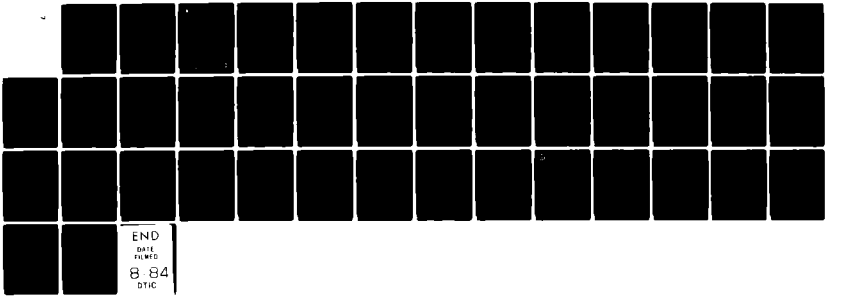
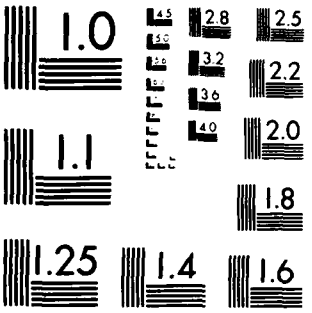
AD-A142 528 THE ARMY NEEDS MORE COMPREHENSIVE EVALUATIONS TO MAKE 1/
EFFECTIVE USE OF IT..(U) GENERAL ACCOUNTING OFFICE
WASHINGTON DC NATIONAL SECURITY AND.. 24 FEB 84
UNCLASSIFIED GAO/NSIAD-84-40 F/G 5/1 NL



END
DATE FILMED
8 84
DTIC



(1)

BY THE U.S. GENERAL ACCOUNTING OFFICE

Report To The Secretary Of The Army

The Army Needs More Comprehensive Evaluations To Make Effective Use Of Its Weapon System Testing

Evaluations of development and operational tests are used to assess the progress of new weapon systems as they proceed through development. Because many Army organizations contribute to the preparation of these evaluations, the results reaching acquisition officials at critical decision points are often fragmented. Seldom do these evaluations adequately interpret the test findings in terms of potential operational consequences. Evaluations need to be broadened and integrated to provide a more meaningful and coherent picture of system development progress and potential operational effectiveness. This will require greater participation by Army analysts in planning evaluations and in determining the technical and operational implications of test findings.

DOD agreed with GAO's findings and has taken steps to address these issues.

GAO NSIAD 84 40
FEBRUARY 24, 1984

UNITED STATES GENERAL ACCOUNTING OFFICE
WASHINGTON, D.C. 20548

NATIONAL SECURITY AND
INTERNATIONAL AFFAIRS DIVISION

B-211509

The Honorable John O. Marsh, Jr.
The Secretary of the Army

Dear Mr. Secretary:

This report discusses improvements that can be made in preparing evaluations of Army major weapon system tests so that decisionmakers are given better information about a system's progress in development and operational effectiveness.

We made our review as part of our continuing assessment of major weapon system acquisition management in the Department of Defense.

This report contains recommendations to you on page 24. As you know, 31 U.S.C § 720 requires the head of a Federal agency to submit a written statement on actions taken on our recommendations to the House Committee on Government Operations and the Senate Committee on Governmental Affairs not later than 60 days after the date of the report and to the House and Senate Committees on Appropriations with the agency's first request for appropriations made more than 60 days after the date of the report.

We are sending copies of this report to the Chairmen of the above Committees and of the House and Senate Committees on Appropriations and on Armed Services and to the Secretary of Defense.

Sincerely yours,

for Frank C. Conahan
Director

GENERAL ACCOUNTING OFFICE
REPORT TO THE SECRETARY
OF THE ARMY

THE ARMY NEEDS MORE COMPREHENSIVE EVALUATIONS TO MAKE EFFECTIVE USE OF ITS WEAPON SYSTEM TESTING

D I G E S T

Army evaluators of major new weapon system tests have not been giving acquisition officials technical and operational evaluations of sufficient scope and breadth to permit them to fully assess the weapons' development progress and potential combat usefulness. This is the preponderant view of a cross section of Army decisionmakers GAO interviewed. In most cases, the evaluations seldom went beyond reciting the test results and did not address critical issues such as the consequences of fielding the system before resolving remaining technical problems.

GAO interviewed scores of decisionmakers, project managers, testers, and evaluators in its study of the use made of test evaluations covering new Army systems during their development. GAO also examined reports by Army test evaluators on four current system acquisitions.

Recommendations on whether to continue a weapon system acquisition after it has begun development are made by the Army Systems Acquisition Review Council. The Council is composed of high level military and civilian officials. Based on the test evaluations and other information it receives from various Army sources, the Council recommends to the Secretary of the Army either that the system be allowed to proceed further into development or production, as the case may be, or that it undergo additional testing or redesign before proceeding to the next acquisition phase. The system's performance in testing, its latest projected cost, and the urgency to field it are some of the principal factors considered. (See pp. 1 to 5.)

NUMEROUS ORGANIZATIONS ARE INVOLVED IN EVALUATING ARMY WEAPON SYSTEMS

The two principal evaluation organizations for major acquisitions are the Army Materiel Systems Analysis Activity, which evaluates

development test results, and the Operational Test and Evaluation Agency, which evaluates operational test results. Development testing verifies how well a weapon system has met performance specifications. Operational testing gauges the system's potential operational effectiveness and supportability in a combat environment.

In addition, other organizations assess the system's demonstrated logistics supportability, cost-effectiveness, performance in a countermeasures environment, and ease of operation and maintenance by troops. (See pp. 6 to 9.)

There has been very little communication and coordination among the evaluators. Consequently, their evaluations lack the impact that a coherent, comprehensive evaluation might have on the decision process. Presently no single agency coordinates the evaluations of the different organizations.

For example, the Operational Test and Evaluation Agency listed many deficiencies it found while testing the Sergeant York air defense gun, but did not analyze the effect of producing and fielding it with the deficiencies not corrected or how much better it would be than the current air defense gun system. The evaluators did not estimate the retrofit costs, the loss in capability to engage and kill enemy aircraft, or the increase in personnel or in logistical support costs that fielding the system would entail. Such matters are studied by other Army organizations but are not brought together in a single overall evaluation of a weapon system. Without addressing these matters, the operational evaluators' report was insufficient to balance the arguments for beginning Sergeant York's production in spite of its shortcomings. (See pp. 9 to 13.)

MORE BALANCED AND COMPREHENSIVE ASSESSMENTS OF WEAPON SYSTEMS ARE NEEDED

A principal concern of numerous GAO studies has been the objectivity and comprehensiveness with which test evaluations are carried out. The importance of objectivity becomes apparent

when one considers the strong pressures to field new systems as quickly as possible, particularly in the Army, which has been in the throes of a weapon system modernization program since the early 1970s.

In assessing the progress of a new weapon system acquisition, decisionmakers must balance the views of system developer and user representatives who would accelerate the replacement of the older, less capable weapon systems, against the views of the test evaluators, who tend to urge more deliberate action to assure that the new system is ready for the next phase.

In the forefront of the system advocates are the Army's Training and Doctrine Command, the user representative that is anxious for a rapid force modernization, and the weapon system developer, the Materiel Development and Readiness Command represented by the project managers, who have a responsibility to keep the acquisition on schedule.

If test evaluators have doubts about rushing the acquisition, they must counter the arguments advanced by the project manager and others who want to press ahead. However, a major shortcoming of their evaluations has been a lack of comprehensive risk assessments of proposed systems.

A comprehensive risk assessment should consider the acquisition cost, schedule, and technical uncertainties in development plus the cost of delay, the military urgency, and the consequences of adopting alternative courses of action. An analysis of the consequences should include the added operating and support costs and decreased military utility which could result from fielding the system with deficiencies. (See pp. 13 to 20.)

RESOURCES TO PERFORM NEEDED COMPREHENSIVE ANALYSES AND EVALUATIONS SHOULD BE BROUGHT TOGETHER

There is no dearth of talent in the Army to produce the type of evaluation that the decisionmakers are seeking. Besides the test evaluation organizations, there are the analysis organizations that make assessments

using analytical models. The three principal analysis organizations are the Concepts Analysis Agency and the Systems Analysis Activity of the Training and Doctrine Command, along with the Army Materiel Systems Analysis Activity (which also does test evaluations). Finally, there are the many Army technical laboratories which make engineering assessments using component tests, experiments, and analytical models.

The resources scattered among the test evaluation and analysis organizations need to be harnessed and their contributions integrated, so that they can address the critical issues of concern to the decisionmakers. These issues include aspects of system effectiveness and suitability in an operational environment, maintainability, reliability, affordability, and potential risks. (See pp. 20 to 23.)

