United States General Accounting Office

GAO

Report to the Chairman, Committee on Armed Services
House of Representatives

February 1986

MEASURING MILITARY CAPABILITY

Progress, Problems, and Future Direction

GAO/NSIAD-86-72

GAO

United States
General Accounting Office
Washington, D.C. 20548

National Security and International
Affairs Division

B-221706

February 24, 1986

The Honorable Les Aspin
Chairman, Committee on Armed Services
House of Representatives

Dear Mr. Chairman:

This responds to the House Armed Services Committee report on the Defense Authorization Act, 1986 which directed the General Accounting Office (GAO) to review various issues related to Department of Defense (DOD) efforts to measure military capability. Our findings and conclusions are summarized below and are discussed in detail, along with our objectives, scope, and methodology, in appendix I. A listing of offices and activities visited during our review is in appendix II.

In recent years there has been much congressional interest in (1) determining what increases in military capability have resulted from previous defense budget increases and (2) identifying what future capability improvements can be expected from planned military expenditures. The Office of the Secretary of Defense (OSD) has formed several different task forces and steering groups to provide responses to various congressional inquiries in this area. Additionally, capability assessment systems have been developed by OSD, the Joint Chiefs of Staff (JCS), and the military services as a normal part of the defense planning, programming, and budgeting process.

Military capability is defined by DOD as consisting of four components—readiness, sustainability, force structure, and modernization. We focused our examination on identifying (1) potential modifications to the Unit Status and Identity Reporting (UNITREP) system which provides one indicator of military capability and (2) initiatives in the OSD, JCS, and the military services to develop analytical systems for assessing changes in military capability.

The UNITREP system is only one indicator of military capability and, therefore, it does not provide a comprehensive capability assessment. For example, UNITREP does not measure units' ability to accomplish combat missions. UNITREP information is also not necessarily comparable from one unit to another or from one reporting period to the next. Guidelines which result in more consistent reports could increase UNITREP usefulness for comparing the status of military units or assessing readiness trends. We are making a few suggestions for

improving UNITREP data for consideration by DOD and the Committee (see pp. 13 through 16).

Since 1978 the Congress has sought to have DOD develop improved techniques for evaluating readiness and sustainability. DOD has devoted considerable activity to this objective. Many of the analytical systems developed thus far focus on the relationship of spare parts to weapon systems availability. More recent efforts have focused on how the availability of other resources such as petroleum, oil, and lubicants (POL); munitions; and maintenance affect readiness and sustainability. However, measuring the effects of the many potentially impacting items has proven difficult, even when the services are trying to measure readiness, much less overall military capability. As stated in the DOD Force Readiness Report, the complex interaction between the various readiness components and the difficulty of quantifying the relative importance of each has precluded finding a suitable quantitative measure of overall force readiness.

DOD's attempts to develop improved tools for evaluating the impact of force structure and modernization changes on overall military capability have also proven a most difficult task. While several models which attempt to measure various aspects of war-fighting capability are being developed, the multiplicity of models and approaches suggests there is no general consensus regarding what constitutes an acceptable measure. The current state of military capability forecasting is such that only inferences can be drawn from the variety of models being used to forecast aspects of military capability.

As indicated in our capability assessment chronology, there has been no consistent focal point for coordinating the numerous activities in this area. Although the Strategic Plans and Resource Analysis Agency was tasked in 1984 with assessing how DOD program and budget proposals impact war-fighting capability, responsibility for developing capability assessments since that time has shifted between other JCS or OSD activities. In view of the numbers of organizations in DOD which are and have been involved in attempting to measure aspects of military capability, it may be useful to identify a focal point for coordinating future activities in the area, as well as for interacting with the Congress.

We conducted our work between September and December 1985 and followed generally accepted government auditing standards. As requested by your office, we did not obtain official agency comments on this

report. However, we have discussed its contents with DOD and service officials and have incorporated their comments where appropriate.

As arranged with your office, unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days from the date of the report. At that time we will send copies to the Senate Committee on Armed Services; the House and Senate Committees on Appropriations; the Secretary of Defense; the Secretaries of the Army, the Navy, and the Air Force; and to other parties upon request.

If we can be of further assistance, please let us know.

Sincerely yours,

Frank C. Conahan
Director

Contents

Abbreviations

AFCAP	Air Force Capability Assessment Program
AFIRMS	Air Force Integrated Readiness Measurement System
AIMVAL/ ACEVAL	Air Intercept Missile Evaluation/Air Combat Evaluation
AMRAAM OUE	Advanced Medium Range Air-to-Air Missile Operational Utility Evaluation
C-rating	Combat Readiness Rating
DOD	Department of Defense
DRB	Defense Resources Board
GAO	General Accounting Office
JCS	Joint Chiefs of Staff
LCMS	Logistics Capability Measurement System
MICAF	Measuring Improved Capabilities of Army Forces
NSIAD	National Security and International Affairs Division
OSD	Office of the Secretary of Defense
POL	petroleum, oil, and lubricants
RSAS	Rand Strategy Assessment System
SPECTRUM	Simulation Package for Evaluation by Computer Techniques of Readiness, Utilization, and Maintenance
S-rating	Sustainability Rating
TAA	Total Army Analysis
TAC PACERS	Tactical Air Command Peacetime Assessment of the Combat Effectiveness of Reparable Spares
TASCFORM	Technique for Assessing Comparative Force Modernization
TFCA	Total Force Capability Assessment
UNITREP	Unit Status and Identity Reporting System
WSMIS	Weapon System Management Information System

Measuring Military Capability— Progress, Problems, and Future Direction

Objectives, Scope, and Methodology

Public Law 98-525 directed the Department of Defense (DOD) to submit a report to the House and Senate Committees on Armed Services on the implementation of a readiness measurement system. DOD's March 1985 response concluded that the Unit Status and Identity Reporting (UNITREP) system provided an adequate basis for the unit readiness reporting system desired by the Congress.

The DOD report to the Congress included some suggestions for modifications to the UNITREP system. However, the House Committee on Armed Services' report on the Department of Defense Authorization Act for 1986 stated the Committee did not believe the DOD report had addressed all the Committee's concerns and asked us to review methods of evaluating military readiness and DOD's response.

As agreed with the Committee staff, the objectives of our review were to

- provide a chronology of recent DOD efforts to assess military capability;
- identify potential modifications to the UNITREP system which could make it a more uniform, comparable, and objective indicator of the readiness status of personnel and equipment; and
- identify and examine the status of major DOD initiatives for measuring changes in military capability.

We developed a chronology of congressional legislation and reports; DOD reports, directives, correspondence, and other documents; and reports by the General Accounting Office (GAO), the Congressional Budget Office, the Congressional Research Service, and various academic and private institutions—all applicable to military capability assessment.

Our review of the UNITREP system began with two DOD task force reports on this system. We also conducted interviews at Army, Navy, Marine Corps, and Air Force active and reserve component activities listed in appendix II. We analyzed the Joint Chiefs of Staff (JCS) and service UNITREP directives to identify similarities and differences in the services' implementation and use of this system.

