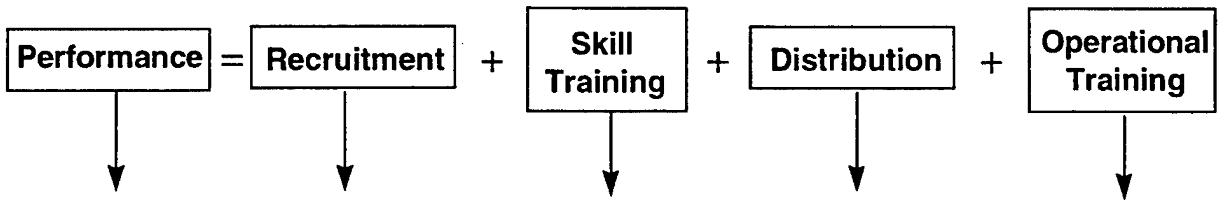


Figure 8. Measures of effectiveness (MOEs) used as predictor and criterion variables in preliminary regression analysis.

Establishing the link between the predictor and the performance variables was a two step process. The first step was to create a composite score for the measures within each process (recruitment, skill training, distribution, and operational training) so that each process could be represented by a single value. For example, operational training is assessed by three separate measures of practice (training time, currency of training, and completed exercises). To combine these three different scores into a single (composite) value, we first converted the training scores into standardized (Z) scores. Then we regressed performance (the average ASW scores) on the standardized scores. Finally, we used the regression coefficients (betas) from the regression analysis as weights for creating a linear combination of the standardized scores. In this particular example, the betas were 1.816 for training time (hours spent performing required ASW exercises), .52 for currency of training (number of days since the last training event), and -1.821 for completed exercises (total number of ASW exercises performed). We derived a composite score for each ship by multiplying the three standardized scores by their respective weights (betas), then added all the scores together. This composite scores then became a single number that reflects the overall level of operational training for each ship. The same basic procedure was conducted for the other three process (recruitment, skill training, and distribution).

Following this first step (creating composite scores), the second step in establishing the link between predictor and performance variables involved another regression analysis. This final regression analysis used the composite scores (one each for recruitment, skill training, distribution, and operational training) as the predictor variables and performance (ASW scores) as the criterion variable. The results of this final regression analysis are shown in Figure 9.



ASW Score = .19(Ability) + .40*(Skill Level) + .53(Billet Fill) + .26(Practice)**

R = .85
R^2 = .72
F (4, 11) = 7.02, p < .01
Beta Coefficients: * p < .05 ** p < .01

	Actual Performance	Predicted Performance	Error
DESTROYERS	2.25	2.32	-.07
FRIGATES	2.30	2.32	- .02
CRUISERS	2.02	1.96	.06

Figure 9. Results of linear regression analysis showing the link between four predictors and one performance measure.

Figure 9 reveals some interesting information. First the strength of the relationship between the predictors and performance is reasonably high ($R^2 = .72$). In other words, 72 percent of the variation in the ASW scores can be explained by the predictors. In social science research, explaining 72 percent of the variance is considered quite good. A second point of interest is that the beta coefficients for two of four predictors achieve conventional levels of significance ($p < .05$). Although it would be far too hazardous to extrapolate from this limited set of data, it is tempting to speculate that perhaps these two predictors (skill training and distribution) are the areas in which policy makers should invest to get the biggest increases in readiness. A third point of interest in Figure 9 is that the ability to predict the performance of the ships is quite good when all of the predictors are used. The difference between the actual performance and the predicted performance (the prediction error) is not large. Again, it is dangerous to generalize, but if these preliminary results hold up under closer examination, it is clear that Navy planners can be confident that their ability to predict readiness with these surrogate measures may be very good.

Other Approaches to Readiness Assessment

Although the standard regression approach used here has many strengths, given some of the unusual characteristics of the readiness process (e.g., multi-level processes, non-linear relationships) there are other techniques that may prove more powerful as readiness assessment tools.

Structural Equation Modeling (SEM)

SEM (sometimes referred to as causal modeling, latent variable models, covariance structural modeling, or linear structural relations [LISREL]) is similar to regression analysis in that it attempts to determine how a set of variables co-vary (Hayduk, 1987; Joreskog and Sorbom, 1987). SEM differs from regression, however, in a number of important ways. First, SEM is a structured technique for determining the plausibility of a model of interconnected variables, such as the model shown in Figure 5. SEM is a formalized technique for testing the "goodness of fit" of the model to the data. In the context of readiness, SEM methods can tell us whether, for example, the model in Figure 6 is a plausible representation of the MPT process given the available data.

A second way in which SEM differs from standard regression analysis is that SEM can test the relationships between latent variables. Many of the predictor and criterion variables in readiness are hypothetical and cannot be measured directly by a single indicator. For example, "mental quality" is a constellation of different components (e.g., Armed Forces Qualifying Test [AFQT] scores, education level) and can not be captured by any single measure. Hypothetical constructs like mental ability, leadership, or unit cohesion are known as latent variables in SEM parlance. SEM has an advantage over regression because relationships between latent variables can be tested as a part of the overall SEM procedure. Latent variable relationships can be tested using regression techniques also, but the procedure is more cumbersome and less precise (as clearly shown in the above attempt to create composite scores). SEM also enables the researcher to test whether the separate indicators of a latent variable converge to form a single, meaningful construct, or whether these indicators form more than one latent variable.