CONCLUSIONS AND RECOMMENDATIONS

While Army test evaluators are generally outspoken in their criticism of the systems, the fact that each involved evaluation organization looks only at a portion of the total picture precludes disclosing the collective impact of what are often serious overall deficiencies. The test evaluators and the analysts need to integrate and interpret their findings and speak with one strong voice, especially at the decision reviews, to provide a better balance between their views and those of the user and the developer.

It would be unrealistic to put all the expertise required for a truly comprehensive evaluation of each new weapon system into one organization. However, one principal evaluation agency should be designated to have access to the information available throughout the Army and be knowledgeable enough to interpret this information, integrate it into one comprehensive evaluation, and provide it to the decisionmakers. By virtue of the background and experience of their personnel any of the three main Army analysis organizations - the Concepts Analysis Agency, the Systems Analysis Activity of the Training and Doctrine Command, or the Army Materiel Systems Analysis Activity - would be a good choice to assume this responsibility.

In a draft of this report, GAO proposed that one of the Army's three main analysis organizations be designated the principal evaluator. Because DOD stated it has adopted a pilot program to test the concept of the Operational Test and Evaluation Agency handling this role, this proposal has been deleted from the final report, pending the results of the test. Nevertheless, GAO is making several recommendations designed to provide more useful weapon system evaluations to Army decisionmakers.

AGENCY COMMENTS

The Department of Defense agreed that Army test evaluation is too fragmented to provide a meaningful picture of a system's progress and that more balanced and comprehensive assessments of weapons are needed. However, the Department did not agree that one of the analysis agencies should make these assessments. Instead, the Army intends to institute a pilot program in which the Operational Test and Evaluation Agency would direct the comprehensive evaluations for a period of about 2-1/2 years.

DOD believes that the Operational Test and Evaluation Agency would be the best agency to coordinate assessments of a system's status at any time in the development cycle and report significant changes to decisionmakers. To accomplish this mission, the Operational Test and Evaluation Agency will get analytical support from the Army's analysis organizations. Although GAO believes the preferable way is to designate one of the Army's three main analysis organizations, it considers the pilot program with this organizational structure to be an acceptable approach. (See pp. 27 to 33.)

Contents

ABBREVIATIONS

AMSAA	Army Materiel Systems Analysis Activity
ASARC	Army Systems Acquisition Review Council
DARCOM	Materiel Development and Readiness Command

DOD	Department of Defense
GAO	General Accounting Office
OTEA	Operational Test and Evaluation Agency
TRADOC	Training and Doctrine Command

CHAPTER 1

INTRODUCTION

Many organizations in the Department of the Army have a role in the weapon system acquisition process, the pivotal player for new major systems [1]/ being the project manager. The project manager controls most actions affecting the system's development and wields considerable influence over the fortunes of the system. The project manager is the principal spokesperson for the system in DOD and before congressional committees.

Test and evaluation organizations, while not having direct control of weapons under development, are nevertheless in a position to influence their fate. The two principal Army test evaluators for major acquisitions are the Operational Test and Evaluation Agency (OTEA) and the Army Materiel Systems Analysis Activity (AMSAA). OTEA, which reports to the Vice Chief of Staff, evaluates operational tests. AMSAA, which reports to the Materiel Development and Readiness Command (DARCOM), is the evaluator for development tests. Although they are a part of the Army, both organizations are sometimes referred to as "independent" evaluators because they plan and prepare their evaluations free of any direction from the project managers. Some other organizations that frequently contribute to the Army's deliberations on weapon systems performance are the Human Engineering Laboratory, the Ballistics Research Laboratory, the Office of Missile Electronic Warfare, and the Logistics Evaluation Agency.

The distinction between "test evaluation" and other kinds of evaluation needs to be clarified because it is not well defined. This report refers to a "test evaluation" as an evaluation based primarily on test results. An example is an evaluation of operational testing by OTEA. Most other evaluations are based on results of analytical or logical modeling of one kind or another, such as computer simulations and war games. Information entered into the models may be combat data, experimental data, assumptions, and data generated by other models.

For purposes of this report, organizations that evaluate test results are referred to as "test evaluators."

---

[1]The Department of Defense (DOD) characterizes as "major" those weapon systems whose development cost is expected to exceed $200 million, or whose production cost is expected to exceed $1 billion.

Organizations that use results from analytical models in their evaluations are referred to as "analysis" organizations.

Testing can be divided into two types: development and operational. Development testing is used to verify how well the system has met performance specifications. Operational testing addresses the operational effectiveness and suitability of a system. It is intended to gauge how well the system can be expected to perform in the operational or combat environment, how it should be employed, and whether the system can be operated and maintained effectively by military personnel.

Test evaluation organizations often appear to be adversarial in their evaluation reports and to be challenging the weapon system. In fact, Army test and evaluation h· two purposes: to help the project manager uncover system oblems for correction and to help the decisionmakers determin f development is progressing satisfactorily and if t' system is likely to meet operational needs.

ACQUISITION CYCLE SETS FRAMEWORK FOR TESTING

To understand the role of testing and evaluation in Army acquisition decisionmaking, it is necessary to know the different phases of the acquisition cycle and where the decision points occur.

The acquisition cycle begins with concept definition and proceeds through two phases of development where the system design is gradually firmed up to prepare for production. The Army Systems Acquisition Review Council (ASARC), the principal Army weapon system decisionmaking body, generally meets at three critical points in the cycle to review the progress of the system's development. ASARC proposals to move the system to the next acquisition phase are subject to review at the Secretary of Defense level by the Defense Systems Acquisition Review Council (DSARC).

A typical cycle for a major system Army acquisition is depicted in figure 1.

Figure 1

Typical Acquisition Cycle for Major System

Phase	Milestone	Activities	Description
1		Concept definition	Explore different alternative designs to meet requirements
	1	ASARC I/DSARC I	Decision to enter Phase 2
2		Demonstration and validation	Component development and early system prototype fabrication
	2	ASARC II/DSARC II	Decision to enter Phase 3
3		Full-scale engineering development	Maturation of components and system prototype
	3	ASARC III/DSARC III	Decision to enter Phase 4
4		Production and deployment	Convert prototype to production version and produce

Development testing and operational testing, and evaluations of each, usually occur before both the second and third decision milestones. At the first milestone, the ASARC occasionally has force development testing and experimentation results provided by the Training and Doctrine Command (TRADOC) and component testing results provided by the developer to help judge the system concept.

PRINCIPAL ARMY DECISIONMAKERS ARE REPRESENTED AT MILESTONE REVIEWS

For major programs, most of the principal decisionmakers are represented on the ASARC. The Army Vice Chief of Staff chairs the ASARC. Members include the Deputy Under Secretary of the Army for Operations Research; the Assistant Secretary of the Army for Research, Development and Acquisition; and the Commanding Generals of DARCOM, TRADOC, and OTEA. Other members are the Deputy Chiefs of Staff for Research, Development and Acquisition, for Personnel, and for Logistics. While not officially a member, the Under Secretary of the Army usually attends the reviews.

Through discussions among these top officials, an Army consensus on the weapon system's merits and progress is reached, and a recommendation to the Secretary of the Army is developed on whether to enter the next phase in the acquisition cycle.

OUR PREVIOUS REVIEWS OF TEST AND EVALUATION FOCUSED ON TESTING

Our previous reviews of test and evaluation have generally emphasized the problems with the testing, not with the evaluation.[2] They stressed that testing is an important tool in assessing the degree of risk inherent in a weapon system acquisition and pointed out common problems in test planning, conduct, and reporting.

For example, in a 1972 report, we concluded that DOD's testing was inadequately planned and conducted and cited instances when test results were not being provided to decisionmakers before important decision points. In a 1979 report on development testing, we reported that development tests were reduced as the systems experienced cost growth and schedule slippages.

In 1979, we specifically reviewed the conduct of operational test and evaluation in the Army. Our report, entitled "Army Operational Test and Evaluation Needs Improvement," called attention to a lack of realism in some of OTEA's tests and the fact that in some instances, timely test results were not provided to the decisionmakers.

OBJECTIVES, SCOPE, AND METHODOLOGY

We began our review in September 1982 and completed it in April 1983. We reviewed the role of Army test and evaluation organizations in the acquisition process because of our concern either that they were not being heard by decisionmakers or that their reservations about some of the weapon systems, if heard, were not being taken seriously enough at the milestone reviews. In our previous examination of several weapon system acquisitions, we observed that the reservations that organizations, like the Logistics Evaluation Agency, AMSAA, and the Human Engineering Laboratory, had about system performance and readiness for production were subordinated in favor of the primary objective of getting the system into the field as soon as possible. For example, the Sergeant York, the Patriot, and the Abrams tank systems moved from development to production

---

[2]Appendix I lists our previous reports on DOD testing.

before major problems revealed by the tests were resolved. In other cases, such as the Viper, test results indicated that the system was only marginally more effective than the existing system, but the decisionmakers, nevertheless, approved production.