To accomplish our third objective, we interviewed Office of the Secretary of Defense (OSD), JCS, and service officials who identified the primary systems which are currently used in assessing military capability. These officials described the capabilities and shortcomings of these systems and identified key developmental projects intended to improve the existing analytical capability. We discussed these systems with contractor and service personnel and, to the extent possible, reviewed

system specifications and other documentation. We did not, however, try to evaluate the ability of the systems to provide accurate, valid information.

What Is Military Capability and How Is It Measured?

DOD defines "military capability" as the ability to achieve a specified wartime objective—for example, win a battle or a war or destroy a target. Military capability is a broad term which cannot be readily quantified; therefore, DOD has divided capability into the following four subsets or pillars:

- Readiness: the ability of the military forces, units, weapon systems, or equipment to deliver the output for which they were designed (i.e., for a tank to move and shoot) in peacetime and at the outset of hostilities. Readiness is measured in terms of manning, equipping, and training the force and is defined to include the force's ability to mobilize, deploy, and employ without unacceptable delays.
- Sustainability: the staying power of military forces, or how long the forces can continue to fight. Sustainability involves the ability to resupply engaged forces during combat operations and is sometimes measured in terms of the estimated number of fighting days for which supplies are available.
- Modernization: the technical sophistication of forces, units, weapon systems, and equipment. Modernization can include new procurement and/or modifications, depending on the service. Assessments of modernization may compare new types of equipment with the items they replaced or may compare equipment in the U.S. inventory with that of potential adversary forces.
- Force structure: the numbers, size, and composition of units constituting the military forces. Force structure is usually described as numbers of divisions, ships, or wings.

The military services are not consistent in the items which are included under each of the four pillars, resulting in some difficulty in establishing clear-cut distinctions between them. For general discussions of military capability, readiness and sustainability are often discussed together, as are force structure and modernization. Developing tools which will measure current military capability, project future capability, and examine the potential impact of applying alternative levels of funding to a given military requirement is a very complex task.

In this report, we focus our discussion of military capability on the following three areas:

- the UNITREP system, which is used by JCS and the military services to monitor the status of resources at the unit level and is an indicator of military readiness;
- analytical systems used by JCS and the services to assess military capability from the perspective of force readiness and sustainability; and
- analytical systems used by JCS, OSD, and the services to assess military capability from the perspective of force structure and modernization.

Unit Status and Identity Reporting System

UNITREP is the DOD-wide system that monitors readiness by reporting the status of critical resources at the unit level. Through this system, about 8,200 of the approximately 55,300 military units[1] in the active and reserve forces report information about the availability and condition of their assigned resources. These units represent the primary combat, combat support, and selected combat service support units in each service.

In UNITREP, the principal indicator of unit readiness is the combat readiness (C) rating—C-1, representing a fully ready condition; C-2, substantially ready; C-3, marginally ready; C-4, not ready; and C-5, service programmed, not combat ready. For each reporting unit, a C-rating is computed for the following four resource areas: equipment and supplies on hand, equipment condition, personnel, and training. Overall C-ratings are assigned to each rated unit based on the lowest C-rating in any of the four resource areas. That is, if a unit's four areas are rated C-1, C-1, C-1, and C-2, the unit's overall rating will be C-2. UNITREP is but one indicator of military readiness and does not measure a unit's ability to accomplish combat missions.

Readiness and Sustainability

The Congress and OSD have asked for the development of improved analytical capabilities to assess readiness and sustainability. The performance of the military in these areas is influenced by the quantities and condition of both personnel and materiel and the ability of the military to maintain, transport, and distribute these resources. Through the 1978 Defense Authorization Act (Public Law 95-79), the Congress tasked the Secretary of Defense to project the effect of appropriations on materiel

[1]Of the total number of military personnel, 57.1 percent of those in active forces and 76.8 percent of those in the reserve components are assigned to units reporting under the UNITREP system.

readiness,[2] i.e., to ". . . connect dollars to readiness." As a result, the annual Defense Guidance[3] since that time has directed the services to develop methods to model the relationship of force readiness to associated manpower and dollar resources.

Force Structure and Modernization

DOD historically has assessed changes in force structure and modernization by tabulating the types and quantities of weapons in each military unit and division and, in some cases, by comparing these tabulations with intelligence estimates of the comparable equipment held by other national forces. Since the mid-1960's, an assortment of DOD and intelligence agencies have tried to find more satisfactory measures of force structure and modernization. These agencies have used various analytical methods to relate inputs in terms of prebattle force structures to outputs, which are derived by analytically projecting the military's performance in executing wartime missions.

Basically, there are three types of analytical models used for this purpose—static models, simulation models, and a hybrid of the first two. In static models, each weapon in the U.S. inventory is assigned a qualitative weight or score which ranks every weapon against other U.S. weapons and also attempts to evaluate enemy forces in the same way. Simulation models involve a dynamic, two-sided combat war game which attempts to replicate combat and to assess how well the U.S. forces could be expected to perform against a specific enemy force. In recent years, there has been an increasing use of a third type of model, which is a combination static simulation. This hybrid model is said to take advantage of certain benefits of both static and simulation techniques.

[2]Because readiness and sustainability are so closely related, from the perspective of capability assessments, they are often treated at the same time. Both terms may not, however, always be used. During this review, we also found that the services are not consistent regarding what factors they include under the readiness and sustainability pillars, further complicating a clear distinction between the two.

[3]The Defense Guidance is the primary OSD guidance document for providing central policy and direction for program development. This document presents the rationale for the defense program and covers military strategic concepts and objectives; planning and programming guidance; force levels; and manpower, support, and fiscal guidance.

Chronology of Military Capability Assessment Activities

For many years, the Congress and OSD have called for the development of better tools for measuring military capability. The first component of capability for which the services developed a quantitative measurement system was force readiness. In the early 1960's, the services reported combat readiness status in the Operational Readiness Report, which evolved into the Force Status and Identity Report in 1968 and the UNITREP system in 1980. DOD provides additional indicators of readiness in the Situation Report; the Force Readiness Report; the Annual Capability Report; and other documents, which are identified in our June 13, 1985, report, Measures of Military Capability: A Discussion of Their Merits, Limitations, and Interrelationships (GAO/NSIAD-85-75).

