REPORT TO THE CONGRESS

The Importance Of Testing And Evaluation In The Acquisition Process For Major Weapon Systems

B-163058

Department of Defense

BY THE COMPTROLLER GENERAL OF THE UNITED STATES

AUG. 7.1972

COMPTROLLER GENERAL OF THE UNITED STATES
WASHINGTON, D.C. 20548

B-163058

To the President of the Senate and the
Speaker of the House of Representatives

This is our report on the importance of testing and evaluation by the Department of Defense in the acquisition process for major weapon systems.

Our review was made pursuant to the Budget and Accounting Act of 1921 (31 U.S.C. 53), and the Accounting and Auditing Act of 1950 (U.S.C. 67).

Copies of this report are being sent to the Director, Office of Management and Budget; the Secretary of Defense; and the Secretaries of the Army, Navy, and Air Force.

Elmer B. Staats

Comptroller General
of the United States

CONTENTS

ABBREVIATIONS

AWACS	Airborne Warning and Control System
DOD	Department of Defense
DSARC	Defense System Acquisition Review Council
GAO	General Accounting Office
OSD	Office of the Secretary of Defense
SRAM	Short-Range Attack Missile

COMPTROLLER GENERAL'S
REPORT TO THE CONGRESS

THE IMPORTANCE OF TESTING AND EVALUATION IN THE ACQUISITION PROCESS FOR MAJOR WEAPON SYSTEMS
Department of Defense
B-163058

D I G E S T

WHY THE REVIEW WAS MADE

Testing new weapons is one of the Department of Defense's (DOD's) key controls in the complex process of acquiring today's multibillion dollar systems. Testing at decisive stages of development shows where problems exist and helps military managers to make sounder decisions affecting future production and purchase of weapons than would otherwise be possible. Therefore the General Accounting Office (GAO) has reviewed the policies and practices of the military services in testing and evaluating weapon systems.

Background

GAO examined systems for which substantial testing history was available. Immediately prior to the start of this review, DOD was changing its policies for the acquisition process. These new policies, which are basically sound and in various stages of implementation, are being applied primarily to new systems entering the acquisition process. Test cases cited in this report therefore are not fully representative of current policies.

There are three basic categories of testing and evaluation.

Engineering testing to demonstrate physically, before a weapon system is accepted for production, that the system will perform as intended.

Acceptance testing to demonstrate that the state and quality of the system fulfill the legal and/or commercial requirements agreed to by the seller and the buyer.

Operational suitability testing to demonstrate that the weapon system, the operating personnel, and the tactical operations can work together to accomplish an established combat mission.

For a description of these basic elements, see pages 5 to 8.

The benefit of testing is accomplished through properly assessing risks and delivering test results to the decisionmaker at key points in the acquisition cycle when final decisions must be made. A breakdown in performing any of the testing steps would lead to a lack of timely, accurate, or complete information, which undoubtedly would handicap the decisionmaker.

AUG. 7,1972

FINDINGS AND CONCLUSIONS

GAO reviewed 13 weapon systems with estimated total costs of more than $46 billion. They include such weapons as the Army's Improved HAWK missile, the Navy's DE-1052 (destroyer escort), and the Air Force's F-15 aircraft.

A list of these weapons and essential information on them are in appendix I.

On the basis of its observations of the pattern of testing performance, GAO concluded that in DOD:

--Practices used to establish objectives for testing generally were adequate.

--Most weapon systems did not have adequate plans for conducting tests.

--Testing and evaluation for most weapon systems was not accomplished in a timely manner.

--Most test reports were adequate, but their value was diminished due to inadequate test planning and actual testing. Some reporting improvements could be made.

--Complete and valid test and evaluation data was not available prior to those times in the acquisition cycle when decisions had to be made. (See ch. 5.)

Each of the three services has longstanding policies that essentially require the completion of engineering testing before production begins. These policies have been waived frequently. For instance, the Department of the Army has such a policy but it also provides for waiving the policy and beginning limited production when the need is urgent, when the risk is low, and when no other system satisfies the requirement.

Most, if not all, of the major weapons procured by the Army in recent years have been procured under this waiver. Similarly, the Mark 48, the F-111, and a number of other weapon systems in the Navy and Air Force have entered production under waivers to the basic policy.

Other examples of the varying patterns of testing performance are in chapter 2.

RECOMMENDATIONS OR SUGGESTIONS

Several key areas of testing and evaluation need continued attention and control. In implementing its new policies and practices regarding testing and evaluation, DOD should continue to emphasize the need for:

--Completion of appropriate testing and evaluation prior to key decision points in the acquisition cycle.

--Adequate controls over the granting of waivers from required testing and evaluation.

--Succinct summary reports to be prepared by the testing agency for all levels of management. (Interested management levels may wish to comment on these summary reports; however, they should not be permitted to change the basic summaries.)

AGENCY ACTIONS AND UNRESOLVED ISSUES

The Director of Defense Research and Engineering stated that the implementation of policies on weapon system acquisition issued by the Office of the Secretary of Defense since May 1970 would correct the deficiencies in testing and evaluation disclosed in this report. The Director stated that these recent policies are now being implemented but cautioned that the process of changing takes time. GAO is currently conducting a review of the new testing and evaluation policies.

He also mentioned that there are many programs which are well advanced and which cannot be completely transferred to the new testing policies at this time due to contracts or other binding agreements; however, these programs are being modified to the extent practical. For the text of the Director's response, see appendix IV.

MATTERS FOR CONSIDERATION BY THE CONGRESS

This report provides information to the Congress on the status of testing and evaluation of weapon systems currently being acquired.

BLANK

CHAPTER 1

INTRODUCTION TO TESTING AND EVALUATION

The General Accounting Office (GAO) has reviewed the military departments' policies and practices in performing tests and evaluations of weapon systems during the various stages of acquisition. Testing is a management technique for controlling activities to ascertain and minimize risk. In this report testing means testing and evaluation in its broadest context throughout the entire acquisition cycle and includes the purposes, the types, the timing, and the performers of the tests.

BASIC TESTING ELEMENTS

There are only three basic categories of testing and evaluation.

Engineering testing to physically demonstrate, before a system is accepted for production, that it will perform as intended.

Acceptance testing to demonstrate that the state and quality of the weapon system fulfill the legal and/or commercial requirements agreed to by the supplier and the customer.

Operational suitability testing to demonstrate that the weapon system, the operating personnel, and the tactical operations can work together to accomplish an established combat mission.

A brief description of the basic elements of these three types of testing is presented below.

Engineering testing

What--All scientific and objective testing done for experimenting with, developing, and proofing a system or its parts.

Where--Performed under controlled conditions to properly assess the physical properties and

characteristics of the item being tested, usually in such places as laboratories, wind tunnels, environmental facilities, and ranges.

Why--Experimentation to physically demonstrate that the item will perform as intended.

When--All testing performed before accepting a system (or item) for production, including testing a part, subcomponent, subsystem, or entire system. In concept formulation, it may involve brass board or model testing. In validation, it may involve mockups of major subassemblies or models. In development, it may involve various types or degrees of prototyping.

By whom--Supervised by the developer; performed by a Government laboratory or contractor; and carried out by scientists, engineers, and technical experts.

How much--An iterative process performed until success is achieved or until the item is discarded. Practicality dictates that success be defined in specific terms, as to both quantity and quality (tolerances).

Acceptance testing

What--That testing done to demonstrate that the State and quality of the item either fulfill the legal and/or commercial requirements agreed to by both the supplier and the customer or are otherwise satisfactory and acceptable to both parties.

Where--Where practicall/ feasible, it is performed on the entire system, i.e., airplane, ship, tank, or missile. As prudence dictates, some testing is performed prior to final assembly (quality-control-type testing during production) under controlled circumstances to technically demonstrate the presence of the contracted state and quality of the system.

Why--That official, arm's-length process by which a customer assures himself that the supplier has, in fact, provided what was previously agreed upon. Unqualified acceptance leaves little or no recourse to a customer from the supplier. In addition, acceptance testing is used to verify the compatibility of subsystems.

When--Usually acceptance testing occurs when an item is selected for production, when it is produced, and before it is deployed.

By whom--The military material command expert personnel or an independent chartered organization would perform acceptance testing at selection and upon completion of production. Upon deployment the using command would also perform acceptance testing when it receives the system from the military material command.

How much--Each item accepted should be tested by the receiving party to the extent necessary to assure itself of the state and quality of the item received. Practicality may justify use of scientific sampling techniques in given situations. Intensive testing of initial items, coupled with adequate contractual guarantee arrangements, may also be used to reduce costs of testing.

Operational suitability testing

What--Testing done to demonstrate that the weapon system can perform the mission as part of an integrated combat operation.

Where--Performed in the field under simulated or actual combat conditions.

Why--To gain assurance that the weapon system, the operating personnel, and the tactical operations can work together to accomplish the mission of the new weapon.

When--Testing should start, if feasible, as soon as the total weapon is first assembled during development as a preproduction prototype or when the first production models are available before full-scale production and at deployment.

By whom--Performed by the using command, i.e., the military personnel who will be operating the weapon.

How much--Testing must continue until an acceptable weapon is established and proven. User must be satisified that weapon fulfills its needs.