The objectives of our review were to

--ascertain how Army test and evaluation organizations are organized and what their roles are in the acquisition process;

--determine the types of information that decisionmakers considered important for inclusion in Army evaluations of test results; and

--if necessary, identify ways to give the decisionmakers more responsive and complete system evaluations.

In the review, we considered the contributions of the major Army organizations involved in development and operational test evaluation. We compared these contributions with the information needs expressed by decisionmakers at all levels in the Army. We concentrated on the decision process for major system acquisitions.

Since previous reviews have covered the importance and adequacy of testing, we focused on the importance and adequacy of evaluation. We felt that improvements were needed in evaluating the test results as well as in the testing itself.

We interviewed a cross section of decisionmakers, testers, and evaluators, mostly in the Army, to obtain data, views and ideas relevant to the three objectives. Among these were the Under Secretary of the Army; the directors of the Army's principal test evaluation and analysis agencies; project managers and system managers of several key major weapon systems; and some past and present member of the ASARC. To complement the interviews, we reviewed many reports and studies by the Office of the Secretary of Defense, the Army, and other defense organizations, as well as some of our previous reports on individual weapon systems and on testing. We also examined Army reports covering testing, evaluation, and acquisition decisions made on four major weapon system programs--the Abrams tank, the Apache helicopter, the Sergeant York air defense gun, and the Army Helicopter Improvement Program. These four programs provide a cross section of major weapons in various stages of development.

Our review was made in accordance with generally accepted government auditing standards.

CHAPTER 2

EVALUATIONS OF TESTS MUST BE INTEGRATED AND BROADENED TO MEET GROWING NEEDS OF ACQUISITION DECISIONMAKING

Many Army test evaluation organizations try to provide useful information to decisionmakers about weapon systems in development, but most decisionmakers we interviewed expressed dissatisfaction with the reporting and scope of the evaluations. Many maintained that the information was fragmented, sometimes contradictory, and usually not complete or convincing. They, therefore, had to rely on the projections of the system project managers as the bases for their decisions.

The decisionmakers said that to gain a broader perspective and to check the assessments of the project managers, they need a more thorough interpretation or evaluation of test findings. They believe these evaluations should address specific critical issues and include an assessment of risks inherent in the program and in alternative courses of actions. The evaluations should go beyond assessing how well the system performed under test to projecting how useful it will be when fielded, compared with alternative systems. Sufficient technical and analytical knowledge and expertise to perform this thorough evaluation is available in a number of Army organizations but is not being adequately used.

At the ASARC meetings, the project managers have the major roles, since they are the most familiar with the system under review. However, the emphasis on staying within cost and schedule limitations and the visibility that occurs when they are exceeded sometimes causes them to understate the importance of system performance shortcomings and to see technical or other obstacles to timely development or production as easy to overcome. This is not to say that the project managers deliberately try to obfuscate the issues. There is, however, a definite tendency to view troublesome areas with undue optimism.

For example, during the summer of 1982, the Patriot project manager stated that program risks had been reduced to the point that operational tests slated for early 1983 were in his opinion no longer necessary. Subsequently, Patriot did poorly in development tests held in December 1982, and the 1983 operational tests had to be stopped not because they were not necessary, but because of the system's poor technical performance.

The TRADOC system manager represents the user rather than developer. He presents the need and cost-effectiveness analysis at the ASARC and is also a strong supporter of the system since he reflects the user's eagerness to field new systems. We found no case where the TRADOC system manager did not recommend moving ahead with a program although several were experiencing serious technical difficulties.

OTEA, which is the test evaluation representative at ASARC meetings, is thus the only real counterbalance to (or check on) the assessments of the project manager and the TRADOC system manager. For this reason, it is important that OTEA's evaluation and presentation to the ASARC be objective, expert, and comprehensive enough to satisfy decisionmakers' needs.

FRAGMENTED ARMY TEST EVALUATION MUST BE OVERCOME TO ENABLE BETTER ASSESSMENTS

Over the years, concerns about the test and evaluation process have led to different arrangements of the many groups involved in Army test and evaluation. This has resulted in a complex organizational structure and in piecemeal presentation of test findings. While considerable money (around three quarters of a billion dollars annually) and effort are spent in accumulating and interpreting test data in the different organizations, the evaluations which are the products of these organizations are not integrated to produce a single comprehensive evaluation for the decisionmakers.

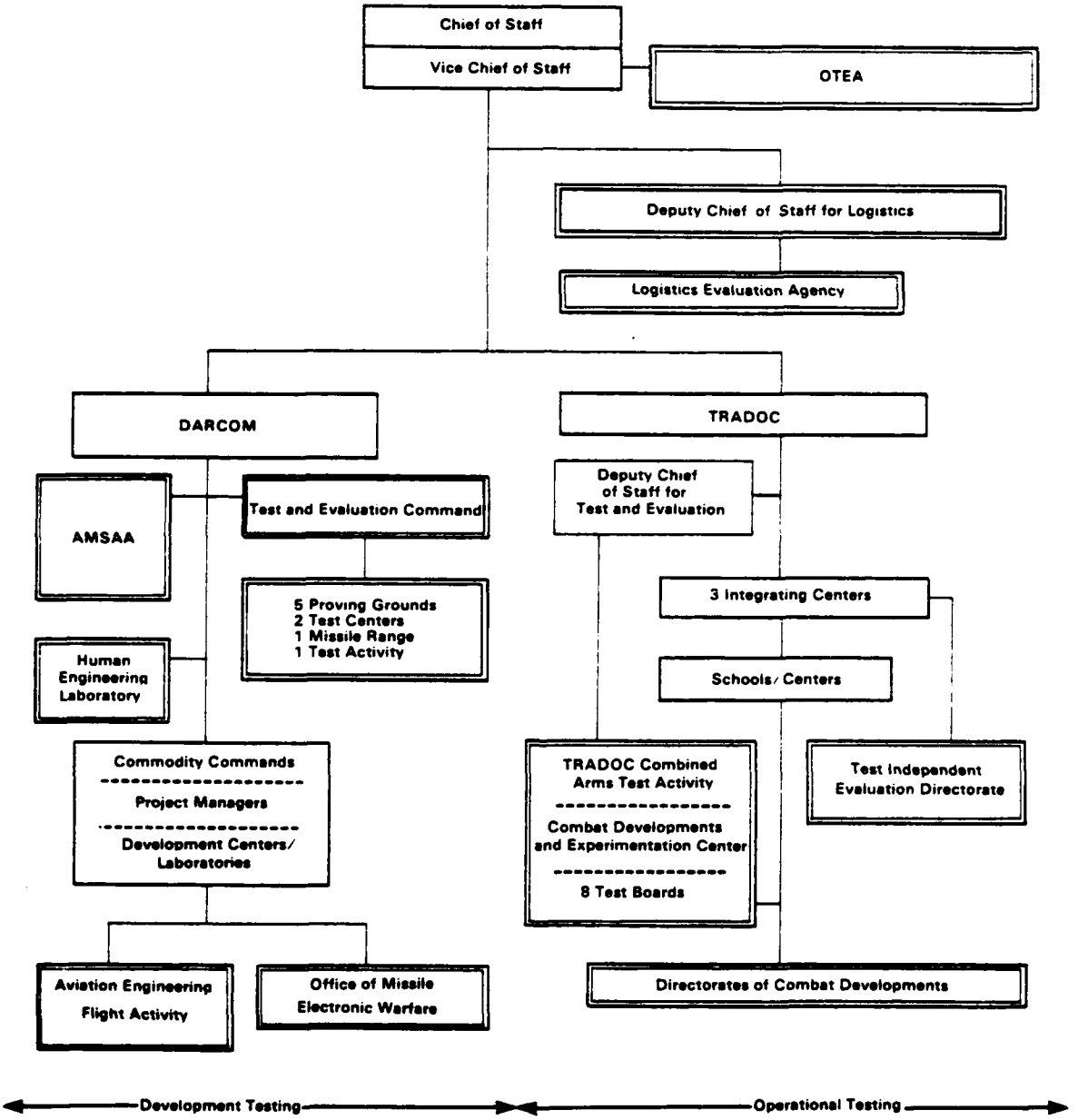
Functions split among numerous agencies

Figure 2 depicts the dispersed Army structure for test and evaluation. OTEA and AMSAA are the two principal Army test evaluators. OTEA reports to the Vice Chief of Staff and is responsible for planning, managing and evaluating operational tests of major systems. AMSAA reports to DARCOM headquarters and performs independent design and evaluation of development tests on major systems.