In January 1984, the Secretary of Defense cited UNITREP data in his 1985 budget report to the Congress, stating that the percentage of U.S. units attaining C-1 or C-2 combat ready ratings had increased by 39 percent. The Congress questioned what this figure meant in terms of actual military improvements. The figure cited represented a combination of all service ratings. Although there had been a large increase in the number of Navy units attaining C-1 or C-2, there had also been a 25-percent decrease in Army units and a 15-percent decrease in the Air Force units reporting a C-1 or C-2 status. In commenting on this fact, the Chairman of the Joint Chiefs of Staff said that UNITREP should not ". . . be used to describe the readiness of the force to the taxpayer. It is an internal management tool." These events prompted much congressional interest and DOD activity in trying to better define and measure military capability, including:

- In 1984, at the direction of the Congress, the Secretary of Defense created the Strategic Plans and Resource Analysis Agency within JCS. This activity was tasked with analyzing how DOD program and budget proposals impact U.S. war-fighting capability and identifying planning issues which could integrate service resource requirements.
- In 1984, a DOD report to the Congress, Improvements in U.S. Warfighting Capabilities, 1980-84, responded to congressional questions about how U.S. war-fighting capability had improved since 1980.
- In April 1984, the Secretary of Defense established a readiness task force chaired jointly by JCS and OSD representatives. The group was directed to try to develop models which relate the capability impacts of alternative resource inputs and logistics support measures and factor in any other meaningful measures of force readiness and overall combat capability which the task force identified. This group decided to first focus on UNITREP. After reviewing it, the task force recommended changes to this system in January 1985.

- Later in 1985, JCS responded to the proposed readiness task force changes and concluded that with one exception, (dealing with determination of Navy aircrew training readiness, see p. 12) the proposals did not significantly contribute to satisfying the Secretary of Defense's objective of clearing up confusion about capability increases in DOD. JCS also said its staff was now developing a new capability report, the Military Status Report (discussed further below).
- In October 1985, the Deputy Secretary of Defense initiated a new steering group under the OSD Director of Program Analysis and Evaluation. This group was tasked with developing a set of measures of military capability. According to steering group officials, their first priority was to review the Military Status Report. The group submitted comments on this report to the Secretary of Defense in January 1986. According to steering group officials, the group now intends to review other capability assessment systems, including the Army's Measuring Improved Capabilities of Army Forces (MICAF) system and the Technique for Assessing Comparative Force Modernization (TASCFORM) which is under development for OSD's Office of the Director of Net Assessment. (See pp. 24 and 28 for descriptions of these systems.) The group wants to determine if these systems have potential for use as capability assessment tools for evaluating all the services.
- The JCS Operations Directorate compiled inputs from the services and formulated the Military Status Report in the fall, 1985. This report is a compilation of various service-selected indicators of capability summarized to show increases in military capability since 1981. The Military Status Report graphically presents various war-fighting categories (e.g., strategic offensive forces) and displays improvements achieved within these categories. Information presented in this report was compiled from various service documents, systems, and reports. The Army included data from its new MICAF system, and the Air Force included analysis from its new Relative Capability Assessment, which calculates changes in force capability against some predefined representative targets or goals.

UNITREP—An Indicator of Military Readiness

UNITREP is currently the only DOD-wide system that measures the status of critical resources at the unit level. It reports the inventory status and condition of both people and equipment. While UNITREP is not intended to measure all variables which affect a unit's readiness—it does not, for example, measure impacts of qualitatively different weapon systems (M1 tanks vs. M60A3 tanks) or intangibles, such as morale and leadership—it is an important readiness indicator and internal management tool.

Differences in UNITREP Implementation

The JCS implemented UNITREP in 1980 to support the National Command Authorities and to help identify which units are ready if a conflict occurs. JCS provided policy guidance and implementation instructions to the services, but gave them latitude to interpret some of these instructions. Now each service maintains its own UNITREP data base and implements UNITREP differently through supplemental regulations (see table I.1 for examples of differences). Because of this lack of uniform implementation, UNITREP should not be used for comparing readiness among the services. For example, JCS instructions ask all services, in determining aircrew training readiness, to measure against their wartime crew level requirement. The Navy, however, has decided to measure against peacetime authorized levels. The Navy believes units should not have to degrade themselves in both the training and personnel ratings if for budgetary reasons aircrews at the wartime levels cannot be provided.

Additionally, intraservice comparisons of UNITREP data may be misleading. For example, Army units which do not have required equipment may substitute existing equipment. Thus, comparing units' equipment-on-hand ratings would not necessarily reveal the units which have shortages of required equipment.

Table I.1 Service Differences in UNITREP Implementation

Issue	Army	Navy	Air Force	Marine Corps
Is the aircrew training C-rating computed based on wartime requirements?	Yes	No	Yes	Yes
Can the commander subjectively change the unit's overall readiness C-rating?	Yes	No[a]	Yes	Yes
Are critical skill levels measured in the personnel ratings?	No	Yes	Yes	Yes
Are training ratings based on objective, quantifiable criteria?	No	Yes	Yes	No[b]
Are supplies/sustainability items included in the equipment and supplies-on-hand resource area C-rating?	No	Yes	Yes	No

[a]Although the Navy does not allow subjective changes, it accomplishes essentially the same thing through an internal process (see p. 14).

[b]While the criteria for aircrews are objective and measurable, criteria for ground units are more subjective.

Service flexibility in implementing UNITREP can impact readiness trend information. For example, until recently, officials said the Army had not extensively used the C-5 category (not combat ready due to special programs) for units in transition or undergoing reorganization, as other services do. Officials said the Army had decided to make more use of the C-5 category to be more consistent. Since C-5 units are considered separately in trend analyses, Army readiness trends may show improvements which are not based on resource improvements. Users lacking information about the cause of the trend improvement could erroneously conclude that it resulted from increased or improved resources.

Although DOD has repeatedly pointed out that UNITREP was never intended to be used as a comparative tool, it is the only common multiservice data system measuring readiness. Therefore, until another multiservice system is developed comparing unit status across the services, there will probably be a tendency to use it for this purpose.

Opportunities for Better UNITREP Reports

We analyzed some of the UNITREP implementation practices to determine their influence on reported readiness and to help identify how DOD might improve UNITREP data.

Linking Reserve Component Readiness Status to That of the Parent Unit

The readiness status of an Army division is a composite of the ratings of its active subordinate units. Some Army divisions depend on reserve components to "round out" their structure in case of conflict. However, under UNITREP, these divisions do not consider the readiness status of their roundout units in determining the division's status.

The Army recently proposed that divisions begin to compute a second set of UNITREP ratings that consider the roundout unit's readiness status. Commanders would provide these ratings, however, only in the Part 2 comments section of the division's UNITREP submission. This section of the report is not widely circulated to decisionmakers above the command level. Therefore, under the Army's proposal, key managers may not be aware of the impact of the roundout unit's readiness on the division's overall status.

Army headquarters officials believe the status of roundout units should not be reflected in divisions' reported ratings. These officials told us they needed to continue to separately assess the readiness of the parent active unit without the roundout, since the active unit would usually

deploy earlier than the roundout and the division had only limited control over the peacetime status of these units. On the other hand, reserve component officials told us combined ratings which include roundouts would give greater attention to reserve component readiness issues.

According to the Total Force Policy,[4] Army roundout units are intended to be an integral part of their parent division. Thus the roundout unit's readiness status is an important aspect of the division's total readiness. An alternative to the Army's proposal would be to provide two ratings, one which includes the roundout units and one which does not.

