

BY THE COMPTROLLER GENERAL

Report To The Congress

OF THE UNITED STATES



LM106104

Operational Testing Of Air Force Systems Requires Several Improvements

Established in 1974, the Air Force Test and Evaluation Center is responsible for managing Air Force operational test and evaluation and providing information on test results to the Secretary of the Air Force and the Air Force Chief of Staff.

GAO found that the test center has improved Air Force operational testing. However, increased test realism, additional test resources, better test criteria, and more dedicated operational testing are needed. This report contains recommendations to the Secretary of Defense on these issues.



PSAD-78-102
JUNE 2, 1978

U 66590



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 report presents our views on improvements needed in Air Force operational testing. A draft of this report was reviewed by agency officials; their comments are incorporated as appropriate.

Established in 1974, the Air Force Operational Test and Evaluation Center has, during its short existence, made positive impacts on the Air Force test and evaluation process. Air Force comments on our draft report outline other improvements being made or considered. However, we believe that additional improvements are required to increase test realism and assure that all important operational test objectives are accomplished before production decisions.

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

We are sending copies of this report to the Director, Office of Management and Budget, and the Secretary of Defense.


Comptroller General
of the United States

COMPTROLLER GENERAL'S
REPORT TO THE CONGRESS

OPERATIONAL TESTING OF
AIR FORCE SYSTEMS REQUIRES
SEVERAL IMPROVEMENTS

D I G E S T

GAO strongly supports the concept of an independent test organization capable of conducting realistic operational testing.

An operational test and evaluation is an assessment of mission performance in a system's intended environment when operated, maintained, and supported by personnel with qualifications similar to those who will perform these functions in the field. (See p. 1.) It consists of two phases--the initial evaluation conducted before the first major production decision and the follow-on evaluation made after the production decision. (See p. 2.)

Established in 1974, the Air Force Test and Evaluation Center is responsible for managing Air Force operational test and evaluation and providing an independent assessment on the test results to the Secretary of the Air Force and the Air Force Chief of Staff. (See pp. 1 and 2.)

GAO found that the test center, during its short existence, has had positive effects on the Air Force test and evaluation process. The Air Force comments on a draft of this report indicate that action is being taken to further improve operational testing during future programs. However, during its review GAO found several limitations in Air Force operational testing that warrant consideration by the Secretary of Defense and the Congress. (See p. 17.)

Operational testing conducted by the Air Force Test and Evaluation Center was generally not performed in an environment that the systems would encounter when deployed, and the systems were not always operated and maintained by personnel of the type and qualifications of expected users. (See p. 4.)

The test center had little control over the resources necessary to conduct initial operational testing:

- Test items and related system hardware were not always available.
- Test ranges often were not suitably equipped.
- Time constraints restricted the test center in accomplishing the test objectives. (See p. 8.)

Most Air Force initial operational test and evaluation programs were combined with development testing to minimize cost. While combined programs have certain advantages, the objectives of the two types of testing are completely different, and it is highly unlikely that all important objectives can be accomplished concurrently. (See pp. 12 and 14.)

The test center's evaluations of operational effectiveness have been limited because they were not based on the most current operational concept. During our review, the Air Force was actively considering a new policy to provide operational concepts that will serve as a "contract" between developers and users and will provide current criteria for test and evaluation. This new policy has subsequently been incorporated into Air Force regulations. (See pp. 15 and 16.)

The Air Force expressed basic agreement with our position and stated that many improvements have been implemented on later programs. For example, the Air Force plans to increase the use of realistic operational test locations and to emphasize the need to provide adequate resources and time for operational testing. The Air Force has also amended its regulations to assure development of better operational test criteria. In addition, the Air Force plans to emphasize separate initial operational test and evaluations where feasible.

GAO believes that these measures, if fully implemented, will improve Air Force operational testing. However, GAO also believes that additional measures are required to assure that a separate dedicated phase of initial operational test and evaluation is conducted on all major acquisition programs before the production decision and that average operational and support personnel are used for operational testing.

RECOMMENDATIONS

Because of the importance of making a realistic assessment of the operational effectiveness and suitability of a weapon system before production and the subsequent deployment of the system to the operational forces, we recommend that, for future test programs, the Secretary of Defense

- monitor the Air Force's implementation of the improvements in operational testing cited in its comments to our draft report,
- direct the Air Force to conduct operational test and evaluation with operational and support personnel of the type and qualifications of those expected to use and maintain the system when deployed, and
- direct the Air Force to conduct a separate dedicated phase of IOT&E to assure that all operational test objectives are accomplished under the appropriate test conditions.