Finally, SEM is well suited for testing mediator and moderator effects among variables. A mediator variable is a variable that accounts for the relationship between two other variables. An example of a mediator variable might be the effects of training (e.g., recruit training, schoolhouse training, fleet training) as shown in Figure 5. For instance, training may account for the relationship between mental quality and job performance (e.g., higher mental quality people may perform better because they are more easily trained). A moderator variable is a variable that changes the relationship between two other variables. An example of a moderator variable might be quality of life, which is shown under Agents/Mechanisms in Figure 5. Suppose that people of higher mental quality perform better than lower mental quality people only when quality of life is good, but there is no relationship between mental quality and job performance when quality of life is bad. Quality of life in this example moderates the mental quality/job performance relationship. SEM can estimate parameters that reflect both mediator and moderator relationships. These parameters are very useful for modeling purposes.

Systematic SEM techniques have not been applied to the readiness process. Fortunately, a group of researchers at Carnegie Mellon University is attempting to apply algorithms for model selection and SEM techniques (embodied in a computer program known as TETRAD II) to the problem of readiness. TETRAD II (see Spirtes, Glymour, & Scheines, 1993) will allow these researchers to explore many of the complex relationships shown in Figure 5 and, perhaps, propose alternative models that are consistent with existing readiness data.

Artificial Neural Networks (ANN)

As powerful as SEM is, it still has one major drawback. SEM assumes that the relationships between dependent and explanatory variables are linear. There are ways of testing for nonlinearity using SEM, but ANN is a technique that will permit flexible, nonlinear modeling of unknown relationships. ANN is also a powerful tool for determining which of a large number of measured attributes are the most important for assessing performance.

The literature on ANN is expanding at a rapid rate, but perhaps the most prominent work is that of Rumelhart, Hinton, and Williams (1986), that propounds the method of "back propagation," and White (1989), that explores the relationships between neural networks and statistics. In these and other ANN approaches, a network is composed of interconnected "units ." Each unit converts the pattern of incoming activities that it receives into a single outgoing activity that it transmits to other units. This conversion is accomplished in two stages. First, the unit multiplies each incoming activity by the weight on its connection and adds together all these weighted inputs to get a total input. Second, a unit uses an input-output ("activation") function to transform the total input into the outgoing activity. The behavior of the neural network depends on both the weights and the activation function. The most common neural networks consist of three layers of units: input units, so-called "hidden" or "intermediate" units, and output units (see Figure 10).

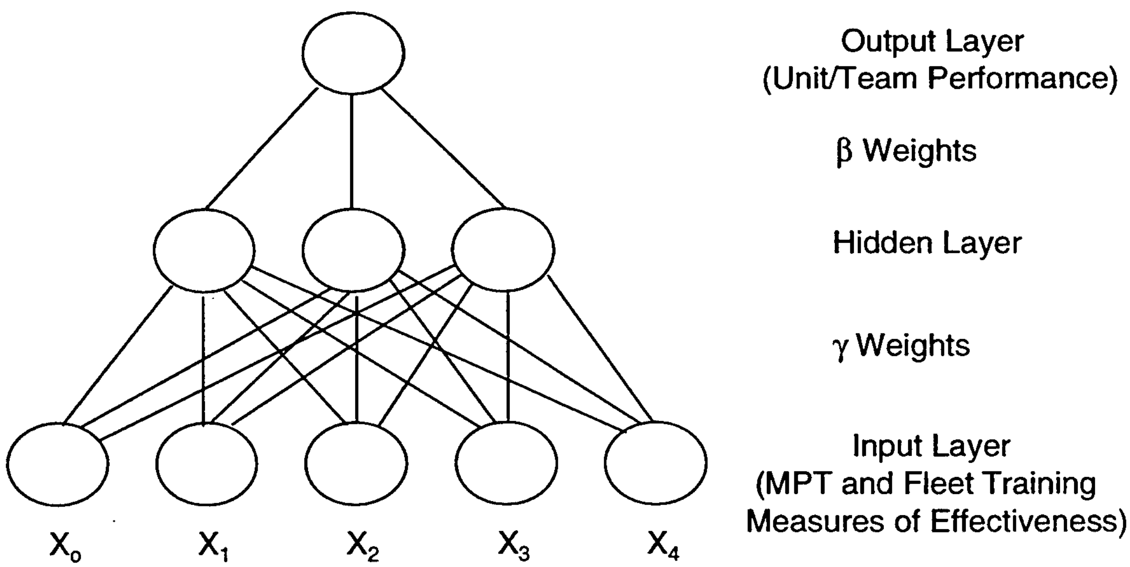


Figure 10. Example of an artificial neural network for predicting unit and team performance.

Such a three-layer network can be trained to perform different tasks through a simple "training" exercise. First, present the network with training examples (each example is a pattern of activities for input units together with the desired pattern of output activities). Then, determine how closely the actual output of the network matches the desired output. Finally, change the weight of each connection so that the network produces better and better approximations of the desired output. The result of this iterative process is an appropriate weight for each connection. A standard procedure for assigning weights is known as "back propagation." This involves assigning weights to units using mathematical calculations that start with the output connection weights and work backwards through the hidden layer connection weights until the input unit connection have been reached. The mathematics allows both linear and nonlinear weighting combinations.