Commanders' Ability to Subjectively Change UNITREP Status

JCS instructions allow a commander to subjectively raise or lower the unit's overall readiness rating if he believes the revised rating is a more accurate reflection of the unit's true readiness. All services except the Navy authorize subjective changes. These services believe the authority is necessary because UNITREP does not measure the readiness impact of intangibles, such as experience, morale, and leadership. Use of this provision, however, could affect the validity of comparisons.

Although the Navy does not authorize subjective changes as such, it accomplishes similar results through an internal process. Units can improve by one the single lowest mission rating in each resource area. Mission ratings are used to determine the unit's four reported resource ratings. The Navy process could impact intra-Navy comparisons. For example, two Navy units might have a C-2 overall rating when one of the units was actually C-3 in one mission rating for each resource area.

Army and Marine Corps officials told us that changes generally resulted in rating upgrades. Air Force officials were unable to generalize regarding the effect of changes. Service officials told us commanders, in making their subjective changes, may be mistaken or overly optimistic in their judgments because of concern that the ratings reflect their individual performance. On the other hand, commanders might be pessimistic because of a desire to highlight perceived problem areas and hopefully obtain additional resources. As a result, readiness trends or comparisons of the status of like units can be misleading. One way to address this issue would be to prepare two ratings, one which includes the commander's subjective assessment and one which does not.

[4]The policy, initiated in 1973, is designed to integrate active, guard, and reserve forces into a total fighting force, increasing the roles and responsibilities of reserve component forces. Guard and reserve forces are to be the initial and primary augmenting forces to the active components.

Substitute Equipment Used in Army's Rating

Since Army units measure their equipment combat readiness ratings against new equipment requirements, the Army has difficulty accurately measuring equipment readiness when it cannot provide the necessary equipment. Units in the other services do not measure against the new requirements until the equipment is available. The Army's practice can result in lower readiness ratings until new equipment reaches the field, and in some cases this could take several years.

To compensate, the Army allows commanders to substitute other equipment for new requirements. In cases where the Army has established a plan to provide a unit its new equipment, the unit's existing equipment is considered substitutable for items on the requirements list. Otherwise, the Army suggests commanders restrict substitution to an approved list of substitutable items. However, the commanders currently can still substitute items not on the list.

Substitution can mask the readiness of units, in particular that of reserve component units. These units substitute more extensively and, in some cases, substitute nonstandard items which may not be supportable during war. Displaying the extent of substitution in UNITREP could help in identifying shortages of required equipment and improving the validity of comparisons of the equipment status of like-type units.

Personnel Qualifications

An important aspect in determining personnel readiness is the availability of personnel trained in their service occupations. The Army is the only service which does not fully measure required skill levels of its personnel in UNITREP ratings. The Army uses a five-digit code representing the career field, skill level, and special qualifications of each soldier's occupational specialty. The fourth digit indicates the soldier's skill level. For UNITREP purposes, however, the Army measures only against the first three digits. This means the Army is measuring only whether it has personnel in the required career field but not whether they have the required skill level.

Determining qualifications to at least the fourth digit of the occupational code would provide this information. Since it is also important to know whether soldiers are qualified at least in their career fields (determined by the first three digits of the code), UNITREP could show two separate ratings: one indicating whether the Army has personnel in the required career field and the other indicating whether it has personnel at the required skill levels.

Objectivity of Training Ratings

The Army and the Marine Corps processes for determining their training ratings are more subjective than those of the other services. Currently, Army and Marine Corps ground unit commanders, based on available information on their units' training performance, estimate how many additional weeks of training are required to be combat ready. These estimates determine the units' training readiness ratings. Both the Army and the Marine Corps have initiatives under way to better standardize a unit's required training activities. However, until these services can establish objective criteria to help commanders assess training status, comparisons of the readiness of like-type units within these two services may not be valid.

The Air Force and the Navy have established standard lists of training tasks and qualifications that serve as criteria for determining the training ratings.

Separating Readiness and Sustainability Items

Three of the joint OSD/JCS readiness task force recommendations dealt with creating a new resource area rating in UNITREP, a sustainability (S-) rating. (See p. 10). Currently, each service differs in whether it counts spares and support items in determining the equipment and supplies inventory readiness rating. OSD task force representatives proposed that the services separate such items from the readiness rating and count them under the new S-rating.

Most of the services did not fully support these recommendations. Navy officials considered certain spares and supplies as necessary to "deliver the outputs for which they were designed" and thus readiness items. The Navy considers resupply of these items to be sustainability. Marine Corps units do not control their own supplies. The Army disagreed that a sustainability measure should be included in a readiness report. Only the Air Force did not object to the proposal because aircraft spares kits could be easily identified and measured under the proposed rating.

The task force recommendation tried to reduce the impact of differences in the way the services count (or do not count) items in UNITREP. With this proposal, however, the services would still have the latitude to define the items to be evaluated under the readiness and sustainability categories. Consequently, it is unlikely that the new rating would accomplish the intended objective. To effectively implement the proposed change would require a more clear delineation and agreement on what resources should be counted in each category.

Evaluating the Impact of Readiness and Sustainability on Military Capability

The resource areas which most directly impact readiness and sustainability are commonly referred to as logistics resources and include spare and repair parts; munitions; manpower; maintenance capability; support and test equipment; training; transportation; distribution capability; and petroleum, oil, and lubricants (POL).

Military managers thus face the question of how defense dollars should be allocated among the various logistics resource areas in order to obtain a desired level of readiness and sustainability. The question which must be addressed is what mix of resources can produce the greatest amount of war-fighting capability, given some finite level of funding.

Providing an answer to this question has proven to be a difficult and complex task. Congress has directed that DOD develop "quantifiable and measurable readiness requirements" and project the effects of appropriations requested for materiel readiness. While DOD has directed that these analyses be accomplished, the services have not yet been able to develop the capability to make such assessments.

Air Force Systems to Assess Readiness and Sustainability

Air Force efforts to develop a means of assessing and reporting on readiness and sustainability began in July 1976, when the Air Force Chief of Staff convened the Constant Readiness Conference. One finding of that conference was that the information available to Air Force management did not support decisionmakers in their efforts to identify and quantify the impact of alternative solutions for resolving readiness problems. The final conference report issued by the Chief of Staff called for the development of a ". . . responsive means of assessing and reporting combat capability." The following Air Force systems are intended to assist Air Force operational units, major command headquarters, and the Air Force Headquarters staff in making resources-to-readiness assessments. These systems have the same basic objectives. They differ in which Air Force command or organization they are designed to support and in the number and types of logistics resources included.

TAC PACERS Provides Data at the Unit Level

The Tactical Air Command Peacetime Assessment of Combat Readiness of Reparable Spares (TAC PACERS) is designed to project the supply level in each unit's War Readiness Spares Kit[5] during the first 30 days of war.

[5]The War Readiness Spares Kit is an air transportable package of war reserve materiel (spares, repair parts, and related maintenance supplies) required to support planned wartime or contingency operations of a weapon or support system for a specified time pending resupply.