TEST MODEL FOR ACQUISITION CYCLE

We are concerned with testing used in the acquisition process. The Department of Defense (DOD) lists five phases in the acquisition process of weapon systems: (1) concept formulation, (2) validation and ratification, (3) development, (4) production, and (5) deployment.

We have formulated a model (see app. II) to depict the role of testing in the acquisition cycle. Our model shows (1) the acquisition phase, (2) the critical decision points for the Defense System Acquisition Review Council (DSARC), (3) testing criteria, (4) responsible parties, and (5) basic testing categories.

The purpose of our model is to reinforce and emphasize certain ideal concepts of testing and evaluation in the acquisition process, as follows:

--Testing and evaluation is an important ingredient throughout the acquisition process.

--The sequential nature of engineering; acceptance; and, to some degree, operational suitability testing and evaluation.

--The responsibility for the success of testing and evaluation throughout the acquisition process lies with the developer, the user, and the contractor in different degrees and at different times.

--To provide a means for DOD and weapon systems program managers to evaluate the testing and evaluation process in their respective programs.

REVIEW METHOD

We established specific overall objectives to assess testing; utilized the aforementioned model (or logical plan) for testing in the acquisition process; and utilized a case study approach of recent, current, and planned testing practices for selected weapon systems. The model, together with basic testing elements, was compared with the testing practices employed in acquiring the selected systems. The results of this comparison, coupled with the results of our specific assessments, provided the basis for developing common threads of good and bad testing practices used by DOD.

SCOPE

The specific objectives of this review were to assess:

--The practices followed in developing test objectives and in relating them to the mission objectives of the system.

--The development of test plans and the means utilized to insure that they would accomplish the test objectives.

--The extent to which tests were accomplished timely and effectively.

--The relationship of test results to mission objectives and the conclusions reported.

--The extent to which test results were effectively utilized in making key management decisions.

A list of the 13 specific weapon systems selected for case studies is shown in appendix I, together with the service, type of system, acquisition phase, and total estimated cost through completion, as of June 30, 1971.

We conducted a review of the pertinent policies, procedures, records, and practices involved in each weapon system, as appropriate, at the involved offices in the Office of the Secretary of Defense (OSD), service headquarters, material commands, program management offices, subordinate commands, and contractors' sites.

Our field review of individual systems was made between February and July 1971.

We examined 13 systems for which there was substantial testing history available, recognizing that DOD's policies for the acquisition process were changing immediately prior to and during our review. These new policies, which are in various stages of implementation, are being applied primarily to new systems entering the acquisition process. Therefore the test histories studied are not fully representative of current policies; however, the results of our review should contribute to improvements in testing and evaluation practices. We plan to further review this subject area as the new policies are implemented.

CHAPTER 2

THE PATTERN OF TESTING PERFORMANCE

The present DOD policies recognize the need for, and the importance of, testing in the decisionmaking process. Comprehensive test standards have existed in the three services for a number of years and have emphasized the necessity for engineering and operational suitability testing in the decisionmaking process. Any comprehensive test standard incorporates a sequential testing process. The test standard must:

--Formulate test objectives to satisfy the mission objectives of the weapon.

--Develop test plans to accomplish test objectives.

--Implement testing on the basis of test plans.

--Evaluate test results and prepare test reports.

--Use test results when making key management decisions.

We have used this normal progression of a weapon system test program to provide a basis for evaluating the effectiveness of the test programs for 13 weapon systems.

The successful completion of a test program involves not only the conduct and evaluation of engineering, acceptance, and operational suitability tests but also the planning, recording, and reporting efforts which precede and follow it. But the real benefit of testing is in properly assessing the risk and in delivering test results to the decisionmaker at key decision points in the acquisition cycle. A breakdown in performing any of the testing steps would lead to a lack of timely, accurate, or complete information, which undoubtedly would handicap the decisionmaker.

Each of the five steps essential to good testing is discussed below, followed by an example of adequate and inadequate application of testing criteria. Although each example is based on an evaluation of the testing practices for a particular weapon system, it is not the purpose of this report

to focus on any particular weapon. The examples cited illustrate that there is good and bad application of testing criteria. The trends we found in testing are shown in the conclusions and recommendations section of this report. (See ch. 5.)

DEVELOPMENT OF TEST OBJECTIVES

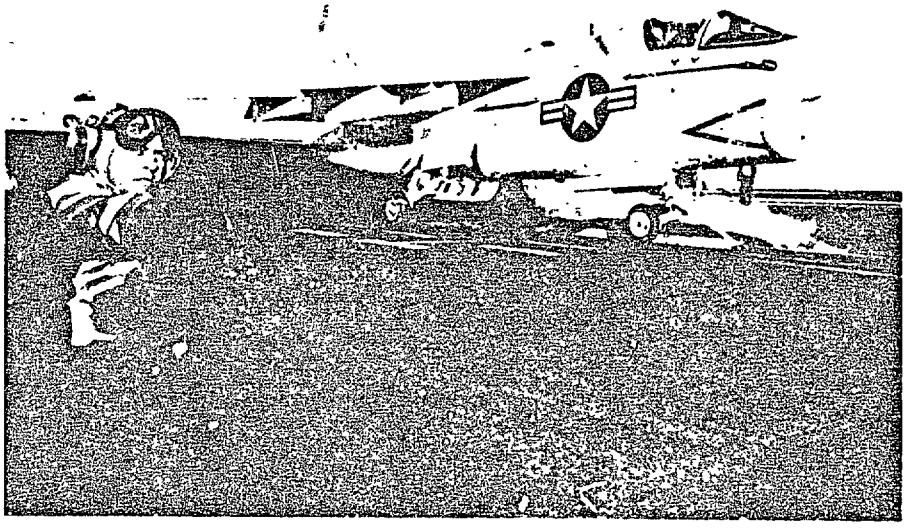
Testing objectives providing the framework for measuring a new weapon's capability to meet its mission requirements should be established early in the acquisition program. Adequate test objectives provide for engineering testing to physically demonstrate that the weapon system will perform as intended, acceptance testing to assure the customer that the supplier has in fact provided the product previously agreed upon, and operational suitability testing to insure that the weapon system and the personnel who operate it can work together to accomplish the mission.

Authorizing a weapon to progress in the acquisition cycle without establishing adequate testing objectives can result in deploying a weapon system which falls short of meeting its mission requirements and which no doubt will require an expensive retrofit program. On the other hand, positive identification of hardware problems through testing will permit the program managers to request the needed money and resources to resolve the problem in its infancy.

Following are examples of satisfactory and unsatisfactory development of test objectives.

Adequate application of criteria

A-7D/E aircraft

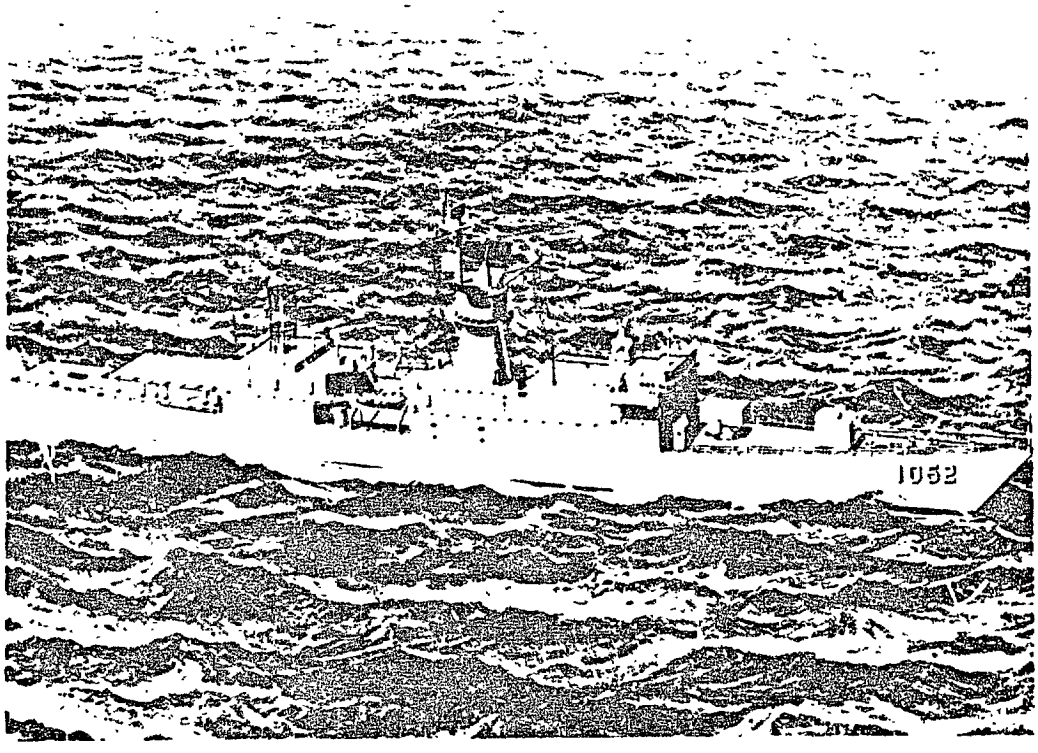


Broad test objectives for the A-7D/E aircraft were established by military specification MIL-D-3708A, (WEP) dated September 1960, pertaining to demonstration requirements for airplanes. These test objectives were further defined in addendums to the military specification prepared by the prime contractor and in test plans prepared by the Air Force and the Navy. The objectives were designed to test the aircraft's performance characteristics prescribed by the contract detail specifications. These specifications, in turn, relate to the A-7's mission of close ground support and interdiction.