The implications of ANN for readiness are striking. If we can supply the desired output (e.g., quantitative measures of naval force readiness as shown in Figure 5) and a large number of inputs (e.g., quantitative measures of recruitment results, classification data, training scores), ANN procedures can find the optimal set of weights for predicting the outputs from the inputs. For example, suppose we have exercise scores for a large number of Naval units (e.g., ships performing gunfire support), and a large number of crew attributes (e.g., NEC distributions, test scores, manning levels). Given this information, ANN procedures can provide us with the optimal set of weights for predicting gunfire support from crew attributes. In addition, ANN analyses can tell us which attributes are most relevant to the prediction and which attributes are unimportant. Identifying relevant variables can be vital for future data collection because it narrows the range of choices. ANN has not been applied to naval readiness, but researchers at the University of California, San Diego, are embarking on such an effort (see related research by White, 1989). Should the effort prove successful, the Navy will acquire a powerful tool for predicting readiness outcomes from multiple inputs.

Multi-Level Analysis (MLA)

As shown in Figure 2, readiness is a layered, or multi-level phenomenon. Such phenomena require special analytical treatment. The concept of levels of analysis has been discussed for some time (Hammond, 1973, Robinson, 1950). It has only been recently, however, that methodological research has progressed to the point that conceptual methods and techniques have been available to help understand and deal with the problems of multi-level research (Burstein, Linn, & Capell, 1978; Dansereau, Alutto, Yammarino, 1984; Hill, 1982; Irwin & Lichtman, 1976; James, 1982; Lincoln & Zeitz, 1980; Mathieu, 1988; 1990, 1991; Mathieu et al., 1993; Mossholder & Bedeian, 1983; Rousseau, 1978; 1985; Yammarino & Bass, 1990; 1991; Yammarino & Markham, 1992).

Level, as it is used here, refers to a hierarchical structure among related elements (e.g., in biological systems the human body as an entity is composed of systems, composed of organs, which in turn are composed of cells). Organizations are hierarchical systems too (e.g., the Navy is composed of fleets, made up of battle groups, composed of ships and squadrons, made up of teams of people). Given rational criteria for distinguishing one level from another, each of these levels can be an appropriate level of analysis. For example, Figure 11 illustrates how readiness at the team level is influenced by properties of the units (e.g., ships) in which the teams operate and the characteristics of the individuals that make up the teams. Similar diagrams could be constructed for readiness at any level (individual, ship, fleet, etc.).

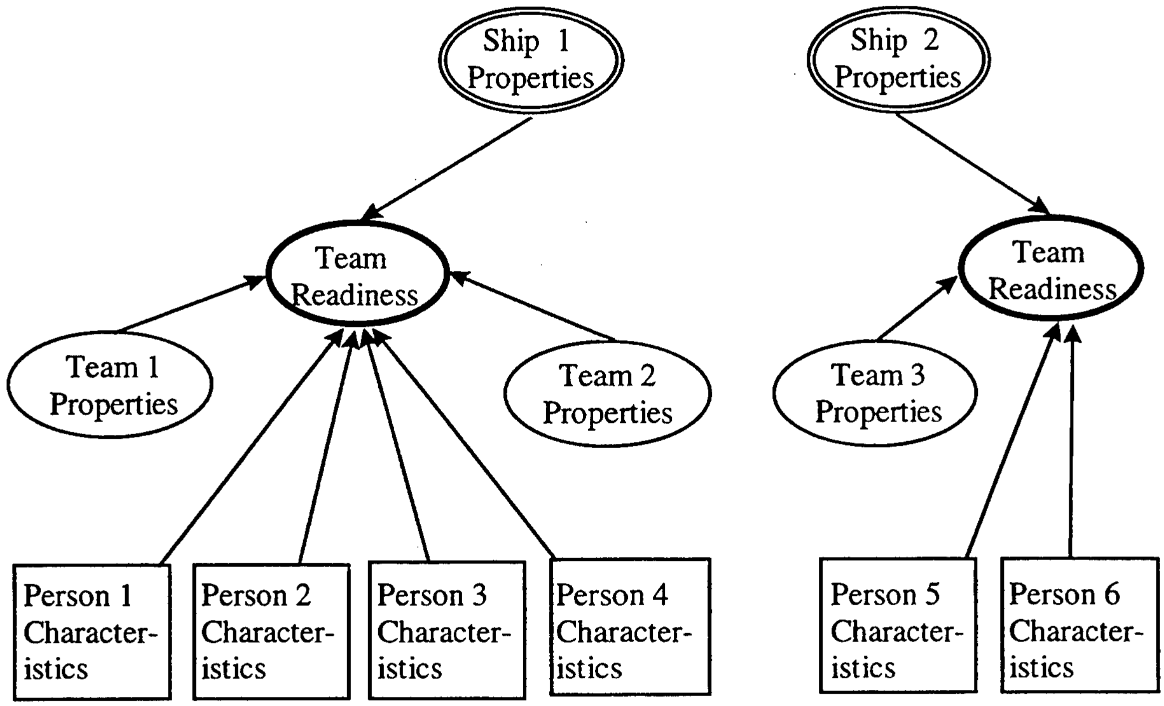


Figure 11. Team readiness influenced by unit (ship) properties and individual (personal) characteristics.