BLANK

C o n t e n t s

	<u>Page</u>
DIGEST	i
CHAPTER	
1 INTRODUCTION	1
Categories of operational testing	2
Objective and scope	2
Difficulty in obtaining access to certain agency records	2
2 OPERATIONAL TEST REALISM IS NOT ALWAYS ACHIEVED	4
Suitability of test sites	4
Qualifications of test personnel	5
APTEC efforts to make its test programs more realistic	7
3 LIMITED TEST RESOURCES HAVE ADVERSELY AFFECTED AIR FORCE OPERATIONAL TESTING	8
Availability of test items and related system hardware	8
Suitability of test ranges	9
Too little time provided for IOT&E	10
4 SEPARATE DEDICATED PHASE OF OPERATIONAL TESTING NEEDED BEFORE PRODUCTION DECISION	12
Problems in combined test programs	12
Air Force position on combined test programs	12
Defense and Air Force guidance on combined-versus-separate test programs	13
5 THE AIR FORCE RECOGNIZES THE NEED FOR BETTER OPERATIONAL TEST CRITERIA	15
6 CONCLUSIONS AND RECOMMENDATIONS	17
Recommendations	18
APPENDIX	
I AGENCY COMMENTS	19

ABBREVIATIONS

AFSC	Air Force Systems Command
AFTEC	Air Force Test and Evaluation Center
DT&E	development test and evaluation
FOT&E	follow-on test and evaluation
IOT&E	initial operational test and evaluation
OT&E	operational test and evaluation
TEWS	tactical electronic warfare system

CHAPTER 1

INTRODUCTION

Testing is a key ingredient in the defense system acquisition process because it provides data for evaluating system development progress. Information on system performance and reliability are provided to management through frequent testing of components and assemblies up through complete system configurations. Management needs adequate and valid test data to make informed program management decisions and to determine the advisability of system production. Two basic kinds of test and evaluation occur during the defense system acquisition process--development test and evaluation (DT&E) and operational test and evaluation (OT&E).

DT&E is to demonstrate whether engineering design and development is complete, design risks are minimized, and the system meets specifications. DT&E is a detailed engineering analysis of system performance (beginning with individual subsystems and progressing through a complete system) where system design is tested and evaluated against engineering and performance criteria by the implementing command.

OT&E is conducted to estimate the system's military utility and operational effectiveness and suitability. OT&E is an assessment of a complete system's mission performance in its intended environment when operated, maintained, and supported by personnel with qualifications similar to those who would normally operate it. In addition, OT&E provides information on organization and personnel requirements as well as the tactics and doctrine to be used when the system is deployed.

Defense Directive 5000.3 establishes policy for conducting test and evaluation of defense systems by military departments and Defense agencies. It requires one major field agency in each military service, separate and distinct from implementing and using commands, responsible for OT&E, that reports directly to the military service chief.

The Air Force established the Air Force Test and Evaluation Center (AFTEC) as its principal field command to manage OT&E. It provides OT&E information to the Secretary of the Air Force and the Air Force Chief of Staff. AFTEC was officially activated in January 1974, and became completely operational in October 1974.

Air Force Regulation 80-14 implemented Defense Directive 5000.3; the regulation outlines Air Force policy and procedures for managing test and evaluation during system development, acquisition, and deployment. It established the management relationships during a system's life cycle among the commands responsible for acquiring defense systems and for managing DT&E programs (normally the Air Force Systems Command (AFSC) or the Air Force Logistics Command), AFTEC, and operating and supporting commands.

CATEGORIES OF OPERATIONAL TESTING

OT&E consists of two phases--initial operational test and evaluation (IOT&E) and follow-on test and evaluation (FOT&E). IOT&E is to determine a system's operational effectiveness and suitability before the first major production decision.

FOT&E is all operational testing after the first major production decision. FOT&E is to refine determinations of a system's operational effectiveness and suitability that were made during IOT&E and to assist in further production decisions or system configuration changes. FOT&E could continue throughout the system's life cycle.

OBJECTIVE AND SCOPE

This review was to evaluate AFTEC independence and effectiveness in managing OT&E. To do this we examined AFTEC management of OT&E on selected aircraft programs, including the B-1, A-10, F-15, F-4G Wild Weasel, and the advanced medium short take-off and landing transport. Work was performed at AFTEC Headquarters, Kirtland Air Force Base, New Mexico, and selected test sites.

We interviewed officials at AFTEC Headquarters and field elements and selected DT&E team members. We also reviewed test file documentation and other information provided by AFTEC.

DIFFICULTY IN OBTAINING ACCESS TO CERTAIN AGENCY RECORDS

Our evaluation was hampered considerably because AFTEC would not provide us access to trip reports and memorandums to the record because, according to the Air Force these documents provided the limited perspective of the writer and not an AFTEC or Air Force position. During previous work at AFTEC, these reports and memorandums were the only documented source of information on AFTEC relationships with other Air

Force entities. Without this source of information, we could not fully evaluate AFTEC independence from implementing and using commands.

This is part of a broader problem that GAO is experiencing regarding access to Air Force records. GAO is attempting to resolve this at the Air Force and Department of Defense levels.