Some of the objectives for the engineering, acceptance, and operational suitability testing for the A-7D/E aircraft were (1) to insure that established requirements were met, (2) to determine the system's actual performance capabilities, (3) to identify deficiencies in time for changes to be incorporated before significant production buildup, (4) to evaluate the safety and reliability of the system, and (5) to assess the manpower needed to support the system.

Inadequate application of criteria

DE-1052



The DE-1052 ships are not being delivered within the time and cost constraints originally planned by the Navy. In addition, they are being delivered without satisfactory equipment and thus without the complete capability planned for them.

A contributing factor to these conditions is that the ships' original testing objectives were directed toward engineering (functional) testing of equipment on the ships but not toward operational suitability testing of the integrated weapon systems' (the ships') ability to perform a specific mission.

Later the Navy's test objectives were changed to include an evaluation of the ships' ability to perform their mission. We were advised by Navy officials that the DE-1052 ships would undergo an operational appraisal to determine, by empirical methods, how well the ships performed their assigned mission, designed tasks, and contingent tasks. However, by the time the operational suitability tests will be performed, many DE-1052 ships will have been manufactured and the value of the testing will have been greatly reduced. Deficiencies disclosed by this type of testing should be corrected early in the production cycle of the acquisition process.

A Navy official stated that had an operational evaluation been included in the test program, it is unlikely that it could have been performed earlier, due to the fleet's need for the ships. He also said that such tests would have delayed the construction program.

DEVELOPMENT OF TEST PLANS

Adequate test plans must be made to insure that the test objectives will be accomplished. Good management practices dictate that, over the cycle of the three types of testing--engineering, acceptance, and operational suitability--adequate test and evaluation planning have these general requirements.

1. Flexible milestones requiring that a weapon meet certain requirements before it can move to a more advanced phase in the acquisition process.

2. Proper timing in the acquisition process so that decisionmakers are provided with testing results at important points in the program.

3. Realism in the testing environment so that there is assurance that the weapon will meet its intended purpose.

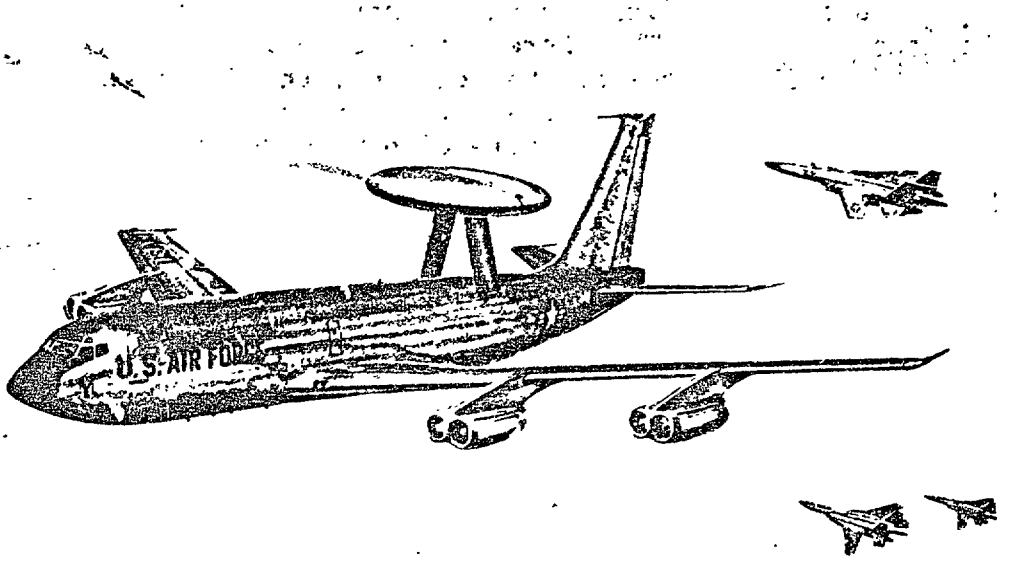
4. Sufficient test weapons to allow for failures and retests.

An adequate test plan is an essential part of a test program. Without an adequate plan it is unlikely that testing will be performed completely or on time.

The following examples illustrate good and bad development of test plans.

Adequate application of criteria

Airborne Warning and Control System (AWACS)[1]



The AWACS program demonstrates adequate test planning, particularly in the area of flexible milestones, by providing, by contract, that certain milestones be met. Until the contractor demonstrates through engineering tests that development of the system has passed certain technical milestones, the Government may delay actions it must take, such as alloting funds or exercising options for additional aircraft. Essentially, at the conclusion of certain engineering testing, the Government has the option either to continue the program with the successful radar or to cancel the entire program. If the demonstration shows unfavorable

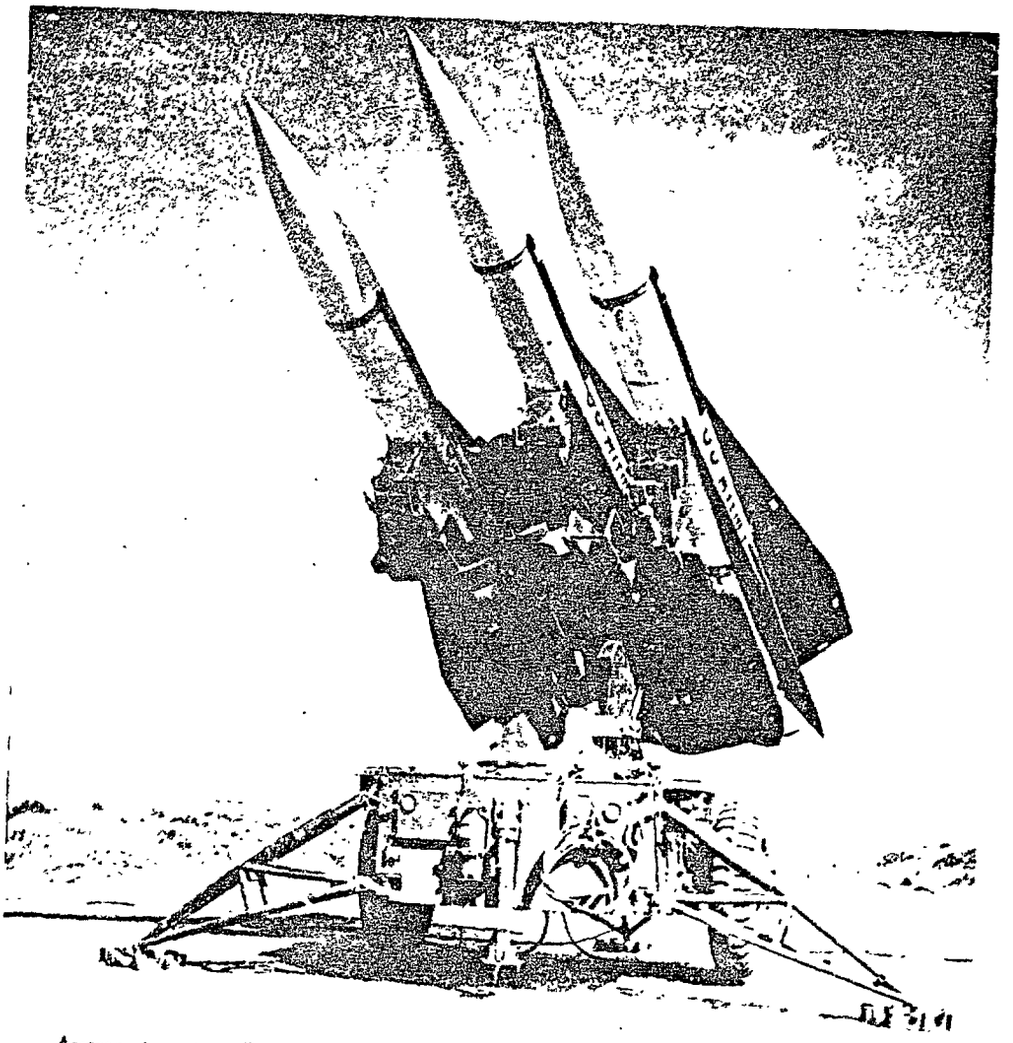
[1]We did not specifically review the testing program for AWACS; however, information about its test plan was included in other recent GAO reviews.

progress, the Government may either require additional development and testing on a cost-reimbursement basis or terminate the program.

The Government has retained the flexibility to defer indefinitely the making of further binding contractual commitments until significant progress has been made in the development of AWACS. Therefore, if trouble is encountered during the early phases of the contract, the Government may elect to continue working toward resolution of the problems encountered without the pressure of having to meet a contractually set schedule date.

Inadequate application of criteria

Improved HAWK missile



Army test plans for the HAWK included engineering tests to determine the technical, performance, reliability, maintainability, endurance, and safety characteristics of the system and included operational suitability tests to determine the fitness of the item for Army use.

These test plans set forth a specific number of test objectives and allowed only one missile for each objective. This planning was not realistic because, by the time that two-thirds of the test missiles were fired, only about 20 percent of the planned testing had been completed. In addition, most of the test missiles fired were of different configurations and this diluted the value of the tests.

The Army planned that the improved HAWK would have a high degree of development and production concurrency. This limited the amount of time available for the testing and evaluation of the missile prior to the start of production.