In addition, other organizations are involved in testing and test evaluation. Most of these operate under DARCOM or TRADOC. Organizations under DARCOM are involved mostly with development testing, while those under TRADOC are involved mostly with operational testing.

FIGURE 2

ARMY STRUCTURE FOR TESTING AND TEST EVALUATION



Note: Double lined boxes represent test and test evaluation organizations.
Single lined boxes show the organizations to whom they report.

Some of the more important of these organizations and their test-related responsibilities are:

Organization	Responsibility
DARCOM agencies:	
Test and Evaluation Command	Plans, conducts, and reports results of development
Human Engineering Laboratory	Examines man/machine interaction in development tests
Office of Missile Electronic Warfare	Assesses systems' vulnerability to electronic counter-measures
TRADOC agencies:	
TRADOC Combined Arms Test Activity	Conducts operational and force development tests with units greater than battalion size
Combat Developments Experimentation Center	Conducts operational and force development tests of weapons in the context of small unit engagements
Test boards (eight)	Conduct operational and force development tests

Also, the Logistics Evaluation Agency serves as the Army's independent logistician and reports to the Deputy Chief of Staff for Logistics. Although it is not directly involved in designing and conducting operational tests, it does evaluate test results for each system from the logistical standpoint. Highlights of this evaluation are presented at the ASARC.

Many more organizations than shown in figure 2 are also involved to some extent in planning, conducting, and evaluating tests. They include the project management offices, the DARCOM laboratories, and the TRADOC system manager offices.

Presentations to decisionmakers not coordinated

The test evaluation organizations do not coordinate their presentations to the decisionmakers. Their assessments are generally considered separately and, therefore, lack the impact that a coherent, comprehensive evaluation of all the available system information might have on the decision process.

Furthermore, the presentation to the ASARC of significant portions of the test evaluations is usually made by parties other than those familiar with the formulation of the evaluations. Therefore, matters requiring further clarification many times are reportedly not addressed.

For example, OTEA presents both the operational and development test findings. This is done in the interest of providing fewer and more coherent presentations. AMSAA evaluators are concerned that OTEA cannot convincingly present their issues. The OTEA evaluator often lacks understanding or expertise in the difficult technical areas of the system's development. The evaluator generally limits the discussion of the implications of system shortcomings found in testing to the most obvious operational implications and avoids any technical or overall risk assessments. For instance, the OTEA evaluator of the Apache cited a reduced number of kills and incorrect missile flight times as two consequences of problems found with the target acquisition and designation subsystem. However, the evaluator did not interpret the reduction in kills in terms of the resulting decrease in cost effectiveness and did not compare the Apache's cost effectiveness with that of alternative systems. Also, instead of assessing the chances of correcting the problems, the evaluator merely stated that no fixes had yet been demonstrated.

In addition to the ASARC general briefings, the decisionmakers can obtain detailed information through less formal channels. In some cases, certain decisionmakers will pay close, personal attention to key acquisitions. The Patriot program, because of its complexity and expense, was closely monitored by the then Assistant Secretary of the Army for Research, Development and Acquisition, before it was approved for limited production. He was particularly concerned with the computer software problems uncovered during multiple aircraft target testing.

However, most systems are not monitored that closely and decisionmakers rely on the formal summaries from the evaluation agencies.

Existing coordination mechanisms are not sufficient to provide a comprehensive evaluation

Communication and coordination among the test evaluation organizations is limited considering the similarity of their objective of providing data and assessments for key acquisition decisions. However, at certain points, the Army does bring together the representatives of the different organizations involved in testing.

A permanent committee allocates test funding, recommends test priorities, and manages operational test resources for all

systems under development. There is also a working group for each system which is concerned with the number of prototypes to be tested, test affordability, the types of test instrumentation and targets, and the models or simulations needed to supplement the tests.

In addition, the main participants in the testing of a major system meet in the operational test scoring conference to analyze test incidents. This conference is a four-member panel, chaired by OTEA, with representatives from TRADOC, AMSAA, and the project manager. The panel assesses the importance of any failures or mishaps during operational testing and determines if the problem should be charged to the weapon itself or to some other cause.

Although these efforts at test coordination are beneficial in reducing duplication, scheduling resources, and scoring test results, no similar efforts are made to coordinate the test evaluations of the different organizations. Each appears to work in a vacuum with only limited knowledge of the others' work. OTEA, in particular, pursues this policy of isolation in the interest of maintaining its independence.

Reluctance to share data may not be the only reason for this lack of interaction. The heavy work loads of the testers and evaluators, the remote locations of some of the organizations, or the difficulty of understanding some of their complex products all lessen the likelihood of extensive coordination.

TEST EVALUATIONS MUST BECOME BROADER AND FOCUS ON KEY ISSUES

Army decisionmakers have stated that merely integrating the fragmented test evaluations would not suffice. The test evaluations the decisonmakers receive now are often not comprehensive enough and leave important questions unanswered. They are limited mainly to test results, listing problems encountered and performance thresholds not met, but rarely going beyond this to analyze the seriousness of these deficiences, individually or collectively. The evaluations should also focus on key issues and be broadened to include risk assessments of systems under review. We believe such evaluations could be done using the expertise available within the Army.

OTEA's independent operational evaluation of the Sergeant York air defense gun offers an example of the narrow type of evaluation decisonmakers have been receiving. It listed the problems found in two combined operational and development tests, including failure to meet requirements in mission performance, survivability, and reliability. However, the evaluation did not analyze the effect of producing and fielding the Sergeant York with the deficiencies not corrected or how much better than the current air defense gun system it would be. It did not estimate the retrofit costs, the loss of capability to engage and kill enemy aircraft, or the increase in

personnel or in logistical support costs that its deployment would require. Since the effectiveness and cost risks were not included, the project manager's argument against delay could not be properly balanced by arguments for delay.

An example of the kind of analysis decisionmakers would find more useful was performed for the Abrams tank program as an add-on to the cost and operational effectiveness analysis. This analysis addressed tank maintenance staffing and logistics because of the concerns generated by the Abrams falling short of the reliability and maintainability standards. It tried to identify for the Abrams tank the relationships between combat effectiveness, operational availability, system reliability and maintainability, spares provisioning, and maintenance staffing. It also estimated the support resources and associated costs required to achieve peacetime and wartime readiness objectives as a function of Abrams reliability and maintainability.

Analytical support was provided by TRADOC's Armor Center, Systems Analysis Activity, and Logistics Center. Data was provided by AMSAA, the Ballistics Research Laboratory, the Abrams project manager's office, the Concepts Analysis Agency, and the Army Missile Command. After considerable effort, a battalion level analysis was issued by the Armor Center and a theater level analysis by the Logistics Center. However, it came too late to be of use to decisionmakers.

Need to better define and focus on critical issues

The decisionmakers we interviewed felt that many of their major concerns were not being addressed by the issues explored in operational tests. Their concerns generally fell into the following categories:

--affordability, compared to competing demands of other systems;

--development risks in meeting cost, schedule, and performance requirements;

--producibility risks in terms of cost, schedule, and quality control;

--combat utility in combined arms teams against enemy threat and countermeasures, and in various environments;

--operational suitability from an operation and maintenance standpoint, including the skill levels required.

The last Apache operational test in 1981 is an example of not addressing the right issues. Considerable effort went into performing a realistic force-on-force test to judge the effectiveness of the Apache attack helicopter as a part of a combined arms team against a Soviet type force. Good data was collected on exposure times and engagements between the competing weapon systems. A number of good analyses were made using these data which compared the contributions of the Apache with those of the Cobra attack helicopter in terms of kills and losses.

However, the Soviet air defense weapons and the Cobra capabilities simulated were mostly those of the 1970s rather than of the late 1980s when the Apache will be introduced in quantity. The real issue of how effective the Apache will be, compared with an advanced Cobra and against the threat of the 1980s, was not considered in the test or in the evaluation. We believe it could have been considered since the Apache cost and operational effectiveness analysis at the time did address this issue.

Evaluations should include risk assessments

Technical risks are those involved in translating a system design into a producible weapon within the constraints of cost, schedule, and performance specifications. Operational risks are those involved in fielding a system that is cost effective, suitable for the troops to operate and maintain, and logistically supportable. Generally, the only formal risk assessments presented to the ASARC principals are those made by the project manager with some assistance from the contractor and the laboratories in the development commands. They emphasize affordability and cost and schedule risks and seldom emphasize the technical and operational risks involved in the program.