CHAPTER 2

OPERATIONAL TEST REALISM IS NOT ALWAYS ACHIEVED

Operational testing is to determine how well a system performs its intended mission in a representative environment when operated and maintained by military personnel typical of those who would normally perform these functions. However, we found that operational testing conducted by AFTEC was generally not performed under such conditions. Therefore, operational test results might not adequately reflect the performance, maintainability, and readiness of the systems in a realistic operational environment.

SUITABILITY OF TEST SITES

Defense Directive 5000.3 and the Air Force implementing Regulation 80-14 state that OT&E should be conducted in as realistic an operational environment as possible. However, aircraft IOT&E was usually performed at Edwards Air Force Base, California, which has environmental characteristics desirable for development testing but not for operational testing. It has desert terrain, almost unlimited visibility, and consistently good weather, whereas many aircraft systems are required to operate in adverse weather conditions.

For example, the A-10 aircraft was designed to provide close air support under adverse weather conditions that frequently exist in Europe, Southeast Asia, and Korea--it is estimated that low cloud cover and limited visibility prevail in these areas about 25 percent of the time. Since A-10 IOT&E was conducted at Edwards, its performance in a low visibility environment was not fully evaluated before the production decision.

Because AFTEC has only existed since 1974 and because of the limited number of aircraft test programs that have been completed under AFTEC direction, we cannot document the adverse effects of conducting aircraft test programs at Edwards Air Force Base. However, we believe that operational test programs conducted in such environments are likely to produce limited test results and allow programs to go into production before there is assurance that the systems can perform as required under realistic conditions.

It is difficult to achieve a realistic operational environment for many reasons. There are few locations within the United States that offer a representative environment suitable for testing a wide variety of systems; most available locations also have constraints such as limited size and close proximity to population centers and/or commercial air

traffic. In addition, most operational locations lack the instrumentation facilities available at development test sites. The Air Force also believes that it would be costly to move the test aircraft, ground support, test equipment, and other support from Edwards to a more representative location.

These factors are important considerations in selecting sites for future IOT&E programs. However, the validity of test conditions is also an important consideration for decisionmakers when determining whether an acquisition should move forward. Therefore, we believe that some testing should be conducted where environmental conditions more closely approximate the environment where the system will be expected to perform its mission. While this may increase the cost of testing, it will also provide more reliable information on the military utility of systems before the production decision.

In commenting on this report, the Air Force stated that increased use of realistic operational test locations will be accorded emphasis in future test programs.

QUALIFICATIONS OF TEST PERSONNEL

Defense Directive 5000.3 calls for OT&E to be performed by operational and support personnel of the type and qualifications of those expected to use and maintain the system when deployed. Air Force Regulation 80-14 states that OT&E is conducted to estimate how well the system performs its mission when operated and maintained by military personnel in the field. However, Air Force systems are not always operated and maintained during IOT&E by personnel with qualifications similar to those who would operate and maintain the systems after they are deployed. In many cases, contractor personnel perform much of the required maintenance.

Clearly above average Strategic Air Command pilots (graduates of test pilot school) were used during the B-1 test program. Pilots used for the A-10 test program were considered to be more qualified than the average Tactical Air Command pilot. In contrast, pilots used for the YC-15 and YC-14 test program were representative of those in the field. Two of the four people assigned were very junior pilots whose experience had been predominantly as copilots.

Service maintenance personnel assigned to AFTEC IOT&E test teams were generally very high caliber with considerable experience working in their specialty area. They were usually much more highly qualified than personnel who will maintain the system after it is deployed.

Much of the maintenance during IOT&E was performed by the contractor because most of the development contracts required contractor maintenance during DT&E--which is generally combined with Air Force IOT&E programs. In fact, AFTEC obtained the majority of its maintenance data by observing maintenance done by contractor personnel. When systems are maintained by the contractor, problems which may be experienced by average Air Force personnel may not surface because contractor maintenance personnel are generally very knowledgeable about the system being tested. There were also generally more maintenance personnel available during IOT&E than could be expected at an operational location. We believe that these two factors can affect the accuracy of maintainability estimates by making the system appear more easily maintainable than it will be once it is deployed.

During the A-10 IOT&E, four of the six available test aircraft were contractor maintained. Although AFTEC performed limited "hands-on" maintenance during the B-1 and YC-15 IOT&E programs, the contractor was primarily responsible.

Maintainability is a very important factor in determining a system's readiness and cost of ownership over an extended period of time. We believe that during IOT&E maintenance should be conducted by personnel with the same qualifications as the expected users in order to obtain accurate estimates of system maintainability. The Air Force stated that it concurred, in principal, with this position. However, the Air Force also stated that personnel with above average qualifications are necessary to assess the military utility of weapon systems (particularly prototype and preproduction systems).