One of the common practices of research on hierarchical systems is to aggregate data from a lower level to a higher one. It is also possible to disaggregate data from a higher level to a lower one. Because aggregation and disaggregation change the variances and covariances of the data, their correlations and regression coefficients are changed as well. As a result, their meaning, and even their character, are possibly changed. For example, at the individual level, we know from census data that higher education is associated with higher incomes. It would be wrong, however, to generalize this relationship to higher levels of aggregation (e.g., organizations). Organizations composed of people with higher education (e.g., colleges and universities) are not necessarily more affluent. As another example, Pedhazur (1997, p. 682) reports a correlation between height and math achievement at the group level (groups of males and females), but no such correlation at the individual level. The seeming paradox is due to the fact that males tend to score higher on math achievement tests and also tend to be taller than females. When males and females are aggregated into groups, the disparity in height and test scores creates a spurious correlation at the group level that does not exist at the individual level.

The basic problem of multi-level analysis is misspecification (Rousseau, 1985). This is when a relationship observed at one level is attributed to another; for example, when team performance is attributed to individual-level results (as with group math achievement noted above) . Although Armed Services Vocational Aptitude Battery (ASVAB) scores may capture individual mental quality, the assumption that aggregated ASVAB scores measure an important team ability may be questionable. It may be true that teams composed of individuals with high ASVAB scores perform better than teams with low ASVAB scores. But, unless ASVAB scores are shown to have construct validity at both the individual level <u>and</u> when aggregated at the team level, we run the risk of misspecification. Only when ASVAB scores are shown to be valid at both the individual and team level are we justified in aggregating ASVAB scores and using these scores as indicators of team performance.

Another significant problem for multi-level analysis is aggregation bias. Generally, aggregation bias is a statistical artifact. Depending on the pattern of within-group and between-group variances, the correlations of aggregated variables at a higher level will be an overestimate of the individual level relationships (Rousseau, 1985). For example, correlations of aggregate variables based on homogeneous groups are higher than their individual-level counterparts. As a result, spurious results can be obtained from aggregated data. This is particularly a problem when attempts are made to infer individual level relationships from aggregated data. The problem is not in aggregating *per se,* but in the incongruity between the level of the conceptual construct and the operational measures used to represent it.

A final problem arising in multi-level analyses is a failure to specify what effects social or physical settings have on a relationship. This problem is referred to as the contextual fallacy. The contextual fallacy is related to our earlier discussion of moderator variables; the context has a moderating effect on a relationship. When moving from one level to another, is the context of a relationship changed such that a relationship at one level is now altered at the new level? Problems occur both when we assume that context makes no difference when it in fact does, or when we assume context has an effect when it does not.

As a prelude to the application of formal MLA methods, more traditional methods can be used to perform the necessary analyses to detect misspecification errors and other biases (Mossholder & Bedeian, 1983). In addition, moderated multiple regression (Cohen & Cohen, 1983) can be an effective tool for performing many of the required analyses to detect context and moderator effects. Once these preliminary analyses are performed, recently developed MLA techniques should be relevant to readiness research efforts. One method called WABA (Within Analysis/Between Analysis) has spawned a substantial book (Dansereau et al., 1984). WABA is the only method currently designed exclusively to conduct multi-level analyses, but innovations is multi-level SEM are beginning to show promise as an analytical technique (see Muthen, 1994).

To summarize, the approach to readiness assessment advocated in this report involves three general steps. The first step is to develop readiness criteria. These criteria should be based on performance measures that directly reflect readiness outcomes (e.g., exercise scores, inspection results). The relevant outcomes should based on the judgments of Naval experts regarding which missions are essential and which standards of performance are appropriate. The criterion measures should also reflect performance at many levels (e.g., individual, team, unit). The second step is to identify the major predictors of readiness and develop reliable and valid measures of these predictors. Also, the second step requires modeling plausible structural relations, such as those shown in Figures 5 through 7. The third step is to establish the empirical relationships between the predictors and the criteria. There are several quantitative tools for investigating these relationship. Besides standard regression analysis, three promising techniques were described above (i.e., SEM, ANN, and MLA).

Summary and Conclusions

There is little doubt that readiness is a top priority among lawmakers and DoD officials. To quote former Defense Secretary William Perry:

"From the moment I took office as Secretary of Defense, my No. 1 priority has been readiness. . . . Every decision I make is driven by the fundamental determination that a ready military force comes first and that all else supports that goal (Perry, 1995, p.9)."

As recently as January 30, 1998, President Clinton, in a speech to the National Defense University, stated that "Readiness remains the number one priority" (Bender, 1998).

With this kind of emphasis, it only makes sense that readiness be clearly defined and accurately measured. Once the definitions and measurement problems are adequately dealt with, it is equally important that we understand the linkages between the resources invested in readiness and the actual results achieved. Past attempts at measuring readiness and linking resources to these readiness measures have not produced remarkable results. The current SORTS readiness reporting system is flawed in many ways, as documented in this report and elsewhere (e.g., CBO, 1994, GAO, 1994). Moreover, the empirical studies to date that have attempted to identify predictors of readiness leave much to be desired. For example, studies of Navy aircraft and ships have found modest relationships between flying hours, steaming days, and operational performance on such critical tasks as carrier landings, bomb drops, and gunfire support (Polich, Winkler, Fernandez, Gotz, & Wild, 1989, pp. 8-17). As important as these studies are, they leave many questions unanswered and the findings are not particularly useful for planning or policy purposes.