Risk assessments should address both the risks of moving too rapidly and the risks of moving too cautiously. For instance, consideration of relative risks may dictate more development and testing before proceeding into the next acquisition phase, or it may dictate going ahead and accepting degraded operational effectiveness in order to field the system quickly. The risks of haste include the higher costs of later correcting system shortcomings and the danger of fielding a less operationally effective system, entailing greater support costs and troop support. The risks of delay include increased development and production costs and delayed fielding of an urgently needed system.

A comprehensive risk assessment should consider the cost, schedule, and technical uncertainties in development plus the cost of delay, the military urgency, and the consequences of success or failure of alternative courses of action. The consequences of failure might include the added operating and support costs and decreased military utility of fielding a system

with deficiencies. The total comprehensive risk assessment must then strike a balance between these differing factors and give the decisionmakers a coherent view of the possible consequences of their decision.

In our studies of Army weapon system acquisitions, we found that risk assessments were usually performed or coordinated by the project manager's office. They seldom elaborated on the magnitude of the technical risk involved. Rarely were any problems rated "high" risk. A few were called "moderate" risk, but most were characterized as either "low" or "no" risk. Such assessments frequently differed from the test evaluators' analyses.

For example, the Sergeant York project manager assessed the many technical problems the system was encountering to be low risk. However, AMSAA officials expressed serious reservations about continuing the program as scheduled, stating that there was an extremely high risk in exercising contract options before completing performance tests. Nonetheless, the Army is exercising these options, even though scheduled tests could not be completed because of system prototype breakdowns.

The most thorough technical risk assessment of the four cases we reviewed was conducted on the Abrams tank. In February 1979, a panel of technical experts from industry and DOD assessed the power train's reliability and durability and the adequacy of the modifications which followed earlier test failures. While concluding that the failure rate was not unusual considering the tank's stage of development, the panel stated that confidence in the success of the modifications could be obtained only by further development and testing.

How much testing is necessary to support an adequate evaluation?

Some of our previous reports have cited serious inadequacies in testing. Some tests were incomplete, and others were deferred or waived before key decision points. Also, tests were generally conducted in unrealistically favorable environments. While the test evaluators argued against omissions and for more complete testing, they did not assess the implications of omitting the tests.

Testers, evaluators, project managers, and ASARC principals all agree on testing's importance in helping to develop a good product. It is in the type and amount of testing needed and the timing of some of the tests that they differ. The testers and evaluators espouse more testing. They feel it is necessary to provide a good evaluation.

On the other hand, the project managers we interviewed pointed out that testing, particularly operational testing, is expensive in time, resources, and money. They said that the

testers often overtest, using up valuable time and resources. They criticized operational tests that were just a replay of development tests with the addition of troops. The project managers wanted the tests directed more at employment and supportability of the system and less at finding specific hardware deficiencies. They would like the test evaluators to determine for them the priority of problems to be corrected from an operational standpoint.

The operational test evaluators counter that they want to do broader operational testing involving system employment and supportability, but that they are not usually able to get the resources to do it right. Many times, system prototypes are tied up in concurrent development testing or are being fixed by the contractor to correct problems found. Those assigned to operational testing often suffer from reliability problems not corrected or not found in development testing, so that their availability for broader testing is generally limited. The same three Sergeant York prototype air defense guns were put through several series of tests because no others were available.

The operational evaluators would like to have large two-sided battles included in tests to measure all the interactions of the new weapon system with friendly as well as with enemy forces. For example, the OTEA evaluator for the upcoming operational test of the Army helicopter improvement program had originally planned a large-scale simulated battle to test the many facets of a scout helicopter working with attack helicopters and ground forces. Program officials said the test plan had to be significantly scaled down because of the reluctance of the Forces Command to commit so many troops and surrogate weapons.

Lacking the means for the broader tests, the test evaluators have often concentrated on providing precise measurements of such things as mean miles driven, or rounds fired, between failures without establishing the significance of the tests that were omitted. Army analysts familiar with both testing and evaluation were concerned that operational testing could become too institutionalized by concentrating on precise measurements and large sample sizes rather than key issues.

The analysts did not believe that large forces were always needed to obtain meaningful employment and supportability data on the new systems. First, they suggested that a good many of the operational issues could better be answered by combat data, war gaming, computer simulations, and analytical studies. Second, operational tests could be designed to specifically provide data on the critical issues for which testing is required, rather than trying to address all issues. Finally, test evaluations could integrate the test data with other data sources to address the critical issues. The analysts said that generally the operational evaluations have not been designed to build on previous operational and development testing and evaluation. Better planning would enable a good deal more relevant data to be obtained for the same testing resources and expenditures.

CHAPTER 3

ARMY ANALYSIS COMMUNITY COULD ENHANCE TEST PLANNING AND EVALUATION BY ASSUMING A LARGER ROLE

Two trends are creating a greater need to complement physical testing with other sources of data for test evaluation. They are the increasing complexity of new systems and the recognition that systems should be evaluated early before their designs are frozen. Realistic and complete operational testing must await production of an adequate number of operationally configured systems, but by this time it is too late to use the results to help in the acquisition process. If operational evaluations of complex systems are to be performed early enough to support decisionmaking, results of physical testing must be complemented by results of simulations, analysis, and war gaming.

Greater use of the evaluation expertise available in the Army's analysis organizations could help produce the type of comprehensive weapon system evaluation needed for acquisition decisions. The analysis organizations can help provide more in-depth evaluations early enough in the acquisition cycle to make a difference. By using their analytical modeling capabilities and synthesis experience, they could provide data which would complement the information derived from test results. Combining this information would produce a comprehensive evaluation of the weapon system under review.

WHAT THE ANALYSIS AGENCIES CAN CONTRIBUTE TO TESTING AND EVALUATION

The analysis community uses mathematical models to evaluate the operational effectiveness of systems. Given the capabilities of a system, the analysts try to predict how well the system will do either by itself or with other friendly systems in a combat situation. With good models, the analysts should be able to determine what factors drive the system's effectiveness. The factors could be the capabilities of the system, such as its weapon lethality, accuracy, or reliability; the capabilities of enemy weapon systems or friendly support systems; or environmental conditions, such as weather or terrain.

The Army analysis organizations have developed large-scale models at the three force levels--theater, corps/division, and battalion. These models are used in cost and operational effectiveness analyses conducted for the decision milestones of major systems. Most of the time the analyses are run concurrently with the development and operational tests so that the data used in the models is not based on testing. Similarly, the

test designers and evaluators do not have the benefit of the insight gained from the analyses. Better scheduling and coordinating of their efforts could result in providing better evaluations for the decisionmakers.

Probably the most important contribution the analysts can make to test evaluation is to run the test results through their models and determine the implications of these results for the system's operational effectiveness and supportability. To do this in time for the decision milestones, the models could be run before the tests with a range of values for each driving factor. Thus, when the actual test results are received, the relationships would already have been developed and the implications could be quickly determined.

Complexity of new systems creates need for more frequent in-depth evaluations

The trend toward more complex weapon systems is a challenge to the developer, tester, evaluator and decisionmaker. Realistic testing of complex systems is more difficult, owing in part to the lack of mature or sufficient prototypes, and the inability to completely duplicate the battlefield, including the threat, for a reasonable cost. Therefore, the evaluators should be able to rely more on information obtained in ways other than tests.

Evaluations of complex weapon systems should, according to several officials we interviewed, assess equipment performance in the changing operating environment. The environment should include not only the usual physical aspects, such as temperature, humidity, and vibration, but also the countermeasure threat, compatibility, and interoperability of both the hardware and software with other friendly equipment, the skill level of the operation and maintenance crews, and logistic support levels.

Innovative testing, using early prototype and surrogate hardware, has been tried in some cases to address the questions of systems requirements, technical feasibility, operational effectiveness and suitability, supportability, and many other issues. Progress has been made in simulating the threat, battle environment, and other factors too costly or too difficult to represent physically in tests. For example, computer simulation of message traffic, noise, and countermeasures is being used in command, control, and communications testing. Even with all these aids providing more realism to field tests, senior analysts we interviewed believe that complementary studies, analyses, and war gaming will be required to fully address the critical issues. In addition, they assert that these tools would make testing more efficient through better planning.

Two DARCOM test evaluation and analysis organizations, the Office of Missile Electronic Warfare and the Human Engineering

Laboratory, have demonstrated that in their areas of responsibility, it is possible to combine simulations, analysis, laboratory experiments,and field testing to achieve comprehensive evaluations of complex systems. Both organizations, like AMSAA, are primarily staffed with experienced experts who are used as the principal evaluators.