TAC PACERS identifies and ranks critical items in the kit which are predicted to affect each fighter squadron's sortie capability[6] and aircraft availability to perform its wartime requirement. The system is designed to be used to make trade-off evaluations at the unit level to determine the effects of various alternative resource decisions on sortie capability. These alternatives may include such options as using partially mission capable aircraft[7] or accelerating the movement of spare parts out of a depot.

Weapon System Management Information System

The Weapon System Management Information System (WSMIS), developed and used by the Air Force Logistics Command, is designed to assess the readiness and sustainability of Air Force systems by identifying logistics support resources which may limit a weapon's combat capability. Currently, this logistics evaluation is limited to spare parts.

The baseline for the system, as with other Air Force systems which measure readiness and sustainability, is the Defense Guidance planning scenario.[8] The information produced is intended to represent wartime mission-related criteria. For example, for readiness, WSMIS will project the number of mission capable aircraft and for sustainability, it will project the number of sorties which can be generated. Using WSMIS, the Air Force intends to develop the capability to compare the contribution of not only spare parts, but also spare engines, POL, and other logistics resources to overall sortie capability.

The WSMIS design includes three different modules. The Readiness and Sustainability modules are operational at two Air Force commands and are undergoing testing at a third. The Get Well module, which is still under development, is supposed to integrate information from the other WSMIS modules and recommend alternative approaches for correcting logistics deficiencies.

[6]A sortie is an operational flight by one aircraft.

[7]A system or equipment is considered mission capable when it is safely usable, and can perform at least one of its assigned missions and fully mission capable if it has all mission-essential subsystems installed and operating as designated by a military service.

[8]The Defense Guidance planning scenario is the Secretary of Defense's guidance to the military services regarding the type and condition of warfare they should use as the basis for developing force requirements. It provides a continuum of events and time frames which the U.S. military forces should be prepared to combat.

Since January 1985, the Air Force has experimented with using data from the WSMIS assessments of sortie capability in the UNITREP system. Currently the UNITREP spares analysis is based on each unit's fill rate of its War Readiness Spares Kit—that is, what percentage of a given spares requirement the unit has on hand. Incorporating a WSMIS assessment in UNITREP provides a projection of the impact of spares on supporting a unit's aircraft availability. Thus, Air Force officials believe, the UNITREP equipment-on-hand C-rating may become a more meaningful evaluation by assessing the relationship of spare parts to combat potential.

Logistics Capability Measurement System (LCMS)

The Logistics Capability Measurement System (LCMS) is being developed as a management information system for the Air Force Deputy Chief of Staff, Logistics. LCMS will identify Air Force logistics requirements to aid that office in obtaining and apportioning funds and overseeing logistics budget execution. The system uses a series of broad-based aggregate analytic models to project the number of aircraft sorties which can occur using varying levels of aircraft spare and repair parts, spare engines, fuel, and munitions.

The LCMS currently consists of three models, two of which are used for assessing spares requirements and a third for evaluating munitions. A fourth model is a prototype and is designed to evaluate a wartime sortie profile to determine where fuel shortages might occur and how best to restructure the logistics system to correct anticipated shortages. A fifth model will try to integrate inputs from the two spares models and the munitions and POL models to produce recommended levels of support in these resource areas to achieve maximum sortie capability.

According to Air Force officials, there is some concern that the broad macro-type assessment which this and similar systems generate may not provide the kind of detailed analysis needed to generate valid logistics requirements. Officials said the integration of LCMS with more detailed analyses provided by simulation models could improve this system and provide a more balanced projection of logistics requirements in supporting wartime conditions.

Air Force Capability Assessment Program

The Air Force Capability Assessment Program (AFCAP) represents a recent redirection of the former Air Force Integrated Measurement System (AFIRMS). Air Force officials told us that although the AFIRMS project had been initiated in 1978, the system development effort had

faltered. An October 1985 internal Air Force assessment revealed significant questions about AFIRMS' concept, scope, and program management.

AFIRMS' objective was to collect information on tactical air wings' resource availability and the status and capability of aircrews, aircraft, maintenance support, and base support facilities. This information would be compared with the wings' combat operation plans to determine how many of the required combat sorties can be generated. Individual wing capabilities would be transmitted and aggregated at the appropriate major command. Resources controlled at the theater level (i.e., POL and munitions) would also be factored into the scenario to form an overall assessment of a given theater's or command's capability.

Air Force officials said that AFCAP will narrow the scope of the AFIRMS program. However, a clear delineation of AFCAP's scope and objectives will not be made until completion of the definition and design phase. The current milestone for the system's full implementation is 1995.

Army Logistics Assessment

The Army Logistics Assessment represents the primary available tool for projecting the readiness and sustainability of Army forces. This assessment indicates the status of 38 categories of logistics resources (for example, strategic lift, repair parts, POL, conventional ammunition, and missiles). Resource requirements in each logistics category are established for each day of a scenario, and Army logisticians manually compute what percentage of the anticipated logistics requirement can be met on each day of the war.

According to Army officials, this process is a useful assessment of Army readiness and sustainability. However, the process is very labor intensive and can be completed only once each year. Army logisticians have recognized the need for an enhanced logistics assessment and have initiated the Army Logistics Assessment-Extended. This program is intended to provide an automated methodology for evaluating the criticality of various resource shortages and the relative resource funding priorities to achieve the greatest increase in war-fighting capability. Current plans provide for using a computer analysis based on an existing Army capability assessment model.

In addition to the readiness and sustainability information made available by the Army Logistics Assessment, the Army's Total Army Analysis, which is primarily used for analyzing changes in force structure,

also provides input to the overall evaluation of readiness and sustainability. For example, one model is designed to determine whether the force can be transported to an area of conflict in a timely fashion and another model tries to determine what support units a combat unit would require to effectively perform its mission. The Total Army Analysis is discussed further on page 24 of this report.

Navy Assessments of Readiness and Sustainability

Navy officials identified three types of assessments currently used as the primary analytical tools for evaluating readiness and sustainability. A fourth assessment technique is under development.

Existing Analytical Tools

Simulation Package for Evaluation by Computer Techniques of Readiness, Utilization, and Maintenance (SPECTRUM) simulates the logistics activities supporting aircraft wings on a carrier or at a naval air station. The purpose is to evaluate alternative support scenarios to help in determining resource requirements. The inputs for the SPECTRUM assessment include the number of aircraft, available test equipment, total flying hours, cost estimates, available funding levels, personnel, and maintenance management policy. Using this mix of resources, SPECTRUM determines how many aircraft combat sorties can be generated and how ready the aircraft are when measured by peacetime mission and fully mission capable standards. SPECTRUM identifies where aviation resource limitations occur. However, it does not develop a preferred mix of resources.

Navy officials said that SPECTRUM is used as needed to assess funding decisions and to develop and justify budget requests. SPECTRUM also is used to evaluate trade-offs between other factors such as staffing levels and aircraft use.