IMPLEMENTATION OF TESTING

Timely and effective tests to measure the development progress of a new weapon are dependent, in large part, on the adequacy of the test plans; that is, if the test plan is inadequate, chances are that the implementation of the tests will also be inadequate.

Basically the same criteria that apply to good test planning apply to test accomplishment. Testing must be timely; that is, key tests must be completed before a system can move to a more advanced phase of the acquisition cycle. If a weapon is moved to a more advanced phase without the completion of key tests, the risk of fielding an unsatisfactory weapon is greatly increased.

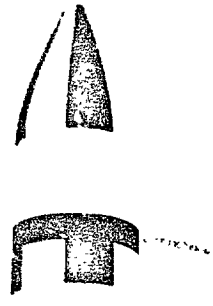
Engineering testing must be performed before accepting a system for production. This type of testing is done to determine whether a system meets specified performance characteristics, such as speed, range, and altitude.

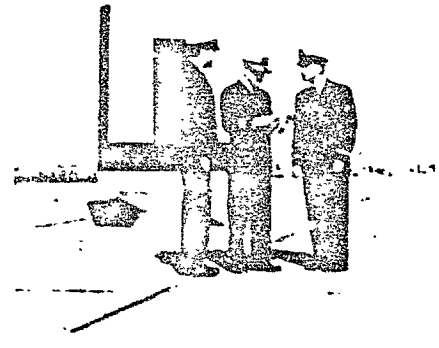
Acceptance testing must be performed at selection and upon completion of initial production by the military command expert personnel. Upon deployment the using command also must perform acceptance testing when it receives the system. This type of testing is necessary to insure that the contractor has, in fact, provided the weapon system previously agreed upon.

Operational suitability testing must start, if feasible, as soon as the total weapon is first assembled during development or when the first production models are available before full-scale production. This testing must be performed by the using command under simulated combat conditions to insure that the weapon system and the men who operate it can work together to accomplish a planned mission.

Adequate application of criteria

POSEIDON missile system



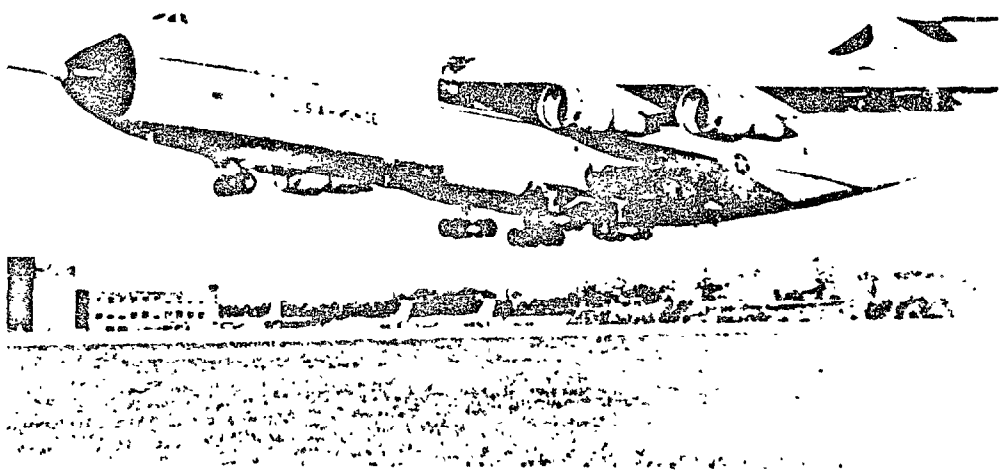


We evaluated the engineering and operational suitability testing for the POSEIDON program. The engineering tests were conducted during the research and development phase to verify that the missile met its development objectives. After these objectives were demonstrated, several pilot production missiles were flight-tested to demonstrate their performance characteristics. During the engineering test program, the tests were conducted by scientific personnel rather than by sailors.

As the POSEIDON progresses through the test program, the test environment is to be more operationally realistic. The operational suitability testing is to be conducted by sailors under simulated combat conditions to insure that the men and the system can work together to accomplish the mission. Some operational suitability testing has been performed to determine how well the total weapon system functions. The testing program requires continuous operational suitability testing throughout the life of the system.

Inadequate application of criteria

C-5A aircraft



To insure an orderly and efficient progression of an aircraft test program, engineering testing must be completed prior to large-scale production. Through engineering testing, the design should be verified and significant problems should be detected and corrected. However, this has not been the case with the C-5A.

Due to the nature of the particular total package procurement contract used to acquire the C-5A, the normal testing process was not possible. By using this procurement concept, DOD relinquished acquisition management and thereby lost control over engineering testing and could not perform acceptance testing or operational suitability testing prior to full-scale production. As a result, acceptance and operational suitability testing could not be performed by the Government until after the contractor delivered production aircraft. Subsequent events proved this practice to be disastrous.

Test schedules have slipped to the point that 81 aircraft--the total number of aircraft to be produced--will be in production assembly or will be completed before the engineering test program is completed. Delays in identifying deficiencies through tests resulted in producing aircraft with faults, such as structural defects and inoperable and unreliable subsystems. The structural weaknesses which were not found in time during early engineering tests are now being encountered in fatigue and static tests. Since these defects are being disclosed after production, retrofit or fixes will be expensive.

EVALUATION AND REPORTING OF TEST RESULTS

Conducting tests for a new weapon system is often a complex process; it may take weeks or even months to thoroughly evaluate the test results. A test's highlights should be reported to management within a few hours after it is conducted. However, the preparation of a detailed test report must await the often time-consuming analysis necessary to draw valid conclusions. During this test evaluation period, management must exercise care to avoid making decisions based on premature data which might be unreliable.

Decisionmakers must have test results available to them for an effective acquisition process. In order for decisionmakers to make informed judgments on the development and production progress of weapons, they must receive test and evaluation reports that:

--Describe the tests' purposes objectively and completely.

--List the basic assumptions and evaluations about the tests.

--State the test results in terms of mission objectives.

--Indicate the tests' limitations and inherent risks.

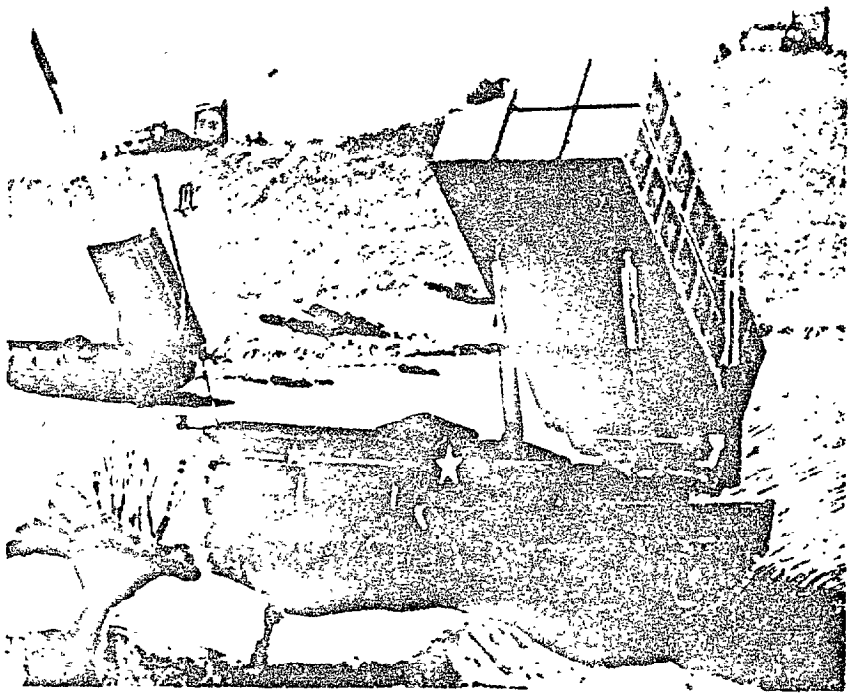
--Are prepared in a succinct, timely way by the testing authority.

These reports should contain a summary of the foregoing data in as nontechnical a way as possible. Furthermore, reports should not be changed, modified, or condensed in any way by intermediate commands. If test results are inadequately reported, decisionmakers will be handicapped by a lack of timely, complete, and valid information on the status of the system.

Following are examples of both adequate and inadequate reporting of test results.