In September 1976 a former commander of the Navy independent operational testing agency commented before the National Security Industrial Association that an important characteristic of operational testing includes the use of "fleet type" personnel,

"* * * not people specially trained, not test pilots, not technicians or engineers, but the average sailor who has all the propensity for making Murphy's law happen that typical officers and men in the fleet do."

We believe that using highly qualified personnel for operational testing is likely to provide an optimistic assessment of a system's operational effectiveness and suitability.

AFTEC EFFORTS TO MAKE ITS TEST
PROGRAMS MORE REALISTIC

AFTEC has tried to make its test programs more realistic. During YC-15 testing, operational mission profiles were devised and flown at Edwards Air Force Base--during these missions, efforts were made to simulate the employment concepts envisioned for a short take-off and landing transport. They provided as realistic a test as practicable in an environment such as Edwards, and included simulated and some actual equipment and personnel air drops.

The B-1 IOT&E included three missions dedicated to operational testing that closely resembled a normal Strategic Air Command exercise. It included base escape, low altitude terrain following, high altitude supersonic penetrations, air refueling, and weapon deliveries.

The test manager stated that A-10 testing during FOT&E, which was conducted after the production decision, was more realistic than during IOT&E because the aircraft participated in training missions using aggressor squadrons, test ranges with threat simulations and targets, and deployment to a representative geographic area (McChord Air Force Base) for low ceiling flying.

CHAPTER 3

LIMITED TEST RESOURCES HAVE ADVERSELY AFFECTED AIR FORCE OPERATIONAL TESTING

AFTEC had very little control over the resources necessary to conduct initial operational testing on the systems that we reviewed. The Air Force systems command had prime responsibility for assuring the availability of test resources. We found that (1) test items and related system hardware frequently were not available, (2) test ranges were often not suitably equipped, and (3) time constraints have affected the ability of AFTEC to accomplish its test objectives. Although we could not determine the effect of these limitations on the effectiveness of the weapon systems that we reviewed, it is clear that they adversely affected the effectiveness of AFTEC in performing its mission as an independent testing agency.

Many test resources, such as aircraft, spare parts, and data, were provided by the development contractor as required by the development contract. Since AFTEC has existed only since 1974, operational requirements for test resources have generally not been reflected adequately in these contracts. AFTEC is trying to become involved early enough in upcoming programs to resolve some of the problems with test resources. In commenting on this report, the Air Force stated that considerable attention is being devoted to resolving the limited test resources problem.

AVAILABILITY OF TEST ITEMS AND RELATED SYSTEM HARDWARE

Lack of operational aircraft, spare parts, and/or other test hardware limited the overall effectiveness and suitability evaluations of the A-10 and F-4G Wild Weasel IOT&E programs as well as the F-15 FOT&E program because important elements could not be fully tested. For example, 10 preproduction aircraft were originally planned for use on the A-10 combined DT&E/IOT&E test program; the first 6 were to be used for DT&E and the last 4 for IOT&E. However, all four IOT&E aircraft were cut from the program. Most A-10 IOT&E testing took place during a 3-month period (from September through November 1975) with two of the six DT&E aircraft. Since these were DT&E aircraft, they were instrumented for DT&E rather than for IOT&E.

Operational testing was limited because the aircraft were not operationally representative and for other reasons.

There was not room for all of the standard operational equipment (such as UHF homing equipment or a strike camera to provide bomb accuracy data) during IOT&E. Also, several items of A-10 hardware, including electronic countermeasure pods and chaff and flare dispensers, were not available for testing during IOT&E. In addition, several weapon systems, such as laser and TV-guided bombs, could not be tested because they had not been certified for use on the A-10 aircraft.

AFTEC planned to evaluate the adequacy and transportability of aerospace ground equipment as part of its evaluation of A-10 logistics supportability. However, not all contractor-furnished ground equipment was available to AFTEC during IOT&E, nor was all the equipment that was available evaluated by AFTEC. Of approximately 265 ground equipment items identified, only 149 production articles were available for evaluation at the test site and only 89 of these were fully evaluated due to changing system designs, lack of technical data, and limited use.

According to the Phase II IOT&E report AFTEC could not fully assess the effectiveness of the GAU-8 gun against close air support targets because of insufficient sample size.

The lack of spare parts in correct configurations significantly limited the maintenance evaluation of the F-4G Wild Weasel program. Although AFTEC believes sufficient data was obtained to provide a data base, we were told that further investigation during FOT&E could produce results that vary by as much as 25 percent from results obtained during IOT&E.

The tactical electronic warfare system (TEWS) for the F-15 was not available for testing during FOT&E, although the aircraft used for testing were production models. TEWS was still being developed when the F-15 was tested; without TEWS, the F-15 capability against enemy ground-to-air defenses could not be flight tested. Therefore, F-15 survivability over hostile territory was not evaluated. Also, testing of the AIM-9L short-range missile with the F-15 was limited due to a shortage of AIM-9L missiles.