The Navy has been using nonnuclear ordnance requirements models since 1975 to determine its annual requirement for two different types of conventional ordnance. One series of ordnance models determines how many torpedoes, antiship cruise missiles, antiair missiles, and surface-to-air missiles are needed. The models also measure effectiveness by estimating the percentage of combatants who do not run out of ordnance before their allocated portion of the threat or enemy targets has been killed. In developing requirements for more common types of conventional ordnance, the Navy uses three models which calculate (1) sorties generated and the number of aircraft available for a defined

scenario, (2) the best weapon system to generate the sorties and aircraft availability rates, and (3) the ordnance required to support the weapon systems.

Baseline Area Assessment, a process which evaluates the Navy's ability to provide timely transportation of logistics support resources to battle groups under defined maritime scenarios, evaluates the ability to transport materiel along the points in the logistics chain, from point of origin to the operational battle groups.

Assessment Techniques in Development

Navy officials said the Navy is attempting to link historical changes in specific resources to changes in readiness. However, that linkage has been difficult to establish. Navy analysts, by developing mathematical formulas, are trying to determine if and how much of a relationship exists between historical changes in resources and changes in readiness. For example, did additional spare parts increase the availability of aircraft and, if so, by how much? If historical relationships can be established, officials hope to use them to predict the best mix of future resources. Officials said the Navy is at the early stages of developing initial formulas.

Horizontal Integration Aims to Improve JCS Assessment Capability

Horizontal Integration is a developmental project for the logistics directorate of the JCS. The project's objective is to analytically evaluate readiness and sustainability goals of combined U.S. military forces and compare the potential contribution of various mixes of logistics resource acquisitions to required war-fighting capability. Four logistics resource areas will be evaluated: munitions, major end items, repairable spares, and mission support equipment. The system is intended to

- determine the mix of logistics resources needed to support the Defense Guidance planning scenario and various other hypothetical operational plans,
- identify resource shortages and priorities for acquiring resources to most cost-effectively meet defense requirements, and
- evaluate the potential logistics capability to be acquired during each year of the Five Year Defense Plan.

This information is expected to be useful to JCS in recommending alternative strategies and positions for issues addressed by the Defense Resources Board (DRB), which is the major decisionmaking body in the DOD resource allocation process. We were told by JCS officials that the

lack of analytical tools to quantify the adverse impact of logistics shortages on potential war-fighting capability had weakened the Chairman's effectiveness in supporting readiness and sustainability issues before the DRB. Horizontal Integration is intended to improve the ability of JCS to accomplish the required analyses. JCS officials said this system could now assess 72 munitions classes. The current estimate for expanding the system to cover other resource areas is the end of fiscal year 1987.

Evaluating the Impact of Force Structure and Modernization Changes on Military Capability

Defense managers must decide what weapon systems to modernize, how many to buy, and what other force structure changes to make to develop an integrated force. Military models are used to support this decision-making process. These models provide input to the resource allocation process to help assess the impact of current decisions on future war-fighting capability. While the models are not expected to be able to predict the future, in theory they can provide a limited but systematic methodology for examining alternative future courses of action. For purposes of this discussion, we refer to military force structure and modernization models and analyses as combat assessments.

Army Combat Assessments

The Army uses computerized war games, combat simulations, and analytical models[9] to help evaluate combat capabilities and determine resource requirements. Two primary analyses which are based on these processes are OMNIBUS—the Army's annual assessment of current capability to meet the threat—and Total Army Analysis (TAA)—the Army's assessment of future combat and support force requirements to meet the threat. The Army uses these studies in its planning, programming, and budgeting process. The Army has also developed an assessment called MICAF, which was specifically developed to provide information to the Congress about recent improvements in the Army's combat capability.

[9]Although the terms "war gaming," "simulation," and "analytical modeling" are sometimes used interchangeably, they have different purposes and normally require different techniques. "War gaming" applies only to the analysis of combat situations in which human players form a part of the decision process. Hence, models are components of war game analysis, but are not war games themselves. War gaming is most often used for training purposes, but also offers information which can be useful in developing combat strategy and tactics and for resource planning. "Simulation" is a type of model which replicates some reasonably well understood process. To use simulation, one must have sufficient knowledge about a process to model it and specify what events are expected to occur and their probability of occurring. "Analytical models" mathematically represent combat processes to calculate combat measures of effectiveness. These models are used when uncertainty exists about the basic process to be modeled or when the objective is specifically to calculate a measure of effectiveness.

Omnibus and Total Army Analysis (TAA)

OMNIBUS and TAA are based on a series of models and analyses which attempt to identify major factors that may affect overall Army combat performance. For example, a transportation model is used to analyze the Army's ability to move reinforcing units, replacement personnel, and resupply items into a combat area, and a logistics model tries to determine what the support requirements will be. The main components of the OMNIBUS and TAA analyses are the models which represent actual combat operations.

To implement a combat modeling process which tries to portray actual combat operations, the Army has sought to develop a series of models—a hierarchical family—which could be used to represent combat operations at the various Army organizational levels. One model would represent battalion or company operations and provide input to the model representing the next higher level of combat operations, the corps or division. Outputs from the latter model would be input to the model representing combat operations at the next higher organizational level—the theater.

The hierarchy concept, however, has proven difficult to implement. The corps/division level model was not designed to accept inputs from the battalion level model or provide outputs which were compatible with the theater level model. Efforts to make the required revisions have been unsuccessful. Army officials said that some progress had been made in developing an improved model to meet the needs of the model hierarchy, but they did not know when it would be available. Additionally, Army officials recognize that the model used to project combat operations at the theater level has many limitations and they are trying to implement an improved version. Army officials told us these and other needed improvements are being sought through the Army Models Improvement Program.

Measuring Improved Capability of Army Forces (MICAF)

MICAF is an analytical process developed for measuring and reporting how the capabilities of active and reserve component forces improve as new items are introduced into the force. MICAF uses a model to quantify the combat potential of a division and the combat contribution of its individual weapons. This model, the Analysis of Force Potential, quantifies the estimated number of targets killed for specific items in a unit and the resources expended in achieving those kills. The inputs can be adjusted to reflect the influence of different combat support and combat service support scenarios.

The model output provides values for each selected weapon and is intended to represent the ratio of targets killed to weapon system losses. The model then adds the value of each weapon system in the division's inventory to estimate the division's overall combat potential. This composite value is intended to reflect the total impact of modernization (improvements in weapons quality) and of force structure (increases in the number of weapons in a division's inventory).

The Analysis of Force Potential is a static measure that does not explicitly deal with the dynamics of warfare. Although Army officials told us that this system is superior to earlier static models, OSD officials told us that such broad-based macro assessments may not be capable of validation and questioned the model's ability to provide meaningful results.

Air Force Mission Area Analysis

The Air Force Mission Area Analysis was designed to assess Air Force war-fighting capabilities and identify limiting factors. The Air Staff Board and the Air Force Council use results in the budget process to analyze force structure and modernization decisions.

The models used in the Mission Area Analysis process compare the relative combat contributions made by current and future weapons systems and support elements. Because of the broad scope of the objectives, the process contains numerous assumptions and limitations. Air Force officials consider this analysis only one of many tools to support decision-making and restrict its use outside the Air Force, because they believe the results may be taken out of context.