Adequate application of criteria

SAM-D missile system



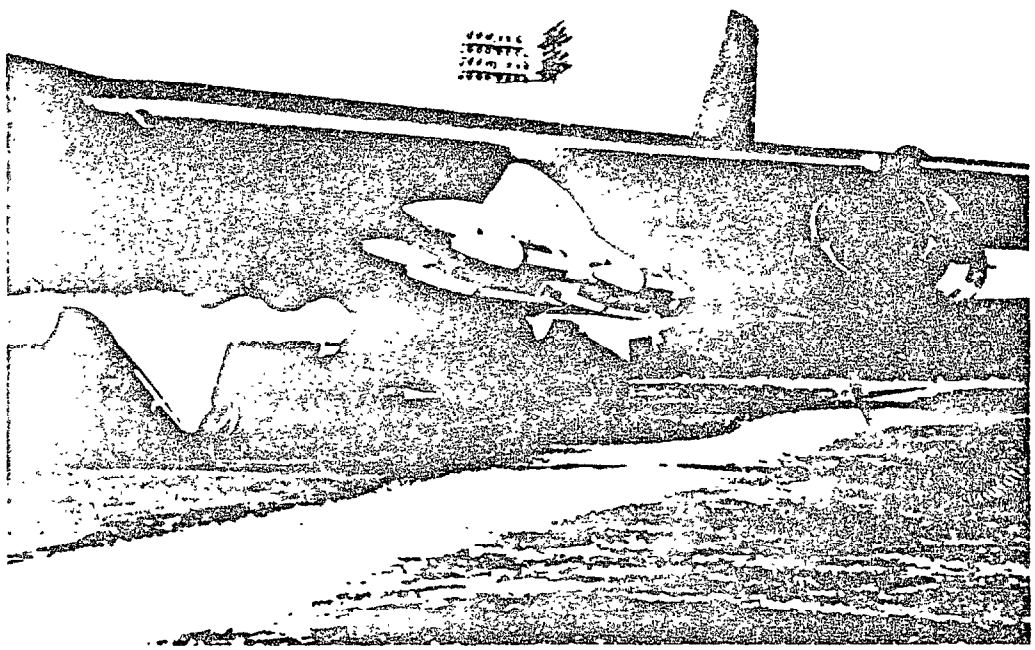
The contractor performed and reported the engineering tests conducted during the early development of the SAM-D. The prime contractor was required to submit all test results to the project manager within 30 days following completion of the test. Within the contractor's organization, the tests were conducted by a design engineer and were monitored by the test engineer, who was responsible for preparing the test report.

The test report was then reviewed by the contractor's project engineer and system design department to insure that the reports were complete and correct and that they satisfied

the test objectives. Next the test reports were submitted to the Army project manager.

Officials of the project manager's product assurance and test division, which had the responsibility for testing the system, reviewed the reports to insure that the test results were properly in line with the testing performed and were related to the accomplishment of test objectives.

Inadequate application of criteria for short-range attack missile (SRAM)



A contractor document dated October 9, 1970, disclosed that the missile's reliability was less than that required in system specifications. However, we were advised that missile reliability was not included in the project office's formal presentation to DOD officials in November 1970 when these officials were determining whether the SRAM should enter production. Project office officials informed us that they had not discussed reliability because, at the time of the product decision, the reliability prediction had

continuously progressed upward and because they had not anticipated any degradation.

Mission objectives were excluded from many test reports whose purpose was to inform higher headquarters of significant testing events or to present a detailed review of the overall program. Thus, these reports did not fully demonstrate the significance of the test results to their readers.

MANAGEMENT USE OF TEST RESULTS

Availability of test results to decisionmakers is a necessary requirement to an effective acquisition process. It is of equal importance that decisionmakers use these test results when making important decisions on the progress of a weapon system acquisition. Test results provide management with information on which to base decisions such as to modify a design approach or to change basic system development plans. These results are particularly important in deciding whether a system should be authorized to proceed to the next phase.

Unless decisionmakers use test results to assure themselves that a system is ready to move to a more advanced phase of the acquisition cycle, such as from development to production, it is likely that defective weapons will be fielded or that expensive modification programs will have to be undertaken.

Following are examples of effective and ineffective management use of test results.

Adequate application of criteria

POSEIDON missile system

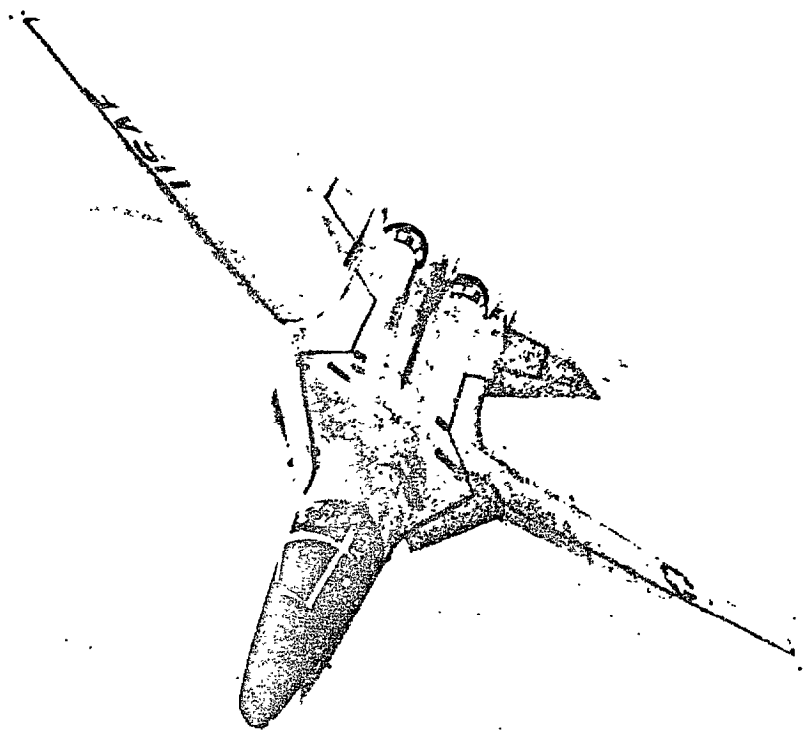
The results of the POSEIDON (engineering) test program were utilized by key project office officials in managing the POSEIDON program.

In one instance, after reviewing the results of flight tests, management decided that a satisfactory level of confidence in the reliability of the system could be maintained by periodic testing of a reduced number of missiles.

In another instance, production evaluation missile flight tests and demonstration and shakedown flight tests disclosed a problem area. Acting on the test results, the project office has applied additional resources to resolve the problem.

Inadequate application of criteria

F-111 aircraft



This example illustrates the consequences of allowing a system to enter production without having corrected the system deficiencies disclosed by engineering testing.

Engineering tests (wind tunnel and early flight tests) conducted in late 1964 and early 1965 disclosed that there was a serious engine stall problem on F-111 at high altitudes and speeds. The problem was attributed to an incompatibility between the aircraft air inlet and the engine. This problem had not been resolved when production of the aircraft was authorized in 1965. The decision to enter production was made by officials who had knowledge of available test results.

The engine stall problem required extensive development effort before it was solved. The complete solution came too late for incorporation in the first 141 F-111 aircraft. However, a significant inlet change was incorporated in these aircraft which eliminated the problem except for portions of the secondary mission. The contractor stated that these 141 aircraft will likely continue to experience engine stalls at speeds over mach 2.

CHAPTER 3

DOD IMPROVEMENT EFFORT FOR THE TESTING PROCESS

Since May 1970, DOD has increasingly emphasized the test and evaluation area and has instituted the following concepts in testing practices, some of which are new and some of which simply reemphasize old concepts. These policies are now being implemented on new weapon systems.

1. On May 28, 1970, the Deputy Secretary of Defense issued certain policies to apply to major systems acquisition. These policies stress selecting the proper form of contract, emphasize the necessity for providing hardware and design before moving into procurement, and call for performance testing during the development cycle.

2. By directive of February 11, 1971, the Deputy Secretary of Defense emphasized the importance of adequate operational test and evaluation and directed that the military departments create strong staff focal points and designate field commands independent of the developer to be responsible for such operational test and evaluation.

3. On April 21, 1971, the Deputy Secretary of Defense issued a directive which formalized the review process now being followed with respect to major weapon system acquisition programs. It directs that DSARC formally meet and review such programs when concept formulation is completed, when validation and ratification is completed, and when full-scale development is completed. At the very beginning of a program, a list of critical questions and issues is to be prepared and included in the appropriate development concept paper, together with test schedules, to eliminate risks at certain critical decision points.

4. On July 13, 1971, DOD Directive 5000.1 was issued. It states, in part, that technical uncertainty will be continually assessed and that progressive commitments of resources will be made only when confidence in program outcome is sufficiently high to warrant continuing. Models, mockups, and system hardware will be used to the greatest extent feasible to increase confidence level, and test and evaluation will commence as early as possible. A determination of operational

suitability, including logistics support requirements, will be made prior to large-scale production commitments and will make use of the most realistic test environment possible and the best representation of the future operational system available. The results of this operational testing will be evaluated and presented to DSARC at the time of the production decision.

5. On August 3, 1971, the Deputy Secretary of Defense issued a further instruction which, among other things, strengthened the responsibilities and authority of the Deputy Director (Test and Evaluation). This directive provides that an initial phase of operational effectiveness and suitability be accomplished for all new weapon systems prior to the first major production decisions. The functions of the Deputy Director (Test and Evaluation) were strengthened in several areas, primarily in making him responsible for submitting his opinions and recommendations at DSARC decision points both to DSARC and to the Deputy Secretary of Defense.

CHAPTER 4

CONGRESSIONAL CONCERN OVER TESTING OF MAJOR ACQUISITIONS

Because of the major problems encountered with weapons programs as a result of inadequate testing, congressional committees have increasingly qualified the authorization of resources assigned to programs until comprehensive testing has been completed. Following are some examples from the fiscal year 1971 and 1972 authorizing appropriations to illustrate this increased interest by the Congress. The restraints applied most often occur during requests by the services to place a major weapon in full-scale production.