Evaluations are needed earlier in the acquisition cycle

Systems fielded without adequate testing and evaluation may later disclose serious operational problems and require costly retrofits. We believe ways must be found to make testing and test evaluation more efficient and responsive if the Army is to have both shorter development cycles and weapons fielded without serious problems.

The period of greatest opportunity to affect operational performance, reliability, maintainability, and ultimate life-cycle cost is the early design phase. Considerations of cost and schedule make system changes in the later stages difficult, if not impossible. A great deal more effort is needed "up front" to prevent serious problems from occurring. This involves more evaluation of trade-offs during the conceptual phase. It involves more emphasis on evaluating planned systems and their interactions with the expected tactical environment and with related systems.

Although decisionmakers (including the project managers), and test evaluators alike, agree that an early evaluation is the most likely to have an effect, the evaluator's involvement in the early stages has usually been minimal. In two recent cases that we examined, early involvement by the test evaluators has been practically nonexistent.

The Sergeant York program featured an accelerated acquisition approach and minimum government involvement, with the purpose of fielding the system in the shortest possible time. The strategy was approved in April 1977, and two prototype development contracts were awarded in January 1978. The test evaluation organizations were not allowed to participate in the program until June 1980, and even then, not extensively, consistent with the Army's belief that this would help achieve early deployment.

In this case, the testing strategy represented an optimistic approach to the acquisition. The test evaluation organizations reluctantly agreed to the strategy, although their subsequent evaluation reports expressed serious reservations about the quality and completeness of the information they were able to provide to the decisionmakers.

In the Army helicopter improvement program, the early involvement of government testers and test evaluators was also limited. The Army believes this was appropriate since the program dealt with an existing aircraft, rather than one that was newly developed, and because of the design effort already accomplished on the mast-mounted sight. The system skipped formal development and operational testing during demonstration and validation, although some technical feasibility tests were conducted. The first government operational tests will be in September 1984. OTEA assessed the mast-mounted sight, based on developer and contractor component engineering testing. While the mast-mounted sight is the program's highest risk component, other important issues, such as the efficacy of the scout as a target designator for the attack helicopter, will not be addressed until the September 1984 testing, when the first prototype will be available. As in the Sergeant York program, the desire to shorten the acquisition cycle kept government test and evaluation to a minimum until late in development, forcing the Army to make important decisions based upon incomplete information.

While these two examples may be extreme, in most cases the operational test evaluator does not become involved in the development program until some tests of the prototype system can be done. The evaluator faces a dilemma: the integrity and realism of the test demand a mature prototype, but the evaluator's ability to have a real effect on the program requires an evaluation before the prototype can be developed.

Realistic logistics support testing is especially difficult to do early in the development cycle because the mature components and test sets are not available. Nevertheless, logistics supportability assessments must be a part of the comprehensive evaluation of the system since logistics will have a large effect on the total life cycle cost of the weapon.

Logistics supportability evaluations often suffer the most from compressed development schedules and shortened acquisition cycles. The Sergeant York development contract does not require the integrated logistics support to be delivered until just before fielding. This late development of the support equipment, training of maintenance personnel, and spare parts provisioning, will make any design changes necessitated by weaknesses in the logistics supportability very expensive, coming as they will when production units have been manufactured and the design is set. Also, the lack of information about logistics supportability deprived the Army of an important tool in its decisionmaking.

Complex weapon systems require continuous as well as early interaction among users, developers, testers, and logisticians. Continuous evaluation of such systems could provide constant, timely, and relevant information both to the decisionmakers and

to the developers. If the evaluators are involved only at certain discreet points in the acquisition cycle, then they might be cut out from influencing the majority of the decisions made throughout the program. Problems may be so entrenched by the time the evaluation is done that they are very difficult to remedy.

The Army's development test evaluation has benefited from assigning the evaluation function to AMSAA, an established analysis organization. AMSAA uses analytical models to predict reliability growth from test data and combines reliability, accuracy, lethality, vulnerability, and environmental test data to estimate kill and survivability capabilities of weapon systems. AMSAA is primarily a civilian-staffed analysis organization. It has a large staff of experienced technical, engineering, and analytical experts that cover nearly every area of interest to the Army.

In view of the quality of the evaluations that the Army obtained after it assigned to AMSAA the independent evaluation of development tests, we looked for opportunities for deriving comparable results in evaluations of operational tests of new systems.

Currently four kinds of Army organizations perform analyses of individual systems:

--The TRADOC schools and centers determine whether each system is needed as well as the performance required of each new development.

--TRADOC's Systems Analysis Activity provides analytical support to the schools and centers, especially in conducting cost and operational effectiveness analyses to determine which alternative system best meets the needs of the Army.

--AMSAA assesses most of the system's performance potential based partly on its analyses of development test results.

--DARCOM laboratories and project managers provide system characteristics to AMSAA.

Besides AMSAA, organizations now directly involved in designing or evaluating tests are the Army Research Institute, the Logistics Evaluation Agency, the Human Engineering Laboratory, the Ballistics Research Laboratory, other DARCOM laboratories, and the Directorate of Combat Developments in each TRADOC school and center. However, of the big three Army analysis organizations--the Concepts Analysis Agency, the TRADOC Systems Analysis Activity, and AMSAA--only AMSAA is significantly involved with designing and evaluating tests.

To perform the type of comprehensive operational evaluation needed by the decisionmakers is a challenge because of the many factors that must be considered in this type of assessment. One way to satisfy this need would be to make greater use of the analytical capabilities that now reside in the Army analysis organizations in planning and evaluating operational tests.

WHO SHOULD BE RESPONSIBLE FOR THE COMPREHENSIVE EVALUATIONS

We found substantial agreement on the need for more comprehensive evaluations to accompany operational testing. However, we found no such consensus on the best mechanism to direct, pull together, and integrate the evaluations and information prepared separately by the respective evaluation agencies.

In our discussions with numerous Army officials, four principal alternative proposals for assigning responsibility to prepare comprehensive evaluations emerged. They were to have the separate evaluations planned, directed and integrated by either TRADOC's Systems Analysis Activity, the Concepts Analysis Agency, AMSAA or OTEA. The advantages and disadvantages of each of these becoming the principal evaluator are discussed below.

TRADOC's Systems Analysis Activity As The Principal Evaluator

If TRADOC's Systems Analysis Activity were given the principal evaluator task, it would, under proposals we have heard, report to the Vice Chief of Staff or someone at that level instead of to TRADOC in order to satisfy the independence requirement. This option has the advantage of having both operational test evaluations and cost and operational effectiveness analyses done by the same agency, expediting a cross-fertilization of the two efforts. The Activity, however, has had only a limited amount of experience in planning, designing, and evaluating tests. Also, its remote location at White Sands might work against a close working relationship with OTEA and other test, analysis and technical organizations.

Concepts Analysis Agency As The Principal Evaluator

The Concepts Analysis Agency has the advantage of being located in the Washington area and of already reporting to the Director of the Army Staff. In the last few years, the Concepts Analysis Agency has increased its capability to analyze all aspects of forces e.g., operations, structure, logistics, manpower, and personnel, and to support with studies the various staff offices in Army headquarters, including those of the Deputy Chiefs for personnel, operations, logistics, and research, development and acquisition.

Because its efforts have been directed more at force level than system level analysis, its experience on particular systems would not be as great as the TRADOC Systems Analysis Activity's. Nevertheless, the Concepts Analysis Agency might bring a broader perspective into the planning, directing and pulling together of the comprehensive evaluations and risk assessments. The capabilities of the Concepts Analysis Agency to perform this function would be enhanced if some of the other Army analysis organizations, such as the Army Research Institute and the Logistics Evaluation agency were teamed with it to coordinate the comprehensive evaluations as well as continue with their other functions for the Army Staff. These two agencies have been involved with test evaluation of individual systems for some time.

AMSAA As The Principal Evaluator

AMSAA is probably the best qualified agency to pull together the overall evaluation because it is already experienced in doing evaluations of development tests. However, having it report to the level of the Vice Chief of Staff in the interest of enhancing its independence would probably remove it too far from the development community to have it still be effective in doing development test evaluations and its other analysis tasks for DARCOM. AMSAA might better be used in adding the technical risk assessments to its current responsibility to do development test evaluations.

OTEA as the principal evaluator

Having OTEA assume responsibility for the single comprehensive evaluation has the advantage of building on an existing independent organization, OTEA, which reports at a high level and is already the spokesperson for the test and evaluation community at the ASARC reviews.

However, OTEA's evaluators are military officers whose tours of duty at OTEA are usually just 3 years. Consequently, they are often not available for the full development cycle. Such a short tour is insufficient time to acquire the skills and gain the experience required to direct comprehensive evaluations.