Mission Area Analysis uses four analytic models representing the strategic offense, defense, theater, and force projection missions. These models provide a framework to define the required war-fighting and support capability for a specified mission objective and to reflect how well the Air Force attains its primary mission objective of achieving and maintaining air superiority.

Air Force officials have told us that because the evaluation is a macro assessment, and includes a number of questionable assumptions, its usefulness for determining force requirements is limited.

Navy Combat Assessments

Studies employing analytical or simulation models of naval conflict are conducted to provide input at major decision points in the Navy planning, programming, and budgeting process. Navy officials have told us

the Navy's approach to combat assessment relies chiefly on fleet exercises and increasingly on war gaming. The Navy has not maintained a continuing modeling capability but does its modeling principally on a decision oriented ad hoc basis. According to Navy officials, many different models are available that analytically represent overall naval war-fighting capability, any of which may be used for a given study or analysis.

The Navy's primary combat assessment activity is performed by the Director of Naval Warfare, who is required each year to prepare Warfare Area Appraisals for all mission areas, such as antisubmarine, amphibious, electronic warfare, and space. The appraisals' objectives are to observe trends in current warfare capability and project capability 5 and 10 years into the future. Shortcomings in capability are identified and solutions—or war-fighting improvements—are recommended. If solutions are not readily identified, the appraisal process may recommend actions to determine how to cope with the threat.

JCS Capability Assessment Processes

In September 1978, the Joint Chiefs of Staff established a requirement to develop analytical procedures which assess U.S. and Allied force capabilities worldwide. To meet this requirement, JCS developed the Total Force Capability Assessment (TFCA). More recently, JCS has initiated several developmental efforts to improve its analytical capability in this area, one of which is the Forces Planning Program.

Total Force Capability Assessment

According to JCS officials, the term "TFCA" actually refers to two different concepts, an annual assessment and the analytical methodology whose objectives are to

- improve analytical support for the Joint Strategic Planning System and assist in other force assessment efforts;
- assess total capability to oppose specific threats; and
- jointly assess land, sea, and air capabilities.

TFCA provides an overview of the forces' balance by individually and collectively assessing sea, land, and air forces; identifying mission overlap; and evaluating the overall war-fighting process, including interactions between the services. Representatives of each military service take part in this war gaming exercise. Each year the TFCA process takes about four months to complete and could not be readily replicated using different resource inputs for comparative analysis.

JCS Forces Planning Program

To improve its analytical capability, JCS has initiated the Forces Planning Program. This system is intended to be a primary analytical tool of the new JCS Strategic Plans and Resource Analysis Agency. DOD established this agency at the request of the Congress to improve JCS' analytical capability and to support the Joint Chiefs' independent judgment of the readiness, logistics, and doctrinal impacts of defense budget and program proposals.

According to JCS officials, DOD has no integrated set of analytical tools to quickly and comprehensively assess trade-offs among military force postures, levels of effectiveness, strategies, and cost. Officials said such information would be useful to the JCS Chairman in providing input to the President and the Secretary of Defense on the capabilities of current military forces and on the recommended composition of future U.S. forces. The JCS Forces Planning Program, which was initiated in 1984, is a 3-year developmental effort designed to provide this data. The primary objective of this effort is to provide a technique to assess the composition, capabilities, costs, strategies, risks, and other factors of alternative military force mixes. The secondary objective is to identify and assess alternative force postures for development of the Defense Guidance.

JCS officials told us that the Forces Planning Program system will consist of two basic models, a force effectiveness model and a cost model. Both are still in development. The force effectiveness model is designed to allow military analysts to make force trade-offs using very general and aggregate data, such as combined firepower scores of various weapons groupings. The cost model is designed to provide the associated costs of weapons systems and logistics support for various force compositions. Officials have said integrating these two models and incorporating logistics impacts are very difficult and have yet to be accomplished. JCS currently estimates the model integration effort will be completed by October 1986.

OSD Net Assessment Sponsors Development of Systems to Assess Combat Capability

The OSD Director of Net Assessment is responsible for advising the Secretary of Defense on U.S. military capability and on how that capability compares with that of other national forces. According to a net assessment official, there is currently no real consensus in the intelligence and DOD communities as to adequate, acceptable measures of effectiveness for making conventional force comparisons. Several systems are being developed by Net Assessment which are intended to improve existing combat assessment capability. These systems include the Technique for

Assessing Comparative Force Modernization (TASCFORM) and the Rand Strategy Assessment System (RSAS).

Technique for Assessing Comparative Force Modernization

The original objective of TASCFORM was to develop a comprehensive technique for comparing the modernization of the conventional weapons of the U.S and the Soviet Union. Tactical aviation forces were chosen as a test case. A model was developed which uses a mathematical function to compare the performance potential of tactical aviation systems in five mission roles (for example, close air support and fighter). The methodology provides a system for scoring individual aircraft and ranking them relative to their mission performance capability. The main application for these scores is to weight force inventories. For instance, when the results of multiplying the aircraft average system performance scores by their inventory levels are totaled for a given force structure, the results represent a basic measure of force potential which can be compared to an opponent's score.

The basic design approach for TASCFORM consists of developing quantified measures of effectiveness which are intended to represent the quality (as represented by weapon system performance) and the quantity (as represented by inventory levels) of each side's operational inventory at specific points in time. Changes in the resultant values over time are then interpreted as relative force modernization rates.

The TASCFORM system is a static measurement tool. System designers say it provides a very macro assessment. Therefore, it cannot be used to predict the outcome of specific combat engagements nor should it be used alone to determine detailed acquisition strategies. OSD officials have said that it may not be effective in defining a global assessment of comparative weapons modernization as it was intended. Navy officials were critical of its representation of Navy combat, but contractor personnel said the Navy portion of the system is being modified.

Rand Strategy Assessment System

Development of the Rand Strategy Assessment System stemmed from a 1978 Defense Science Board Summer Study conclusion that improvements were needed in comparative analyses of strategic nuclear capability. The concept was later expanded to include conventional as well as nuclear warfare. The system's current capability provides for evaluating ground and air combat in central Europe. However, modeling interfaces to represent the logistics functions required to support these

combat operations have not been completed. Naval combat modules are planned, but have not yet been developed.

The RSAS is based on a new evaluative technique that combines war gaming and analytic modeling. According to OSD officials, human-interactive war games are slow and cannot be reproduced. RSAS tries to make war gaming more efficient and analytical by using computer models to replace some or all of the human element. Although people can still participate in the exercise, the intent is to capture most of the so called "expert judgment" so that a single analyst can use the system for exploring a range of combat scenarios.

Prototype development will not be completed until late 1986. However, an initial prototype is operational and will soon be installed at three DOD components (the Joint Analysis Directorate of the JCS, the Army Concepts Analysis Agency, and the National Defense University) as an initial testbed for government use and as an aid in the system development process. Military analysts we interviewed in OSD are optimistic about this system's potential, although they pointed out that many developmental and testing activities remain to be accomplished.