This report attempts to improve our understanding of readiness by clarifying what is meant by readiness, discussing vital measurement issues, and reporting some preliminary findings using what appears to be a promising research strategy. A conceptual model is presented that shows the existing manpower, personnel, and training systems that underlie current Naval readiness. Three new approaches for analyzing readiness data (i.e., SEM, ANN, MLA) are outlined. Given the magnitude of the problem, there is no guarantee that any of these emergent measurement technologies will improve the Navy's ability to assess readiness. But, the track records of each of these technologies are impressive in non-military settings (e.g., education, economics, artificial intelligence), so the prospects are good that substantial improvements in readiness assessment can be made.

If we can show strong relations between resource inputs and readiness outcomes, these relationships can form the foundation for several future developments that could benefit the Navy enormously. One immediate possibility is that new, improved measures of readiness can be generated. These new measures could take several forms. New measures could simply be a re-formulation of old, established measures. For instance, SORTS currently reports information on the completion of training exercises. This information is useful, but could be made more valuable if it were supplemented with information on the level of success. Units are loath to report such "sensitive" information (few commanding officers want to run the risk of looking bad), but if the data were aggregated, and no individual units were identified, readiness could be measured without putting anyone "on report." Although unit level information would be lost when aggregating, much more would be gained by having more accurate information.

New measures could also take the form of substituting direct, but difficult to obtain measures, with indirect, more easily assessable measures. This can happen, for example, when there is a strong relationship between a resource input and a performance output. Substituting an easily obtained resource input (e.g., manning level) for a more difficult to obtain output measure (e.g., fleet exercise performance) is justified if we know that the input (manning level) accurately predicts the output (exercise). Some years ago, a study by Holzbach and Williams (1976) showed a strong relationship between manning levels and gunnery scores for carriers. These results suggest that the relatively easily obtained data on manning levels could serve as a surrogate for the more arduous task of conducting gunnery exercises. Of course, this does not mean that the Navy should eliminate gunnery exercises, but it does mean that manning levels can serve in lieu of exercise data, in some cases, when assessing readiness. Current SORTS reporting uses manning levels to assess personnel readiness, but there is no empirical justification for this. As Holzbach and Williams (1976) showed, the value of manning level data depends on type of ship, paygrades, ratings, and functional areas, and these relationships must be systematically mapped out before surrogate measures can be proposed.

Finally, a comprehensive study of a large set of readiness variables may result in measures never before contemplated. It is difficult to imagine what genuinely "new measures" would look like, but a few possibilities can be suggested. One thing that might be learned from sophisticated analyses such as SEM, ANN, and MLA is that expert judgment may be an excellent predictor of readiness when compared to more objective measures (e.g., rounds on target, successful missile launches, submarine tracking scores). Powerful techniques for collecting and organizing expert judgments currently exist (e.g., the Nominal Group Exercises, the Delphi Technique), and these techniques might be applied to make accurate assessments of readiness. Another possibility is that innovative techniques might be developed for aggregating individual level data up to higher levels (teams and units). Perhaps data from the sophisticated embedded training that exists on AEGIS class ships could be combined in ways that would predict the performance not only of the individual crew members, but also entire teams, ships, or fleets.

Leaving new measures aside, another benefit of studying the relationships among readiness variables is that these relationships may, at some point, be expressed in mathematical terms and used to forecast future readiness and conduct accurate cost accounting. Investigations of readiness have concluded that the current system of readiness reporting does a poor job of forecasting future readiness, and has limited usefulness for budgeting and life-cycle cost accounting (CBO, 1994; DSB, 1994). If strong relationships can be discovered between resource dollars and readiness outcomes, and these relationships can be expressed mathematically, defense planners will have a powerful tool for analyzing and justifying their budgets and accounts.

There are many more benefits from this research strategy that could be mentioned (e.g., data could be made available for strategic planning, information could be used to improve program results, reports could be used for goal setting and feedback). Suffice it to say that increasing our understanding of readiness, and developing better ways to assess readiness, is potentially advantageous to many factions: defense planners could have improved planning tools, policy makers might make better decisions, operational units may perform better, and the taxpayer should save money. Ultimately, the biggest benefit could be improved readiness. The old adage "you can't improve what you can't measure," applies with equal force to readiness as it does to any other complex system. Improving readiness assessment technology should, in the final analysis, improve our warfighting capability and our national defense.

Recommendations

1. Support ongoing development of readiness databases to be used for analyses employing a variety of measurement technologies. The current effort to create a readiness database for the Navy is making good progress, and this effort should continue. The structure of the database should be compatible with joint service data, but will include information unique to the Navy.

2. Support research efforts in the areas of Structural Equation Modeling, Artificial Neural Networks, and Multi-Level Analyses. These are promising measurement approaches because they are capable of handling many of the complexities that exist within the readiness process. The ability to deal with complex sets of variables increases the likelihood that these tools will more accurately predict readiness outcomes and make more reasonable long-term forecasts than do existing systems.