In Public Law 91-441, dated October 7, 1970, the Congress stated that:

> "Of the total amount authorized to be appropriated by this Act for the procurement of the F-111 aircraft, $283,000,000, of such amount may not be obligated or expended for the procurement of such aircraft until and unless the Secretary of Defense has (1) determined that the F-111 aircraft has been subjected to and successfully completed a comprehensive structural integrity test program, (2) approved a program for the procurement of such aircraft, and (3) certified in a written report to the Committees on Armed Services of the Senate and the House of Representatives, that he has made such a determination and approved such a program, and has included in such written report the basis for making such determination and approving such program."
> (Underscoring supplied.)

In House Report 92-232 on authorization appropriations for fiscal year 1972, SRAM and MAVERICK were discussed.

> "The AGM-69A SRAM is an air launched air-to-surface missile for planned use on the B-52G/H and FB-111 aircraft. The SRAM missile is equipped with a nuclear warhead designed to attack targets defended by sophisticated defense

systems. The fiscal year 1972 program provides for proceeding to full-scale production considering completion of testing.

"MAVERICK is an air-to-surface Air Force missile, electro-optically guided for use against small hard targets such as tanks and bunkers. Last year the Congress, on the recommendation of the Committee on Armed Services, denied requested procurement funds for MAVERICK and directed that the program be continued in research and development to avoid concurrency and to allow more reliable development and test results prior to initiating procurement."
(Underscoring supplied.)

The Armed Forces Appropriation Authorization Act for 1972 (Public Law 92-156, dated November 17, 1971) further exemplified the increased congressional interest in testing. The act contains a section (sec. 506) requiring DOD to report to the Congress each year, beginning in calendar year 1973, on operational testing and evaluation for each weapon system for which procurement funds are requested.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

CONCLUSIONS

Testing and evaluation is a key management technique for controlling the acquisition of major weapon systems. The three basic categories of testing--engineering testing, acceptance testing, and operational suitability testing--can provide management with vital information about the workability, acceptability, and utility of a major weapon system. Timely and complete testing and evaluation results must be made available to decisionmakers before key decisions are made in the acquisition process in order to ascertain and minimize risk.

Over the years DOD has developed numerous policies, procedures, and organizations for the testing and evaluation of major weapon system acquisitions. The current emphasis is on more hardware proofing through the use of prototypes in the development of a system. Recent policy promulgations have reiterated the need for early testing and evaluation and for the determination of operational suitability, including logistic support requirements, prior to large-scale production commitments. DOD is making increased efforts to assess technical uncertainty and to control the progressive commitments of resources to programs.

Our general observations regarding the development of test objectives and of test plans, the implementation of testing and evaluation, the reporting of results, and the management use of testing and evaluation results for the 13 systems reviewed are as follows:

--The practices used in establishing test objectives were generally adequate.

For the few systems which did not have adequate testing objectives, the primary deficiency was that operational suitability testing either was not provided for or was not stressed.

--Most weapon systems did not have adequate test plans.

More often than not, this situation ocurred because the overall acquisition plan called for concurrent development and production, which made it impossible to devise an adequate test plan. In general, the test plans were unduly optimistic and success oriented. As a result, when problems developed, it was difficult for management to cope with them.

For those weapons with deficient test planning, we found that:

1. Test schedules were predicted on the assumption that minimal problems would be encountered during testing, and the schedules did not provide for contingencies and fallback positions that would allow management to modify acquisition decisions.

2. Not enough weapons were tested to draw valid conclusions.

3. Inadequate testing environments were planned to prove the weapon's military utility. The services generally underemphasized the need for demonstrating such military utility.

4. Operational suitability testing was planned too late in the acquisition cycle to be of real value.

--Testing and evaluation for most weapon systems was not accomplished timely and effectively.

We found a definite correlation between the adequacy of test plans and the timely and effective accomplishment of the tests. On the older weapon systems, test completion dates were set on an inflexible basis (being tied to calendar dates rather than to completion of an event) but were seldom achieved. When one date was missed, subsequent dates tended to be missed by an even wider margin.

To save time in the acquisition process, acceptance testing and operational suitability testing were sometimes waived. Although this practice resulted in some immediate

timesaving, it was frequently disastrous in the long run. Numerous ineffective weapons were fielded which required costly modifications or fixes.

Engineering testing was not completed before production began. It sometimes continued through the deployment of the weapons to the users. Acceptance testing was impossible under such circumstances. A similar condition existed regarding operational suitability testing; it was done too late, if at all.

Users have been accepting the weapons as they are provided by the developing commands, rather than insisting that the weapons be capable of fulfilling their expected missions and threat requirements under simulated combat conditions.

Many test programs lacked operational realism because of target limitations. Target availability was identified as a persistent problem in all three services.

--Most test reports were adequate; however, due to inadequate test planning and implementation, their value was diminished.

This situation was caused primarily by concurrency of development and production. For example, if engineering tests disclosed weapon system deficiencies after production was underway, the repairs would be more expensive than if they had been determined prior to production.

In those instances where testing was reported inadequately, we found that the test reports

--did not fully report the limitations under which the tests were made and generally indicated a higher degree of success than was actually demonstrated and

--were often highly technical in nature and lacked concise statements of how well the weapon systems met the test objectives.

Informal reporting often was used to report test results. All services were downgrading the value of a

formal reporting system of test results. Some of this was caused by the lateness of the tests themselves, plus the additional time required to make a formal evaluation of the test data. The danger in this approach is the possibility of bias from the enthusiasm of the system advocates, which could distort the true test results.

--Complete and valid test data was not available to decisionmakers prior to key decision points in the acquisition cycle.

Previously, we stated that availability of test results to decisionmakers is a necessary requirement to an effective acquisition process. It is of equal importance that decisionmakers use these test results when making important decisions on the progress of a weapon system acquisition. We found that management generally did give consideration to the available test results when making key management decisions.

Complete and valid test data is necessary for making sound decisions concerning the suitability of a weapon to advance to the next step in its life cycle. However, because of breakdowns in the testing process prior to decision points--such as poor planning, poor implementation of plans, and sometimes poor reporting of results--the information given to decisionmakers often was of diluted value.

Since July 1969, the Deputy Secretary of Defense has issued a series of policy statements which in their entirety set forth the framework for obtaining an improved acquisition process, including such goals as reducing the extent of concurrent development and production. However, we observed a number of instances in our study in which decisions to advance weapon systems to some stage of production had been made before completion of engineering testing.

Each of the three services have longstanding policies that essentially require engineering testing to be completed before production begins, but these policies have been frequently waived. For instance, the Department of the Army has such a policy but also provides for waiving the policy and beginning limited production under certain exceptional circumstances (i.e., when the need is urgent, when the risk is low, and when no other system satisfies

the requirement). However, most, if not all, of the major weapon systems procured by the Army in recent years have been procured under this waiver. Similarly, MARK 48, F-111, and a number of other weapon systems in the Navy and Air Force have entered production under waivers to the basic policy. The extent to which weapon systems have entered into production under policy exceptions in recent years raises serious doubts as to what the real policy has been.

We believe the recent policy statements of OSD which are intended to reduce the extent of concurrent development and production are basically sound and recognize that concurrency probably cannot be completely eliminated. However, we believe DOD needs to examine the criteria that the services have applied in the past for granting exceptions to the basic policy with a view to substantially reducing them.

RECOMMENDATIONS

To insure the most efficient and economical acquisition of major weapon systems, we recommended that DOD, in implementing its new policies and practices regarding testing and evaluation, continue to emphasize the need for:

--Completion of appropriate testing and evaluation prior to key decision points in the acquisition cycle.

--Adequate control over granting waivers from required testing and evaluation.

--Succinct summary reports to be prepared by the testing agency for all levels of management. Interested management levels may wish to comment on these summary reports; however, they should not be permitted to change the basic summaries.

The Director of Defense Research and Engineering said that, although it is true that many acquisition programs in existence cannot be changed, modifications are being made when practical. It was DOD's intention that the establishment of DSARC and DOD Directive 5000.1 dated July 13, 1971, would provide the means for implementing the necessary improvements in testing and evaluation, but the process of changing takes time.

We plan additional reviews in the testing and evaluation area to determine the success of these new policies in reducing the incidence of deficiencies.