SUITABILITY OF TEST RANGES

Test ranges are not completely suited for operational testing due to the lack of instrumentation and threat simulators as well as the placement of range equipment. These problems reduce realism, restrict tactics, limit scenarios, and result in more qualified data.

One of the two ranges AFTEC used for F-15 air combat maneuvering testing could not provide time/space position data and a real time display for engagements. This information was needed to plot the engagements, verify tactics, and validate shots. The data had to be obtained during crew debriefings after each test rather than by instrumentation. In addition, limited air space and the proximity of nonparticipating aircraft at Luke Air Force Base limited F-15 tactical scenarios and flight employments during FOT&E.

Simulators were not available to flight test some F-15 components against some of the newer threats. This precluded making conclusive evaluations on the aircraft's survivability.

The location of bomb scoring towers affected A-10 testing at Edwards Air Force Base. The towers restricted certain bombing approaches and thus prevented realistic tactical maneuvers.

TOO LITTLE TIME PROVIDED FOR IOT&E

Time constraints affect IOT&E programs. In cases where DT&E took longer than planned, IOT&E was generally compressed to a shorter time period to meet the scheduled production decision.

In some cases, assessments of operational suitability were qualitative (subjective opinion) rather than quantitative (numerically expressible) because sufficient testing could not be accomplished before the production decision. For example, the evaluation of the A-10 close air support capability at night was limited because sufficient testing could not be accomplished before the production decision.

The six preproduction A-10 aircraft were not available throughout the combined test program due to delivery delays of up to 2-1/2 months and shortages of spare parts for those aircraft available. Also, DT&E testing was extended, which caused delays and cancellations of some IOT&E flights. According to the A-10 test manager, AFTEC was not able to fully evaluate

- general operational suitability,
- military utility and operational effectiveness during performance of certain ground attack missions,
- survivability/vulnerability,
- reliability/maintainability,

--logistics supportability, and

--gun performance.

According to AFTEC logistics personnel, sufficient data could not be obtained during the B-1 IOT&E program to support a statistically sound evaluation of maintainability, reliability, or personnel requirements because of the limited time available to gather such data.

CHAPTER 4

SEPARATE DEDICATED PHASE OF OPERATIONAL TESTING NEEDED BEFORE PRODUCTION DECISION

At the time of our review most AFTEC-managed IOT&E programs were combined with DT&E to minimize cost. As of October 1976 30 of 36 AFTEC-conducted IOT&E programs were combined with DT&E. While this practice probably reduces the cost of testing and provides AFTEC with early and sustained involvement during system development, we believe that many of the problems discussed in chapters 2 and 3 could be overcome if the final phase of IOT&E was separate from DT&E.

PROBLEMS IN COMBINED TEST PROGRAMS

The objectives of DT&E and IOT&E are considerably different. DT&E is a detailed engineering analysis of a system's performance whereas IOT&E is an assessment of a complete system's mission performance in its intended environment when operated, maintained, and supported by personnel with similar qualifications of those in the field. Although many objectives of the two types of testing can probably be accomplished concurrently, we believe that a separate dedicated phase of IOT&E is essential to assure that all operational test objectives are accomplished. In commenting on this report, Air Force officials stated that the next revision of Air Force Regulation 80-14 will emphasize separate IOT&E phases where feasible.

In Air Force combined DT&E/IOT&E programs, development test objectives appear to take precedence over operational test objectives. As shown in chapters 2 and 3, combined programs are usually conducted in a development testing environment, aircraft are likely to be instrumented for DT&E, contractor personnel frequently perform much of the maintenance, and time constraints often result in less-than-planned coverage of IOT&E objectives. We believe that these problems are likely to continue in the absence of a separate dedicated phase of IOT&E.

AIR FORCE POSITION ON COMBINED TEST PROGRAMS

At the time of our review the AFTEC commander favored combining DT&E and IOT&E primarily because he believed that this practice is less costly than separate testing. He stated that another advantage is AFTEC early and sustained involvement in testing during system development that increases AFTEC familiarity with the system and enables it to

provide feedback of operational concerns to the development agency. The earlier these operational concerns are surfaced, the easier it is to alter the system to make it more operationally suitable. We were also told that AFTEC early involvement allows it to be sure that OT&E concerns are addressed in pertinent program documents. It should be noted that, although AFTEC favored combined DT&E/IOT&E, it also favored a revision to current Defense directives that would assure that OT&E is allocated some separate dedicated tests.

There is some benefit with AFTEC becoming involved early in a system's development--an increase in its familiarity with the system and early communication of operational concerns to the developing agency. However, we believe that the final phase of IOT&E should be separated from DT&E to assure that all operational test objectives are accomplished before the production decision and that the proper relationship is maintained between the developing agency and the independent testing agency. We believe that, in the absence of a separate dedicated IOT&E phase, any cost savings of combined-versus-separate test programs could be more than offset by the costs of correcting deficiencies after the system enters production--deficiencies that should have been disclosed in IOT&E before the production decision was made.