3. Support efforts to create new readiness measures. This may be accomplished in several ways: (a) find creative ways to use existing measures (e.g., combine existing measures that have been demonstrated empirically to predict readiness), (b) find easily obtained substitutes for current (difficult to obtain) measures, and (c) develop new sets of measures based on new measurement technologies (e.g., expert judgment, data aggregation).

4. Expand current efforts in mathematical modeling (e.g., broader scope, additional techniques) of the resource-to-readiness link. Quantitative representations of the relationships between resource inputs and readiness outcomes are vitally needed by budget analysts and cost accountants who must manage complex military systems.

5. If the current effort demonstrates that readiness outcomes can be accurately predicted, and long-term trends can be reasonably assessed, the research should transition into advanced technology development. Advanced technology development should focus on creating a prototype of an integrated information system that uses some combination of SEM, ANN, and MLA as the underlying statistical model. A prototype system should be evaluated against several criteria (e.g., reliability, accuracy, cost, ease of use, acceptability) to determine its efficacy and superiority to existing readiness measurement approaches.

References

Bender, B. (1998, January). Clinton: Readiness will remain top priority in DoD Budget. *Defense Daily, 1*.

Betts, R. K. (1995). *Military readiness: Concepts, choices, consequences*. Washington, DC: The Brookings Institute.

Burba, E. H., Chambers, A., Chambers, L. C., Gray, A. M., Metcalf, J., Russ, R. D., & Vermilyea, C. L. (1994). *Training readiness in the department of defense* (IDA document D-1517). Alexandria, VA: Institute for Defense Analysis.

Burstein, L., Linn, R. L., & Capell, F. J. (1978). Analyzing multilevel data in the presence of heterogeneous within-class regressions. *Journal of Educational Statistics, 3*(4), 347-383.

Chief of Naval Operations (1994). Memorandum for the Vice Chief of Naval Operations dated 8 July, 1994, (Subj: *Predictive Measures of Future Readiness*).

Cohen, J. & Cohen, P. (1983*). Applied multiple regression and correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Commander Naval Surface Force (1995, June). *Surface force training manual (COMNAVSURFLANTINST 3502.2B, COMNAVSURFPACINST 3502.2B)*. Norfolk, VA and San Diego, CA: Author.

Congressional Budget Office (1994). *Trends in selected indicators of military readiness, 1980 through 1993*. Washington, DC: Author.

Cronbach, L. J. (1960). *Essentials of psychological testing*. New York: Harper and Row.

D. Appleton, Inc. (1993). *Corporate information management: Process improvement methodology for DoD Functional Managers*. Fairfax, VA: Author.

Dansereau, F., Alutto, J. A., & Yammarino, F. J. (1984*). Theory testing in organizational behavior: The variant approach*. Englewood Cliffs: Prentice-Hall, Inc.

Defense Science Board (1994). *Report of the Defense Science Board Task force on readiness* (Reference # 94-S-2378). Washington, DC: Author

General Accounting Office (1994, October). *Military readiness: DOD needs to develop a more comprehensive measurement system*. (GAO Publication GAO/NSIAD-95-29). Washington DC: Author.

Hammon, C. P., & Horowitz, S. A. (1987). *Relating personnel and training resources to unit performance* (IDA Paper P-2023). Alexandria, VA: Institute for Defense Analysis.

Hammond, J. H. (1973). Two sources of error in ecological correlations. *American Sociological Review, 38*, 764-777.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL.* Baltimore, MD: The John Hopkins University Press.

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 3*, 517-539.

Holzbach, R. L. & Williams, H. L. (1976). *Enlisted manning levels and ship performance* (NPRDC TN 76-9). San Diego, CA: Navy Personnel Research and Development Center.

Horowitz, S. A. (1986). *Evaluating Navy manpower, personnel, and training policies in terms of performance.* (IDA Paper P-1919). Alexandria, VA: Institute for Defense Analysis

Irwin, L., & Lichtman, A. J. (1976). Across the great divide: Inferring individual level behavior from aggregate data. *Political Methodology, 3*, 411-439.

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*(2), 219-229.

Joint Chiefs of Staff (1986). *The dictionary of military and associated terms* (JCS Publication 1). Washington, DC: Author

Joreskog, K. & Sorbom, D. (1987). New developments in LISREL. *Data Analyst, 4*, 1-22.

Lincoln, J. R., & Zeitz, G. (1980). Organizational properties from aggregate data: Separating individual and structural effects. *American Sociological review, 45* (June), 391-408.

Mathieu, J. E. (1988). A causal model of organizational commitment in a military training environment. *Journal of Vocational Behavior, 32*, 321-335.

Mathieu, J. E. (1990). A test of subordinates' achievement and affiliation needs as moderators of leader path-goal relationships. *Basic and Applied Social Psychology, 11*(2), 179-189.

Mathieu, J. E. (1991). A cross-level nonrecursive model of the antecedents of organizational commitment and satisfaction. *Journal of Applied Psychology, 76*(5), 607-618.