Improvements Needed in Military Modeling

Besides the unique systems-related problems discussed previously, there are a number of generic problems which impact the ability of models to reasonably predict war-fighting capability. These problems include selecting valid and realistic modeling assumptions, accomplishing model validation testing, and evaluating intangible factors.

Modeling Assumptions and the Need for Validation Testing

The credibility of model results is, to a large extent, influenced by the various assumptions which are made in developing the model. Despite the correctness of what is modeled, inaccurate or incomplete assumptions can invalidate the effectiveness of the modeling output.

A 1985 OSD analysis of the status of theater level war gaming included a quote by the Director of Rand's Strategy Assessment Center regarding the primary reason that contemporary combat simulations lack credibility. The reason cited was that:

"... model results are driven by assumptions about: (a) which forces are counted; (b) the scores given to each force; (c) the attrition rates; and (d) the rates-of-advance assumption. It is possible to get drastically different results depending on assumptions."

The following two examples show how model assumptions affect the results of capability assessment analysis.

- The Army Materiel Systems Analysis Activity analyzed selected engagements of the 1973 Arab-Israeli war. A military analyst told us that using an existing evaluation model and data base, this organization had tried to determine whether its assessment model could replicate the actual outcomes of several battles. Evaluators found that certain model assumptions were invalid, creating inaccurate results. When the model was modified to correctly represent actual conditions, the actual combat outcome could be replicated with reasonable accuracy.
- At the Technical Center, Supreme Headquarters, Allied Powers, Europe, two studies of the same combat situation using different models produced different outcomes. One analysis assumed the defender could maintain a coherent position at the forward edge of the battle area so the opposing force could not break through. The second study used a model which did not make this assumption, and the opposing force achieved a breakthrough which cut off a large portion of the defending force—reversing the ultimate outcome of the battle.

Model validation, if properly accomplished to portray actual military operations, will generally identify erroneous model assumptions. However, OSD officials told us that the various weapons-scoring systems currently being used within DOD have not been adequately validated. We were told that not only had empirical validation of such methodologies seldom been tried but also that there had been few comparisons to determine the compatibility of weapons effectiveness scores developed by different models for assessing the same equipment. Furthermore, when comparisons were tried, major differences emerged. For example, in one recent case where two such scoring systems were applied to historical North Atlantic Treaty Organization and Warsaw Pact ground forces in central Europe, discrepancies in the vicinity of one or more allied corps emerged.

Air Force officials suggested that one approach to model validation is to compare the results of similar analyses to determine if they are compatible. For example, in the Air Force, the WSMIS and LCMS systems make similar assessments to support different Air Force organizations. They said that while comparisons of this nature are not the complete answer to the validation problem, they would help.

DOD and contractor officials told us that a primary reason for inadequate model validation is that necessary data is often unavailable or is of

questionable validity. These officials indicated that additional data could be acquired from records of historical combat and results of testing activities and training and operational exercises.

Evaluating Intangible Factors

The so-called intangible factors, such as leadership, morale, and skill, are generally not treated in existing capability assessment models. However, in the case of air-to-air and ground combat, extensive empirical test data indicate that human interactions are statistically more important than aircraft or combat vehicle performance, avionics, weaponry, or other test variables included in combat models.

Evaluating the human impact on combat outcome has been the focus of some analysis in recent years. Major sources of data for these analyses are the Air Intercept Missile Evaluation/Air Combat Evaluation (AIMVAL/ACEVAL) tests in 1978 and the Advanced Medium Range Air-to-Air Missile Operational Utility Evaluation (AMRAAM OUE) in 1981 and 1982.

AIMVAL/ACEVAL was a series of mock battles which tested air combat tactics, as well as the capabilities of fighter aircraft and air-to-air missiles. In these exercises, pilots flying the smaller, less sophisticated aircraft found that by using a combination of unorthodox tactics and superior numbers, it was possible to defeat top rated Air Force and Navy fighters.

The AMRAAM OUE took place in a simulator facility in 1981 and 1982. Over 20,000 sortie equivalents were "flown." The simulated sorties pitted fighters with varied weapons and avionics against various numbers of threat aircraft with fixed capabilities. The test results indicated that human interactions are statistically more important than aircraft performance, avionics, weaponry, or any other test variables.

Analysis of a data base of about 200 ground combat battles indicated similar findings. This study concluded that for 60 to 80 percent of the battles analyzed, the numerically superior force was victorious. Furthermore, for the great majority of the exceptions, success of a numerically inferior force was due to one or more of the following: defensive posture; surprise; or qualitative superiority due to leadership or better troops. This data base has been expanded and is currently being further analyzed in the Combat History Analysis Study Effort at the Army Concepts Analysis Agency.

List of Activities Visited

Office of the Secretary of Defense, Washington, D.C.

Office of the Joint Chiefs of Staff, Washington, D.C.

Headquarters, United States Air Force, Washington, D.C.

Headquarters, United States Army, Washington, D.C.

Headquarters, United States Marine Corps, Washington, D.C.

Headquarters, United States Navy, Washington, D.C.

Headquarters, United States Air Force Reserve, Washington, D.C.

Headquarters, United States Army Reserve, Washington, D.C.

Headquarters, United States Marine Corps Reserve, Washington, D.C.

The National Guard, Washington, D.C.

Reserve Forces Policy Board, Washington, D.C.

Headquarters, United States Navy Atlantic Fleet and the Naval Surface Force, United States Atlantic Fleet, Norfolk, Virginia

Headquarters, Air Force Tactical Air Command and First Tactical Fighter Wing, Langley Air Force Base, Virginia

Headquarters, Air Force Logistics Command, Wright-Patterson Air Force Base, Ohio

Headquarters, United States Army Forces Command, Fort McPherson, Georgia

Headquarters, First United States Army, Fort Meade, Maryland

Army Concepts Analysis Agency, Bethesda, Maryland

Army Inventory Research Office, Philadelphia, Pennsylvania

Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio

Analytic Sciences Corporation, Arlington, Virginia

The Brookings Institution, Washington, D.C.

Center for Strategic and International Studies, Georgetown University, Washington, D.C.

Decision Science Consortium, Inc., Falls Church, Virginia

Historical Evaluation and Research Organization, Fairfax, Virginia

Logistics Management Institute, Bethesda, Maryland

Military Operations Research Society, Alexandria, Virginia

Northrop Corporation, Hawthorne, California and Washington, D.C.

Rand Corporation, Santa Monica, California and Washington, D.C.

Synergy, Inc., Washington, D.C.

Veda, Inc., Las Vegas, Nevada

Requests for copies of GAO reports should be sent to:

U.S. General Accounting Office
Post Office Box 6015
Gaithersburg, Maryland 20877

Telephone 202-275-6241

The first five copies of each report are free. Additional copies are $2.00 each.

There is a 25% discount on orders for 100 or more copies mailed to a single address.

Orders must be prepaid by cash or by check or money order made out to the Superintendent of Documents.

33803

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

Bulk Rate
Postage & Fees Paid
GAO
Permit No. G100