DEFENSE AND AIR FORCE GUIDANCE ON COMBINED- VERSUS-SEPARATE TEST PROGRAMS

Defense Directive 5000.3 states that operational testing should be separate from development testing. However, development testing and early phases of operational testing may be combined where separation would cause delays that would create unacceptable military risks or that would cause an unacceptable increase in the system acquisition cost. When combined testing is conducted the necessary test conditions and test data required by both the developing agency and OT&E agency must be realized. In addition, the OT&E agency must be assured that the combined test is so planned and executed that it provides the necessary operational test information.

Air Force Regulation 80-14 states that the implementing command (usually Air Force Systems Command), in coordination with the participating test agencies, will develop test and evaluation alternatives early in program planning to determine whether separate or combined testing will save time and money, and assure adequate testing.

In commenting on our report, the Air Force stated that the next revision of Air Force Regulation 80-14 will emphasize separate IOT&E phases where it is feasible.

We believe that it is extremely important to conduct a separate dedicated phase of IOT&E on all major acquisition programs. As stated on page 12 of this report, the objectives of DT&E and IOT&E are considerably different. Because of these differences, we do not believe that all important test objectives can be accomplished concurrently.

A former commander of the Navy's independent operational testing agency stated that

"* * * a clear cut distinction between DT&E and OT&E is that DT&E tests a weapon, while OT&E tests a weapon system (and very seldom does the development program involve the whole system). In DT&E, the people who are in the test organization should be primarily technical. For OT&E, the people in operational test organizations should definitely not be technical. They should be thoroughly operational. I can cite only a few examples, maybe half a dozen, of successful concurrent testing--testing in which DT&E and OT&E could be conducted during the same test period."

CHAPTER 5

THE AIR FORCE RECOGNIZES THE NEED FOR

BETTER OPERATIONAL TEST CRITERIA

According to Air Force Headquarters representatives, operational test criteria have not been well defined early in acquisition programs and operational concepts have not been continually updated. As a result, operational tests tended to be subjective and oriented toward contract specifications, even though the specifications may not have been operationally important. Traditionally, the users have not provided developers with updated information that expressed the intended use of new or improved systems. This lack of information could result in the development of systems that do not fulfill the user's current requirements.

There has been a lack of baselines for operational testing of Air Force systems because program documents have not provided updated operational criteria. As a consequence, evaluations of operational effectiveness were not based on the most current operational concept.

For example, thrust performance requirements for the A-10 were evaluated by AFTEC without an updated operational concept provided by the user. This resulted in a major disagreement between AFTEC and the system program office on the adequacy of the aircraft's thrust. Air Force Headquarters personnel stated that there would not have been a disagreement if the using command had provided an updated operational concept. Also, maintainability criteria for the B-1 aircraft were not fully established by the user before IOT&E; therefore, AFTEC could not determine whether the aircraft was meeting all of its maintainability requirements.

Due in part to AFTEC recognition of this problem, the Air Force was actively considering a new policy on developing and updating operational concepts when we completed our review. The operational concept is intended to be a contract between the developer and the user to stabilize system design before full-scale development. It will be used, among other purposes, to establish criteria for test and evaluation. Air Force Headquarters, supporting commands, and AFTEC will be expected to participate along with the developers and users in formulating and updating the operational concept. This policy was implemented in November 1977.

We believe that the new policy will provide a better means for evaluating operational test results against current operational requirements. It will also provide for a better understanding between the user, developer, supporter, AFTEC, and Air Force Headquarters, and will improve the user's ability to influence the system design during development to fulfill their needs. However, in implementing this policy safeguards should be established to assure that the operational concept is based on current threat assessments, is updated during the acquisition process, does in fact reflect user requirements, and provides a basis for establishing operational test criteria.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

Operational testing is performed to determine a system's military utility and operational effectiveness and suitability. It should be an assessment of a complete system's mission performance in its intended environment when operated, maintained, and supported by personnel with similar qualifications to those in the field. Development contracts and test plans should provide for adequate resources, including time, to complete a separate, indepth operational evaluation of the system before the production decision. The OT&E agency should not be dependent on the development agency for essential resources. Also, program documents such as development contracts and test plans should provide sufficiently detailed and current operational requirements to be used as performance criteria during operational test programs.

We believe the practice of performing IOT&E in a development environment, using predominantly more-qualified-than-average user personnel to operate the system and having development contractors provide most of the maintenance and support, could deprive decisionmakers of relevant operational information needed to make informed production decisions. Limitations such as unavailable and unsuitable test resources, limited time, and inadequate operational test criteria have affected the effectiveness of AFTEC in performing its mission. AFTEC dependence on the developer for test resources during IOT&E could have an adverse effect on its independence. A separate dedicated phase of IOT&E is needed to assure that all operational test objectives are accomplished before the production decision and to assure that the proper adversary relationship is maintained between AFTEC and the development agency.