Mathieu, J. E., Rousseau, D. M., Yammarino, F. J., Hofmann, D. A., Griffin, M. A., & James, L. R. (1993 May 1). *Multi-level theories, methodologies, and analyses: A comparative and interactive session.* In Anonymous. San Francisco, CA: Society for Industrial/Organizational Psychology.

Moore, S. G., Stockfisch, J. A., Goldberg, M. S., Holroyd, S. M., & Hildebrandt, G. G. (1991). *Measuring military readiness and sustainability* (R-3842-DAG). Santa Monica, CA; RAND.

Mossholder, K. W., & Bedeian, A. G. (1983). Cross-level inference and organizational research: Perspectives on interpretation and application. *Academy of Management Review, 8*(4), 547-558.

Muthen, G. O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376-398.

Nunnally, J. C. & Burnstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction.* Orlando, FL: Harcourt Brace.

Perry, W. J. (1995, June). In military, readiness is happiness. *Richmond Time-Dispatch*, 9.

Pexton, P. (1994, April). Ragged edge of readiness. *Navy Times*, 12-13.

Polich, J. M., Winkler, J. D., Fernandez, J. C., Gotz, G. A., & Wild, W. G. (1989). *Operational training and the performance of combat units* (WD-4472-FMP). Santa Monica, CA: RAND.

Prina, L. G. (1994, October). Bipartisan concern over decline in defense readiness. *Sea Power*, 33-38.

Readiness at risk (1997, September 8). *Denver Post*, p. B8.

Robinson, S. W. (1950). Ecological correlations and the behavior of individuals. . *American Sociological Review, 15*, 351-357.

Rosenberger, J. D. (1995). The key to joint readiness, *U.S. Naval Institute Proceedings, 121*(9), 58-61.

Rousseau, D. M. (1978). Characteristics of departments, positions, and individuals: Contexts for attitudes and behavior. *Administrative Science Quarterly*, 23 (December), 521-540.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*, 1-37.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In E. E. Rumelhart & J. L. McClelland (Eds.). *Parallel distributed processing: Explorations in microstructures of cognition (Volume 1).* Cambridge, MA: MIT Press

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Lecture notes in statistics: Causation, prediction, and search.* New York: Springer-Verlag.

Systems Research and Applications (1995). *Final technical report of manpower, personnel, and training "resources to readiness"* (Contract # MDA903-91-D-0061), Arlington, VA: Author.

White, H. (1989). Neural network learning and statistics. *AI Expert, 4*, 48-52.

Wright, L. W., Crosby, M. E., & Bidgood, M. D. (1993). *Output measures of training readiness.* Report prepared for by OGDEN/ERC government Systems for the Program Resources Appraisal Division, OP-81, Department of the Navy (Contract No. N00014-91-C-0051). Washington, DC.

Yammarino, F. J., & Bass, B. M. (1990). Transformational leadership and multiple levels of analysis. *Human Relations, 43*(10), 975-995.

Yammarino, F. J., & Bass, B. M. (1991). Person and situation views of leadership: A multiple levels of analysis approach. *Leadership Quarterly, 2*(2), 121-139.

Yammarino, F. J., & Markham, S. E. (1992). On the application of within and between analysis: Are absence and affect really group-based phenomena? *Journal of Applied Psychology, 77*(2), 168-176.

Glossary

ACTEP	Aegis Combat Training Evaluation Plan
AEGIS	A class of Naval combatants (not an acronym)
AFQT	Armed Forces Qualification Test
AIRPAC	Air Force, Pacific
ANN	Artificial Neural Networks
ASVAB	Armed Services Vocational Aptitude Batter
ATGLANT	Afloat Training Group, Atlantic
ATGMIDPAC	Afloat Training Group, Middle Pacific
ATGPAC	Afloat Training Group, Pacific
ATSG	Aegis Training Support Group
AW	Air Warfare
BGCTT	Battle Group Commander Team Training
BGEX	Battle Group Exercise
BUPERS	Bureau of Naval Personnel
CAG	Combat Air Group
CART I	Command Assessment Readiness Training (review training, material, and personnel readiness before the end of a major deployment)
CART II	Command Assessment Readiness Training (identify required training and assess proficiency in each mission area prior to major deployment)
CASREPS	Casualty Report (equipment not people)
CBO	Congressional Budget Office
CIC	Combat Information Center
CINCLANTFLT	Commander-in-Chief, Atlantic Fleet
CINCPACFLT	Commander-in-Chief, Pacific Fleet
CMTQ	Cruise Missile Tactical Qualification
CNA	Center for Naval Analysis
CNET	Commander, Naval Education and Training
CNRC	Chief of Naval Recruiting Command
CNO	Chief of Naval Operations
CO	Commanding Officer
COMNAVAIRPAC	Commander Naval Air Force, Pacific
COMNAVSUBPAC	Commander Naval Submarine Force, Pacific
COMNAVSURFPAC	Commander Naval Surface Force, Pacific
COMPTUEX/ITA	Complex Training Unit Exercise/ Intermediate Training Assessment
CPO	Chief Petty Officer
C-RATING	Overall SORTS rating for a unit.
CSA	Combat System Assessment
CSSQT	Combat System Ship Qualification Test Trial