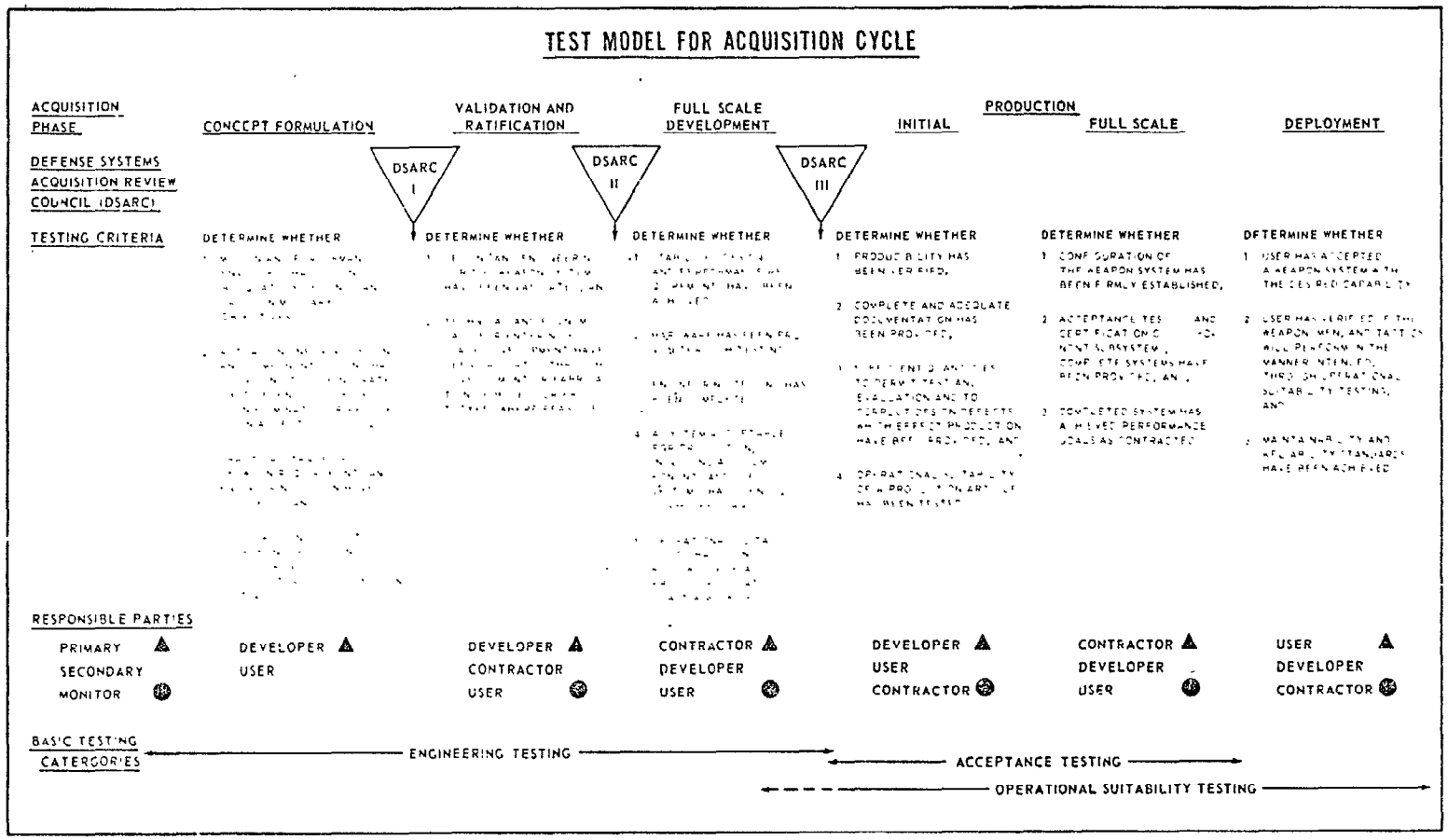
APPENDIX I

SCHEDULE OF MAJOR SYSTEMS
SELECTED FOR INDIVIDUAL CASE STUDIES

System	Acquisition phase	Estimated total program costs and additional costs (note a) (millions)
	(As of June 30, 1971)	
ARMY:		
Improved HAWK missile	Production	$ 846.8
SAM-D missile	Validation and ratification	3,930.3
DRAGON missile	Full-scale development	312.5
NAVY:		
A-7D/E aircraft	Production	3,933.5
Poseidon missile	do.	6,678.2
DE-1052 ship	do.	1,415.8
AIR FORCE:		
C-5A aircraft	do.	4,881.5
F-111 aircraft	do.	7,571.3
SRAM missile	do.	1,752.8
Minuteman III missile	do.	6,188.4
F-15 aircraft	Full-scale development	8,144.8
Maverick missile	do.	383.7
777 COMSAT communication satellite	do.	137.8
	TOTAL	$46,177.4

[a]As defined in DOD Instruction 7000.3, dated June 12, 1970.

TEST MODEL FOR ACQUISITION CYCLE



APPENDIX III

BIBLIOGRAPHY OF PERTINENT DIRECTIVES

REGULATIONS, STUDIES, AND PRIOR GAO REPORTS

OFFICE OF THE SECRETARY OF DEFENSE

Department of Defense Directive 5000.1, Director of Defense Research and Engineering, Office of the Secretary of Defense, Washington, D.C., 13 July 1971.

Establishes policy for major defense system acquisition in the military departments and defense agencies.

DEPARTMENT OF THE ARMY

Army Regulation 70-10, Test and Evaluation during Development and Acquisition of Materiel, effective 15 September 1971.

Current policies and procedures for Army test and evaluation during research and development of materiel.

Army Regulation 71-3, User Field Tests, Experiments and Evaluations, 19 March 1968.

This regulation outlines objectives, policies, responsibilities, and procedures for conduct of user field tests, experiments, and evaluations, which include troop tests, confirmatory tests, field experiments, field evaluations, and combat evaluations. AR 71-3 is under revision to implement new procedures for operational test and evaluation.

DEPARTMENT OF THE AIR FORCE

Air Force Regulation 80-14, Test and Evaluation of Systems, Subsystems and Equipment, 24 February 1967 (revised 5 May 1972).

This regulation states the objectives, policies, and responsibilities for U.S. Air Force test and evaluation activities which support Air Force research and development; acquisition of operational and support systems, subsystems, and equipment; and technical and engineering service programs and projects.

APPENDIX III

Air Force Regulation 55-31, Operational Test and Evaluation, 31 March 1970.

This regulation states the objectives and policies for U.S. Air Force operational test and evaluation activities.

DEPARTMENT OF THE NAVY

OPNAV Instruction 3960.1 D, Prosecution by the Operating Forces of CNO Assigned RDT&E Projects, 4 December 1967.

This includes the functions and specific responsibilities of commands and activities engaged in such research, development, test, and evaluation which involve the participation of the operating forces.

STUDIES, REPORTS, AND OTHER REFERENCES

"Concept of a Program of Strategic Low-Altitude Penetration Tests (U)," Supplement 1 to WSEG Report 74, Weapons Systems Evaluation Group, DOD June 1964 (Classified).

"Concepts for a Program of Tactical Low-Altitude Penetration Test (U)," WSEG Report 74, Weapons Systems Evaluation Group, DOD, May 1964 (Classified).

Provides a concept and integrated test program for weapon systems tests.

"A Prototype Strategy for Aircraft Development (U)," Memorandum RM-5597-PR, by Robert L. Perry, RAND Corporation, April 1968 (Classified).

An examination of the conditions that warrant the use of a prototype strategy in the development of military aircraft. Examples are given on the basis of recent experience both in the U.S. and abroad. Attention is given to the management approach that is most appropriate to a prototype strategy.

"Determination of USAF Testing Policies and Concepts Which Best Help To Achieve Operationally Effective Weapons and Equipment," by Col. George Lutz, June 1959.

A thesis-form discussion of the comparison of testing methodologies of various foreign military aeronautical organizations as well as U.S. civil aviation authority.

"Operational Test and Evaluation of Tactical Air-to-Air Missile Systems (U)," Institute for Defense Analyses, May 1970.

A comprehensive study of the operational test and evaluation methods and procedures for tactical air-to-air guided missile systems. Examines the causes of the less than desired success of air-to-air guided missiles, assesses the adequacy of operational test and evaluation methods and procedures, and recommends needed improvements.

"Report to The President and the Secretary of Defense on the Department of Defense" by the Blue Ribbon Defense Panel, 1 July 1970.

The panel's report offers recommendations in a number of areas, including organization, management of material resources, management procedures, personnel management, and conflicts of interest.

GENERAL ACCOUNTING OFFICE REPORTS

Evaluation of Two Proposed Methods For Enhancing Competition In Weapons Systems Procurement (B-39995, 14 July 1969).

This article is a letter to Senator James B. Allen in response to a request of a further discussion on prototype development as well as GAO's proposed new procurement methods for weapons systems.

Need for Management Improvement of Expediting Development of Major Weapon Systems Satisfactory for Combat Use (B-163058, 17 November 1969).

This report contains a series of actions recommended by GAO for application to current and future development programs to increase management effectiveness and to deploy acceptable weapon systems sooner.

APPENDIX III

Status of The Acquisition of Selected Major Weapon Systems (B-163058, 6 February 1970).

This report provides information to the Congress concerning the status of 54 major weapon systems and an evaluation of the selected acquisition reporting system of DOD.

Deployment of Weapon Systems Before Sufficiently Developed and Tested (B-160877, 20 March 1970).

This report consists of a letter sent to Congressman Sidney R. Yates in response to his letter requesting all available unclassified information concerning concurrent development and production.

Adverse Effects of Large-Scale Production of Major Weapons Before Completion of Development and Testing (B-163058, 19 November 1970).

This report was to resolve significant problem areas resulting from concurrent development and production.

Analysis and Alternatives: The AGM-65A Maverick Missile System (B-160212, 31 December 1970) (Classified).

This report was to identify those weapons in procurement and development which respond the same or which have very similar mission requirements and to identify and analyze the reversible causes of unnecessary duplication among tactical missiles.

Acquisition of Major Weapon Systems (B-163058, March 18, 1971)

This report is GAO's appraisal of the factors most closely related to effective performance in procuring major weapons.

Letter Report to the Secretary of Defense, dated 16 September 1971, Review of the Operational Test and Evaluation of the fast Automatic Shuttle Transfer System and Other Systems.

APPENDIX III

This letter reports GAO findings at the Navy's Operational Test and Evaluation Force and GAO suggestions for earlier operational evaluation to permit timely consideration of any risks related to urgent requirements.