Thus, if OTEA were to become the principal evaluator, some changes should be considered in its staffing. OTEA's ability to perform the comprehensive evaluation may hinge on its acquiring the essential staff continuity and greater technical and analytical expertise.

CHAPTER 4

CONCLUSIONS, RECOMMENDATIONS, AGENCY COMMENTS AND OUR EVALUATION

CONCLUSIONS

Army evaluations of weapon systems prepared for use in assessing a weapon's potential for combat have not been comprehensive enough to satisfy decisionmakers' needs. Their major drawback has been that they do not relate the significance of test findings to the weapon system's effectiveness and its supportability. Also, because these evaluations do not adequately present the operational and technical risks of proceeding to the next phase without correcting problems uncovered in testing, decisionmakers are forced to rely, principally, on the judgment of the project managers who tend to understate the gravity of performance problems and to emphasize the risks of delay.

Although the need for a better and more comprehensive evaluation is evident, the current Army test evaluation structure cannot satisfy it. The Army test evaluation community is too fragmented to provide a coherent and meaningful picture of the system's progress. While the testers and test evaluators are generally outspoken in their criticism of the systems, the fact that each organization involved looks at only a small part of the total picture precludes disclosing the collective effect of what are often serious overall deficiencies. Other Army organizations which perform analytical evaluations, such as cost and operational effectiveness analyses, generally do not use actual test results in their evaluations. The test evaluators and the analysts need to integrate and interpret their findings and speak with one strong voice, especially at the decision reviews, to provide a better balance between their views and those of the user and the developer.

The importance of the evaluations, and the widespread dissatisfaction with their present usefulness, warrant strengthening the present structure to emphasize and upgrade the planning, execution, and presentation of these evaluations. While it would be unrealistic to concentrate in one organization all the expertise required for a truly comprehensive evaluation of each new weapon system, one principal evaluation agency should be designated that would have access to the information prepared separately by the respective evaluation agencies and be knowledgeable enough to interpret this information, integrate it into one comprehensive evaluation, and provide it to the decisionmakers. Any of the three main Army analysis organizations - TRADOC's Systems Analysis Activity, the Concepts

Analysis Agency, or AMSAA - by virtue of the background and experience of their personnel, appear to be logical choices to assume this responsibility.

The analysis organizations should be able to contribute significantly to planning tests and evaluating the test results, particularly in identifying critical issues to be examined and in analyzing the impact of emerging problems on operational effectiveness, cost, supportability and suitability. In this arrangement we see OTEA retaining its function of managing the conduct of the operational tests.

RECOMMENDATIONS

Regardless of which agency is designated by the Army as principal evaluator, we recommend that this organization

--seek the ASARC decisionmakers' concurrence on critical issues that are to be evaluated for each proposed weapon system;

--transmit these issues to the pertinent Army test evaluation organizations, and the analysis organizations, so that they can be addressed in the test reports and analyses which they will be providing for inclusion in the overall evaluation of new systems;

--prepare, for critical major acquisition decision points, a comprehensive evaluation of the weapon system under review by integrating the reports and analyses of the individual test evaluation and analysis organizations, and,

--assume the role of spokesman for the test evaluation and analysis organizations at the ASARC and before other decisionmaking bodies.

To lend more objectivity to the formulation of technical risk assessments of major systems we further recommend that AMSAA, being the evaluator for development testing, also be made responsible for providing these assessments to the decisionmakers.

AGENCY COMMENTS AND OUR EVALUATION

DOD concurred with our findings on the need for more comprehensive evaluations.

In a draft of this report, we proposed that one of the three principal Army analysis organizations assume responsibility for coordinating the comprehensive evaluation's preparation. DOD did not agree. While concurring with us on the need for more involvement by the Army's analytic community, DOD stated that the Army has begun taking steps to have OTEA, rather than one of the analysis organizations, fulfill this role. The Army proposes to increase OTEA's staff with individuals who could give OTEA increased analytical support. The Army will initiate a pilot program early in 1984, to provide a comprehensive evaluation of five weapon systems, using many data sources in addition to the traditional test results. The results will be evaluated in mid-1986, when a decision to modify or to fully implement the concept will be reached. If proven successful during the pilot program, OTEA will receive additional staff when the concept is fully implemented. The staff will have the skills and expertise necessary to prepare a comprehensive evaluation. We believe the pilot program is an acceptable aproach to address the issues identified during our review. We have deleted this proposed recommendation pending the results of the Army's pilot program.

If, following the pilot program's completion, OTEA continues in the role of principal evaluator we believe the Army should consider staffing OTEA's management and evaluator positions with civilian rather than military personnel. If this is not done it may perpetuate the disruption in the test evaluation caused by frequent staff rotations. It would also prevent OTEA's evaluators from acquiring the expertise and experience that comes with uninterrupted service. Staffing the management and evaluator positions with civilians would not require changing the military nature of OTEA's command structure.

DOD agreed with our recommendation to have AMSAA furnish assessments at decision reviews beginning in fiscal year 1984 on a system-by-system basis.

DOD's comments appear in Appendix II.

SOME PREVIOUS GAO REPORTS ON TESTING

Date	Title/subject	Report number
03/29/78	Navy Operational Test and Evaluation: A Valuable Tool Not Fully Utilized (CONFIDENTIAL)	PSAD-78-77
06/02/78	Operational Testing of Air Force Systems Requires Several Improvements (UNCLASSIFIED)	PSAD-78-102
07/25/78	Letter Report to the Secretary of Defense on operational test and evaluation of foreign built systems (UNCLASSIFIED)	PSAD-78-131
10/19/78	Letter Report to the Secretary of Defense on follow-on operational test and evaluation (UNCLASSIFIED)	PSAD-79-1
03/08/79	Need for More Accurate Weapon System Test Results to be Reported to the Congress (SECRET)	PSAD-79-46
06/25/79	Letter Report to the Secretary of Defense on development test and evaluation of six systems (CONFIDENTIAL)	PSAD-79-86
11/13/79	Army Operational Test and Evaluation Needs Improvement (CONFIDENTIAL)	C-PSAD-80-2
08/06/82	Letter Report to the Secretary of Defense on use of the design for testability concept in the development and acquisition of major weapon systems (CONFIDENTIAL)	GAO/C-MASAD-82-38
06/23/83	Better Planning and Management of Threat Simulators and Aerial Targets Is Crucial to Effective Weapon System Performance (UNCLASSIFIED)	GAO/MASAD-83-27

APPENDIX II

OFFICE OF THE UNDER SECRETARY OF DEFENSE

WASHINGTON D C 20301

10 NOV 1983

RESEARCH AND
ENGINEERING

Mr. Frank Conahan, Director
National Security and
International Affairs Division
U.S. General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Conahan:

This is the Department of Defense (DoD) response to your draft report, "The Army Needs More Comprehensive Evaluations to Make Effective Use of Its Weapon System Testing" dated September 13, 1983 (GAO Code 951727/OSD Case No. 6353).

The Department considers the report to be thorough and well presented. Steps are already being taken to align the Army evaluation procedure with the findings and recommendations. DoD considers the logical agency to coordinate the various inputs to be the Operational Test and Evaluation Agency. The rationale for this decision is found in DoD's detailed comments set forth in the enclosure. The opportunity to comment on this draft report is appreciated.

Sincerely,

James P. Wade, Jr.
Principal Deputy Under Secretary
of Defense for Research & Engineering

Enclosure

APPENDIX II

GAO DRAFT REPORT - DATED SEPTEMBER 1983
(GAO CODE NO. 951727) - OSD CASE NO. 6353

"THE ARMY NEEDS MORE COMPREHENSIVE EVALUATIONS TO MAKE EFFECTIVE USE OF ITS WEAPON SYSTEM TESTING"

DEPARTMENT OF DEFENSE COMMENTS

* * * * *

FINDINGS

- **FINDING A:** Fragmented Army Test Evaluations Must Be Overcome to Enable Better Assessment. GAO found that over time, concerns about the test and evaluation process have led to a dispersed Army organizational structure for test and evaluation. Additionally, GAO found that the various evaluations generated by the different organizations are not integrated into a single comprehensive evaluation for a given system. GAO further found that communication and coordination between test evaluation organizations is limited. While noting that Army efforts in test coordination are beneficial in reducing duplication, scheduling resources, and in scoring test results, GAO found, that no similar efforts are made to coordinate the test evaluations of the various organizations. GAO concluded that the Army test evaluation is too fragmented to provide a coherent and meaningful picture of a system's progress. (pp. 7-11, and p. 23)

  [See GAO note, p. 33]