DEP	Delayed Entry Pool (pool of recruits whose entry into boot camp is delayed for various reasons--e.g., waiting to finish high school)
DMDC	Defense Manpower Data Center
DSB	Defense Science Board
EMR	Enlisted Master Record
ECERT	Engineering Certification
ET	Electronics Technician (Rating)
FAIS	Fleet Acquisition Information System (tracks prosecutions in different mission areas for the battle group)
FC	Fire Controlman (rating)
FEP	Final Evaluation Problem
FITREPS	Fitness Report (Officers)
FLTEX	Fleet Exercise
GAO	General Accounting Office
GENDET	General Detail (where personnel are assigned after boot camp if they are not assigned to an A-School--technical school--involving general duties on ship or shore)
GRE	Graduate Record Exam
GTMO	Guantanimo Bay
ICAM	Integrated Computer Aided Manufacturing
IDA	Institute for Defense Analysis
IDEF	ICAM Definition
JCS	Joint Chiefs of Staff
JTFEX	Joint Training Fleet Exercise
LAMPS MK III	Light Airborne Multi-Purpose System Mark III
LOA	Light Off Assessment
MCA	Mid-Cycle Assessment (engineering)
METLS	Mission Essential Task Lists
MLA	Multi-Level Analysis
MOE	Measure of Effectiveness
MPT	Manpower, Personnel, and Training
M-RATING	Mission area readiness rating from SORTS (shows the unit's capability to perform in any given mission area--e.g., mobility, anti-air, amphibious)
NEC	Navy Enlisted Classification
NETPDTC	Naval Education, Training, Professional Development and Technology Center
NGFS	Naval Gunfire Support
NPRDC	Navy Personnel Research and Development Center
NWAD	Naval Warfare Assessment Division

OMF	Officer Master File
OPTEMPO	Operating Tempo (summation of days when a unit is in homeport and not in homeport)
PERSTEMPO	Personnel Tempo of Operations (summation of days when personnel are in homeport and not in homeport)
PLOT	Pre Light Off Training (engineering)
PQS	Personnel Qualification Standards
READYEX	Readiness Exercise
REFTRA	Refresher Training
ROC	Required Operational Capabilities
SEM	Structural Equation Modeling
SFARP	Strike-Fighter Advanced Readiness Program
SORTS	Status of Resources and Training System
SUBPAC	Submarine Force, Pacific
SURFLANT	Surface Force, Atlantic
SURFPAC	Surface Force, Pacific
SUW	Surface Warfare
SWATSCOL	Seabased Weapons and Advanced Tactics School
TORPEX	Torpedo Exercise
TPPH	Transient, Prisoner, Patient, and Holdee
TORPEX	Torpedo Exercise
TRACKEX	Tracking Exercise (submarine)
TRAINTRACK	Enlisted Training Tracking File
TRAPAC	Training Command, Pacific
TRENDS	A data base that tracks prosecutions in different mission areas for the battle group
TSTA	Tailored (or Total) Ship Training Availability
UCSD	University of California, San Diego
USW	Undersea Warfare
VP	Community of fixed-wing patrol aircraft (e.g., P-3 aircraft)
WABA	Within Analysis Between Analysis
WTT	Warfare Team Training

Distribution List

Under Secretary of Defense (P&R)
Assistant Secretary of Defense (FM&P)
Joint Chiefs of Staff (J-1), (J-3)
Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Deputy Assistant Secretary of the Navy (Manpower and Reserve Affairs)
Assistant Secretary of the Navy (Manpower and Reserve Affairs) (OASN) (M&RA)
Chief of Naval Operations (M&P) (N1), (N091), (Logistics) (N4), (N16), (N869), (N813), (N71)
Commanding Officer, Naval Command, Control and Ocean Surveillance Center, Information Management RDTE Division (Code 04)
Commanding Officer, Naval Command, Control and Ocean Surveillance Center, RDTE Division
Commanding Officer, Naval Command, Control and Ocean Surveillance Center, RDTE Division (Code 0141)
Commander, Naval Surface Warfare Center
Commanding Officer, Navy Air Warfare Center, Training Systems Division (Technical Library) (Code 72) (5)
Commanding Officer, Navy Air Warfare Center (NAWC-00)
Office of Naval Research (Code 03), (Code 34), (342)
Chief of Naval Education and Training (Code 00), (Training Performance Evaluation Board)
Chief of Naval Personnel (PERS-00), (PERS-00B), (PERS-05)
Director, Defense Personnel Security and Education Center
Commanding Officer, Afloat Training Group, Atlantic
Commanding Officer, Afloat Training Group, Pacific
Commander, Naval Surface Force, U.S. Atlantic Fleet (N-81)
Commander, Naval Surface Force, U.S. Pacific Fleet (N-13), (N81D)
Commander, Naval Air Force, U.S. Pacific Fleet (N-82)
Commandant of the Marine Corps, HQMC (Assistant DC/S (M&RA))
Director, ARI, Alexandria, VA (PERI-ZT)
AISTA (PERI II), Army Research Institute
Human Resources Directorate, Technical Library, AL/HR-SDKL, Brooks Air Force Base, TX
Director of Research, U.S. Naval Academy
Program Manager, Manpower Research and Advisory Service, Smithsonian Institution
Institute for Defense Analyses, Science and Technology Division
Center for Naval Analyses, Acquisition Unit
Pentagon Library
Defense Technical Information Center (DTIC) (4)