Despite these limitations, we believe that AFTEC, during its short existence, has had positive effects on the Air Force test and evaluation process by adding realism to test programs and by recognizing the need for improved operational test criteria to make better operational effectiveness and suitability evaluations. We strongly support the concept of an independent test organization and believe that it should be able to conduct realistic testing.

In commenting on our report, the Air Force basically agreed with our position (see app. I) and cited several measures being undertaken to resolve the problems. The Air Force stated that increased use of realistic operational test locations will be emphasized in future test programs.

The availability of test resources is receiving considerable attention within the Air Force. For example, AFTEC has established advanced planning to provide for (1) an adequate number of test articles, (2) unique instrumentation, and (3) the time needed to accomplish operational testing. The Air Force also stated that it has initiated action to provide high-level visibility to OT&E resource requirements through development of a resource management system that includes a semiannual review and approval of OT&E resources by a general officer committee. The Air Force also amended its regulation (AFR 57-1) to include a policy for developing operational concepts to provide, among other things, better operational test criteria. In addition, the Air Force plans to emphasize separate IOT&E where feasible.

We believe that these measures, if fully implemented, will improve Air Force operational testing. However, we believe that additional measures are required to assure that a separate dedicated phase of IOT&E is conducted on all major acquisition programs before the production decision and that average operational and support personnel are used for operational testing.

RECOMMENDATIONS

Because of the importance of making a realistic assessment of the operational effectiveness and suitability of a weapon system before production and the subsequent deployment of the system to the operational forces, we recommend that, for future test programs, the Secretary of Defense

- monitor the Air Force's implementation of the improvements in operational testing cited in its comments to our draft report,
- direct the Air Force to conduct operational test and evaluation with operational and support personnel of the type and qualifications of those expected to use and maintain the system when deployed, and
- direct the Air Force to conduct a separate dedicated phase of IOT&E to assure that all operational test objectives are accomplished under the appropriate test conditions.



RESEARCH AND
ENGINEERING

THE UNDER SECRETARY OF DEFENSE
WASHINGTON, D.C. 20301

14 FEB 1978

Mr. R. W. Gutmann
Director, Procurement and
Systems Acquisition Division
U.S. General Accounting Office
Washington, D. C. 20548

Dear Mr. Gutmann:

This is in reply to your letter, dated November 30, 1977, to the Secretary of Defense regarding your report titled, "Air Force Operational Testing Needs Improvement," OSD Case #4770, GAO Code #951319.

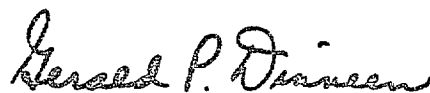
Enclosed is a letter from the Air Force which adequately details their basic agreement with the subject report and is fully supported by OSD. They point out that most of the programs selected for review by the GAO had already entered full-scale development at the time AFTEC was established. Therefore, these established test programs had to be modified to incorporate AFTEC desires and do not reflect the planning and conduct of operational testing that is occurring today. Many of the improvements cited have been implemented on the later programs.

OSD has recently shifted the management of operational testing from USDRE to ASD(P&E). This action should strengthen operational testing in all of DoD. A new DoD Directive 5000.3 is being written but no major policy changes are anticipated. DoDD 5000.3 is a broad directive that establishes operational test and evaluation policy for the Military Departments and other Defense agencies. The GAO report has taken these broad guidelines for operational testing and somewhat narrowly applied them to Initial Operational Test and Evaluation (IOT&E) programs, the majority of programs reviewed. DoDD 5000.3 does not explicitly differentiate between OT&E and IOT&E policy, except with regard to test timing. It must be recognized, however, that different policy is applied to the two OT&E phases. The value gained from early OT&E (particularly combined DT&E/IOT&E) is well worth the necessary limitations imposed by required personnel qualifications, environment, and the limited test time available. These limitations are removed as the system progresses into subsequent OT&E, thereby meeting the requirements of DoDD 5000.3. Further, DoDD 5000.3

states that adequate OT&E must be accomplished only to estimate, not determine, operational effectiveness and suitability. This distinction is often lost when discussing the adequacy of OT&E.

Management tradeoffs are examined on a program-by-program, and often test-by-test, basis to determine if the value of information obtained is worth the cost to obtain it. We shall continue to endeavor to achieve the proper balance of adequate testing at the minimum cost.