  DoD COMMENT: Concurs

- **FINDING B:** Test Evaluations Must Become Broader and Focus on Key Issues. GAO found that most decisionmakers it interviewed are dissatisfied with the reporting and scope of the test evaluations provided. GAO further found that these decisionmakers would like evaluations broadened to include specific critical issues, and include an assessment of risks inherent in the program, and in alternative courses of action. In this connection, GAO found that generally the only formal risk assessments presented to the Army Systems Acquisition Review Council (ASARC) principals are performed or coordinated by the project manager and focus on affordability, cost and schedule risks, but seldom emphasize the technical and operational risks involved in the program. GAO concluded that Army evaluations of weapons systems have not been comprehensive enough to satisfy the decisionmakers' needs. GAO further concluded that the major drawback of these evaluations has been that they do not relate the significance of test findings to the weapon systems effectiveness and supportability. GAO also concluded that, because these evaluations do not adequately present the operational and technical risks of proceeding to the next

phase without correcting problems uncovered in testing, the decisionmakers are forced to rely principally on the judgement of the project managers. Finally, GAO concluded that project managers tend to understate the gravity of performance problems and to emphasize the risks of delay. (pp. 11-15 and 23) [See GAO note, p. 33]

DoD COMMENT: Concurs

- **FINDING C: Future Test Evaluation Must Depend More On Analytical Modeling.** GAO found that the period of greatest opportunity to affect operational performance, reliability, maintainability and the ultimate life cycle cost is in the early design phase. Further, GAO found that these evaluations, according to senior Army officials, should periodically assess equipment performance in the changing operating environment. GAO concluded that if operational evaluations of complex systems are to be performed early enough to support decisionmaking, results of physical testing must be complemented by results of simulations, analysis and wargaming. GAO also concluded that the complexity of weapon systems requires a change in the frequency of test evaluations and the user must be involved on an ongoing basis--i.e., that it is nearly impossible to establish so far in advance a realistic system requirement that recognizes technological bounds and specific equipment performance in a future operating environment. Finally, GAO concluded that greater involvement by the Army's analysis organizations in test and evaluation could produce the type of comprehensive weapon system evaluation needed for acquisition decisions. (pp. 16-19) [See GAO note, p. 33]

DoD COMMENT: Concurs

- **FINDING D: Army Analysis Organizations Should Be Linked To Test Evaluation.** GAO found that the Army development test evaluation has benefited from assigning the evaluation function to the Army Materiel Systems Analysis Activity (AMSAA), an established analysis organization. GAO concluded, based on the good results that the Army obtained in testing and evaluation by assigning AMSAA the independent evaluation of development tests, that similar action is necessary to improve the operational test evaluations. GAO further concluded that to perform the type of comprehensive operational evaluation needed by decisionmakers is a challenge because of the many factors that must be considered in this type of assessment and that one way to satisfy this need would be to directly involve established Army analysis organizations in that planning and evaluation. (pp. 20-21) [See GAO note, p. 33]

DoD COMMENT: Concurs

- **FINDING E: What The Analysis Agencies Can Contribute To Testing And Evaluation.** GAO found that by finding the factors to which the system is the most sensitive, the analysts can help in defining critical issues and assist the test evaluators with designing tests and evaluations to highlight these factors. GAO further found that if these factors are difficult, expensive or even impossible to include in tests, the analysts can also help in finding ways to simulate these factors either in the tests or in separate experiments, computer models, or studies of combat data. GAO concluded that probably the most important contribution the analysts can make to test evaluation is to run the test results through their models and determine the implications of these results to the systems operational effectiveness and supportability. GAO further concluded that the analysis organizations should be able to contribute significantly to planning tests and evaluating the test results, particularly in identifying critical issues to be examined and in analyzing the impact of emerging problems on operational effectiveness, cost, supportability and suitability. (pp. 20-21) [See GAO note, p. 33]

DoD COMMENT: Concurs

- **FINDING F: Who Should Be Responsible For Coordinating The Comprehensive Evaluations.** Noting there is substantial agreement on the need for more comprehensive evaluations to accompany testing, GAO found there emerged five principal alternative proposals for organizing to coordinate comprehensive evaluations. These were: OTEA (with TRADOC assuming OTEA's current testing function), TRADOC's Systems Analysis Activity (reporting at Vice Chief of Staff level), Concepts Analysis Agency, AMSAA (reporting at Vice Chief of Staff level) and a working group drawn from existing test evaluation and analysis organizations. While recognizing it would be unrealistic to concentrate in one organization all the expertise required for a truly comprehensive evaluation of each new weapon system, GAO concluded that one principal operation evaluator should be designated within the Army with access to all information available throughout the Army. GAO further concluded (1) that any of the three main Army analysis organizations appear to be logical choices to manage test planning and evaluation, (2) that this organization should report directly to the Army Staff and (3) that the Army Office of Test and Evaluation (OTEA) should regain its function of managing the conduct of the operational tests. (pp. 21-22) [See GAO note, p. 33]

DoD COMMENT: Concurs with need but feels OTEA is the agency to coalesce evaluations and make report (see discussion under Recommendation 1).

RECOMMENDATIONS

- **RECOMMENDATION 1.** GAO recommended that the Secretary of the Army designate one of the major analysis organizations--TRADOC's Systems Analysis Activity, the concepts Analysis Agency, or AMSAA--to be the Army's principal evaluator of major weapons systems. GAO further recommended that as principal evaluator, the functions of this organization should be to:

  -- identify the critical issues that are to be tested and evaluated for each proposed weapon system, based on prior consultation with the ASARC decisionmakers;

  -- transmit those issue to the pertinent Army test evaluation organizations, and the analysis organizations, so that they can be addressed in the test reports and analysis which they will be providing for inclusion in the overall evaluation of new system;

  -- prepare, for critical major acquisition decision points, a comprehensive evaluation of the weapon system under review by integrating the efforts of the individual test evaluation and analysis organization, and

  -- Assume the role of spokesman for the test evaluation and analysis organizations at the ASARC and before other decision making bodies. (p. 25)
  [See GAO note, p. 33]

DoD COMMENT: DoD concurs with the need to centralize the evaluation function but does not agree with assigning the comprehensive evaluation mission to an Army analytical agency. Instead the Army has assigned the mission of continuing and comprehensive evaluation to the Operational Test and Evaluation Agency (OTEA) with increased analytical support expected from the GAO recommended agencies.

The concept of continuous evaluation calls for a broad analytic approach to the evaluation of developmental systems extending from earliest concept formulation through initial fielding. OTEA will be able to give an assessment of the system's status in the development cycle at any time and report significant status changes based on best evidence to date to the Vice Chief of Staff of the Army and other decision makers.

A system's assessment involves multiple data sources. It is a synthesis of requirements analysis, studies, tactical and logistical modeling, surrogate and mock-up testing, contractor tests, DT, OT, and FDTE, and post-fielding data surveys, into a single plan of evaluation to provide comprehensive assessment of a developing system's ability to meet the stated need:

a. at its current state of development

b. in a mature configuration

The Test, Evaluation and Analysis Master Plan (TEAMP) prepared by OTEA will be the management document used to coordinate this assessment. OTEA would utilize inputs from program managers, AMSAA, CAA, LEA, and TRADOC combat development activities, including TRASANA, to conduct this comprehensive evaluation.

The extension of this concept, to include data sources outside the realm of classic DT/OT, significantly broadens the base of analysis. It facilitates a continuous interaction between the material developer and the user representative, with the evaluator serving as a catalyst by reporting on trends taking into account the maturity of the system and its concepts.

Five candidate systems have been selected and a pilot program has been using his concept. It is envisioned this concept will require some internal reorganization within OTEA as well as an increase in the number of personnel authorized. OTEA will continue to have the mission of managing operational testing. A modest increase in the authorized strength for supporting analytical organizations may also be required to accomplish the additional workload. The pilot program will begin early in CY 84 and its status will be reviewed semi-annually until a decision to modify or to fully implement the concept is reached. A major review is planned for mid-1986.

DoD is in agreement with assigning major systems technical risk assessment to the Army Materiel Systems Analysis Activity (AMSAA). AMSAA already performs most of this function in assisting the program managers in accomplishing the Best Technical Approach. The AMSAA function will be expanded to produce a report at Milestone I and II that will be provided separately to decision reviews and makers. The process will begin in early FY 1984 on a system-by-system basis.

- **RECOMMENDATION 2.** GAO recommended that to lend more objectivity to the formulation of technical risk assessments of major systems, that AMSAA, being the evaluator for development testing, also be made responsible for providing these assessments to the decisionmakers. (p. 34, GAO Draft Report)

DoD COMMENT: Concurs

GAO note: Page references in this appendix have been changed to correspond to those in the final report.

(951727)