APPENDIX IV

DIRECTOR OF DEFENSE RESEARCH AND ENGINEERING
WASHINGTON D C 20301

Mr. Hassell B. Bell
Associate Director, Defense Division
United States General Accounting Office
Washington, D. C. 20548

Dear Mr. Bell:

Thank you for your letter of December 17, 1971, forwarding for our review and comment copies of your draft report on "Testing and Evaluation in the Acquisition Process for Major Weapons Systems." The report is of great interest to us. We have had it reviewed carefully by both the appropriate Assistant Secretaries of Defense and the Military Department Secretaries. As a result I have certain comments to make. (OSD Case #3389)

The report is based on the analysis of 13 specific systems. All of these entered advanced development in the time frame December 1962 to March 1969 and entered procurement between July 1964 and July 1971. Of these 13, two are programs which were carried out under the total package procurement approach. The report concludes that there were certain weaknesses in our testing approach as evidenced by these particular examples. It reaches certain conclusions and gives certain recommendations as summarized in Chapter 5 of the draft report, pages 32 through 37 inclusive.

The report does state briefly on page 3 and in more detail in Chapter 3 pages 27 through 29 inclusive, that the Department of Defense has directed certain improvements. However, it does not cover these in detail. Yet if the report is properly to reflect the current situation and thereby be of greatest benefit to the Congress, there should be added detail relative to the recently adopted procedures. I recommend that you point out both in Chapters 2 and 5 that the systems studied were systems which are not representative in full of current practices. I recommend further that you summarize what are the currently directed practices in Chapter 3 and in Chapter 5 make appropriate references as I shall elaborate on later. To be more specific as to currently directed policies and procedures:

a. On May 28, 1970, the Deputy Secretary of Defense enunciated certain policies and practices to apply for major weapon systems acquisition. These constituted a rather major change from the earlier practices. I have attached a copy of that directive at Enclosure 1. As you will note, it lays stress on selecting the proper form of contract (ruling out the earlier tendency toward total package procurement), emphasizes the necessity for proving of hardware and design before moving into procurement and calls for performance testing during the development cycle. These policies were later repeated in large measure in DoD Directive 5000.1 dated July 13, 1971, but became the policy for new weapon system acquisition programs of the DoD in mid-CY 1970.

b. The Deputy Secretary of Defense by directive of February 11, 1971 (Enclosure 2) emphasized the importance of adequate operational test and evaluation and directed that the Military Departments create strong staff focal points and designate field commands independent of the developer to be responsible for such operational test and evaluation. This structure has been created now and is actively functioning in the manner desired.

c. On April 21, 1971, the Deputy Secretary of Defense issued a directive (Enclosure 3) which formalized the review process now being followed with respect to major weapon system acquisition programs. As you can see it directs that the Defense Systems Acquisition Review Council (DSARC) will formally meet and review such programs at least three times: first, at Milestone I when they enter advanced development; second, at Milestone II when they enter full engineering development; and third, at Milestone III when a first major production decision is to be taken. At the very beginning of the program, a list of critical questions and issues will be prepared and included in the appropriate Development Concept Paper together with test schedules to provide for the elimination of risks by certain critical decision points. The Deputy Director (Test and Evaluation), ODDR&E, then soon to be appointed, was to have responsibility for commenting to the DSARC on the questions and issues and at Milestone III be responsible to submit his individual assessment of the adequacy of the test and evaluation to that date. He was to have access to all available test plans and test data.

d. On August 3, 1971, the Deputy Secretary of Defense issued a further elaborating policy directive (Enclosure 4) which among other things strengthened the responsibilities and authorities of the Deputy Director (Test and Evaluation). As you will note, this directive provides that an initial phase of operational effectiveness and suitability will be accomplished for all new weapon systems prior to the first major production decision. In the case of well-advanced systems where this is not possible, such initial operational test and evaluation will still be accomplished as early as practical. The strengthening of the functions of the Deputy Director (Test and Evaluation) were in several areas. A primary one, however, was in making him responsible at DSARC decision points for submitting his opinions and recommendations both to the DSARC and to the Deputy Secretary of Defense. You will note too that the directive stated that the same principles with regard to initial operational test and evaluation, participation by user personnel and by the Service operational test and evaluation structure were to be applied on all weapon system acquisition programs, not just on the major programs subject to DSARC review.

All the principles that I have spoken of above are now in the process of implementation. It is naturally true that there are many programs in existence which are well advanced and cannot be completely transferred over to the new desirable form at this time due to contracts or other binding arrangements. However, these are being modified to the degree practical. I must stress though that the process of changing is one which takes time.

You do speak in your paper to three different forms of test and evaluation, - engineering testing, acceptance testing, and operational suitability testing. In our directives we have divided all testing into two categories, - one, that accomplished by the developer/producer and, second, that by the agency representative of the user. Your acceptance testing falls for the most part in the former category, that of developer/producer operation, which we classify under development testing, but does fall in part in the area of operational test and evaluation. On looking over your pages 5 and 6, I find that the difference is one of terminology and not one of intent. Either breakout could be used. The overall objectives stated in your pages 9 through 10 are embraced in those of current DoD instructions.

In view of the above, I recommend also that you include in your Chapter 5 under each conclusion and recommendation a reference to what has now been directed with respect to that particular conclusion or recommendation. As example:

a. "The practices used in establishing test objectives were generally adequate."

Early identification of critical questions and issues and scheduling the tests to resolve should effect even more improvement in this regard.

b. "Most weapons systems did not have adequate test plans."

Again the matter of identifying early what testing is required and reviewing at major decision points what has been accomplished and learned should do much to improve this. Even more importantly the DD(T&E) now has responsibility, and adequate staff, to review test plans and submit any necessary recommendation.

c. "Testing and evaluation for most weapons systems was not accomplished in a timely and effective manner."

For major programs, the identification of the critical questions and issues, a scheduling of the necessary tests, the review of test plans and the DSARC review of major progress are all designed to accomplish the needed improvement in this area.

d. "Most test reports were adequate; however, due to inadequate test planning and implementation, their value was diminished."

For programs subject to DSARC review, the identification of the critical questions and issues, a scheduling of the necessary tests, the review of the test plans and the analysis made of the test reports at both Services and OSD levels prior to the detailed DSARC reviews and the DSARC reviews are all designed to insure that required tests are properly performed, test results fully analyzed and critical issues satisfactorily resolved. The new DoD policies require that these actions be completed prior to a particular weapon system being authorized to proceed to the next DSARC milestone.

e. "Complete and valid test data was not available to decision-makers prior to key decision points in the acquisition cycle."

As outlined in comments on the previous conclusion, the DSARC system requires that test data be available for review prior to key decision points at DSARC Milestone II (Ratification Decision) and Milestone III (Production Decision).

i. "Completion of appropriate testing and evaluation prior to key decision points in the acquisition cycle."

Current DoD directives now require that for major programs testing and evaluation be completed, results be reviewed at Service and OSD levels and a decision be made by the Secretary of Defense prior to the program proceeding to engineering development (Milestone II) or major procurement (Milestone III). The Secretary of Defense decision is based upon the DSARC's recommendation as well as an independent recommendation by DD(T&E).

g. "Stringent control over the granting of any waivers from required testing and evaluation."

For major weapon system programs, DoD Directive 5000.1, dated July 13, 1971, requires that when a DoD component is sufficiently confident that engineering is complete and that a commitment of substantial resources to production and deployment is warranted, it will request Secretary of Defense decision to proceed. Prior to the DSARC review, the intensive appraisal by the Services, including the independent evaluation of test results by an agency separate from the developer as well as the DCP Coordination Process, establishes a stringent control over the granting of waivers from required testing and evaluation for major weapon system programs. DoD Directive 5000.1 further states that the management principles outlined are applicable to all programs. The increased emphasis on operational test and evaluation within the Services due to the establishment of separate Headquarters staff elements and the performance of OT&E by an agency which is separate and distinct from the developing command will serve to bring requests for waivers to required test and evaluation under close scrutiny which will subject them to stringent control. In his guidance to the Service Secretaries and others concerned contained in the memorandum subject: "Test and Evaluation in the System Acquisition Process" of August 3, 1971, the Deputy Secretary of Defense stated that the principles of early operational test and evaluation before production decision, participation

by user personnel, and participation by the Service operational test and evaluation structure should generally apply for weapon system acquisition programs not subject to DSARC review. All the Services are implementing these principles by appropriate regulations or directives to include other programs than those under DSARC review. These will provide that any deviation from established policy, such as waiver of required test and evaluation, must be approved at Service Headquarters level.

h. "Succinct summary reports to be prepared by the testing agency for all levels of management. Interested management levels may wish to comment on these summary reports, however, they should not be permitted to change the basic summary."

The distribution of test plans and reports to interested management levels is required and the modification of the test agency's findings by intervening authority is not in consonance with current policies.

There are certain specific corrections required in the report that have been pointed out to me. Enclosure 5 lists these recommended changes.

I feel, however, that modification of the draft report to indicate the current status of the test and evaluation function in the DoD major weapon system acquisition programs as indicated by me above would place the report in proper and more meaningful context.

Thank you again for your kindness in furnishing me the report. The DD(T&E), ODDR&E, is available should you or your staff desire further discussion on any of the points above. I know he has had extensive contact with various members of the General Accounting Office to date and they are extremely interested in the progress being made.

Sincerely,

John S. Foster, Jr.

Enclosures 5