Sincerely,



Gerald P. Dineen

Principal Deputy

Enclosure

DEPARTMENT OF THE AIR FORCE
WASHINGTON, D.C. 20330

OFFICE OF THE ASSISTANT SECRETARY

6 JAN 1978

MEMORANDUM FOR ASSISTANT SECRETARY OF DEFENSE (PROGRAM
ANALYSIS AND EVALUATION)

SUBJECT: GAO Draft Report, Air Force Operational Testing
Needs Improvement (OSD Case #4770) (GAO Code #
951319)

The Air Force has been requested to provide comments to your office on the subject report.

We agree with the goals that are embodied in the GAO recommendations. Fiscal constraints, however, will have an impact on how many expensive test articles we can procure solely for IOT&E and how much funding we can apply toward improving the operational environment at test ranges.

Even before the initiation of the GAO review in October 1976, many of these recommendations were recognized by the Air Force and have been implemented fully or in part. Most of the planning for programs selected for review by the GAO had already entered full scale development by the time AFTEC was established in 1974. Programs were modified to the extent possible to accommodate operational considerations requested by AFTEC. Programs that were not yet structured or contractually committed now reflect the trend toward full consideration of OT&E resource requirements, dedicated test events, and operational deployments prior to the production decision.

We support the first, and in principle, the fourth GAO recommendation (1) that more initial operational testing be conducted at locations that more closely approximate the expected operating environment and (2) that the test plans provide for a separate dedicated phase of IOT&E. Increased use of realistic operational test locations is Air Force policy and this practice will be accorded continued emphasis. Because of dollar, time, and resource constraints, it has not been practical to have separate, dedicated IOT&E phases in every program. Where practical, we have conducted

separate IOT&E phases. The F-4G WILD WEASEL and EF-111A programs are examples. Also, the EF-111A has a separate IOT&E phase scheduled for Mountain Home AFB, Idaho, where the system will eventually be deployed. Other systems such as the GBU-15, COBRA DANE, and ATEC will use as the IOT&E site the planned deployed location. The next revision of AFR 80-14 will emphasize separate IOT&E phases where feasible.

We concur in principle with the second GAO recommendation that operation and support personnel selected for test teams should be of the type and qualifications of those expected to use/support the system once it is deployed. However, we have found that personnel with above average qualifications are necessary to assess the military utility particularly of prototype and preproduction weapon systems. Adequate assessments of the system's effectiveness, maintainability, deficiencies, and need for modifications require highly qualified personnel.

The third GAO recommendation that test documentation should assure that adequate OT&E resources are provided is receiving considerable attention. AFTEC has established an advanced planning function that enhances interface with the developing agency from program inception to ensure that OT&E requirements are included early in the test program, to include provision of an adequate number of OT&E test articles, unique instrumentation, and the time needed to accomplish operational testing. In addition, HQ USAF has initiated action to provide high level visibility to OT&E resource requirements through development of an OT&E Resource Management System (ORMAS). ORMAS includes a semi-annual review and approval of OT&E resources by a general officer committee.

GAO's fifth recommendation that regulations include the proposed policy for developing operational concepts to provide better operational test criteria was implemented by a November 1977 interim message change to AFR 57-1. The change delineates responsibilities to ensure that operational concepts are developed in a timely manner; are compatible with current and emerging doctrine, long range planning, strategy, and force structure, and based on current threats; are updated at appropriate intervals during the acquisition process; and that they provide a basis for operational test criteria and reflect the using command's requirements.

The GAO commented on page 5 that they were denied access to trip reports and memos for the record on the basis that they were not official documents. Access to trip reports and memos for the record were denied on the basis that these documents provided the limited perspective of the writer and not a corporate AFTEC or Air Force position. The GAO had requested total unrestricted access to AFTEC files. AFTEC denied such blanket access; withholding such documents which are not releasable through mutual agreement between the Air Force and GAO (i.e., accident investigations and IG reports), and some documents that did not reflect an AFTEC position such as trip reports and memos for the record.

The essence of AFTEC's independence lies in how tests are assigned, designed, conducted, analyzed, evaluated, and reported. Documents which describe AFTEC's relationship with other Air Force agencies include regulations, test plans, management directives, test reports, program introductions, official letters, messages, etc. Relevant issues raised in trip reports or memo for the record are reflected as an AFTEC position in official letters and messages which were available to the GAO. In addition, the GAO conducted scores of interviews, not only with AFTEC personnel, but operating and implementing command personnel as well.

Although, as stated above, many actions have been taken that are in consonance with the GAO proposals, several caveats are in order. Trade-offs must be made between system acquisition dollars and test equipment/environmental dollars as well as in cost versus schedule. Trade-offs between acquiring more test assets, extrapolation of data from other sources, and waiving particulars of a test must be made based on both program and range capability constraints. Finally, the trend toward separate IOT&E and away from combined testing must not be carried so far that ineffective duplication, which we can ill afford, is the result. We will continue to exercise these management judgments on a program by program basis so that acquisition actions are in the best interests of the government.

John Martin

JOHN J. MARTIN
Assistant Secretary
Research, Development
and Logistics